

THE 27TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 21-23 2015
BERGEGGI, ITALY



EDITED BY
MICHAEL AFFENZELLER
AGOSTINO G. BRUZZONE
EMILIO JIMÉNEZ
FRANCESCO LONGO
YURI MERKURYEV
LIN ZHANG

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2015

ISBN 978-88-97999-48-5 (Paperback)
ISBN 978-88-97999-57-7 (PDF)

© 2015 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 978-88-97999-48-5 (Paperback)

ISBN 978-88-97999-57-7 (PDF)

THE 27TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 21-23 2015, BERGEGGI, ITALY

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 2015 INDUSTRIAL SPONSORS



CAL-TEK SRL



LIOTECH LTD



MAST SRL



SIM-4-FUTURE

I3M 2015 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



EUROMERCI



DE GRUYTER
INTERNATIONAL JOURNAL OF FOOD ENGINEERING

EDITORS

MICHAEL AFFENZELLER

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

AGOSTINO BRUZZONE

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

EMILIO JIMÉNEZ

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

FRANCESCO LONGO

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

YURI MERKURYEV

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

LIN ZHANG

BEIHANG UNIVERSITY, CHINA
zhanglin@buaa.edu.cn

THE INTERNATIONAL MULTIDISCIPLINARY MODELING AND SIMULATION MULTICONFERENCE, I3M 2015

GENERAL CO-CHAIRS

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

PROGRAM CO-CHAIRS

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

THE 27TH EUROPEAN MODELING & SIMULATION SYMPOSIUM, EMSS 2015

GENERAL CO-CHAIRS

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

PROGRAM CO-CHAIRS

MICHAEL AFFENZELLER, *UPPER AUSTRIAN UNIVERSITY OF APPLIED SCIENCES, AUSTRIA*
LIN ZHANG, *BEIHANG UNIVERSITY, CHINA*

EMSS 2015 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*
HIPÓLITO CARVAJAL, *UNIVERSITY OF ORIENTE, CUBA*
ALESSANDRO CHIURCO, *UNIVERSITY OF CALABRIA, ITALY*
FABIO DE FELICE, *UNIVERSITY OF CASSINO, ITALY*
MARIA PIA FANTI, *POLYTECHNIC UNIVERSITY OF BARI, ITALY*
IDALIA FLORES, *UNIVERSITY OF MEXICO, MEXICO*
SERGIO GALLO, *UNIV. OF MODENA AND REGGIO EMILIA, ITALY*
WITOLD JACAK, *UPPER AUSTRIA UAS, AUSTRIA*
EMILIO JIMÉNEZ, *UNIVERSITY OF LA RIOJA, SPAIN*
ANDREAS KÖRNER, *VIENNA UNIVERSITY OF TECHNOLOGY, AUSTRIA*
CARLOS JAVIERRE LARDIÉS, *OTRI UNIVERSITY OF ZARAGOZA, SPAIN*
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*
MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*
RICCARDO MELLONI, *UNIV. OF MODENA AND REGGIO EMILIA, ITALY*
YURI MERKURYEV, *RIGA TECHNICAL UNIVERSITY, LATVIA*
SAURABH MITTAL, *DUNIP TECHNOLOGIES, USA*
TANYA MORENO CORONADO, *UNAM, MÉXICO*
MIGUEL MÚJICA MOTA, *AMSTERDAM UNIVERSITY OF APPLIED SCIENCES, NETHERLANDS*
TERESA MURINO, *UNIVERSITY OF NAPLES FEDERICO II, ITALY*
GASPER MUSIC, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
LETIZIA NICOLETTI, *CAL-TEK SRL, ITALY*
ELEONORA PANTANO, *MIDDLESEX UNIVERSITY, UK*
MERCEDES PÉREZ, *UNIVERSITY OF LA RIOJA, SPAIN*
SIMONLUCA POGGI, *MAST SRL, ITALY*
CHUMMING RONG, *UNIVERSITY OF STAVANGER, NORWAY*
LIBERATINA SANTILLO, *UNIVERSITY OF NAPLES FEDERICO II, ITALY*
PEER OLAF SIEBERS, *UNIVERSITY OF NOTTINGHAM, UNITED KINGDOM*
CHRYSOSTOMOS STYLIOS, *TECHNOLOGICAL EDUCATIONAL INSTITUTE OF EPIRUS, GREECE*
FEI TAO, *BEIHANG UNIVERSITY, CHINA*
ALBERTO TREMORI, *NATO STO CMRE, ITALY*
WALTER UKOVICH, *UNIVERSITY OF TRIESTE, ITALY*
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*
SALLY BRAILSFORD, *UNIVERSITY OF SOUTHAMPTON, UK*
GEORGE GEORGIOULAS, *TEIEP, GREECE*
DANIEL GUIMARANS, *NICTA, AUSTRALIA*
GORAD KARER, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
PETROS KARVELIS, *TEIEP, GREECE*
STAVROS KOLIOS, *TEIEP, GREECE*
YUANJUN LAI, *BEIHANG UNIVERSITY, CHINA*
FENG LI, *BEIHANG UNIVERSITY, CHINA*
VITO LOGAR, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
GARY MAYER, *SOUTHERN ILLINOIS UNIVERSITY, USA*
PASCAL PEREZ, *UNIVERSITY OF WOLLONGONG, AUSTRALIA*
JEAN-FRANÇOIS SANTUCCI, *UNIVERSITY OF CORSICA, FRANCE*
CHUGMAN SEO, *ARIZONA CENTER FOR INTEGRATIVE MODELING, USA*
XIAO SONG, *BEIHANG UNIVERSITY, CHINA*
ANN WELLENS, *UNAM, MEXICO*
YINGNIAN WU, *BEIJING INFORMATION SCIENCE TECHNOLOGY UNIVERSITY, CHINA*
BORUT ZUPANCIC, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
TUNCER ÖREN, *UNIVERSITY OF OTTAWA, CANADA*

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; SIMONLUCA POGGI, MAST SRL, ITALY

SIMULATION OPTIMIZATION APPROACHES IN INDUSTRY, AVIATION, SERVICES AND TRANSPORT

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

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

CHAIRS: ALBERTO TREMORI, NATO STO CMRE, ITALY; 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 A. 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

GENERAL CO-CHAIRS' MESSAGE

WELCOME TO EMSS 2015!

On behalf of the Organization Committee, we would offer our warmest thanks to all those who have contributed to the 27th European Modeling and Symposium (EMSS 2015). After so many editions, EMSS 2015 carries on with the long tradition started in 1988 bringing new ideas, solutions, approaches and plenty more. Its key to success is in its ability to save its unique nature being able, at the same time, to integrate new perspectives and raise new challenges.

EMSS means high-quality research, relevant scientific contributions and much more. As a result of the joint effort of many scientific excellences involved in the EMSS program committee, a network of well-established relations has shaped up over time and ceaselessly grows thanks to all those that, year by year, join the Symposium and provide their contributions. From this point of view, EMSS is an opportunity to meet and discuss and, as usually happens, constructive debate makes people open up, brings food for thoughts as well as new ways to explore encouraging multi-disciplinary cooperation and collaborations. Thus, the heart of EMSS is in all the people that, at various levels, are involved in it with special regard to authors, reviewers, chairs, partners and sponsors who deserve our unconditional appreciation and gratitude.

Thanks to the hard work that has been done so far, EMSS has an undisputed and widely acknowledged value within the scientific arena bringing together researchers, scientists and practitioners, coming from industry and academia, which are concerned with Modeling and Simulation in Industry and other related sectors.

EMSS is a traveling conference that year by year has in its background a different European city. This edition is held in Bergeggi in the wonderful Italian Riviera offering the opportunity to enjoy Italian beauties and traditions. Renewing our commitment for the EMSS 2016, we cannot but wish to all the attendees a superb conference experience and a memorable stay in Bergeggi!



Francesco Longo
DIMEG, University of
Calabria, Italy



Emilio Jiménez
University of La Rioja,
Spain

ACKNOWLEDGEMENTS

The EMSS 2015 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 2015 IPC would like to thank all the authors as well as the reviewers for their invaluable work.

A special thank goes to all the organizations, institutions and societies that have supported and technically sponsored the event.

I3M 2015 INTERNAL STAFF

AGOSTINO G. BRUZZONE, *DIME, UNIVERSITY OF GENOA, ITALY*
MATTEO AGRESTA, *SIMULATION TEAM, ITALY*
TERESA BARBIERI, *CAL-TEK SRL, ITALY*
CHRISTIAN BARTOLUCCI, *SIMULATION TEAM, ITALY*
LUIGI BRUNO, *DIMEG, UNIVERSITY OF CALABRIA*
ALESSANDRO CHIURCO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
MARIO COSENTINI, *CAL-TEK SRL, ITALY*
RICCARDO DI MATTEO, *SIMULATION TEAM, ITALY*
CATERINA FUSTO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
FRANCESCO LONGO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
GIANLUCA MAGLIONE, *CAL-TEK SRL, ITALY*
MARINA MASSEI, *LIOPHANT SIMULATION, ITALY*
LETIZIA NICOLETTI, *CAL-TEK SRL*
ANTONIO PADOVANO, *DIMEG, UNIVERSITY OF CALABRIA, ITALY*
ALBERTO TREMORI, *SIMULATION TEAM, 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

Positive and stable periodic time-varying continuous-time linear systems, Tadeusz Kaczorek	1
A generation algorithm of latin hypercube sampling based on original sampling points Zhizhao Liu, Wei Li, Ming Yang, Jun Xu, Duzheng Qing	7
Stochastic programming for train distribution in a metro transportation system Jorge A. Garcia, Kumaraswamy Ponnambalam	13
An integrated platform for simulation, validation, analysis and optimization of electromagnetic railgun system Yuchen Zhou, Ping Ma, Ming Yang	20
Synergy of matlab and modelica in thermal flows control in buildings Borut Zupančič	25
Web simulation with support of mobile agents Jan Voracek	32
Simulators sustainability assessment Dace Aizstrauta, Egils Ginters, Artis Aizstrauts	36
Dynamic and kinematic simulation of kawasaki manipulator industrial robot using solidworks and matlab simmechanics Zennir Youcef, Makbouche Adel, Souames Hamza	46
Design and development of a mesoscopic simulator specialized in investigating capacities of railway nodes Roman Diviš, Antonín Kavička	52
Ontology and its role in domain-specific modeling Ning Zhu, Yonglin Lei, Xiaobo Li, Yifan Zhu	58
Result validation of a rudder simulation model Lina Meng, Xiaochao Qian, Haiying Wang, Wei Li	63
Web simulation as a platform for training software application Stepan Kartak	70
Modeling and simulation based on finite element method for electron beam welding process Zhengkun Feng, Henri Champlaud	79
Searching the longest common subsequences in distorted data Tomáš Kocyan, Jan Martinovič, Kateřina Slaninová, Daniela Szturcová	84
Simulative analysis of performance shaping factors impact on human reliability in manufacturing activities Valentina Di Pasquale, Salvatore Miranda, Raffaele Iannone, Stefano Riemma	93
Cost-based assessment of partitioning algorithms of agent-based systems on hybrid cloud environments Chahrazed Labba, Narjès Bellamine Ben Saoud	103

Rapid prototyping and verification of simulating systems applying ABASIM architecture Emil Řezanina, Antonín Kavička	113
Model of a railway infrastructure as a part of a mesoscopic traffic simulator Radek Novotny, Antonin Kavicka	120
Urban transport infrastructure: a survey Idalia Flores, Miguel Mújica Mota, Salvador Hernández	126
Ambulance location using linear programming: the case of the national autonomous university of mexico (UNAM) Esther Segura Pérez, Idalia Flores, Alexander Vindel	137
A cloud service based modeling & simulation framework Xiao Song, Feng Li, Lin Zhang, Cheng'en Wang	147
Towards humans issues in cloud manufacturing paradigm Yuqiuge Hao, Rayko Toshev, Petri Helo	151
Cybernetics applied to a six-legged Robot, Michael Bogner, Thomas Haspl, Franz Wiesinger	159
Conducting experimental design and optimization on an innovative car rental business Natalia P. Basán, Mariana E. Cóccola, Carlos A. Méndez	166
Modular approach for modelling an airport system Paolo Scala, Miguel Mujica Mota, Nico de Bock	172
Comparison of optimization methods efficiency in different dimensions of the search space of simulation models Pavel Raska, Zdenek Ulrych	181
Utilization of computer simulation for the localization of rolling stock with utilization technology oracle spatial and dynamic database views Jan Fikejz, Jan Merta	191
Various aspects of multi-method modelling and its applications in modelling large infrastructure systems like airports Barbara Glock, Niki Popper, Felix Breitenecker	197
Performance gain and energy savings using multi-threaded software on an embedded asymmetric chip multiprocessor Michael Bogner, Christof Steps, Franz Wiesinger	207
Pipeline as high level parallel composition for the implementation of a sorting algorithm Mario Rossainz López, Manuel Capel Tuñón, Patricia Domínguez Mirón, Ivo Pineda Torres	213
Heat treatment process parameter estimation using heuristic optimization algorithms Michael Kommenda, Bogdan Burlacu, Reinhard Holecek, Andreas Gebeshuber, Michael Affenzeller	222
Flow simulations in the human cardiovascular system under variable external	228

conditions	
Reece Adragna, Radu C. Cascaval, Maria Pia D'Arienzo, Rosanna Manzo	
Discrete-event simulation modelling of prefabricated wall production line	234
Mohammed Sadiq Altaf, Hexu Liu, Youyi Zhang, Mohamed Al-Hussein, Ahmed Bouferguene	
AMEBA – Structural evolutionary optimization: method and toolbox development	240
Marko Corn, Maja Atanasijević-Kunc	
A fuzzy cognitive maps model to develop a risk analysis model through the identification of critical human factors	250
Fabio De Felice, Antonella Petrillo, Agostino Bruzzone, Francesco Longo	
NumEquaRes - Online simulation tool for education and academic research	258
Stepan Orlov, Nikolay Shabrov	
Extending the retsim simulator for estimating the cost of fraud in the retail store domain	266
Edgar Alonso Lopez-Rojas	
Naval operations' assessment through a multi-purpose HLA simulation framework	272
Davide Tozzi, Federica Valdenazzi, Aldo Zini	
Connection of microcontroller and microcomputer to distributed simulation	282
Martin Jakeš, Josef Brožek	
Simulation-based set-up time optimization using Sim# and Heuristiclab	289
Johannes Karder, Andreas Scheibenpflug, Andreas Beham, Stefan Wagner, Michael Affenzeller	
Hardware libraries for online control of interactive simulations	295
Josef Brozek, Martin Jakes	
A dynamic multicommodity network flow problem for logistics networks	301
Sebastian Raggl, Judith Fechter, Andreas Beham	
Modelling a clustered generalized quadratic assignment problem	307
Judith Fechter, Andreas Beham, Stefan Wagner, Michael Affenzeller	
Combination of learning approaches in modelling and simulation	314
Maja Atanasijević-Kunc, Marko Papić, Borut Zupančič, Janez Bešter	
Sustainable performance management model of manufacturing companies according to sampled methods from performance management system	324
Veronika Burešová, Lilia Dvořáková	
Use of delphi method in obtaining requirements for the game structural masonry	334
Ailton Soares Fereire, Leiliane Santana Souza, Daniela Matschulat Ely, Antônio Edésio Jungles	
Towards an extension of promela for the modeling, simulation and verification of discrete-event systems	340

Aznan Yacoub, Maamar Hamri, Claudia Frydman , Chungman Seo , Bernard P. Zeigler

Validation and approval of PN ciphered subnets: security in PN sharing and storage using PNML, XMLSecurity and XMLSignature	349
Iñigo León Samaniego, Mercedes Perez-Parte, Juan-Ignacio Latorre-Biel, Emilio Jiménez-Macías	
LCA software for environmental impact assesment of injected moulded plastic parts	359
Ana Elduque, Daniel Elduque, Carmelo Pina, Carlos Javierre, Ángel Fernández	
Simulation-based assessment of the influence of requests for information on industrial module construction piping work	368
Amin Amini Khafri, Wenying Ji, Seungjun Ahn, Simaan AbouRizk	
Environmental assessment tool to analyze the presence of critical and valuable raw materials	374
Patricia Gómez, Daniel Elduque, Carmelo Pina, Judith Sarasa, Isabel Clavería, Carlos Javierre	
Supervisory FCM simulation in tunnel construction	382
Limao Zhang, A. J. Antony Chettupuzha, Seungjun Ahn, Stephen Hague, Chrysostomos D. Stylios, Simaan M. AbouRizk	
Investigation of thermal effects on elastic constants to simulate correctly the thermal stress in nano transistors	389
Abderrazzak El Boukili	
Modeling, simulation and optimization of logistics management of a cans packaging line	395
Victoria G. Achkar, Luciana S. Picech, Carlos A. Méndez	
A proposal of using DEVS model for process mining	403
Yan Wang, Grégory Zacharewicz, David Chen, Mamadou Kaba Traoré	
An accredited engineering curriculum for modeling and simulation	410
Frederic D. McKenzie	
DEVS-based framework for message dissemination in multi-layer networks	416
Youssef Bouanan, Mathilde Forestier, Judicael Ribault, Gregory Zacharewicz, Bruno Vallespir	
A specialized library for simulation of coal mining in flat-lying coal seam	425
Victor Okolnishnikov, Sergey Rudometov, Sergey Zhuravlev	
Modelling and simulation examples based on computer algebra systems	430
Stefanie Winkler, Andreas Körner, Andreas Bauer, Martin Bicher, Felix Breiteneker	
The impact of procurement and inventory management in operative performance in a supply chain	436
Jorge Luis García-Alcaraz, Aidé Aracely Maldonado-Macías, Emilio Jiménez Macías, Julio Blanco Fernández	
Cerebellum Function For MSAAS	445
Erdal Cayirci, Hakan Karapinar, Lutfu Ozcakir	

Chemical risk evaluation: application of the Movarish methodology in an industry of the textile sector	451
Vincenzo Duraccio, Gianpaolo Di Bona, Domenico Falcone, Alessandro Silvestri, Antonio Forcina	
Nonlinear scaling in smart adaptive modelling	457
Esko Juuso	
Alternatives aggregation petri nets applied to modular models of discrete event systems	465
Juan Ignacio Latorre Biel, Emilio Jiménez Macías, Jorge Luis García-Alcaraz, Juan Carlos Sáenz Díez Muro, Mercedes Pérez de la Parte	
Inside Virtual: a new app for interactive and intelligent cultural heritage fruition	471
Francesco Longo, Letizia Nicoletti, Gaetano Florio, Marco Vetrano, Luigi Bruno, Lorenzo Caputi	
Synthetic environment for interoperable advanced marine simulation	479
Agostino G. Bruzzone, Marina Massei, Alberto Tremori, Riccardo di Matteo, Giovanni Luca Maglione	
Author's Index	487

POSITIVE AND STABLE PERIODIC TIME-VARYING CONTINUOUS-TIME LINEAR SYSTEMS

Tadeusz Kaczorek

Bialystok University of Technology

kaczorek@isep.pw.edu.pl

ABSTRACT

The positivity and stability of periodic time-varying continuous-time linear systems are addressed. Necessary and sufficient conditions for the positivity and stability of the system are established. The proof of conditions is based on the Lyapunov transformation of time-varying systems to time-invariant linear systems. Examples of positive and stable linear periodic systems are presented.

Keywords: positive, linear, periodic, continuous-time, time-varying, system, stability, test.

1. INTRODUCTION

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

The Lyapunov, Bohl and Perron exponents and stability of time-varying discrete-time linear systems have been investigated in Czornik et. all 2012, 2013, 2014. The positivity and stability of fractional time varying discrete-time linear systems have been addressed in Kaczorek 2014, 2015b, 2015c, 2015d and the stability of continuous-time linear systems with delays in Kaczorek 2009. The fractional positive linear systems have been analyzed in Kaczorek 2008, 2011, 2015a. The positive electrical circuits and their reachability have been considered in Kaczorek 2011d, 2011f and the controllability and observability in Kaczorek 2011a. The stability and stabilization of positive fractional linear systems by state-feedbacks have been analyzed in Kaczorek 2010. The normal positive electrical circuits has been introduced in Kaczorek 2014. The positivity and stability of time-varying continuous-time linear systems and electrical systems has been addressed in Kaczorek 2015f.

In this paper the positivity and stability of periodic time-varying continuous-time linear systems will be addressed.

The paper is organized as follows. In section 2 the Lyapunov reduction of the periodic time-varying

continuous-time linear systems to time-invariant linear systems is recalled. The positivity of periodic time-varying linear systems is considered in section 3. Necessary and sufficient conditions for the stability of positive periodic time-varying linear systems are established in section 4. Concluding remarks are given in section 5.

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

2. REDUCTION TO TIME-INVARIANT OF TIME-VARYING LINEAR SYSTEMS BY LYAPUNOV TRANSFORMATION

Consider the time-varying linear system

$$\dot{x} = A(t)x \quad (2.1)$$

where $x = x(t) \in \mathfrak{R}^{n \times 1}$ and $A(t) \in \mathfrak{R}^{n \times n}$ is a piecewise continuous-time time-varying and bounded matrix.

Let $X = X(t) \in \mathfrak{R}^{n \times n}$ be a matrix satisfying the equation

$$\dot{X} = A(t)X \quad (2.2)$$

and the initial condition $X(t_0) = X_0 = I_n$.

Definition 2.1. The matrix $L(t) \in \mathfrak{R}^{n \times n}$ satisfying the conditions:

- 1) there exists continuous-time $\dot{L}(t) = \frac{dL(t)}{dt}$ in the interval $[t_0, +\infty)$,
- 2) the norms of $L(t)$ and $\dot{L}(t)$ are bounded in the interval $[t_0, +\infty)$,
- 3) $|\det L(t)| > c > 0$ for some constant c ,

is called the Lyapunov transformation matrix.

It is well-known Gantmacher 1959 that the inverse matrix $L^{-1}(t)$ exists and it is also the Lyapunov matrix.

Let

$$X = L(t)Z \quad (2.3)$$

where $Z = Z(t) \in \mathfrak{R}^{n \times n}$.

Using (2.2) and (2.3) we obtain

$$\dot{X} = A(t)X = \dot{L}(t)Z + L(t)\dot{Z} = A(t)L(t)Z \quad (2.4)$$

and

$$\dot{Z} = B(t)Z \quad (2.5)$$

where

$$B(t) = L^{-1}(t)[A(t)L(t) - \dot{L}(t)]. \quad (2.6)$$

The time-varying system (2.1) can be reduced to time-invariant system

$$\dot{Z} = BZ \quad (2.7)$$

by the use of the Lyapunov transformation matrix if there exists matrix $L(t)$ satisfying the equality

$$B = L^{-1}(t)[A(t)L(t) - \dot{L}(t)]. \quad (2.8)$$

For given matrices $A(t)$ and B the matrix $L(t)$ can be found from the equation

$$\dot{L}(t) = A(t)L(t) - L(t)B \quad (2.9)$$

The equation (2.9) follows from (2.8).

The solution of (2.7) has the form

$$Z = Z(t) = e^{Bt} \quad (2.10)$$

for $Z(t_0) = I_n$.

Substituting (2.10) into (2.3) we obtain

$$X = X(t) = L(t)e^{Bt}. \quad (2.11)$$

It is well-known (Gantmacher 1959, Kaczorek 1998) that if the periodic matrix $A(t)$ satisfy the condition

$$A(t+T) = A(t) \quad (2.12)$$

where T is the period of the matrix, then the time-varying system (2.2) can be reduced by the Lyapunov transformation to the time-invariant system (2.5).

In this case from (2.2) and (2.11) we have

$$\dot{X}(t+T) = A(t+T)X(t+T) = A(t)X(t+T) \quad (2.13)$$

and the matrices $x(T)$ and $x(t+T)$ are related by

$$X(t+T) = X(t)V \quad (2.14)$$

where $V \in \mathfrak{R}^{n \times n}$ is a time-invariant nonsingular (constant) matrix since from (2.14) we have

$$\begin{aligned} \dot{V} &= X^{-1}(t)[\dot{X}(t+T) - \dot{X}(t)V] \\ &= X^{-1}(t)[A(t)X(t+T) - A(t)X(t)V] \\ &= X^{-1}(t)A(t)[X(t+T) - X(t)V] = 0. \end{aligned} \quad (2.15)$$

From (2.11) we have

$$L(t) = X(t)e^{-Bt}. \quad (2.16)$$

The matrix $L(t)$ satisfies the condition

$$L(t+T) = L(t) \quad (2.17)$$

since by (2.17) and (2.13)

$$\begin{aligned} L(t+T) &= X(t+T)e^{-B(t+T)} = X(t)Ve^{-BT}e^{-Bt} \\ &= X(t)e^{-Bt} = L(t) \end{aligned} \quad (2.18)$$

From (2.14) and (2.11) it follows that

$$V = e^{BT} \quad (2.19)$$

and

$$B = \frac{1}{T} \ln V. \quad (2.20)$$

From (2.19) it follows that the eigenvalues v_1, \dots, v_n of the matrix V and eigenvalues $\lambda_1, \dots, \lambda_n$ of the matrix B are related by

$$v_k = e^{\lambda_k T} \text{ for } k=1,2,\dots,n. \quad (2.21)$$

From (2.21) we have $|v_k| < 1$ for $k=1,2,\dots,n$ if and only if $\text{Re } \lambda_k < 0$ for $k=1,2,\dots,n$.

Therefore, the following Lemma has been proved.

Lemma 2.1. The time-variant system with the matrix $A(T)$ satisfying the condition (2.12) is asymptotically stable if and only if

$$|v_k| < 1 \text{ for } k=1,2,\dots,n. \quad (2.22)$$

3. POSITIVE TIME-VARYING LINEAR SYSTEMS

Consider the time-varying linear system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \quad (3.1a)$$

$$y(t) = C(t)x(t) + D(t)u(t) \quad (3.1b)$$

where $x(t) \in \mathfrak{R}^n$, $u(t) \in \mathfrak{R}^m$, $y(t) \in \mathfrak{R}^p$ are the state, input and output vectors and $A(t) \in \mathfrak{R}^{n \times n}$,

$B(t) \in \mathfrak{R}^{n \times m}$, $C(t) \in \mathfrak{R}^{p \times n}$, $D(t) \in \mathfrak{R}^{p \times m}$ are real matrices with entries depending continuously on time and $\det A(t) \neq 0$ for $t \in [0, +\infty)$.

Definition 3.1. The system (3.1) is called positive if $x(t) \in \mathfrak{R}_+^n$, $y(t) \in \mathfrak{R}_+^p$, $t \in [0, +\infty)$ for any initial conditions $x_0 \in \mathfrak{R}_+^n$ and all inputs $u(t) \in \mathfrak{R}_+^m$, $t \in [0, +\infty)$.

Theorem 3.1. Let $A(t) \in \mathfrak{R}^{n \times n}$, $t \in [0, +\infty)$. Then

$$\Omega_{t_0}^t(A) = I_n + \int_{t_0}^t A(\tau) d\tau \quad \text{for } t \geq t_0 \quad (3.2)$$

$$+ \int_{t_0}^t A(\tau) \int_{t_0}^{\tau} A(\tau_1) d\tau_1(\tau) d\tau + \dots \in \mathfrak{R}_+^{n \times n}$$

if and only if $A(t) \in M_n$, $t \in [0, +\infty)$.

Proof is given in Gantmacher 1959, Kaczorek 1998.

Corollary 3.1. If the matrix $A(t) \in \mathfrak{R}^{n \times n}$ satisfies the condition

$$A(t_1)A(t_2) = A(t_2)A(t_1) \quad \text{for } t_1, t_2 \in [t_0, t] \in [0, +\infty) \quad (3.3)$$

then

$$\Omega_{t_0}^t(A) = e^{\int_{t_0}^t A(\tau) d\tau}. \quad (3.4)$$

Proof is given in Gantmacher 1959.

Corollary 3.2. Let $A \in \mathfrak{R}^{n \times n}$ be a matrix with constant entries independent of time t . If $A(t) = A$ then

$$\Omega_{t_0}^t(A) = e^{A(t-t_0)}. \quad (3.5)$$

Proof is given in Gantmacher 1959.

Lemma 3.2. The solution of the equation (3.1a) with given initial condition $x_0 = x(t_0) \in \mathfrak{R}^n$ and input $u(t) \in \mathfrak{R}^m$ has the form

$$x = \Omega_{t_0}^t(A)x(t_0) + \int_{t_0}^t K(t, \tau)B(\tau)u(\tau) d\tau \quad (3.6a)$$

where

$$K(t, \tau) = \Omega_{t_0}^t(A)[\Omega_{t_0}^\tau(A)]^{-1}. \quad (3.6b)$$

Proof is given in Gantmacher 1959.

Theorem 3.3. The time-varying linear system (3.1) is positive if and only if

$$A(t) \in M_n, \quad B(t) \in \mathfrak{R}_+^{n \times m},$$

$$C(t) \in \mathfrak{R}_+^{p \times n}, \quad D(t) \in \mathfrak{R}_+^{p \times m}, \quad t \in [0, +\infty). \quad (3.7)$$

Proof is given in Kaczorek 1998.

Definition 3.2. The matrix $L(t) \in \mathfrak{R}_+^{n \times n}$ is called monomial if only one entry in each row and in each column is positive for $t \in [0, +\infty)$ and the remaining entries are zero.

Theorem 3.4. The time-varying linear system (2.1) with periodic matrix $A(t)$ satisfying the condition (2.12) is positive if and only if the matrix $L(t) \in \mathfrak{R}_+^{n \times n}$ is monomial and $B \in M_n$.

Proof. It is well-known that $Z = e^{Bt} \in \mathfrak{R}_+^{n \times n}$ for $t \in [0, +\infty)$ if and only if $B \in M_n$. From (2.3) it follows that $X(t) \in \mathfrak{R}_+^{n \times n}$ for $t \in [0, +\infty)$ if and only if $L(t) \in \mathfrak{R}_+^{n \times n}$ is monomial matrix. \square

Theorem 3.5. The positive time-variant linear system (3.1) with matrices satisfying the condition

$$A(t+T) = A(t), \quad B(t+T) = B(t),$$

$$C(t+T) = C(t), \quad D(t+T) = D(t) \quad (3.8)$$

is positive if and only if:

- 1) the matrix $L(t) \in \mathfrak{R}_+^{n \times n}$ is monomial and $B \in M_n$,
- 2) $B(t) \in \mathfrak{R}_+^{n \times m}$, $C(t) \in \mathfrak{R}_+^{p \times n}$,
 $D(t) \in \mathfrak{R}_+^{p \times m}$ for $t \in [0, +\infty)$. (3.9)

Proof of the first condition follows immediately from Theorem 3.3. Proof of (3.5) is standard and similar to the one given in Gantmacher 2011f. \square

Example 3.1. Consider the system (2.1) with the matrices

$$A(t) = \begin{bmatrix} -\frac{4+2\sin t - \cos t}{2-\cos t} & 0 \\ \frac{2+\sin t}{2+\sin t} & \frac{-2+\sin t + \cos t}{2-\cos t} \end{bmatrix}, \quad (3.10)$$

It is easy to see that the matrix (3.10) is periodic and its period is $T = 2\pi$.

In this case the Lyapunov transformation matrix has the form

$$L(t) = \begin{bmatrix} 0 & 2+\sin t \\ 2-\cos t & 0 \end{bmatrix}, \quad (3.11)$$

and its period is also $T = 2\pi$.

Using (2.5), (2.6) and (2.7) we obtain

$$\begin{aligned}
B &= L^{-1}(t)[A(t)L(t) - \dot{L}(t)] \\
&= \begin{bmatrix} 0 & \frac{1}{2 - \cos t} \\ \frac{1}{2 + \sin t} & 0 \end{bmatrix} \\
&\times \left\{ \begin{bmatrix} -\frac{4 + 2\sin t - \cos t}{2 + \sin t} & 0 \\ \frac{2 + \sin t}{2 - \cos t} & \frac{-2 + \sin t + \cos t}{2 - \cos t} \end{bmatrix} \right. \\
&\quad \times \left. \begin{bmatrix} 0 & 2 + \sin t \\ 2 - \cos t & 0 \end{bmatrix} - \begin{bmatrix} 0 & \cos t \\ \sin t & 0 \end{bmatrix} \right\} \\
&= \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix},
\end{aligned} \tag{3.12}$$

The equation (2.7) has the form

$$\dot{Z} = \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} Z \tag{3.13}$$

and its solution

$$Z = Z(t) = e^{Bt} = \begin{bmatrix} e^{-t} & e^{-t} - e^{-2t} \\ 0 & e^{-2t} \end{bmatrix}. \tag{3.14}$$

The matrix of solution of (2.2) with (3.10) is given by

$$\begin{aligned}
X &= X(t) = L(t)Z(t) \\
&= \begin{bmatrix} 0 & 2 + \sin t \\ 2 - \cos t & 0 \end{bmatrix} \begin{bmatrix} e^{-t} & e^{-t} - e^{-2t} \\ 0 & e^{-2t} \end{bmatrix} \\
&= \begin{bmatrix} 0 & e^{-2t}(2 + \sin t) \\ e^{-t}(2 - \cos t) & (2 - \cos t)(e^{-t} - e^{-2t}) \end{bmatrix}.
\end{aligned} \tag{3.15}$$

It is easy to check that every column of the matrix (3.15) is a solution of the equation (2.2) with (3.10).

Example 3.2. Compute the solution of the equation (3.1a) with the matrix $A(t)$ given by (3.10) and

$$B(t) = \begin{bmatrix} 2 + \sin t \\ 2 - \cos t \end{bmatrix} \tag{3.16}$$

for given initial condition $x(0) = x_0$ and an input $u(t)$.

The matrices (3.10) and (3.16) are periodic with the same period $T = 2\pi$.

Using (3.6b), (3.11) and (3.14) we obtain

$$\begin{aligned}
K(t, \tau) &= X(t)[X(\tau)]^{-1} = L(t)Z(t)[L(\tau)Z(\tau)]^{-1} \\
&= L(t)Z(t)[Z(\tau)]^{-1}[L(\tau)]^{-1} \\
&= \begin{bmatrix} 0 & 2 + \sin t \\ 2 - \cos t & 0 \end{bmatrix} \begin{bmatrix} e^{-t} & e^{-t} - e^{-2t} \\ 0 & e^{-2t} \end{bmatrix}
\end{aligned} \tag{3.17}$$

$$\begin{aligned}
&= \begin{bmatrix} e^{-\tau} & e^{-\tau} - e^{-2\tau} \\ 0 & e^{-2\tau} \end{bmatrix}^{-1} \begin{bmatrix} 0 & 2 + \sin \tau \\ 2 - \cos \tau & 0 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} 0 & e^{-2t}(2 + \sin t) \\ e^{-t}(2 - \cos t) & (e^{-t} - e^{-2t})(2 - \cos t) \end{bmatrix} \begin{bmatrix} \frac{e^\tau - e^{2\tau}}{2 + \sin \tau} & \frac{e^\tau}{2 - \cos \tau} \\ \frac{e^{2\tau}}{2 + \sin \tau} & 0 \end{bmatrix} \\
&= \begin{bmatrix} \frac{e^{2(\tau-t)}(2 + \sin t)}{2 + \sin \tau} & 0 \\ \frac{(e^{\tau-t} - e^{2(\tau-t)})(\cos t - 2)}{2 + \sin \tau} & \frac{e^{\tau-t}(\cos t - 2)}{\cos \tau - 2} \end{bmatrix}
\end{aligned} \tag{3.17}$$

Taking into account that in this case

$$\Phi_{t_0=0}^t(A) = X(t) = \begin{bmatrix} 0 & e^{-2t}(2 + \sin t) \\ e^{-t}(2 - \cos t) & (2 - \cos t)(e^{-t} - e^{-2t}) \end{bmatrix} \tag{3.18}$$

and using (3.6a) and (3.17) we obtain the desired solution of the equation (3.1a) with (3.10) and (3.16) in the form

$$\begin{aligned}
X(t) &= X(t)x(0) + \int_0^t K(t, \tau)B(\tau)u(\tau)d\tau \\
&= \begin{bmatrix} 0 & e^{-2t}(2 + \sin t) \\ e^{-t}(2 - \cos t) & (2 - \cos t)(e^{-t} - e^{-2t}) \end{bmatrix} x_0 \\
&\quad + \int_0^t \begin{bmatrix} \frac{e^{2(\tau-t)}(8\sin t - \cos 2t + 9)}{2(2 + \sin \tau)} \\ \frac{e^{\tau-t}(\cos t - 2)^2}{\cos \tau - 2} - \frac{(e^{\tau-t} - e^{2(\tau-t)})(\cos t - 2)(2 + \sin t)}{2 + \sin \tau} \end{bmatrix} u(\tau)d\tau.
\end{aligned} \tag{3.19}$$

4. STABILITY OF POSITIVE PERIODIC SYSTEMS

Consider the time-varying linear system (2.1) with periodic matrix $A(t)$ satisfying the condition (2.12).

Definition 4.1. The positive periodic time-varying system (2.1) is called asymptotically stable if

$$\lim_{t \rightarrow \infty} \|X(t)\| = 0 \text{ for all } X_0 \in \mathfrak{R}_+^n. \tag{4.1}$$

It is assumed that there exists the Lyapunov transformation matrix $L(t)$ satisfying the conditions of Definition 2.1 that reduces the system (2.1) to (2.7).

Theorem 4.1. The positive periodic time-varying system (2.1) is asymptotically stable if and only if the coefficients a_k of the characteristic polynomial of the matrix B of the reduced system (2.7)

$$\det[I_n s - B] = s^n + a_{n-1}s^{n-1} + \dots + a_1 s + a_0, \tag{4.2}$$

are positive, i.e. $a_k > 0$ for $k = 0, 1, \dots, n-1$.

Proof. It is well-known Kaczorek 2014 that the positive linear time-invariant system is asymptotically stable if and only if the coefficients of the characteristic

polynomial of its Metzler matrix $B \in M_n$ are positive. From (2.3) it follows that for Lyapunov transformation matrix $L(t)$ we have

$$\lim_{t \rightarrow \infty} \|X(t)\| = 0 \text{ if and only if } \lim_{t \rightarrow \infty} \|Z(t)\| = 0. \quad (4.3)$$

This completes the proof. \square

Example 4.1. (Continuation of Example 3.1). From the matrix (3.10) it follows that the matrix B is given by (3.12) and its characteristic polynomial

$$\det[I_2 s - B] = \begin{vmatrix} s+1 & -1 \\ 0 & s+2 \end{vmatrix} = s^3 + 3s + 2, \quad (4.4)$$

has positive coefficients: $a_0 = 2$, $a_1 = 3$.

Therefore, the positive system (2.7) with the matrix (3.12) is asymptotically stable. The Lyapunov transformation matrix is given by (3.11) and it is easy to verify that (4.3) holds.

Remark 4.1. Note that for checking the asymptotic stability of the positive periodic time-varying linear systems all tests given in Kaczorek 2014 can be applied.

5. CONCLUDING REMARKS

The positivity and stability of the periodic time-varying linear systems have been addressed. Necessary and sufficient conditions for the positivity and asymptotic stability of the system have been established. Using the Lyapunov transformation of the periodic time-varying linear systems to time-invariant linear systems. The considerations have been illustrated by examples of periodic time-varying linear systems. The consideration can be easily extended to discrete-time periodic linear systems.

ACKNOWLEDGMENTS

This work was supported under work S/WE/1/11.

REFERENCES

- Czornik A., Klamka J., Niezabitowski M., 2014. On the set of Perron exponents of discrete linear systems, Proc. of World Congress of the 19th International Federation of Automatic Control, Kapsztad, 11740-11742.
- Czornik A., Newrat A., Niezabitowski M., Szyda A., 2012. On the Lyapunov and Bohl exponent of time-varying discrete linear systems, 20th Mediterranean Conf. on Control and Automation (MED), Barcelona, 194-197.
- Czornik A., Newrat A., Niezabitowski M., 2013. On the Lyapunov exponents of a class of the second order discrete time linear systems with bounded perturbations, Dynamical Systems: an International Journal, 28, (4), 473-483.
- Czornik A., Niezabitowski M., 2013. Lyapunov Exponents for Systems with Unbounded Coefficients, Dynamical Systems: an International Journal, 28, (2), 140-153.
- Czornik A., Niezabitowski M., 2013. On the stability of discrete time-varying linear systems, Nonlinear Analysis: Hybrid Systems, 9, 27-41.
- Czornik A., Niezabitowski M., 2013. On the stability of Lyapunov exponents of discrete linear system, Proc. of European Control Conf., Zurich, 2210-2213.
- Farina L., Rinaldi S., 2000. Positive Linear Systems; Theory and Applications, J. Wiley, New York.
- Gantmacher F.R., 1959. The Theory of Matrices, Chelsea Pub. Comp., London.
- Kaczorek T., 2011a. Controllability and observability of linear electrical circuits, Electrical Review, 87, (9a), 248-254.
- Kaczorek T., 2015a. Fractional standard and positive descriptor time-varying discrete-time linear systems, Progress in Automation, Robotics and Measuring Techniques, Advances in Intelligent Systems and Computing, 350, 101-112.
- Kaczorek T., 2008. Fractional positive continuous-time linear systems and their reachability, Int. J. Appl. Math. Comput. Sci., 18, (2), 223-228.
- Kaczorek T., 2011b. New stability tests of positive standard and fractional linear systems, Circuits and Systems, 2, 261-268.
- Kaczorek T., 2014. Normal positive electrical circuits, IET Control Theory and Applications, doi: 10.1049/iet-cta.2014.0183.
- Kaczorek T., 2015e. Positive descriptor time-varying discrete-time linear systems and their asymptotic stability, DOI: 10.12716/1001.09.01.10.
- Kaczorek T., 2011c. Positive electrical circuits and their reachability, Archives of Electrical Engineering, 60, (3), 283-301.
- Kaczorek T., 2011d. Positive linear systems consisting of n subsystems with different fractional orders, IEEE Trans. Circuits and Systems, 58, (6), 1203-1210.
- Kaczorek T., 2011e. Positivity and reachability of fractional electrical circuits, Acta Mechanica et Automatica, 5, (2), 42-51.
- Kaczorek, T., 2015b. Positivity and stability of fractional descriptor time-varying discrete-time linear systems, Submitted to AMCS.
- Kaczorek, T., 2015c. Positivity and stability of time-varying discrete-time linear systems, Intelligent Information and Database Systems, Lecture Notes in Computer Science, 9011, 295-303.
- Kaczorek T. 2015d. Positive and stable time-varying continuous-time linear systems and electrical circuits, Submitted to Conf. SPETO.
- Kaczorek T. 2002. Positive 1D and 2D Systems, Springer Verlag, London.
- Kaczorek T., 2015f. Positive time-varying continuous-time linear systems and electrical circuits', Submitted to Con. Logitrans.
- Kaczorek T., 2011f. Selected Problems of Fractional System Theory, Springer Verlag.

- Kaczorek T., 2009. Stability of positive continuous-time linear systems with delays, *Bull. Pol. Acad. Sci. Techn.*, 57, (4), 395-398.
- Kaczorek T., 2010. Stability and stabilization of positive fractional linear systems by state-feedbacks, *Bull. Pol. Acad. Sci. Techn.*, 58, (4), 517-554.
- Kaczorek T., 1998. *Vectors and Matrices in Automation and Electrotechnics*, WNT, Warszawa.

AUTHORS BIOGRAPHY

Tadeusz Kaczorek (born 1932, Poland) received his M.Sc., Ph.D. and D.Sc. degrees from the Faculty of Electrical Engineering at Warsaw University of Technology in 1956, 1962 and 1964, respectively. He began his professional career in 1954 at Warsaw University of Technology and continued to work there till 2003. On his retirement he has worked at the Faculty of Electrical Engineering of Bialystok University of Technology till now. From 1968 to 1969 he was the Dean of the Faculty of Electrical Engineering and from 1970 to 1973 he was the Prorector of Warsaw University of Technology. Since 1971 he has been a professor and since 1974 a full professor at Warsaw University of Technology. In 1986 he was elected a corresponding member and in 1996 a full member of the Polish Academy of Sciences. From 1988 to 1991 he was the director of the Research Centre of the Polish Academy of Sciences in Rome. In June 1999 he was elected a full member of the Academy of Engineering in Poland. In May 2004 he was elected honorary member of the Hungarian Academy of Sciences. He was awarded a degree honoris causa by the 11 Universities in Poland.

His research interests cover the theory of systems and the automatic control systems theory, especially singular multidimensional systems, positive multidimensional systems, fractional systems and singular positive 1D and 2D systems. He has initiated research in the field of singular 2D, positive 2D linear systems and positive fractional 1D and 2D systems. He has published 25 books (8 in English) and over 1100 scientific papers.

A GENERATION ALGORITHM OF LATIN HYPERCUBE SAMPLING BASED ON ORIGINAL SAMPLING POINTS

Zhizhao LIU ^(a), Wei LI ^(b), Ming YANG ^(c), Jun XU ^(d), Duzheng QING ^(e)

^{(a), (b), (c)}Control and Simulation Center, Harbin Institute of Technology, Harbin 150080, China

^{(d), (e)}Science and Technology on Special System Simulation Laboratory, Beijing Simulation Center, Beijing 100854, China

^(a)liuzhizhao2007@hotmail.com, ^(b)fleehit@163.com, ^(c)myang@hit.edu.cn
^(d)xujun711@sina.com, ^(e)qing_dz@163.com

ABSTRACT

In order to utilize original sampling points as many as possible to construct a Latin Hypercube Sampling (LHS), a generation algorithm of LHS is proposed. The original sampling can be an arbitrary sampling and the new sampling is a LHS strictly. As the high structure of LHS, only part of original sampling points can be used for LHS. There are several problems including how to select the reserved original sampling points and generate the new ones if necessary. Matrices are used to describe the original sampling points and the original sampling points are reserved based on an adjacent matrix. The new sampling points are generated by a subtraction rule. Numerical experiments show that the generation algorithm is effective and saves a lot of running time for simulation systems.

Keywords: Latin Hypercube Sampling; generation algorithm; adjacent matrix; subtraction rule.

1. INTRODUCTION

Design and analysis of experiments is an important method of science and technology. For simulation systems, we need to apply different experimental design methods for different analysis goals such as sensitivity analysis (Borgonovo Castaings Tarantola 2005, Storlie Swiler Helton and et al. 2009), variance analysis (Huang Lo and Lu 2013), uncertainty analysis (Zouaoui and Wilson 2003, Helton Davis and Johnson 2005, Braton Nelson and Wei 2013) and so on. All these procedures need sampling, i.e., running simulation system. However, simulation experiments may be time-consuming to execute. There is a popular problem that one analysis cannot apply the sampling points of other analyses, which wastes a lot of original sampling points and needs too long time to run simulation system. For example, factorial design is used for sensitivity analysis at first and LHS is needed for uncertainty analysis. A new sample means a lot of simulation runs and LHS based on original points could reduce the running time. So this paper discusses how to utilize original sampling points for a new analysis.

In 1979, Mckay proposed a stratified sampling algorithm named Latin Hypercube Sampling (LHS)

(Mckay Beckman and Conover 1979). It is one of the most popular sampling methods, which has some advantages such as space filling effect, robustness and good convergence character. There have been some researches about the extension of LHS. They start with a LHS and generate a new LHS of a larger size that remains the original sampling points. Some multiple extension algorithms were proposed during the past decade (Tong 2006, Sallaberry Helton and Hora 2008, Wu and Chen 2011). Under this circumstance, the new LHS can contain all the original LHS points. There were two general extension algorithms in 2003 and 2012 (Wang 2003, Wei 2012). But the general extension algorithm of LHS achieve a similar LHS finally, not a strict one. Therefore, recent researches focus on the generation algorithm of LHS based on a special sample, i.e. LHS. In this paper, we consider a general generation algorithm that achieves a strict LHS and the original sample can be any sample, such as a Monte Carlo sampling, a factorial design, an orthogonal array and so on. This algorithm can be applied for more circumstances, including normal analysis based on original sampling points and sequential simulation experiments.

The remainder of this paper is organized as follows. In section 2, the generation problem and the generation algorithm of LHS are described. The locations of original sampling points in LHS are expressed by several matrices and the relationship of original sampling points in LHS is expressed by an adjacent matrix. Section 3 gives two applications of generation algorithm, Section 3.1 for illustration of algorithm and Section 3.2 for metamodeling. Section 4 gets the conclusion and the future research.

2. LHS AND GENERATION ALGORITHM

Before discussing the generation problem, the procedure of classical LHS is described in Section 2.1. The generation problem and correlation studies are in Section 2.2. Then, we give the steps of generation algorithm of LHS based on original sampling points in Section 2.3.

2.1 Procedure of LHS

The input variables are noted as x_1, x_2, \dots, x_r and the range of x_i is $u_i = [\alpha_i, \beta_i]$, $i = 1, 2, \dots, r$. Then, LHS can be got as the steps below, which can ensure that each input variable has all intervals among its range.

Step 1. divide the range u_i into n equiprobable intervals $u_{i1}, u_{i2}, \dots, u_{in}$, $i = 1, 2, \dots, r$. So the intervals satisfy $u_{i1} \cup u_{i2} \dots \cup u_{in} = u_i$, $u_{ij} \cap u_{ik} = \emptyset$ and $P(x \in u_{ij}) = 1/n$, where $j, k = 1, 2, \dots, n$.

Step 2. for the j th interval of variable x_i , the cumulative probability can be obtained as:

$$Prob_{ji} = (j-1)/n + r_{ji}/n \quad (1)$$

where r_{ji} is a uniform random number ranging from 0 to 1. So all the probability values can be noted as $Prob = (Prob_{ji})_{n \times r}$.

Step 3. transform the probability into the sample value x_{ji} by the inverse of the distribution function $F(\cdot)$:

$$x_{ji} = F_i^{-1}(Prob_{ji}) \quad (2)$$

Then, the sample matrix is

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1r} \\ x_{21} & x_{22} & \dots & x_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nr} \end{bmatrix}.$$

Step 4. the n values of each variables are paired randomly or in some prescribed order with the n values of the other variables. Then the sample matrix of LHS can be written as:

$$X' = \begin{bmatrix} x'_{11} & x'_{12} & \dots & x'_{1r} \\ x'_{21} & x'_{22} & \dots & x'_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ x'_{n1} & x'_{n2} & \dots & x'_{nr} \end{bmatrix}$$

where each row is a sample point.

2.2 Research on key problems

Assume that there is an original sampling noted as $A = \{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_m\}$, where \vec{a}_i is a sampling point, $i = 1, 2, \dots, m$. The objective is to get LHS of size n noted as $B = \{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_n\}$, where $n \geq 1$ and n is an integer. B satisfies the constraint condition as shown in Eq. (3).

$$\max \text{card}(A \cap B) \quad (3)$$

where $\text{card}(\cdot)$ denotes the number of elements in set. It means that the new sampling is a LHS strictly and maintains the most original sampling points.

The first problem is how to indicate the location of original sampling points in the expected LHS, which is the base of selecting original points and adding new ones. In this paper, we use some matrices including

sampling matrix $X_{m \times r}$, cumulative probability matrix $U_{m \times r}$ (Every element of $U_{m \times r}$ is the cumulative probability of the corresponding element of $X_{m \times r}$, that is the $Prob_{ji}$ in Eq. (1)), order matrix $L_{m \times r}$ (Every element of $L_{m \times r}$ is the interval number of the corresponding element of $X_{m \times r}$), random matrix $R_{m \times r}$ (Every element of $R_{m \times r}$ is the random location among interval of the corresponding element of $X_{m \times r}$).

Another problem is how to remain the most original points or delete the least ones. It is difficult because the relation of original sampling points is hard to express. We use an adjacent matrix $A_{m \times m}$ to solve the problem. The element a_{ij} of $A_{m \times m}$ notes whether a variable value of the i th original point is in the same interval as the one of the j th original point, 1 for the existence and 0 for the other circumstance. Therefore, how to select original points is based on $A_{m \times m}$. If $A_{m' \times m'} = \vec{0}$, there is no need to delete any remaining sampling points.

The finally problem is how to add new sampling points if the remaining original ones are not enough. In this paper, we follow the subtraction rule (Wu and Chen 2011). At the beginning, we generate a LHS randomly. And then, delete the interval number which is the same as remaining original sampling points. The rest of interval number is randomly combined and the location among the corresponding interval is also achieved randomly, which can ensure that the remaining points and the new points can construct a LHS that meets the requirement.

2.3 Generation algorithm

In Section 2.2, we solve the main problems of generation algorithm. This section mainly shows the step of generation algorithm of LHS based on original sampling points.

Step 1. Compute the location of original sample points in LHS.

a. Compute the cumulative probability matrix $U_{m \times r}^{(1)}$ and the order matrix $L_{m \times r}^{(1)}$ as shown in Eq. (4) and Eq. (5).

$$u_{i,j}^{(1)} = \frac{x_{i,j}^{(old)} - \beta_j}{\alpha_j - \beta_j} \quad (4)$$

$$l_{i,j}^{(1)} = \begin{cases} 1 & , x_{i,j}^{(old)} = \beta_j \\ \left\lceil \frac{x_{i,j}^{(old)} - \beta_j}{\alpha_j - \beta_j} \cdot n \right\rceil & , \beta_j < x_{i,j}^{(old)} \leq \alpha_j \end{cases} \quad (5)$$

where $i = 1, 2, \dots, m$, $j = 1, 2, \dots, r$, $\lceil \cdot \rceil$ is a bracket function to get the maximum integer number which is more than the element inside.

b. Get the random matrix $R_{m \times r}^{(1)}$ as shown in Eq. (6).

$$r_{i,j}^{(1)} = u_{i,j}^{(1)} \cdot n - l_{i,j}^{(1)} + 1 \quad (6)$$

where $i=1,2,\dots,m$, $j=1,2,\dots,r$.

Step 2. Get the maximum remaining sample matrix $X_{m' \times r}^{(2)}$ of the original Sampling.

a. Get the adjacent matrix $A_{m \times m}$ of the original sampling based on $L_{m \times r}^{(1)}$ as shown in the equation bellow. If $A_{m \times m} = \vec{0}$, there is no need to delete more sample points and go to *step 3*. Otherwise, go to *step 2-b*.

$$a_{ij} = \begin{cases} 1, & \exists q \in [1 m], q \in Z, l_{iq}^{(1)} = l_{jq}^{(1)} \\ 0, & \text{otherwise} \end{cases}$$

b. Get all the combinations of the original points, which is noted as $D = [D_1 \ D_2 \ \dots \ D_q]^T$ where every row of D is a rank (If the original point is deleted, the element value is 1. Otherwise, the element value is 0) and the number of 0 in B_i is not less than B_j where $1 \leq i < j \leq m$. Set $p=1$ and delete the point(s) corresponded with the element(s) of D_p , then judge whether the remain adjacent matrix A' is equal to $\vec{0}$. If $A' = \vec{0}$, record the order of delete point(s) as D'_p that is the non-zero element of D_p and the dimension number of A' is noted as m' which is the number of remain points. Otherwise, set $p = p + 1$, delete the points corresponded with the elements of D_p and follow the deleting procedure of *step 2-b*.

c. Delete the rows of $L_{m \times r}^{(1)}$, $R_{m \times r}^{(1)}$, $U_{m \times r}^{(old)}$ and $X_{m \times r}^{(old)}$ according to the element of D'_p . The remain matrices are noted as $L_{(m-m') \times r}^{(2)}$, $R_{(m-m') \times r}^{(2)}$, $U_{(m-m') \times r}^{(2)}$ and $X_{(m-m') \times r}^{(2)}$. If $m' \geq n$, select n sampling points randomly note as $L_{n \times r}^{(new)}$, $R_{n \times r}^{(new)}$, $U_{n \times r}^{(new)}$ and $X_{n \times r}^{(new)}$. $X_{n \times r}^{(new)}$ is the sampling matrix of LHS. If $m' < n$, go to *step 3*.

Step 3. Generate the new sample matrix $X_{(n-m') \times r}^{(3)}$.

a. Generate a new order matrix $L'_{n \times r}$ of size $n \times r$ and every column of $L'_{n \times r}$ is a random rank of the integer number from 1 to n .

b. For every column, delete the element that is the same as the corresponding column in $L_{m' \times r}^{(2)}$ and get the matrix $L_{(n-m') \times r}^{(3)}$.

c. Generate a random matrix $R_{(n-m') \times r}^{(3)}$ where r_{ij} is a uniform random number ranging from 0 to 1.

d. Compute the cumulative probability matrix $U_{(n-m') \times r}^{(3)}$ as shown in Eq. (7).

$$u_{i,j}^{(3)} = \frac{l_{i,j}^{(3)} - 1 + r_{i,j}^{(3)}}{n} \quad (7)$$

e. Compute the sample matrix $X_{(n-m') \times r}^{(3)}$ as shown in Eq. (2).

Step 4. Get the order matrix $L_{n \times r}^{(new)}$, the random matrix $R_{n \times r}^{(new)}$, the cumulative probability matrix $U_{n \times r}^{(new)}$ and the sample matrix $X_{n \times r}^{(new)}$ of LHS.

$$L_{n \times r}^{(new)} = \begin{bmatrix} L_{m' \times r}^{(2)} \\ L_{(n-m') \times r}^{(3)} \end{bmatrix}, R_{n \times r}^{(new)} = \begin{bmatrix} R_{m' \times r}^{(2)} \\ R_{(n-m') \times r}^{(3)} \end{bmatrix},$$

$$U_{n \times r}^{(new)} = \begin{bmatrix} U_{m' \times r}^{(2)} \\ U_{(n-m') \times r}^{(3)} \end{bmatrix}, X_{n \times r}^{(new)} = \begin{bmatrix} X_{m' \times r}^{(2)} \\ X_{(n-m') \times r}^{(3)} \end{bmatrix}$$

It is easy to be proved that the matrix $X_{n \times r}^{(new)}$ is a sample matrix of LHS.

3. NUMERICAL EXPERIMENTS AND RESULTS

The generation algorithm of LHS based on original sampling points is illustrated by a simple example in Section 3.1. The effectiveness is demonstrated by an application in Section 3.2.

3.1 Illustration of algorithm

Assume that we have a Monte Carlo sampling of size 7 and there are two variables noted as x_1 and x_2 . The sample matrix is

$$X_{7 \times 2}^{(old)} = \begin{bmatrix} 18.5813 & 33.9815 \\ 14.1254 & 37.8181 \\ 15.4463 & 39.1858 \\ 13.3452 & 30.9443 \\ 16.4734 & 22.7725 \\ 13.5714 & 22.9859 \\ 15.0299 & 25.1502 \end{bmatrix}$$

x_1 has a triangular distribution on [10, 20] with mode at 15 and x_2 has a uniform distribution on [20, 40]. The distribution function $F(\cdot)$ is

$$F_1(x_1) = \begin{cases} \frac{(x_1 - 10)^2}{50}, & 10 \leq x_1 \leq 15 \\ 1 - \frac{(20 - x_1)^2}{50}, & 15 < x_1 \leq 20 \end{cases}$$

$$F_2(x_2) = 0.05 \times (x_2 - 20)$$

The inverse of the distribution function $F(\cdot)$ is

$$x_1 = \begin{cases} 10 + \sqrt{50 \times F_1(x_1)}, & 0 \leq F_1(x_1) \leq 0.5 \\ 20 - \sqrt{50 \times (1 - F_1(x_1))}, & 0.5 < F_1(x_1) \leq 1 \end{cases}$$

$$x_2 = 20 \times F_2(x_2) + 20$$

The objective is to get a LHS of size 5. The steps are as follows.

Step 1. Compute the distribution of original sample points in LHS.

a. Compute cumulative probability matrix $U_{m \times r}^{(1)}$ and the order matrix $L_{m \times r}^{(1)}$ as shown in Eq. (4) and Eq. (5). Fig. 1 shows the location of original sampling.

$$U_{7 \times 2}^{(1)} = \begin{bmatrix} 0.9597 & 0.6991 \\ 0.3404 & 0.8909 \\ 0.5853 & 0.9593 \\ 0.2238 & 0.5472 \\ 0.7513 & 0.1386 \\ 0.2551 & 0.1493 \\ 0.5060 & 0.2575 \end{bmatrix}, L_{7 \times 2}^{(1)} = \begin{bmatrix} 5 & 4 \\ 2 & 5 \\ 3 & 5 \\ 2 & 3 \\ 4 & 1 \\ 2 & 1 \\ 3 & 2 \end{bmatrix}$$

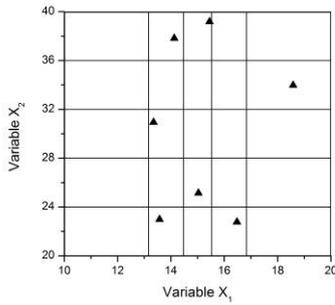


Fig. 1 Distribution of original sampling.

b. Get the random matrix $R_{m \times r}^{(1)}$ as shown in Eq. (6).

$$R_{7 \times 2}^{(1)} = \begin{bmatrix} 0.7987 & 0.4954 \\ 0.7019 & 0.4545 \\ 0.9263 & 0.7965 \\ 0.1191 & 0.7361 \\ 0.7563 & 0.6931 \\ 0.2755 & 0.7465 \\ 0.5298 & 0.2875 \end{bmatrix}$$

Step 2. Get the maximum sample matrix $X_{m' \times r}^{(2)}$ of the original Sampling.

a. Compute the adjacent matrix $A_{7 \times 7}$.

$$A_{7 \times 7} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

b. Get D .

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Set $p=1$ and delete the first sampling point. Then,

$$A' = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \neq \bar{0}$$

Set $p = p + 1 = 2$, delete the second sampling point. Then,

$$A' = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \neq \bar{0}$$

Repeat the step 2-b. When $p = 46$, $A' = \bar{0}$. At this time,

$$D_p = [0 \ 1 \ 1 \ 0 \ 0 \ 1 \ 0], D'_p = [2 \ 3 \ 6] \text{ and } m' = 4.$$

c. Get the remain matrices are noted as $L_{4 \times 2}^{(2)}$, $R_{4 \times 2}^{(2)}$, $U_{4 \times 2}^{(2)}$ and $X_{4 \times 2}^{(2)}$.

$$L_{4 \times 2}^{(2)} = \begin{bmatrix} 5 & 4 \\ 2 & 3 \\ 4 & 1 \\ 3 & 2 \end{bmatrix}, R_{4 \times 2}^{(2)} = \begin{bmatrix} 0.7987 & 0.4954 \\ 0.1191 & 0.7361 \\ 0.7563 & 0.6931 \\ 0.5298 & 0.2875 \end{bmatrix},$$

$$U_{4 \times 2}^{(2)} = \begin{bmatrix} 0.9597 & 0.6991 \\ 0.2238 & 0.5472 \\ 0.7513 & 0.1386 \\ 0.5060 & 0.2575 \end{bmatrix},$$

$$X_{4 \times 2}^{(2)} = \begin{bmatrix} 18.5813 & 33.9815 \\ 13.3452 & 30.9443 \\ 16.4734 & 22.7725 \\ 15.0299 & 25.1502 \end{bmatrix}$$

$m' = 4 < n = 5$, so go to step 3.

Step 3. Generate the new sample matrix $X_{1 \times 2}^{(3)}$.

a. Generate a new order matrix $L'_{5 \times 2}$.

$$L'_{5 \times 2} = \begin{bmatrix} 4 & 1 \\ 5 & 2 \\ 3 & 3 \\ 2 & 5 \\ 1 & 4 \end{bmatrix}$$

b. Get the matrix $L_{1 \times 2}^{(3)}$.

$$L_{1 \times 2}^{(3)} = [1 \ 5]$$

c. Generate a random matrix $R_{1 \times 2}^{(3)}$.

$$R_{1 \times 2}^{(3)} = [0.3507 \ 0.9390]$$

d. Compute the cumulative probability matrix $U_{1 \times 2}^{(3)}$ as shown in Eq. (7).

$$U_{1 \times 2}^{(3)} = [0.0701 \ 0.9878]$$

e. Compute the sample matrix $X_{1 \times 2}^{(3)}$ as shown in Eq. (2).

$$X_{1 \times 2}^{(3)} = [11.8722 \ 39.7560]$$

Step 4. Get the order matrix $L_{5 \times 2}^{(new)}$, the random matrix $R_{5 \times 2}^{(new)}$, the cumulative probability matrix $U_{5 \times 2}^{(new)}$ and the sample matrix $X_{5 \times 2}^{(new)}$ of LHS. The LHS is shown as Fig. 2.

$$L_{5 \times 2}^{(new)} = \begin{bmatrix} L_{4 \times 2}^{(2)} \\ L_{1 \times 2}^{(3)} \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ 2 & 3 \\ 4 & 1 \\ 3 & 2 \\ 1 & 5 \end{bmatrix},$$

$$R_{5 \times 2}^{(new)} = \begin{bmatrix} R_{4 \times 2}^{(2)} \\ R_{1 \times 2}^{(3)} \end{bmatrix} = \begin{bmatrix} 0.7987 & 0.4954 \\ 0.1191 & 0.7361 \\ 0.7563 & 0.6931 \\ 0.5298 & 0.2875 \\ 0.3507 & 0.9390 \end{bmatrix},$$

$$U_{5 \times 2}^{(new)} = \begin{bmatrix} U_{4 \times 2}^{(2)} \\ U_{1 \times 2}^{(3)} \end{bmatrix} = \begin{bmatrix} 0.9597 & 0.6991 \\ 0.2238 & 0.5472 \\ 0.7513 & 0.1386 \\ 0.5060 & 0.2575 \\ 0.0701 & 0.9878 \end{bmatrix},$$

$$X_{5 \times 2}^{(new)} = \begin{bmatrix} X_{4 \times 2}^{(2)} \\ X_{1 \times 2}^{(3)} \end{bmatrix} = \begin{bmatrix} 18.5813 & 33.9815 \\ 13.3452 & 30.9443 \\ 16.4734 & 22.7725 \\ 15.0299 & 25.1502 \\ 11.8722 & 39.7560 \end{bmatrix}$$

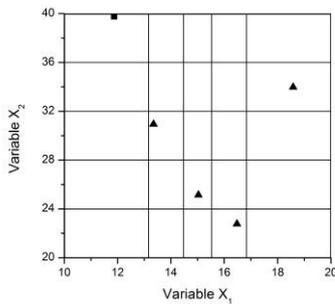


Fig. 2 Distribution of generated LHS

Key: ▲ Initial sample
■ New sample

3.2 Application for metamodeling

In this section, the generation algorithm is applied for metamodeling. There is a HLA-based electromagnetic gun simulation system, whose running

time cannot be ignored. There is a sample of size 40, which is used for sensitivity analysis. We would like to use a metamodel-based simulation optimization method to optimize the system parameters for the best performance. In order to construct a metamodel, we want a LHS of size 90 and another LHS of size 40. LHS of 90 is the training set for constructing metamodel and evaluating the fitting effect. LHS of 40 is the predicting set for evaluating the predicting effect. For saving running time of simulation system, we choose the algorithm proposed in this paper to generate the training set.

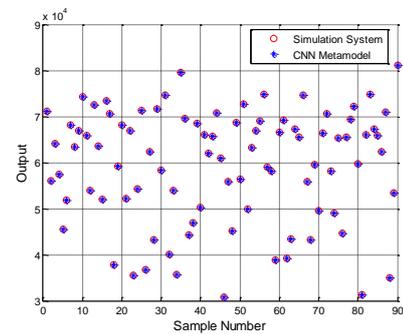
The algorithm is realized by MATLAB 2008a. There are 7 original sampling points to delete and 33 original sampling points to remain. Another 57 new sampling points are got by running the electromagnetic gun simulation system.

The evaluating index of metamodel is Multiple Correlation Coefficient (R^2) as shown in Eq. (8). R^2 is belonged in [0 1] and the more R^2 is, the better metamodel is.

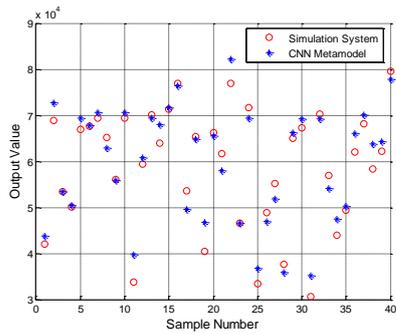
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

Artificial Neural Network (ANN) is the metamodel and the result is shown in Fig. 3. $R^2=1$ for training set and $R^2=0.9487$ for predicting set.

A contrast experiment is carried out, which use a classical LHS instead. In this case, $R^2=1$ for training set and $R^2=0.9443$ for predicting set. Therefore, the algorithm proposed in this paper achieves a LHS that have the similar result of classical LHS and remains 33 original sampling points. It means LHS by proposed algorithm have a good space filling effect the same as classical LHS and shorten 36.7% time for getting training set, which is because that the algorithm gets a strict LHS and utilizes the most original sampling points.



a. training set



b. predicting set

Fig.3 output values of simulation system and metamodel

4. CONCLUSION

As many simulation systems are costly to execute, the original sampling points might be a precious resource of a new analysis. A generation algorithm of LHS based on original sampling points gives a method to construct LHS based on original sampling points. The original sampling can be a Monte Carlo, a LHS, an orthogonal array, a factorial design or any other sampling. Actually, the algorithm proposed in this paper is an expansion of LHS extension algorithm, which does not require the original sampling to be a LHS. Therefore, the generation algorithm can be applied in the same fields of the LHS extension algorithm such as adaptive metamodel building and sequential experiments design and analysis.

For the future research, there are at least two suggestions. One is the improvement of the proposed algorithm. *Step 2-b* is an ergodic method and need lots of time to select remaining sampling points. When there are many original sampling points and variables, the time cost cannot be ignored. So a fast generation algorithm of LHS based on original sampling points is worth to research. The other suggestion is the research on the generation algorithm of other sampling method. Many analysis methods need special sampling method and generation algorithm of LHS is just one of them.

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

Borgonovo E., Castaigns W., Tarantola S., Model emulation and moment-independent sensitivity analysis: An application to environmental modeling. *Environmental Modelling and Software*, 34(6), pp. 105-115(2005).

Storlie C. B., Swiler L. P., Helton J. C. and et al., Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering and System Safety*, 94(11), pp. 1735-1763(2009).

Huang S. C., Lo Y. L., Lu C.N., Non-Technical Loss Detection Using State Estimation and Analysis of Variance. *IEEE Transactions on Power Systems*, 28(3), pp. 2959-2966(2013).

Zouaoui F., Wilson J. R., Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions*, 35(9) pp. 781-792(2003).

Helton J. C., Davis F. J., Johnson J. D., A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. *Reliability Engineering and System Safety*, 89(3), pp. 305-330(2005).

Braton R. R., Nelson B. L., Wei X., Quantifying Input Uncertainty via Simulation Confidence Intervals. *INFORMS Journal on Computing*, 14(6), pp. 74-87 (2013).

Mckay M. D., Beckman R. J., Conover W. J., A comparison of three algorithms for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), pp. 239-245(1979).

Tong C., Refinement strategies for stratified sampling algorithms. *Reliability Engineering and System Safet*, 91(11), pp. 1257-1265(2006).

Sallaberry C. J., Helton J. C., Hora S. C., Extension of Latin hypercube samples with correlated variables. *Reliability Engineering and System Safety*, 93(7), pp. 1047-1059(2008).

Wu Y. Z., Chen L. P., Simulation optimization algorithm of multi-domain physical system. *Science Press*, pp. 186-189(2011).

Wang G. G., Adaptive Response Surface Algorithm Using Inherited Latin Hypercube Design Points. *Journal of Mechanical Design*, 125(2), pp. 210-220(2003).

Wei X., Research of Global Optimization Algorithm Based on Metamodel. *Doctoral Dissertation of Huazhong University of Science and Technology*, pp. 70-75(2012).

AUTHORS BIOGRAPHY

ZHIZHAO LIU is a Ph.D. student at Harbin Institute of Technology (HIT), and received the B.S., M.E. from HIT in 2010 and 2012. His research interests include simulation optimization and operational effectiveness evaluation. His email is liuzhizhao2007@hotmail.com.

WEI LI is the corresponding author. He is an associate professor at HIT, and received the B.S., M.E. and Ph.D. from HIT in 2003, 2006 and 2009 respectively. His research interests include simulation evaluation, simulation data analysis, and distributed simulation. His email is fleehit@163.com.

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.

STOCHASTIC PROGRAMMING FOR TRAIN DISTRIBUTION IN A METRO TRANSPORTATION SYSTEM

Jorge A. Garcia^(a), K Ponnambalam^(b)

^{(a),(b)} University of Waterloo, Ontario, Canada

^(a) ja4garci@uwaterloo.ca, ^(b) ponnu@uwaterloo.ca

ABSTRACT

We present a mixed integer linear programming formulation with an application for the optimal train distribution problem. The objective was to assign trains to metro routes so that the sum of deficit capacity costs (i.e., passengers that cannot be carried) and the overcapacity costs (i.e., the number of empty seats) in a metro transportation system were minimized. We selected Mexico City's Metro Transportation System (MTS) as a study case. We narrowed our model to the operation and demand of peak hours. First, we solve a deterministic model using mean values for the parameters. Then, we built a two-stage stochastic model to include uncertainty into the parameters and we calculated estimates for mean and variance using maximum likelihood estimation. We discuss the results and compare the solutions for the four cases that we created. Finally, we propose an extension of the model that includes a time index.

Keywords: stochastic programming, train distribution, transportation system, integer programming

1. LITERATURE REVIEW

Train distribution is related to frequency optimization since there is a direct connection between number of trains and frequencies of rides. The use of linear models for frequency optimization in transportation has been widely studied; Yoo (2010) proposed an approach for a multi-modal transportation in which the problem is divided into two sub-problems, in the first they solve the route-selection problem from the passenger perspective, and for the second they find the optimal frequency from the operator perspective. Zhou et al. (2005) proposed a bi-level formulation in which the expected profit is maximized. Fernandez et al. (2008) proposed a methodology to deal with public transportation design and showed an application for the city of Santiago, Chile. Something that has been disregarded is the trade-off associated with non-uniform demand along the route, i.e., to have a higher demand in some particular stations in one direction than the rest of the line. We believe that measuring the overcapacity under such conditions would be useful for policy decision-making.

2. INTRODUCTION

Collective transport systems are a better mobility solution for medium to large distances than private cars within highly populated metropolis. Mexico City has more than 10 million people that are moving every day within the city. Its Metro Transport System (MTS) moves more than five million passengers every day. A key factor in the operation is the correct distribution of trains to meet the demand at every station; however, there is a fixed capacity in the whole system. So, the problem is how to assign trains to subsystems (i.e., routes) so that the operation is optimized.

We would like to present the following example to show the approach we used for the train distribution problem. Presented here is a five-station metro route with uneven flows between stations/directions in the studied time period.

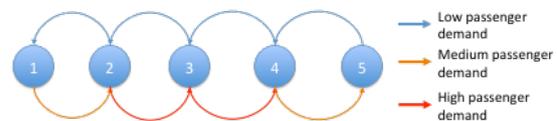


Figure 1. Example of a five-station metro route

Demand from station one to five is higher than that in the opposite direction (i.e., from five to one), also, stations two and three have a higher demand. This situation is common in transportation systems: to have a high demand of service in one direction in a certain period while the opposite direction has a low demand. One question that arises from the example: *is it worthy to have a high and fixed capacity in the whole route just to meet the demand of two stations?* Although it is true that the passenger flow changes over the day, the traffic demand is always concentrated in a subset of stations. As described above, we find compelling the formulation of a model that not only penalizes the unsatisfied demand at some station, but also the capacity not used on the rest. We believe this could be an interesting approach to evaluate the tradeoffs of having a fixed capacity on both sides of the route. All the routes we considered (routes 1 to 9) have trains with the same characteristics (e.g. pneumatic wheels, size, etc.).

3. METHODOLOGY

We followed the next methodology:

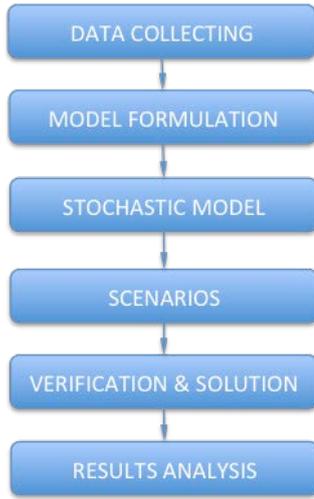


Figure 2. Methodology to address the optimal distribution problem

3.1 Data Collecting

Inflow passenger data per station and the number of trains were obtained from the Mexico's MTS site, which has data from 2012 to date, and it is available for public consultation.

3.2 Model Formulation

The process of model formulation started from the creation of one simple model and gradually we added more variables. First we present the deterministic model: objective function, decision variables, etc.

3.3 Stochastic model

We show the parameters that have uncertainty and the procedure we followed to handle their randomness. Passenger inflow per station and service capacity were the main two parameters we estimated. Below we comment about the process and assumptions made.

3.4 Scenarios

For the stochastic model, we tried different number of scenarios. Using Monte Carlo sampling method we produced realization of the demand series. Since we are interested in the trade-offs between satisfying demand and overcapacity service we created four cases: the first two consist in finding the optimal distribution without penalizing the exceeding capacity; the other two, do penalize the exceeding capacity. The scenarios are explained in detail below.

3.5 Verification & Solution

In this step we first selected the software to write and solve the model. We chose Matlab due to its easy code writing and data retrieval utilities.

3.6 Result analysis

We discuss and compare the optimal solutions for both of the models, the deterministic and the stochastic.

4. MODEL FORMULATION

4.1 Decision variables

x_{ijk} is the number of passengers being transported from station i to station j on route k in the given time period. It includes the passengers from the previous station ($i-1$) plus the current passengers at station i .

v_k is the number of trains assigned to k route. It is the main decision variable since it determines the distribution of trains over the routes. It is an integer variable.

4.2 Auxiliary variables

d_{ijk} is the number of passengers not served at station i in direction to station j on route k in the given time period. It is used for quantifying the deficit of the service.

e_{ijk} is the number of empty seats at station i to station j on route k in the given time period. It is an auxiliary variable used to quantify the exceeding capacity.

4.3 Objective function

We used a linear cost function

$$\text{Min } z = \pi_1 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m d_{ijk} \right) + \pi_2 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m e_{ijk} \right) \quad (1)$$

The objective function minimizes the total sum of the deficit capacity cost (π_1) and the exceeding capacity cost (π_2) occurred at every station and route of the metro system. By penalizing both the deficit (d_{ijk}) and the overcapacity (e_{ijk}) we are forcing the model to find an optimal tradeoff between exceeding the demand and exceeding the capacity. We set $\pi_1=1$ and $\pi_2=0$ for A and B scenarios, and $\pi_1=6$ and $\pi_2=1$ for III and IV scenarios, respectively.

4.4 Restrictions

The first set of restrictions is in the form

$$b_{ijk} + p_{ij} * x_{i-1ik} = x_{ijk} + d_{ijk}, \quad \forall i=1,2,\dots,n-1, \quad \forall k \quad (2)$$

The equation states that the quantity of passengers arriving at station i (represented by b_{ijk}) plus a percentage p_{ij} of passengers travelling from the previous station ($i-1$) is equal to the number of passengers being transported from i to j plus the deficit variable d_{ijk} .

$$x_{ijk} \leq C_k * v_k, \quad \forall i, j, k \quad (3)$$

Equation (3) ensures that the rate of passengers travelling from i to j at k route is less or equal to the capacity of the train (C_k) times the number of trains (v_k). We show how C_k is obtained below.

$$C_k * v_k - x_{ijk} = e_{ijk}, \quad \forall i, j, k \quad (4)$$

Equation (4) assigns the exceeding capacity to the overcapacity variable (e_{ijk}). From the restriction (3) we know that x_{ijk} is less or equal to $C_k * v_k$, so e_{ijk} will be zero or positive.

$$\sum_{k=1}^m v_k = T \quad (5)$$

Restriction (5) states that the sum of all assigned trains must be equal the total number of trains available (T).

4.5 Deterministic model

The complete deterministic model is

$$\text{Min } z = \pi_1 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m d_{ijk} \right) + \pi_2 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m e_{ijk} \right)$$

$$b_{ijk} + p_{ij} * x_{i-1ik} = x_{ijk} + d_{ijk} \quad \forall i = 1, 2, \dots, n-1, \quad \forall k$$

$$x_{ijk} \leq C_k * v_k \quad \forall i, j, k$$

$$C_k * v_k - x_{ijk} = e_{ijk} \quad \forall i, j, k$$

$$\sum_{k=1}^m v_k = T$$

$$x_{ijk} > 0 \quad \forall i, j, k$$

$$v_k \in Z^+$$

The deterministic model has a total of 819 variables and 811 constraints.

5. STOCHASTIC MODEL

The deterministic model does not consider that demand and capacity are variable. Demand in one day may be different from that on another day. The capacity of the route depends on the number of train loops, which are not always the same due to technical reasons or delays. Stochastic models introduce randomness into parameters. Below we discuss how uncertainty was introduced into the model's parameters.

5.1 Demand (b_{ijk})

This parameter represents the number of passengers at station i moving to station j along k route. The first step was to calculate the mean and standard deviation.

From the historical data, we did an exploratory graphical analysis. We discovered a positive trend in the demand series (Figure 3). We removed the trend by differentiating one time (Figure 4).

The histogram of the differentiated series resembles a Normal distribution (Figure 5), so we decided to assume the demand behaves in that manner. We calculated an estimate for the mean and variance using Maximum Likelihood estimation (See Appendix B) for every route.

Then, we studied the correlation in the series using the sample autocorrelation function and the sample partial autocorrelation function. We noticed that the series cuts

off after lag 1 in the autocorrelation function, implying that there is a negative correlation only between time t and $t+1$ (Figure 6). The partial autocorrelation function shows that the process is dying off, which confirms a negative correlation (Figure 7).

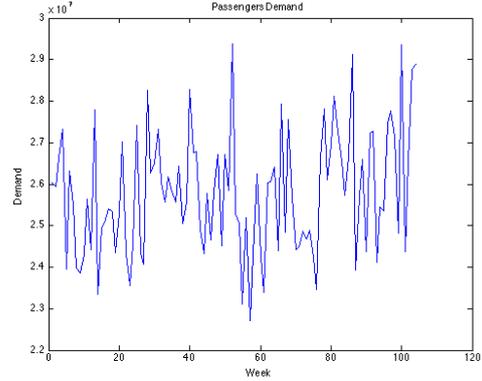


Figure 3. Weekly passengers demand from January 2012 to December 2013

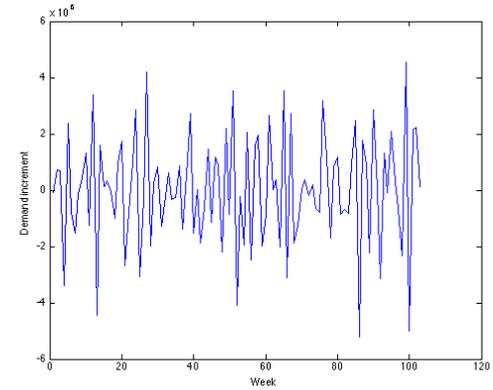


Figure 4. Differentiated demand series

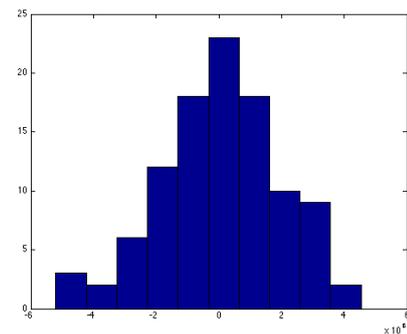


Figure 5. Histogram of the demand after differentiating

Now, we have data about the number of passengers entering at station i in route k , but we do not know how many of them are moving in which direction. We determined these two numbers by using the linear combination:

$$b_{ik} = \lambda b_{ijk} + (1 - \lambda) b_{rik}, \quad 0 \leq \lambda \leq 1$$

b_{ik} is the known number of passenger inflow in the station i , route k ; b_{ijk} is the portion of b_{ik} moving in the direction of j ; b_{hi} is that moving in the direction of h . We assume λ decreases linearly from 1 in station $i=1$, to 0 in station $i=n$ (terminal).

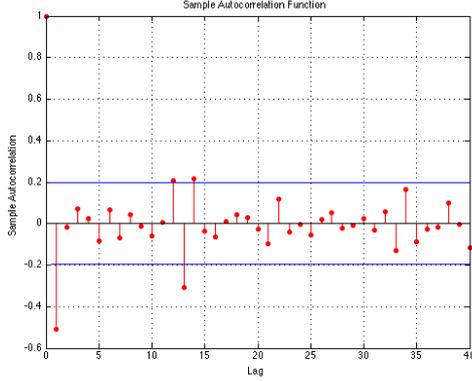


Figure 6. Sample auto correlation function

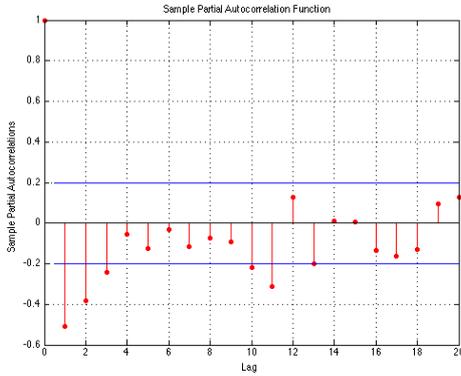


Figure 7. Sample partial autocorrelation function

5.2 Changes in the objective function

Following the formulation of the two-stage stochastic programming, we added two more variables δ_{ijk}^s and ε_{ijk}^s to absorb the deficit or surplus of demand for every scenario s . We penalized those values with the same weight we gave to the deterministic variables for deficit (π_1) and surplus (π_2) times $1/S$; S is the total number of scenarios:

$$\frac{\pi_1}{S} \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{s=1}^S \delta_{ijk}^s \right) + \frac{\pi_2}{S} \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{s=1}^S \varepsilon_{ijk}^s \right)$$

Since it is a minimization model, these costs have a positive sign on the objective function.

5.3 Changes in restrictions

The only set of restriction that changed is the first one. We added the two-stage variables of deficit (δ_{ijk}^s) and surplus (ε_{ijk}^s):

$$b_{ijk} + p_{ij} * x_{i-1ik} + \varepsilon_{ijk}^s - \delta_{ijk}^s = x_{ijk} + d_{ijk}$$

$$\forall i = 1, 2, \dots, n-1, \quad \forall k, s$$

5.4 Service Capacity (C_k)

This parameter represents the capacity of service per unit of time and per train. In this work we considered the unit of time as one hour, so for every station we calculated C_k by a function of the length and the number of stations. A comparison was made between the results obtained using this formula and those provided by the City Metro – Map & Route Planner App (See Appendix A).

The time of one circuit loop for every train is

$$t = 2(\alpha * s + \beta * l)$$

where s is the number of stations and l is the total length of the route in meters. α and β are random variables for waiting time and speed, respectively. We assume they have a uniform distribution over the interval $[0.8\mu, 1.2\mu]$ where μ is the mean value. We also assume that the mean value for α is 1.5 minutes, and that for β is 35.5 km/hour. So, we calculate the expected time for every route and then, use that value to obtain the expected capacity service:

$$E[t_k] = 3s_k + \frac{6}{1775} l_k \quad (6)$$

So, equation (6) is the formula to calculate the expected time of one circuit for every route k .

Now we proceeded to calculate C_k and do the dimensional analysis:

$$\frac{60[\text{min}/\text{hour}]}{t_k[\text{min}]} * tc \left[\frac{\text{passengers}}{\text{train}} \right] = \left[\frac{\text{passenger}}{\text{hour} - \text{train}} \right] = C_k$$

Finally, we have

$$C_k = \frac{35,500 tc_k}{1775 s_k + 2 l_k} \quad (7)$$

The capacity (tc_k) is 1,020 passengers per train in routes four and six, and 1,530 for the rest of the routes. The results are shown in Table 1.

We used the $C_k^{0.98}$ value for restrictions 2 and 3, to ensure the solution will satisfy the capacity requirement 98 out 100 cases.

Table 1. Summary data of the C_k (capacity)

Route	Stations s	l (m)	t_k	C_k (mean)	$C_k^{0.98}$ (98% conf.)
1	20	16,804	116.8	786	646
2	24	20,863	142.5	644	529
3	21	21,428	135.4	678	556
4	10	9,513	62.1	984	809
5	13	14,435	87.8	1,045	857
6	11	11,434	71.6	854	701
7	14	17,011	99.5	923	756
8	19	17,679	116.7	786	646
9	12	13,033	80	1,147	940

5.3 Two-Stage Stochastic Model

The complete two-stage stochastic model is

$$\begin{aligned} \text{Min } z = & \pi_1 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m d_{ijk} \right) + \pi_2 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m e_{ijk} \right) \\ & + \frac{\pi_1}{S} \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{s=1}^S \delta_{ijk}^s \right) + \frac{\pi_2}{S} \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{s=1}^S \varepsilon_{ijk}^s \right) \\ & b_{ijk} + p_{ij} * x_{i-lik} + \varepsilon_{ijk}^s - \delta_{ijk}^s = x_{ijk} + d_{ijk} \quad \forall i=1, \dots, n-1, \quad \forall k, s \\ & x_{ijk} \leq C_k * v_k \quad \forall i, j, k \\ & C_k * v_k - x_{ijk} = e_{ijk} \quad \forall i, j, k \\ & \sum_{k=1} v_k = T \\ & x_{ijk} \geq 0 \quad \forall i, j, k \\ & \delta_{ijk}^s, \varepsilon_{ijk}^s \geq 0 \quad \forall i, j, k \\ & v_k \in \mathbb{Z}^+ \end{aligned}$$

The number of variables and restrictions depends on the number of scenarios; for the largest case we tried (35 scenarios), it has a total of 19,719 variables and 9,991 restrictions.

6. SCENARIOS

We used Monte Carlo sampling method to produce s scenarios (realizations), each of them with probability of $1/S$. We assumed the demand has a normal distribution; the mean and the variance were obtained from the historical data (see Appendix B). We solved the model for 2, 5, 10, 20, 25, 30, and 35 scenarios. From the tenth onwards, the optimal solution did not change.

In order to compare the solutions of the deterministic and stochastic models, we produced four cases:

1. Deterministic model; assigning $\pi_1=1$, $\pi_2=0$, and average values in demand and service capacity.

2. Stochastic model; assigning $\pi_1=1$, $\pi_2=0$, and using 35 scenarios for the demand and 98% of confidence in service capacity.
3. Deterministic model; assigning $\pi_1=6$, $\pi_2=1$, and average values in demand and service capacity.
4. Stochastic model; assigning $\pi_1=6$, $\pi_2=1$, and using 35 scenarios for the demand and 98% of confidence in service capacity.

In III and IV cases we are establishing that one occupied seat is six times more important than one available space (i.e., $\pi_1=6$ and $\pi_2=1$).

7. RESULTS

The results for the optimal train distribution are shown in Table 2. To make comparable the solutions, we associated a cost for the unmet demand (Table 3). The total cost is equal to the sum of the costs: $c*(capacity-demand)$, for every route, we set $c=1$.

Table 2. Distribution of trains for every route

Route	Current	Optimal solution (trains)			
		I	II	III	IV
1	30	29	29	35	24
2	24	37	42	38	40
3	29	33	33	30	28
4	8	4	3	4	9
5	15	12	7	9	10
6	9	10	5	9	13
7	19	12	12	15	15
8	18	18	18	15	15
9	17	14	20	14	15

Table 3. Passengers capacity per hour

Route	Optimal solution (passengers)			
	I	II	III	IV
1	22,794	22,794	27,510	18,864
2	23,828	27,048	24,472	25,760
3	22,374	22,374	20,340	18,984
4	3,936	2,952	3,936	8,856
5	12,540	7,315	9,405	10,450
6	8,540	4,270	7,686	11,102
7	11,076	11,076	13,845	13,845
8	14,148	14,148	11,790	11,790
9	16,058	22,940	16,058	17,205
Cost	65,725	66,102	65,977	64,163

We can see that the minimal cost occurs on the case IV, i.e., solving the two-stage stochastic model for 35 scenarios.

The mayor changes are in the route 2 and seven. All the cases have at least 13 more trains in the former and at least four less in the latter. The current train distribution seems to be uneven in regards to meeting the demand.

8. CONCLUSIONS

In this paper we proposed a two-stage stochastic mixed integer program to assign train to routes in a

transportation system, minimizing a cost function, which penalizes both, deficit and exceeding capacity. We found the function to be a compelling idea to assess the trade-offs associated to a non-uniform distributed demand in the system.

9. EXTENSIONS OF THE MODEL

One problem with the proposed model is that it is based on hour-average operation parameters. If we add a time index to x , d and e variables, then we can change the size of the time interval and make it arbitrarily small to obtain a more accurate result; however, the number of variables increases notably. So, we have for the stochastic model:

$$\begin{aligned} \text{Min } z = & \pi_1 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{t=1}^T d_{ijkt} \right) + \pi_2 \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{t=1}^T e_{ijkt} \right) \\ & + \frac{\pi_1}{S} \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{s=1}^S \sum_{t=1}^T \delta_{ijkt}^s \right) + \frac{\pi_2}{S} \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^m \sum_{s=1}^S \sum_{t=1}^T \varepsilon_{ijkt}^s \right) \\ b_{ijkt} + p_{ijt} * x_{i-1ikt} + \varepsilon_{ijkt}^s - \delta_{ijkt}^s = & x_{ijkt} + d_{ijkt} \quad \forall i=1, \dots, n-1, \quad \forall k, s, t \\ x_{ijkt} \leq & C_k * v_k \quad \forall i, j, k, t \\ C_k * v_k - x_{ijkt} = & e_{ijkt} \quad \forall i, j, k, t \\ \sum_{k=1} v_k = & \text{Total} \\ x_{ijkt} \geq & 0 \quad \forall i, j, k, t \\ \delta_{ijkt}^s, \varepsilon_{ijkt}^s \geq & 0 \quad \forall i, j, k, t \\ v_k \in & Z^+ \end{aligned}$$

ACKNOWLEDGMENTS

We would like to thank the economical support from CONACYT and the Science, Technology and Innovation Office (SECITI) of Mexico City to make this work possible.

APPENDIX A. TRAIN LOOP TIMES

In Figure 8 we show a comparison between our formula t_k and the times calculated by a commercial application.

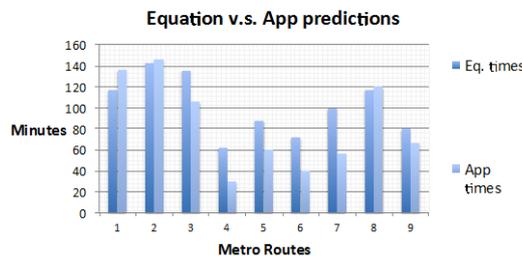


Figure 8. Comparison of one-loop times

The highest relative difference is on route 4 where the formula estimation is 107% higher than the App's. We believe that App predictions underestimate loop times in low-demand routes (4 to 7) possibly because it uses

passenger demand to calculate the times and delays disregarding distance.

APPENDIX B. MONTHLY AVERAGE DEMAND

The estimates for mean and standard deviation were calculated from the historical data using the maximum likelihood estimation. The monthly results are shown in the following table:

Table 4. Monthly mean and standard deviation

Route	Mean	SD	CV
1	21,594,861	1,044,094	4.8%
2	24,901,980	851,348	3.4%
3	20,105,945	588,261	2.9%
4	2,413,326	166,630	6.9%
5	6,611,076	376,188	5.7%
6	4,184,297	481,499	11.5%
7	7,610,587	596,027	7.8%
8	11,471,053	397,340	3.5%
9	9,657,134	586,501	6.1%

In the box plot of the differentiating total demand series, we can see that it is highly concentrated around the median. The interquartile range has length of 2.5 millions, which represents approximately 28% of the range. It has long extreme values though. The median is around zero. Two outliers are shown in red.

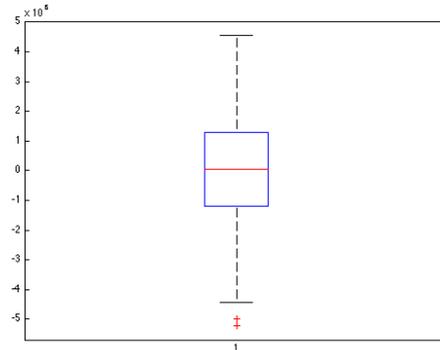


Figure 9. Boxplot of the observations after differentiating

REFERENCES

- Fernandez E. et al. 2008, Demand Responsive Urban Public Transport System Design: Methodology and Application. Transportation Research Part A Vol. 42 No. 10: p 951-972
- Yoo g., et al. 2010, Frequency Design in urban Transit Networks with Variable Demand: Model and Algorithm. Transportation Engineering Vol. 14 No 3: p. 403-411
- Zhou J., Lam, W., Heydecker B. 2005, The Generalized Nash Equilibrium Model for Oligopolistic Transit Market with Elastic Demand. Transportation Research B Vol. 39 No. 6: p. 519-544, 2005

AUTHORS BIOGRAFY

Jorge A. Garcia is a PhD student of Systems Design Engineering, University of Waterloo, working in the Design Optimization Under Uncertainty group.

K. Ponnambalam is a professor of Systems Design Engineering, University of Waterloo, and works in the areas of design optimization under uncertainty.

AN INTEGRATED PLATFORM FOR SIMULATION, VALIDATION, ANALYSIS AND OPTIMIZATION OF ELECTROMAGNETIC RAILGUN SYSTEM

Yuchen Zhou ^(a), Ping Ma ^(b), Ming Yang ^(c)

Control and Simulation Center, Harbin Institute of Technology
Harbin, P. R. China

^(a) zhouyuchen-01@163.com, ^(b) pingma@hit.edu.cn, ^(c) myang@hit.edu.cn

ABSTRACT

There exist coupled mechanical, thermal, and electromagnetic diffusive processes with high-speed moving conductors in electromagnetic railgun (EMRG). Simulation can validate related theories and reduce physical experiments. Besides, post-simulation analysis reveals the influence of various uncertain factors on the system performance and supports the improvement of the system performance. The Electromagnetic launch technology advanced rapidly and it put forward higher requirements for the simulation tools. Thus, an integrated electromagnetic railgun simulation platform, which has multi-functions including simulation, model validation, post-simulation analysis and optimization, is needed urgently. Furthermore, the simulation platform should be capable of strengthening the management of simulation resources. An integrated EMRG simulation platform is designed and implemented consisted of HLA-based EMRG model, model validation tool, experimental design & data analysis tool and optimization tool. The application result shows that the platform can support the simulation, validation, analysis and optimization of EMRG simulation system.

Keywords: integrated simulation platform, electromagnetic railgun, model validation, optimization

1. INTRODUCTION

The electromagnetic railgun is a launcher using the Lorenz force to accelerate the projectile and consisted of Pulsed Power Supply (PPS), rail, armature and projectile. During the launching, PPS form a trapezoidal current pulse at megampere levels. Armature and rail work in extreme conditions. Furthermore, there exists ultra-high speed friction between rail and launch package. Meanwhile, mechanical, thermal and electromagnetic diffusive processes coupled intensely. Compared to the traditional guns, EMRG is characterized by faster response, higher damage effectiveness, lower demands for logistical support (Fair, 2007; Li, Yan and Yuan, 2014).

As the theoretical research of electromagnetic launch technology continuous improved, the EMRG model was built and simulation experiments were carried out on the computer model. Based on the post-simulation analysis, the structural parameters of electromagnetic railgun

were optimized and engineering prototype was fabricated. The simulation model and theory can be examined and validated through physical tests. For the problems discovered in the experiments, theoretical research method, simulation method and prototype experiment approach can be combined to promote the development of electromagnetic launch technology.

Post-simulation analysis reveals the influence of various uncertain factors on the system performance and reduces the research and development (R&D) fund. Moreover, simulation can accelerate the EMRG research progress and improve the system performance and effectiveness significantly.

The Institute for Advanced Technology (IAT) designed and developed the EMAP3D program to solve the internal ballistics simulation of EMRG, which was a 3D Lagrangian finite element program for coupled mechanical, thermal, and electromagnetic diffusive processes with high-speed moving conductors (Hsieh, 1995). The EMAP3D code continues to be improved by introducing hybrid finite element-boundary element (FE-BE) algorithm (Hsieh, 2006) and developing novel split-domain iteration scheme to shorten the computation time (Hsieh and Thiagarajan, 2008; Tzeng and Hsieh, 2014).

Stephan Hundertmark in the French-German Research Institute of Saint-Louis (ISL) used the COMSOL program to set up a transient 3-D simulation of a small caliber railgun (Hundertmark and Roch, 2012). Furthermore, the muzzle velocity is compared to the experimental value and proved the validity of the simulation model under permissible error.

Antonino Musolino proposed a hybrid finite element method-method of moments (FEM-MOM) formulation capable of taking into account the velocity skin effect (Musolino, 2004). Nevsan Sengil applied the experimental design method to the EMRG simulation and some critical model parameters, like magnetic flux density and inductance gradient, were calculated through Monte Carlo method (Sengil, 2012).

As to the optimization of EMRG, the interaction between the launcher and pulsed power supply (PPS) system was studied and a numerical EMRG model powered by capacitor based PPS module was designed to find the optimal triggering configuration for a target muzzle velocity (Pitman, Ellis and Bernardes, 2014;

Yan, Chao, Ma, Yang, Hu, 2014). The University of Texas at Austin Center for Electromechanics (UT-CEM) make great effort on the optimization of integrated launch package (ILP) (Kitzmilller and Driga, 2003; Kitzmilller, Pratap, Aanstoos, Cook, Kuenast, Murphy and Perkins, 1995).

Different methods or tools have been used to study the electromagnetic railgun simulation, analysis and optimization from various perspectives. The purpose of EMRG simulation research is to improve the system performance and provide decision support for the development of the engineering prototype. Thus, an integrated simulation platform is indispensable for the further study of electromagnetic launch technology.

Integrated simulation platform is a set of simulation model and support tools, which can support the calculation of simulation model and solution of a series of application problems (Henriksen, 1983; Cui, Xiao, Yang, Xie and Li, 2004). Besides, it can effectively manage the simulation model, algorithm, data and other simulation resources.

An integrated electromagnetic railgun simulation platform, which has multi-functions including EMRG simulation, EMRG model validation, post-simulation analysis, and simulation-based performance optimization, will play a significant role in the EMRG research.

Based on the EMRG model established in the literature (Ma, Yang and Hu, 2012) and the previous works on the EMRG model validation and optimization (Hu, Ma, Yang and Wang, 2012), the requirement of the integrated EMRG simulation platform is discussed and the architecture is designed; then, the detail design of the integrated simulation platform is proposed. Finally, the implantation and application of the platform are introduced to demonstrate that the platform can satisfy the demands.

2. DESIGN OF INTEGRATED EMRG SIMULATION PLATFORM

2.1. Requirement Analysis

Software requirement analysis is the first step in software design and it is critical to the success of simulation platform. As to the functional aspect of EMRG simulation platform, the calculation of EMRG model should be the primary function of simulation platform. The EMRG system model includes the internal ballistic model and exterior ballistic model.

Furthermore, to solve the problems encountered in the EMRG simulation, which include model validity, post-simulation analysis, and performance optimization, various supporting tools should be designed. These functions will be the primary functions of the integrated EMRG simulation platform.

- EMRG Simulation: simulate the accelerate process of projectile within the tube and the flight in the atmosphere.

- Model validation: provide suitable validation methods to analyse the similarity of the EMRG simulation data and test data.
- Experimental design: generate experimental design plan and save it to certain format which can be directly used by the EMRG simulation system.
- Post-simulation analysis: analyse the EMRG simulation data to discover the influence of uncertain factors on EMRG system.
- Performance optimization: optimize the discharge time series of capacitors and structural parameters to achieve better fire performance.

As to the structural aspect, the integrated EMRG simulation platform should have the ability to manage the simulation resources effectively for which we reference the partitioning design of OSI model. At the same time, to improve the reusability of the functional modules, the connections between different modules should be reduced. All the algorithms will be packaged to dynamic link library and data structure of the platform will be designed. The encapsulation of algorithms and reasonable design of data structure is the key to strengthen the independence of each module.

2.2. Architecture

For the reason that the integrated simulation platform has multiple functions and there is requirement for function expanding and software updating, it is necessary to design an extensible architecture. The architecture design should guarantee the reusability of simulation resources. The EMRG architecture is divided into Application Layer, Tools Layer, Resources Layer, and Infrastructure Layer.

- Application layer: the integrated simulation platform can satisfy the requirement of EMRG simulation, model validation, post-simulation analysis and simulation-based optimization.
- Tools layer: the integrated simulation platform consists of the HLA-based EMRG simulation system, model validation tool, experimental design & data analysis tool and optimization tool.
- Resources layer: the simulation resources can be managed effectively and reused easily in other software through resources layer. The projects, simulation data, algorithms and analysis reports will be saved to the designated database respectively.
- Infrastructure layer: the infrastructure layer includes the software and hardware which is essential to the establishing of the integrated simulation platform.

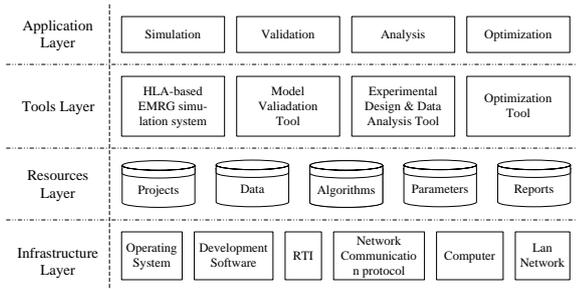


Figure1. Integrated simulation platform architecture

The HLA-based EMRG simulation system has specially interface which allows to read the experimental design project and perform simulation experiment. The model validation tool and experimental design & data analysis tool can access the database to obtain the simulation data generated by the EMRG simulation system. The optimization tool calls the EMRG simulation model to accomplish the performance optimization.

2.3. Design of EMRG Simulation Model

The high-level architecture (HLA) is a standardized architecture for distributed simulation systems. The interactions in HLA-based simulation systems are managed by run-time infrastructure (RTI). This mechanism separates the simulation model development from run-time management and communication, which tremendously enhance the flexibility and extendibility of the simulation system.

The HLA-based EMRG simulation system is composed of electromagnetic railgun federate, target federate, detection federate, threat assessment federate console federate and data collection federate. The electromagnetic railgun federate includes the internal ballistic model and 6 degree-of-freedom (DOF) exterior ballistic model.

The console federate is designed to configure the model parameters and strengthen the simulation process management. The data collection federate can collect the data generated in the simulation and save it to the configured paths.

2.4. Design of Model Validation Tool

Simulation model validation is defined as determining that the model's output behavior has sufficient accuracy for the model's intended purpose over the domain of model's intended applicability (Sargent, 2009). The credibility of the EMRG model decides whether the simulation experiment result and the data analysis result can be accepted.

Data consistency analysis is a common approach to validate simulation model. The functions of model validation tool include data input, data consistency analysis and validation report generation.

According to the characteristic of EMRG simulation data, the validation tool provides Theil's inequality coefficient (TIC) and Grey Relational Analysis (GRA) for the data consistency analysis. TIC, firstly used to validate the missile systems (Naim, Willard, 1978), is

an classical method providing insight into the validity of the EMRG simulation system.

The EMRG model validation includes the consistency analysis of discharge current, projectile velocity and projectile displacement. The discharge current is the most important evaluation index. The discharge current curve can be spitted into ascent stage, stable stage and decline stage and proper method should be chose to analyse each stage of the current curve with the aid of the validation tool.

2.5. Design of Experimental Design and Data Analysis Tool

There are numbers of uncertain factors in the EMRG model. Experimental design can be used to reduce the scale of simulation experiment in the studying of the influence of those factors on the simulation result. Multi-factors statistical methods provide a quantitative analysis method for the post-simulation analysis.

The functions of experimental design & data analysis tool include factors design, methods selection, experimental design table generation, and multi-factor analysis. There are three experimental design methods (Monte Carlo experimental design, uniform experimental design and orthogonal experimental design) and two multi-factor analysis methods (variance analysis and regression analysis). The experimental design table can be saved to the Access database or saved in XML format.

2.6. Design of Optimization Tool

The electromagnetic railgun launches projectile by the Lorenz force generated by the discharge current. Due to the drag is far less than the thrust, according to $F=L'I^2/2$, the stability of acceleration process mainly depends on the stability of discharge current. The stable current can reduce the damage to the rail, caused by high speed sliding contact between the armature and the rail.

Hu, Ma, Yang, Lu and Wang (2014) proposed a multi-stage optimization strategy based on improved particle swarm optimization (PSO) to optimize the discharge current form wave. The optimization tool realizes this algorithm using C++. Optimization tool function includes optimization index selection, constraint conditions setup and optimization.

3. IMPLEMENTATION AND APPLICATION

The EMRG simulation platform was developed with VC++ 6.0 and it works on a single computer installed windows operating system or on a distributed system composed of several computers. We illustrate the application of EMRG simulation platform.

Figure 2 shows the data visualization function of HLA-based EMRG simulation system and the dynamic curves displayed on the right are the capacitor discharge current curve and the total current curve. Then the GRA method is used to validate EMRG model and the data similarity analysis results proved that the model has a high credibility (Figure 3).

As is shown in figure 4, variance analysis approach is used to study the uncertain factors' effects on firing accuracy. The uncertain factors include projectile mass, muzzle velocity and so on. Figure 5 shows the optimization of the discharge current curve.

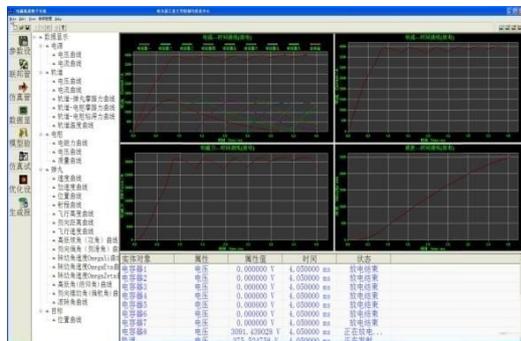


Figure2 HLA-based EMRG simulation system

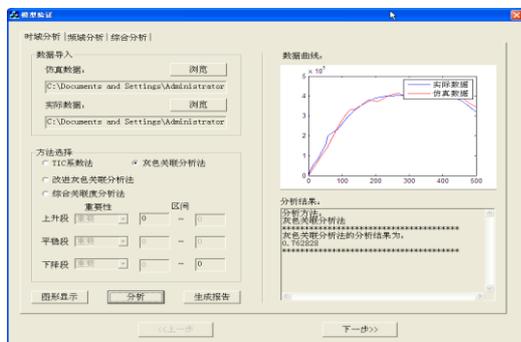


Figure3 Model Validation Tool



Figure4 Experimental Design and Data Analysis Tool

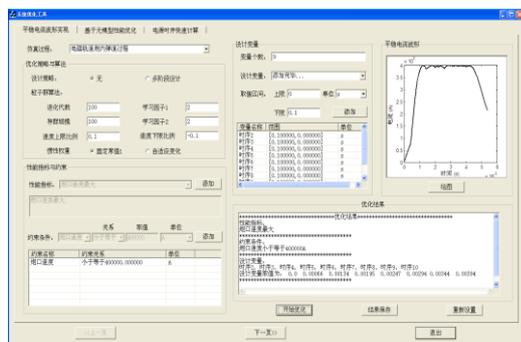


Figure5 Optimization Tool

The simulation and analysis result shows that, the integrated simulation platform of electromagnetic railgun achieved the designated objective and can be used to the electromagnetic railgun simulation, validation, post-simulation analysis and optimization.

4. CONCLUSION AND FUTURE WORK

An integrated EMRG simulation platform consisted of a whole-trajectory model, model validation tool, experimental design & data analysis tool and optimization tool is designed and implemented. The application result indicates that the platform can support the EMRG simulation, validation, analysis and optimization. Due to the repeatability of simulation approach, the simulation experiments can replace the physical tests over the model's intended applicability domain. Therefore, the simulation method can accelerate the research and development of electromagnetic launch technology.

Presently, the study of EMRG optimization has been concentrated on the optimization of the discharge sequence to obtain higher muzzle velocity, which is still a constrained single objective optimization. Actually, the efficiency of the PPS in continues firing should not be ignored. Thus, the constrained multi-objective optimization of the current will be considered. The tools of Integrated EMRG Simulation Platform are so loosely integrated that we plan to extend the platform to a more flexible development environment.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation of China (No. 61374164).

REFERENCES

Cui H. L., Xiao H. K., Yang Y. C., Xie Z. F. and Li Z. S., 2004. Design and implementation of universal integrative simulation platform software. Journal of System Simulation, 16(3),436-440.

Fair H. D., 2007. Progress in Electromagnetic Launch Science and Technology. IEEE Transactions on Magnetics, 43(1): 93-98.

Henriksen J. O., 1983. The Integrated Simulation Environment. Simulation Software, 31(6), 1053-1073.

Hsieh K. T., 1995. A Lagrangian formulation for mechanically, thermally coupled electromagnetic diffusive processes with moving conductors. IEEE Transactions on Magnetics, 31(1), 604-609.

Hsieh K. T., 2006. Hybrid FE-BE implementation on electromechanical systems with moving conductors. Proceedings of 14th Symposium on Electromagnetic Launch Technology (EML). May 22-25. Potsdam, Germany.

Hsieh K. T. and Thiagarajan V., 2008. A Novel, Split-Domain Iteration Scheme for Solution of Electromagnetic Diffusion Problems Modeled by the Hybrid Finite Element-Boundary Element Formulation, Proceedings of 14th Symposium on

- Electromagnetic Launch Technology (EML) , pp. 1-4. June 10-13. Victoria, British Columbia.
- Hundertmark S. and Roch M., 2012. Transient 3-d simulation of an experimental railgun using finite element methods. Proceedings of 17th Symposium on Electromagnetic Launch Technology (EML). May 15-19. Beijing, China.
- Hu Y. W., Ma P., Yang M. and W Z. C., 2012. Validation and optimization of modular railgun model. Proceedings of 17th Symposium on Electromagnetic Launch Technology (EML). May 15-19. Beijing, China.
- Hu Y. W., Ma P., Yang M., Lu L. Y. and Wang Z. C., 2014. Multi-stage optimization strategy of discharge sequence of pulsed power supply for electromagnetic launcher system based on improve PSO. System engineering and electronic, 36(7), 1232-1237.
- Li J., Yan P. and Yuan W. Q., 2014. Electromagnetic gun technology and its development. High Voltage Engineering, 40(4), 1052-1064.
- Ma P., Yang M. and Hu Y.W., 2012. A Whole Trajectory Simulation for the Electromagnetic Rail Gun. Proceedings of the fourth International Conference on Advance in System Simulation, pp. 92-96. Lisbon, Portugal.
- Naim A.k., Willard M. Holmes, 1978. On validating simulation models of missiles systems. Journal of Simulation.
- Kitzmiller J. R. and Driga M. D., 2003. An optimized double ramp integrated launch package design for railguns. IEEE Transactions on Magnetics, 39(1): 62-65.
- Kitzmiller J. R., Pratap B. A., Aanstoos T. A., Cook K. G., Kuenast R. A. Murphy B. T. and Perkins D. E., 1995. Optimization and critical design issues of the air core compulsator for the cannon caliber electromagnetic launcher system (CCEML). IEEE Transactions on Magnetics, 31(1), 61-66.
- Musolino A., 2004. FEM/MOM formulation for the analysis of current distribution in rail launchers. Proceedings of 13th Symposium on Electromagnetic Launch Technology (EML), pp. 363-368. May 25–28, Snowbird. USA.
- Pitman R. K., Ellis R. L. and Bernardes J. S., 2004. Iterative transient model for railgun electromechanical performance optimization. Proceedings of 13th Symposium on Electromagnetic Launch Technology (EML), pp. 96-99. May 25–28, Snowbird. USA..
- Sargent R. G., 2009. Verification and validation of simulation models. Proceedings of Winter simulation Conference, pp.166-183. New Jersey, USA.
- Sengil N., 2012. Implementation of monte carlo method to an electromagnetic launcher simulator. Proceedings of 17th Symposium on Electromagnetic Launch Technology (EML). May 15-19. Beijing, China.
- Tzeng J. T. and Hsieh K.T., 2014. Electromagnetic field effect and analysis of composite structure. Proceedings of 17th International Symposium on Electromagnetic Launch Technology (EML). July 7-11. San Diego, USA.
- Yan Y., Chao T., Ma P., Yang M., Hu Y.W., 2014. Optimization of EM railgun based on orthogonal design method and harmony search algorithm. Proceedings of 17th International Symposium on Electromagnetic Launch Technology (EML). July 7-11. San Diego, USA.

AUTHORS BIOGRAPHY

Yuchen Zhou is a Ph.D. student at Harbin Institute of Technology (HIT). His research interests include simulation system evaluation and optimization. His email is zhoyuchen-01@163.com.

Ping Ma a professor of control and simulation center at HIT. Her research interests include system simulation theory and simulation system optimization. Her email is pingma@hit.edu.cn.

Ming Yang is the corresponding author. He is a professor and the director of control and simulation center at HIT. His research interests include system simulation theory and VV&A. His email is myang@hit.edu.cn.

SYNERGY OF MATLAB AND MODELICA IN THERMAL FLOWS CONTROL IN BUILDINGS

Borut Zupančič

University of Ljubljana, Faculty of Electrical Engineering
Tržaška 25, 1000 Ljubljana, Slovenia
borut.zupancic@fe.uni-lj.si

ABSTRACT

The paper briefly describes the Modelica model of a cubic shaped room with one window. The 'physical' model was then implemented as a Modelica (Dymola) block in Matlab-Simulink environment. Simulink was used for the realisation of different control schemes, which were 'manually' and 'automatically' optimized. The experiments show that the synergetic combination of Matlab-Simulink and Dymola-Modelica environments is an efficient and powerful approach giving the possibility to realize several important goals: realisation preserving modelling in Modelica, efficient simulation with Simulink and many possibilities for control system design and optimization using basic Matlab and appropriate Matlab toolboxes. However the experiences with Modelica modelling taught us that Modelica models become rather complex and therefore model reduction techniques in order to obtain usable and efficient models are desired. The last part of the paper briefly describes some research activities in this area and also our contributions.

Keywords: OO modelling, control design, model reduction, Modelica

1. INTRODUCTION

Modelling in control is very important in many phases: for the design of new control methods and algorithms, for the implementation of a control algorithm (e.g. model based control systems), for the design of a concrete control system solution but also on higher CIM levels dealing with supervision, fault detection and diagnosis, production supervision, coordination and optimization. The conventional modelling and simulation approach was based on causal block oriented tools, e.g. Matlab-Simulink and before on the so called CSSL languages.

However due to many disadvantages of this approach new modelling techniques were developed in nineties, e.g. Bond graphs and OO acausal and multi domain modelling which preserve the realisation aspects of the systems being modelled. The result was the Modelica language (Fritzson, 2004), (Modelica, 2010) and also the de-

velopment of several environments (Dymola, Math Modelica, Open Modelica, MapleSim, ...) (Cellier, 1991), (Dymola, 2014).

These modelling techniques were used in our long term activities in modelling of thermal and radiation flows in buildings. We started with this area 15 years ago in cooperation with the Faculty of Civil Engineering, University of Ljubljana. Our first simulator was developed in Matlab-Simulink environment (Škrjanc et al., 2001), (Lah et al., 2005). A miniature test building (cubic shaped, 1m, 1 window) with which we were able to validate the model was also developed. However several disadvantages were noticed: the approach itself was never properly accepted by people from the Civil engineering department, because models in Simulink were difficult to understand. The documentation in Simulink is very problematic and not very transparent. Then we also learned that it is not possible to build the library of reusable components in Simulink. Namely when we wanted to use our one room model for a several rooms model, it was simply not possible. Every new configuration demanded the design almost from the scratch.

Due to these disadvantages we switched to the Dymola-Modelica environment and a new simulator using the standard Modelica library and some components developed by ourselves were developed from the scratch. The results of these developments were published in (Zupančič and Sodja, 2008), (Sodja and Zupančič, 2009), (Zupančič and Sodja, 2013b).

2. MODELLING OF THERMAL AND RADIATION FLOWS IN BUILDINGS IN MODELICA

The basic idea of implementation in Dymola-Modelica is to decompose the system into components that are as simple as possible and then to start from the bottom up, connecting basic components (classes) into more complicated classes, until the top-level model is achieved. The model of the room was built from the prepared model classes. Mostly the model classes from the standard Modelica library for one dimensional thermal processes were used (e.g. `HeatCapacitor`, `ThermalConductor`,

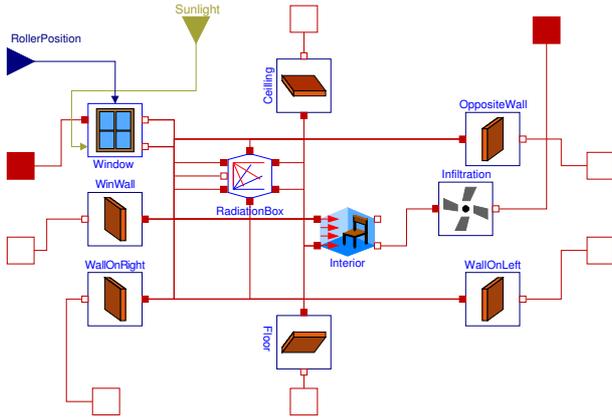


Figure 1: Modelica Model of the Room

Convection, Body Radiation). The standard connector Heatport was also used with heat flow and temperature interface variables.

The appropriate model scheme of the room is shown in Fig. 1. We can notice the model classes for walls, window, floor, ceiling, interior (furniture), etc. All phenomena of thermal flow transfers are taken into account: conduction, convection and radiation. Solar radiation through the window and radiations among the walls and window represent the most complicated modelling parts.

3. PREPARATION OF THE MODELICA MODEL FOR MATLAB-SIMULINK

Dymola-Modelica is an extremely powerful tool for true physical modelling. However for complex experimentations (e.g. optimization, linearization, steady state calculation, etc.), for results presentation it is far from Matlab possibilities. So we decided to use Dymola-Modelica just for the 'physical' part and Matlab-Simulink for all other needs: Simulink for control systems description and Matlab with some toolboxes for making experiments. We prepared a top level Modelica model which can be used as a Dymola (Modelica) block in the Matlab-Simulink environment. Actually we had to prepare appropriate connectors, which are compatible with other Simulink blocks. Such top level Modelica model is shown in Fig. 2. We prepared five inputs (outdoor temperature, roller blind position, direct solar radiation, diffuse solar radiation and artificial heating-cooling) and one output (indoor temperature). Then we prepared Simulink environment to accept Dymola block. This block has to be compiled within Simulink before the simulation is started.

4. CONTROL SYSTEMS OPTIMISATION IN MATLAB

Of course there is no need to use the Matlab environment for pure simulation runs as these can be performed efficiently also in Dymola. However Matlab is efficient if

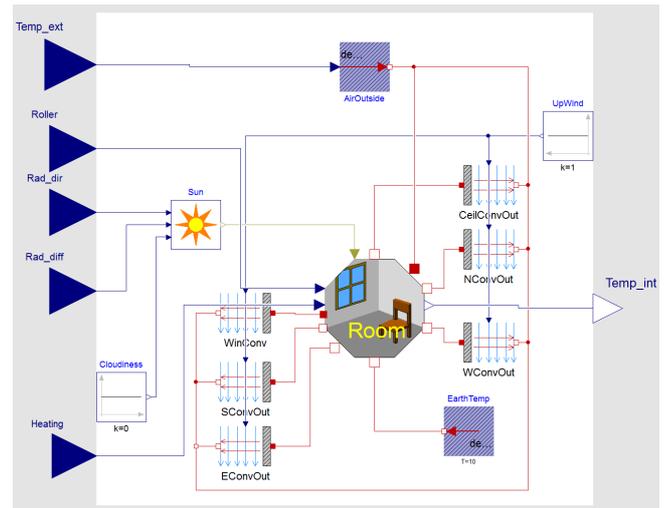


Figure 2: Top Level Modelica Model Intended for the Use within Simulink

we program more sophisticated experiments using Toolboxes. In the design of control systems we can determine the PID controller parameters K_P , T_I , T_D

$$u(t) = K_P \left(e(t) + \frac{1}{T_I} \int e(t) dt + T_D \frac{de(t)}{dt} \right) \quad (1)$$

$u(t)$ is the control variable and $e(t)$ is the error or the difference between reference and actual room temperature. Optimisation scheme is shown in Fig. 3. The main Matlab programme actually after initialization calls the optimization function which is supplied also with the special Matlab function for criterion evaluation. Criterion function is evaluated by the help of control systems simulation using Simulink-Dymola model.

Optimization toolbox and unconstrained optimization with the function `fminsearch` were used.

5. EXPERIMENTS WITH P AND PI CONTROLLERS

Although the basic goal was to harmonise the thermal and also radiation flows which influence temperatures and illuminations, we started with more basic experiments to control the internal temperature with additional heating/cooling. Fig. 4 shows the appropriate Simulink diagram. The controller minimises the error between the desired and the actual room temperature. Prior to that, we also performed a number of open loop experiments (Zupančič and Sodja, 2013a). We used a variety of test signals: constants, the step changes as well as signals derived from actual measurements on the test room. Beside usual controller inputs - reference temperature and actual temperature, we added additional input - the signal of direct solar radiation. With this input we intend to improve the control with appropriate feed forward control. The scheme includes the calculation of the criterion functions by means of which an effective manual or auto-

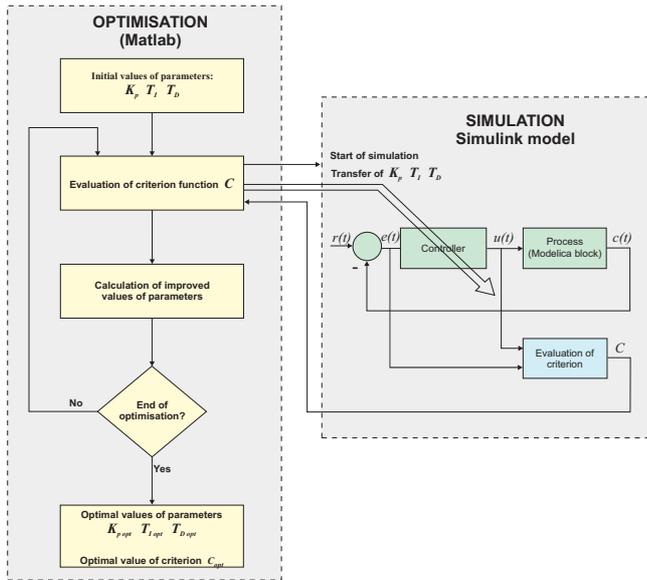


Figure 3: Optimisation Using Matlab, Simulink and Modelica

matic tuning (see Fig. 3) is performed. In the first experiment we used a P controller with manually tuned gain $k_p = 50$. With the block FromWorkspace the changeable reference room temperature was set. Due to P control a steady state error appeared which was decreased by a higher gain value. The steady state error was also decreased with a superposition of a feedforward signal 20W to the control signal.

Using PI (proportional-integral) controller it was possible to significantly decrease the controller gain. The optimisation calculated the gain $k_p = 2$ and the integral time constant $T_i = 10$. Fig. 5 shows the heating/cooling signal and the indoor temperature when the reference temperature changes. The steady state error is small (the biggest value app. 0.6°C). The control signals are also significantly smaller as with P controller (max. 90W). Fig. 6 depicts signals in the environment of real measurements: direct (Rad_dir) and diffuse radiation (Rad_diff) and outdoor temperatures (Temp_ext) were recorded in the period of five days using our pilot set-up. PI controller (with the same parameters as before) was used for the control of indoor temperature (Temp_int). In some time instances we also made changes in the roller blind opening (Roller). The last two diagrams in Fig. 6 show the heating signal (Heating) and the indoor temperature (Temp_int). The reference temperature was changed from 15°C to 20°C , 25°C and again to 20°C .

6. REALISATION-PRESERVING MODEL REDUCTION OF MODELS IN MODELICA

Beside described examples we used Modelica with Matlab in many other applications. We learned that OO and multi-domain modelling approach is very efficient especially in model definition phase, but unfortunately not so

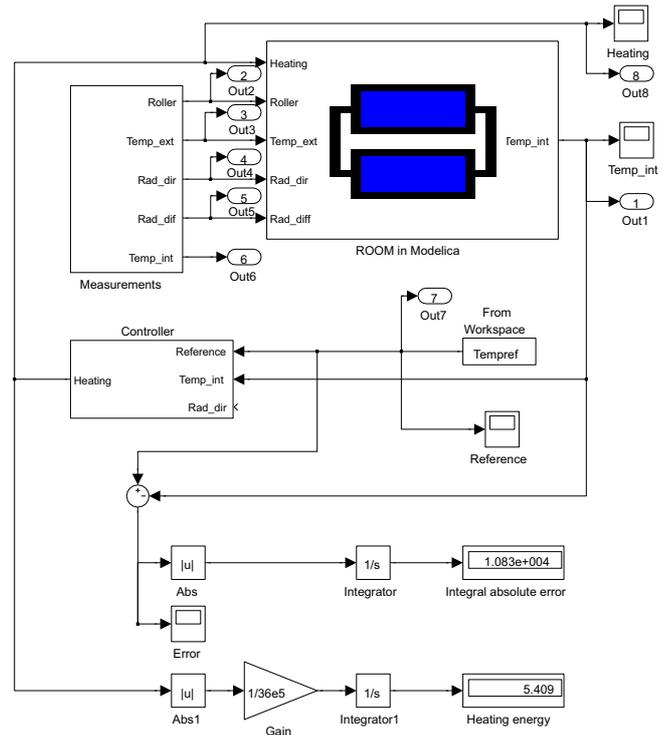


Figure 4: Simulink Scheme with Modelica Model for the Control Experiments

much in model execution. Namely under the surface of very transparent models very complex structures for execution are obtained. If we use well tested components it does not mean that the model will produce accurate results when many components are put together into a model. If one room model performs accurate results it does not assure that the model with several rooms is also accurate and usable. A simplification and/or model reduction is therefore very important in each modelling application. It is a well-known guideline that a model should not be more complex as necessary for a given purpose. Models satisfying this requirement, i.e. having proper complexity, are often designated as proper models (Wilson and Stein, 1995). However, contemporary component-based modelling approach often yields very detailed models from the beginning and the obtained models can be too complex for many intended tasks. Therefore, automatic model reduction techniques are active research topic and so far numerous automatic model reduction methods have been developed (Ersal, 2007), (Sodja and Zupančič, 2012), (Sodja, 2012). In some fields, e.g., integrated circuits design, they reached a stage when they became an indispensable part of system analysis and hence provided as a part of designated modelling environments (Ugryumova, 2011). The most successful methods, for example, those based on projection techniques, are not realisation-preserving (Ersal, 2007)-the reduced model retains input-output behaviour of the system, but loses physical interpretability of its structure

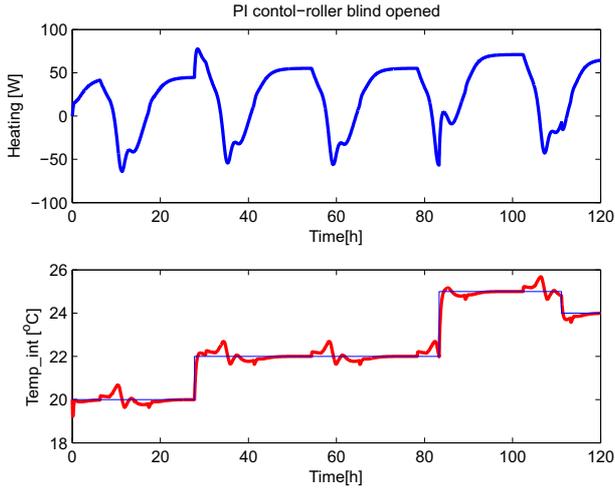


Figure 5: PI Control, Fully Opened Roller Blind

and parameters. In some cases it may be no longer possible to simulate the reduced model with the simulator of the modelling environment which was used at design of the full model. Although preservation of realisation is a very desirable property, realisation-preserving reduction methods are mostly neglected in the literature, mostly due to their bad efficiency. Furthermore, most of existing methods are limited to a certain type of models, e.g., RC circuits (Sheehan, 1999). There are no realisation-preserving model reduction methods known to the author that could adequately handle multi-domain models implemented in contemporary object-oriented modelling languages such as Modelica. Models in Modelica are usually decomposed into several hierarchical levels. At the bottom of the hierarchy, differential-algebraic equations are used for the component description, while on higher levels, model is described by connecting acausal objects (components). This is often done graphically and resulting schematics are called object diagrams (Modelica, 2010). In order to preserve the organisation of original model a combination of model reduction methods is needed. Furthermore, for some tasks, e.g., model verification (Sodja and Zupančič, 2011), only a part of the model might be desired to be reduced.

6.1. Realisation-preserving reduction at object-diagram level

The simplest procedure for reducing models represented with a scheme (graph) is to remove connections (edges) or components (nodes) estimated to have insignificant effect on salient dynamics of the system. Very intuitive approach to determine these connections or components is to use energy and power related metrics. Most energy-related metrics were developed to reduce bond graphs (Louca, 1998), (Ye and Youcef-Youmi, 1999). Bond graphs are object-oriented modelling formalism based on energy and energy exchange and hence very appropriate for energy-based model reduction methods. A power associated with each component is easily obtainable by

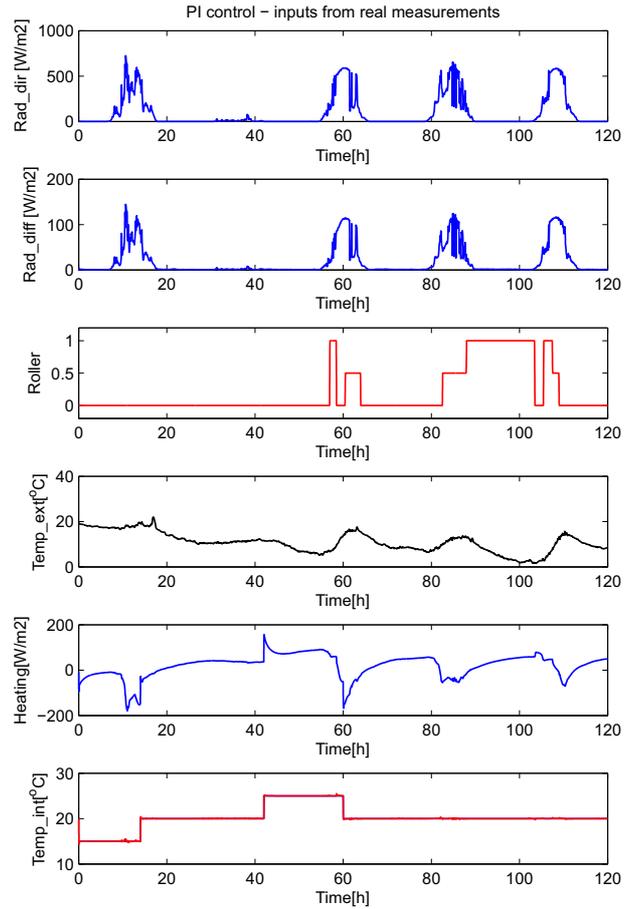


Figure 6: PI Control: Input Signals are Real Measurements

multiplying variables of the associated bond. (Louca, 1998) introduced *activity* of elements, an integral of absolute value of all energy the element (submodel) has exchanged with its surroundings within a given time interval $[t_1, t_2]$:

$$A_i = \int_{t_1}^{t_2} \left| \sum_j \dot{e}_j(t) \right| \cdot dt \quad (2)$$

In Eq. 2 $\dot{e}_j(t)$ designates the j -th energy flow through the boundary of an element. Activity has a physical meaning, it namely represents the amount of energy that flows through the element within a given time interval. It differs from the total RMS energy-flow of an element by putting less weight on the peak values since it uses maximum norm instead of the square averaging which is also in use.

Before element ranking, activities of all elements should be normalised, what means that they are divided by a sum of all elements' activities (total activity of the system)

$$AI_i = \frac{A_i}{\sum_{j=1}^n A_j} \quad (3)$$

so that a time independent measure is obtained. Normalised activity measure is dubbed *activity index* (AI)

(Louca, 1998).

However, the bond graphs are not the prevalent modelling methodology anymore. Energy, which a component exchanges with its environment, is not so explicitly available in Modelica as in bond-graph formalism (Sodja and Zupančič, 2011), (Sodja and Zupančič, 2012), (Sodja, 2012). However, it can be obtained by inspecting the connections of the components. There are only few different types of physical interactions and therefore types of connections, so if a connector is defined appropriately, a list of rules for calculating power of each connection-type is generated and power associated with a component is calculated as the sum of powers of its connections. Elimination of low ranked components (or connections) in Modelica is even more difficult, because components usually can't be classified in generalised inductance, capacitance and resistance as in case of bond graphs. After ranking of the component is done, it can be whether left to the user to decide how to reduce the model (which is adequate in some cases) or the rules for proper removal of components are derived by automatic manipulation of underlying equations.

6.2. Example: Ranking components in the room model

It was mentioned in (Sodja, 2012) that for each connector of Modelica Standard Library it is possible to determine associated energy-flow considering only information provided by connector's definition. Nevertheless, some connectors are not very appropriately defined for the usage with energy-related metrics. Such an example is the connector for 1-dimensional heat transfer found in library `Modelica.Thermal.HeatTransfer - Interface.HeatPort`. It consists of the effort variable, which is the temperature T , and the flow variable which is the heat-flow rate Q_{flow} . Therefore the energy flow is in this case equal to the flow variable Q_{flow} :

$$\dot{e} = Q_{flow} \quad (4)$$

Namely an extensive (flow) variable of connector for 1-dimensional heat transfer is heat (energy) flow itself, so it disregards the information conveyed by the intensive (potential) variable. The more consistent solution would make us the possibility to obtain heat flow by multiplying the flow and potential variable in the connector. Consider the model presented in section 2. that we developed for the thermal behaviour of the test room. It uses almost exclusively connectors for 1-dimensional heat transfer. In Table 1, components of the room submodel are listed (object diagram is shown in Fig. 1) and sorted according to their *activities*, which were calculated (Eq. 3) for a simulation experiment using measured data for three autumn days. The modelled room has a cubic shape with equal walls, so it was expected that activities of the walls are roughly the same. The results at the bottom of the Table 1 where components `RadiationBox` and `Infiltration` have allegedly zero activity are more surprising. That is because these two components only

Table 1: Ranking of the Room-Model Components According to the Activity Metric

Element	Activity [J]	Relative [%]	Accumulated [%]
window	$1.38 \cdot 10^7$	22.32	22.325
OppositeWall	$7.76 \cdot 10^6$	12.59	34.91
WinPort	$7.55 \cdot 10^6$	12.25	47.17
WallOppositePort	$5.95 \cdot 10^6$	9.65	56.82
Ceiling	$3.68 \cdot 10^6$	5.97	62.79
WallOnLeft	$3.63 \cdot 10^6$	5.88	68.67
WallOnRight	$3.63 \cdot 10^6$	5.88	74.56
WinWall	$3.44 \cdot 10^6$	5.59	80.14
CeilingPort	$3.23 \cdot 10^6$	5.24	85.39
Floor	$2.77 \cdot 10^6$	4.49	89.88
WallOnRightPort	$1.40 \cdot 10^6$	2.28	92.15
WallOnLeftPort	$1.40 \cdot 10^6$	2.28	94.43
WinWallPort	$1.39 \cdot 10^6$	2.26	96.69
FloorPort	$1.16 \cdot 10^6$	1.88	98.57
Interior	$8.08 \cdot 10^5$	1.31	99.88
OutsideAir	$7.35 \cdot 10^4$	0.12	100.00
RadiationBox	0.01	0.00	100.00
Infiltration	0.00	0.00	100.00

transfer heat without storing it. Therefore sum of all energy flows on their borders is zero at any time instant. Choice of connector variables where extensive variable is energy flow thus causes that only energy-storing components are considered while transfer-only components are ignored what is by no means acceptable. (Sodja, 2012) used entropy generation rate in Eq. 2 (in place of \dot{e}) to evaluate activity metric for a component instead of using heat flow. However the order of components was the same as in Table 1 but with nonzero but still small values of the last two components.

Of course the main question is, what to do with the Table 1. Of course we can not just eliminate the components with low activity, because some classes can not be directly compared. But nevertheless we can find sometimes a very useful information: e.g. the window is very important, the walls have similar importance - perhaps some walls can be modelled with one unified wall etc. Of course if one component between several similar components has much lower activity, we can think how to eliminate this component from the model. In the next section an example will show how the represented ranking can be used for model verification.

Using ranking for model verification

Energy-related metrics were first used for visualisation of dynamic systems modelled with bond graphs (Rosenberg and Ermer, 1995). Analytical models are usually derived from the principle of energy conservation, so it is very intuitive tool for model verification, because it is easy to estimate energy levels of the submodels already in the phase of model design.

Consider a scenario where the component `WallOnLeft`

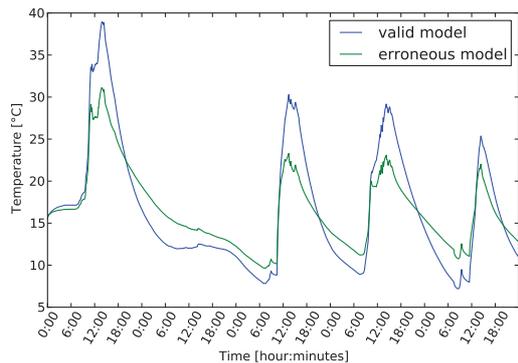


Figure 7: Response of the Erroneous Model Compared to the Response of the Valid Model

Table 2: Component Ranking of the Erroneous Room Model – Component `WallOnLeft` Has the Parameter `BrickworkConductance` Set to Ten Times Higher Value.

Element	activity [J]	relative [%]	accumulated [%]
window	$1.34 \cdot 10^7$	21.21	21.21
WallOnLeft	$1.09 \cdot 10^7$	17.25	38.46
WinPort	$7.50 \cdot 10^6$	11.83	50.29
OppositeWall	$6.82 \cdot 10^6$	10.75	61.05
WallOppositePort	$5.91 \cdot 10^6$	9.32	70.37
CeilingPort	$3.14 \cdot 10^6$	4.95	75.32
Ceiling	$3.06 \cdot 10^6$	4.82	80.14
WallOnRight	$2.64 \cdot 10^6$	4.17	84.31
WinWall	$2.39 \cdot 10^6$	3.77	88.08
...

of the model in Fig. 1 has a parameter set to a wrong value. In particular, the conductivity of the brickwork layer is ten times higher as it should be (i.e., it can be caused by a typing error).

Fig. 7 shows the response of this erroneous model in comparison with response of the model having all parameters set correctly. The erroneous model has a distinctly different behaviour than the valid model, so it is easy to detect the presence of error. On the other hand simulation results provide little information to locate the cause of the error. Component rankings, listed in Table 2, is much more elaborate. The component `WallOnLeft` has higher activity as other components also representing walls with same dimension and composition. Therefore, component `WallOnLeft` is probably the cause of error.

6.3. Realisation-preserving reduction at equation level

There are already commercially tools available (Sommer et al., 2008) for reduction and simplification of a general set of differential-algebraic equations. The method combines various algebraic manipulations and approximation techniques, for example, deletion of a single term in an

equation, replacement of a term with a constant, deletion of a variable or its derivative, etc. Simplification/reduction operations are ranked according to estimated discrepancies of reduced- and full-model trajectories. The method gives good results for algebraic set of equations, while efficient extension to differential-equation systems is more difficult.

7. CONCLUSIONS

As modelling and simulation is very important in many control design phases it is clear that most recent approaches and tools are desired. In comparison with our former implementation of the model of the room in Matlab-Simulink the OO approach with Modelica significantly improves modelling possibilities. The time for the model development is shortened, the models are more transparent and it is easier for a control engineer to work with area professionals as they better understand Modelica models. In the first part of the contribution we wanted to show the efficacy of this approach modelling a control system for the harmonization of thermal flows in buildings. The combination of Matlab-Simulink and Dymola-Modelica was extremely efficient. Unfortunately with this and even more with some other applications we noticed that such approach has also a limitation due to a huge complexity which appears, when a complex hierarchical structure is flattened for the efficient execution. Therefore some model reduction techniques are even more important in such modelling approaches as in traditional ones. Some methods were tested, developed and also built into a Modelica environment. However all methods are still far from being automatically used. Based on selected metrics a ranking table is obtained. User must carefully analyse the information and try to perform one or more reduction steps. All reductions must be properly verified. By now we actually did not succeed to make some efficient and automatized reductions in our complex applications but we were able to obtain some good results in more simple test examples. So there are many possibilities for the future work: to develop new or improved methods for reduction of object diagrams and equations but also for the implementation of appropriate procedures in modelling compilers.

REFERENCES

- Cellier, F. E., 1991. *Continuous System Modeling*, Springer Verlag, New York.
- Dymola, (2014). *Dymola, Dynamic Modeling Laboratory, User's Manual*, Dassault Systèmes.
- Ersal, T., (2007). *Realization-preserving simplification and reduction of dynamic system models at the graph level*, PhD thesis, University of Michigan.
- Fritzson, P., 2004. *Principles of Object Oriented Modeling and Simulation with Modelica 2.1*, IEEE Press, John Wiley & Sons, Inc., USA. Publication.
- Lah, M. T., Zupančič, B. and Krainer, A., 2005. Fuzzy control for the illumination and temperature com-

- fort in a test chamber, *Building and environment*, 40 (12), 1626–1637.
- Louca, L. S., (1998). *An Energy-based Model Reduction Methodology for Automated Modeling*, PhD thesis, University of Michigan.
- Modelica, (2010). *Modelica Specification, version 3.2*. www.modelica.org/documents/ModelicaSpec32.pdf.
- Rosenberg, R. C. and Ermer, G. E., 1995. A bond graph visualization tool to improve engineering system design, *Systems Analysis, Modeling, Simulation*, 18–19, 173–178.
- Sheehan, B. N., 1999. TICER: Realizable reduction of extracted RC circuits, *IEEE/ACM International Conference on Computer-Aided Design, Digest of Technical Papers*, pp. 200–203.
- Sodja, A., (2012). *Object-oriented modelling and simulation analysis of the automatically translated models*, PhD thesis, University of Ljubljana, Fac. of El. Eng.
- Sodja, A. and Zupančič, B., 2009. Modelling thermal processes in buildings using an object-oriented approach and Modelica, *Simulation modelling practice and theory*, 17 (6), 1143–1159.
- Sodja, A. and Zupančič, B., 2011. On using model approximation techniques for better understanding of models implemented in Modelica, *Proceedings of the 8th International Modelica Conference*, Dresden, Germany, pp. 697–703.
- Sodja, A. and Zupančič, B., 2012. Realisation-preserving model reduction of models in Modelica, *Proceedings of the 7th Vienna Conference on Mathematical Modelling*, Vienna, Austria, pp. 322–328.
- Sommer, R., Halfmann, T. and Broz, J., 2008. Automated behavioral modeling and analytical model-order reduction by application of symbolic circuit analysis for multi-physical systems, *Simulation Modelling Practice and Theory*, 16, 1024–1039.
- Ugrumova, M. V., (2011). *Applications of Model Order Reduction for IC Modeling*, PhD thesis, Eindhoven University of Technology.
- Škrjanc, I., Zupančič, B., Furlan, B. and Krainer, A., 2001. Theoretical and experimental fuzzy modelling of building thermal dynamic response, *Building and environment*, 36 (9), 1023–1038.
- Wilson, B. H. and Stein, J. L., 1995. An algorithm for obtaining proper models of distributed and discrete systems, *ASME Journal of Dynamic Systems, Measurement and Control*, 117, 534–540.
- Ye, Y. and Youcef-Youmi, K., 1999. Model reduction in the physical domain, *Proceedings of the American Control Conference*, San Diego, CA, USA, pp. 4486–4490.
- Zupančič, B. and Sodja, A., 2008. Object oriented modelling of variable envelope properties in buildings, *WSEAS transactions on systems and control*, 3 (12), 1046–1056.
- Zupančič, B. and Sodja, A., 2013a. Analysis and control design of thermal flows in buildings: efficient experimentation with a room model in matlab-modelica environment, *Proceedings of the 8th Eurosim Congress on Modelling and Simulation*, Cardiff, UK, pp. 155–160.
- Zupančič, B. and Sodja, A., 2013b. Computer-aided physical multi-domain modelling: some experiences from education and industrial applications, *Simulation modelling practice and theory*, 33 (6), 45–67.

AUTHORS BIOGRAPHY

BORUT ZUPANČIČ received his Ph.D. and became a full professor at the Faculty of Electrical Engineering, University of Ljubljana in 2000. His major research interests are: control systems, multi-domain and object oriented modelling and simulation, continuous and hybrid control systems design, harmonization of thermal and flows in buildings. He is the author of more than 200 conference papers and 50 papers in scientific journals, co-author of one international book (published by Prentice Hall Inc.) and author or co-author of several books in Slovene language. He was the president of EUROSIM - the Federation of European Simulation Societies 2004-07 and the president of SLOSIM - the Slovene Society for modelling and simulation. Currently he is the secretary of EUROSIM and the Head of the Laboratory for Modelling, Simulation and Control at the University of Ljubljana, Faculty of Electrical Engineering. His Web page can be found at <http://msc.fe.uni-lj.si/Staff.asp?person=2>.

WEB SIMULATION WITH SUPPORT OF MOBILE AGENTS

Jan Voracek^(a)

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

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

ABSTRACT

This paper solves the problem of mobile agents in distributed simulation. Its main focus is on the method of implementation of transferring mobile agents in the JavaScript language, which does not support marshalling. The secondary focus is on the general suitability of agent-oriented architectures in JavaScript. All of this with the use of WebRTC, a peer-to-peer communications technology.

Keywords: web simulation, mobile agent, WebRTC

1. INTRODUCTION

Computer simulation is an effective approach to study the behaviour of various systems over time. Distributed simulation works by distributing the calculation to a cluster of computing nodes that are interconnected in a communication network. This approach has many benefits, e.g., the simulation calculation can scale easily or, in some cases, more cost-effective resources can be used, etc. The more complex the problem is, the more advantageous it is to distribute it to more computing nodes (Fujimoto, 1999). Another frequent reason for introducing distributed solution is using multiple operators.

Some of the problems, which have to be solved with running distributed simulation, are (of course among implementation of the simulation model itself) the setup of environment which the simulation will be running in and the interconnection of computing nodes for communication. This interconnection becomes nontrivial when the computing nodes are not on the same computer network.

For the simulation of more complex systems, it is also advantageous to use the agent-oriented architecture. This architecture allows to divide the complex system into several individual logical units which complexity is substantially lower (Macal 2005).

2. THE INTERNET AS A PLATFORM FOR COMMUNICATION

As told in the introduction, the distributed simulation works by distributing the calculation to a cluster of computing nodes. For the needs of distributed simulation, it is often sufficient to communicate in a private network. However, we can be in a situation where we need to run the simulation between two or

more dislocated departments which are not connected to a private network. Alternatively, we want to link individual computing stations around the world, where the building of a private network is not possible for economic or pragmatic reasons. In this case, it is convenient to use the Internet, the global computer network.

This approach, however, has some drawbacks. When communicating using the Internet it cannot be guaranteed which way the data will be transferred or who will be able to access it. Therefore as a security precaution, it is preferred to encrypt all the communication (Mitrovic 2014).

Another difficulty is a problematic direct communication of computing nodes. The stations performing the simulation are typically connected to a private network behind a router and do not have assigned a public IP address.

3. THE WEB AS A RUNTIME ENVIRONMENT

In the introduction, there was also told that one of the problems of running distributed simulation was putting a runtime environment into operation. It is necessary to install this environment on all the stations involved in the simulation calculation. A self-executable simulation model gives us an alternative way to run the simulation. However, both of these approaches are more or less dependent on the hardware (processor architecture) or software (operation system) platform.

Proposed solution uses, as its runtime environment, a web browser, which is available not only for personal computers but also for tablets, smartphones and smart TVs (Mitrovic 2014). The fact that web browsers are (unlike the majority of specialized environments) free may be also considered as a benefit.

A solution using a web browser is not so much dependent on a particular platform as usual approach, but in combination with various types of communication it may be dependent on specific browsers or their versions (Kartak 2014).

After mentioning the communication, it would be appropriate to outline what are the possibilities of communication on the web. It is possible to divide the communication options into two groups:

- native – used from JavaScript,

- third-party products – Java Applets, Flash, Silverlight (Kartak 2014).

Solutions that use third-party products were evaluated as unsuitable as they used to be unwieldy and brings a lot of their own problems – especially in terms of safety (Kartak 2014).

As for the native forms of communication, there are several technologies to use:

- XmlHttpRequest known as AJAX,
- WebSockets,
- WebRTC.

The AJAX is available in all commonly used web browsers while WebSockets and WebRTC come as part of an HTML5 specification and their availability in popular web browsers is listed in Table 1.

Table 1: Availability of HTML5 communication technologies in popular web browsers by their version and its release date

Web Browser	WebSockets	WebRTC
Chrome	14 (09-2011)	23 (11-2012)
Firefox	6 (08-2011)	22 (06-2013)
Internet Explorer	10 (08-2012)	–
Opera	12.10 (11-2012)	22 (06-2014)
Safari	6 (07-2012)	–

Adapted from Kartak 2014

3.1. AJAX

It is a standard client-server communication based on HTTP requests. The communication is always initiated by the client (the web browser), and the server only responds to the requests. It cannot initiate the connection (Zehe 2013). For the needs of distributed simulation, it is not ideal because the communication is centralized, and the server can easily become a bottleneck of the entire simulation calculation.

3.2. WebSockets

WebSockets also offers a client-server communication, however, compared to AJAX, the connection remains open, and the data can flow in both directions. Therefore, the server can initiate a data transfer and send a message to a client at any time (Green 2012). However, the communication is still centralized.

3.3. WebRTC

WebRTC significantly extends the capabilities of web browsers in the field of communication. The main benefit of the distributed simulation is the support of peer-to-peer communication, which does not require any central element, through which all the data flow (Barnes, 2014). In addition, the WebRTC implements an ICE protocol (Interactive Connectivity Establishment), which allows connecting two stations in separate private networks without requiring any configuration of network elements. The proposed

solution uses WebRTC despite the limited support in web browsers.

WebRTC Security

As told, the encryption of communication is one the most important aspects of communicating in a public network. Also because of that, the WebRTC standard specifies that all its implementations have to support the DTLS-SRTP (Data Transport Layer Security – Secure Real-time Transport Protocol). This protocol works like HTTPS only it is based on the UDP communication protocol instead of TCP (Barnes 2014).

4. AGENT-ORIENTED SIMULATION

The agent-oriented simulation consists in dividing of a comprehensive system to autonomous logical units capable of interaction (a.k.a. agents). Combined actions of individual agents then simulate the behavior of the modeled system.

Agents can be classified into different groups. For example, in terms of complexity of their autonomous behavior, they can be classified as (Nwana 1996):

- intelligent – with the ability to meet their goals using logical deduction,
- reactive – with the ability to respond to stimuli,
- deliberative – with the ability to plan their actions and to affect their environment to gain an advantage,
- cognitive – with the ability to deduce logical conclusions from observing their environment, they can learn and create a knowledge base,
- rational – they have all of the characteristics.

In a distributed system, the agents can be classified in terms of variability of place of execution of their code as:

- stationary – the agent is allocated on a single computation node, where its code is executed until its termination or until the end of simulation,
- mobile (migrating) – agent may move its code and state to another computation node by itself (without human intervention), where it continues with performing its task (Pinsdorf 2002, Zehe 2013).

4.1. Mobile agents in web environment

The main focus of proposed solution is an implementation of mobile agents in JavaScript. Unfortunately, this language does not support marshalling (the serialization of data structures representing the agent into a form suitable for transmission over the network), which is an essential prerequisite for moving the agent to another computation node.

5. JAVASCRIPT SERIALIZATION

As mentioned above, unlike most currently used languages, JavaScript does not support a complete transform of objects and functions into a form suitable for transmission over the network, or for storing. It supports only conversion into JSON, which, unfortunately, supports only these data types:

- *Number*,
- *String*,
- *Boolean*,
- *Array*,
- *Object*,
- *Null*,

where the *Object* data type is only a map and can only contain string keys and values with data types listed above (see RFC 7159 and ECMA-404).

5.1. Serialization libraries

Besides the built-in functionality for converting objects to JSON, there are also libraries, which tries the full serialization. Each of them has a greater or lesser restrictions. According to search results, libraries *GSerializer* and *JASON* are the most used. Both project, however, are no longer maintained, and they do not work with the current version of JavaScript. Also, for example, the *GSerializer* does not support serialization of parametric methods.

5.2. Proposed solution of serialization

The proposed solution is slightly inspired in the Dart language and the way this language solves conversion to JSON and back without losing information. Dart allows each class to define methods *toJSON* and *fromJson* whose implementation is completely up to the developer. Support of a similar approach can also be found in other languages. E.g. in Java, there are methods *writeObject* and *readObject* in C# method *GetObjectData* and a special private constructor. The main taught is that there are two methods. One method can extract the object state to a simpler data structure (e.g. those which can already be converted into JSON or XML) and the second one is able to restore the object to its original state based on the data. These methods are called *getState* and *setState* in the proposed solution.

For example, assume an instance of the class *Car*, which has properties *color* and *topSpeed* and several methods representing the behaviour of this object. The methods are the same for all objects. What differs is the current state of individual objects represented by values of their properties. To create any instance of the *Car* class, we therefore, need a prescription of this class and mentioned values of its properties.

The prescription of a class is represented by its source code, and the state is obtained using method *getState*, as shown in Figure 1.

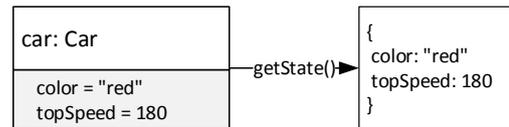


Figure 1: Obtaining the state of an object

To restore the original object just create any (presumably “empty”) instance and use method *setState*, as shown in Figure 2.

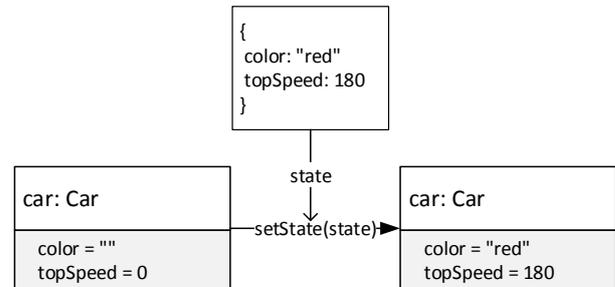


Figure 2: Restoring the object

The problem will occur when a part of the state is an object shared with other serialized object (shown in Figure 3). This may be the subject of further research.

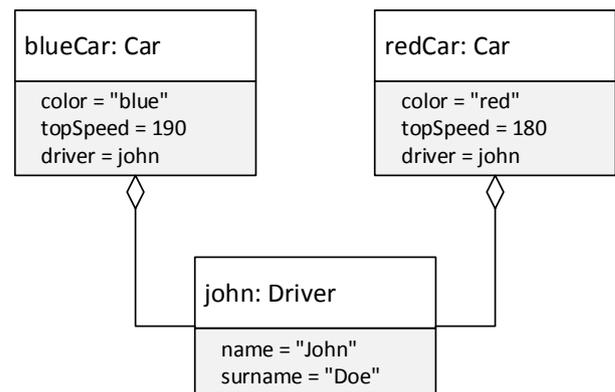


Figure 3: Two objects sharing third one

6. CASE STUDY

As a case study was chosen distributed strategic game codenamed “devconquest”. The target is to conquer the opponent’s castle while protecting their own. Each player builds a castle from available material and programs his military units to defend his castle and conquer the opponent’s one.

The player can choose from several different materials on the construction of fortifications, where more solid are of course more expensive. He can also build loopholes up on the walls into which can be placed archers.

Regarding the military units, the game offers several basic types:

- soldier – offensive / defensive unit with a sword,
- archer – offensive / defensive unit with a bow,
- scout – reconnaissance unit a dagger.

Each unit has a basic offensive and defensive number (its attack and defence capabilities), visibility range and maximum velocity of movement. The player can modify these abilities, but only so that their sum does not exceed a fixed maximum. Each unit is needed to program its behaviour so that player has reached the victory.

During the game, the military units turn in performing their action – moving over the game field, attacking an enemy unit in range, exploring the condition of walls, communicating with other units, etc. First, all units of one player perform their turns, and then continue the second one's and so on.

At the beginning of the game the player has covered the opponent's part of the map and gradually reveals it as his units move over the opponent's field. Opposing units are visible only if some player's unit has them in sight. For a better imagination, there is a game in progress shown in Figure 4.



Figure 4: Game in progress

6.1. Utilization of mobile agents

Every military unit represents one agent who can travel between the connected stations. It uses the services provided by the hosting system for communication with its environment. The agent can, for example, send messages to the other agents, explore the game field or interact with it (attack enemy units, walls).

The agents' logic can be implemented directly in the browser. The application contains simple code editor with highlighting, as you can see in Figure 5.



Figure 5: Agent editor

7. CONCLUSION

The article introduces the reader to the possibilities of implementing a distributed simulation on the Web, especially using the WebRTC technology. WebRTC

brings the possibility of creating distributed simulation without central node (server). The article extends the distributed simulation with a support of mobile agents. To implement mobile agents has to be solved the problem of missing serialization of objects in JavaScript, which eventually became a major problem that this article addresses. The proposed solution is verified through a case study represented by a strategic, conquering game, where mobile agents represent different military units. The proposed solution provides space for further research – especially in the sharing object between multiple agents.

REFERENCES

- Barnes R. L., Thomson M., 2014. Browser-to-Browser Security Assurances for WebRTC. *Internet Computing, IEEE* 18 pp. 11–17.
- Fujimoto R. M., 1999. *Parallel and Distribution Simulation Systems*. John Wiley & Sons, Inc.
- Green I., 2012, *Web Workers: Multithreaded Programs in JavaScript*. O'Reilly Media.
- Kartak S., Kavicka A., 2014. WebRTC Technology as a Solution for a Web-Based Distributed Simulation. In *Proceedings of the European Modeling and Simulation Symposium*, pp. 343–349, Genova: Università di Genova.
- Macal C. M., North M. J., 2005. Tutorial on agent-based modeling and simulation. In *Proceedings of the 37th conference on Winter simulation*, pp. 2–15, Orlando, Florida.
- Mitrovic D., Ivanovic M., Budimac Z., Vidakovic M., 2014. Radigost: Interoperable web-based multi-agent platform. *Journal of Systems and Software* 90 pp. 167–178.
- Nwana H. S., 1996. Software agents: an overview. *The Knowledge Engineering Review* 11 pp 205-244.
- Pinsdorf, U., Roth, V., 2002. Mobile agent interoperability patterns and practice. In *Proceedings of the 9th IEEE International Conference on Engineering of Computer-based Systems*, pp. 238–244.
- Zehe D., Ayt H., Lees M., Knoll A., 2013. JavaScript Distributed Agent Based Discrete Event Simulation. In *Proceedings of the 2013 IEEE/ACM 17th International Symposium on Distributed Simulation and Real Time Applications (DS-RT '13)*. IEEE Computer Society, Washington, DC, USA, 21-29.

SIMULATORS SUSTAINABILITY ASSESSMENT

Dace Aiztrauta^(a), Egils Ginters^(a), Artis Aizstrauts^(a)

^(a)Sociotechnical Systems Engineering Institute of Vidzeme University of Applied Sciences, Cesu street 4, Valmiera LV-4201, Latvia

^(a)dace.aiztrauta@va.lv, ^(a)egils.ginters@va.lv, ^(a)artis.aizstrauts@va.lv

ABSTRACT

It is no secret that simulation plays an ever increasing role in solving various challenges outside the military domain.

Using simulation the policy planners can examine a variety of complex solutions before their implementation as well as forecast the impact of decisions on the attainable goal in general (Silva et al 2009).

Several simulators are being developed, however the lack of unified standards, task diversity and complexity raise costs for the development of a comprehensive simulator. These costs are significantly higher than in traditional software development.

This begs the question of sustainability regarding a new or already absorbed simulator as costs must be justified. How to determine technology sustainability? This question preoccupies users as well as investors.

It is still impossible to get a convincing answer. A few years ago the term “sustainability” was almost exclusively linked to environmental protection, and even today it is very rarely used in relation to technology sustainability.

More than five years ago the authors started developing the sustainability assessment methodology IASAM (Barkane et al 2010, Aiztrauta and Ginters 2013), which is based on system dynamics simulation. After several years of validation the work continues with the second version IASAM2, which left behind the Venkantesh UTAUT acceptance model in favour of Rodger’s diffusion ideas.

Keywords: simulation, system dynamics, sustainability assessment

1. INTRODUCTION

The decision making support tools are typical sociotechnical systems reflecting the wishes of the decision maker and aligning these with the technical possibilities of the developer, thereby promoting project clearness and giving the beneficiary an opportunity to validate the offered solutions and to repair conceptual mistakes in the early stages of software design.

Designing policy decision-making tools consumes time and funding therefore it is reasonable to forecast the

sustainability of the simulators thereby avoiding redundant investments in short-term solutions.

There are several theories that reflect the issues of technology and solution acceptance or sustainability research, but none of them give a full understanding of the combined factors influencing acceptance and sustainability. Theories such as Technology Acceptance Model (TAM) (Davis et al 1989), Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh 2003), Expectation-Confirmation Theory (ECT) (Bhattacharjee 2001) focus mainly on the exploitation stage and deal with prediction and model the behaviour of users who make the decision to adopt the approach or reject it. These methodologies do not seem to fully answer the question how to evaluate technology acceptance and sustainability at any given point in time of the technology lifecycle and forecast the chances of technology to attract users and achieve the aims of its developers. And what are the main elements and factors that influence the acceptance and sustainability of technology?

The Integrated Acceptance and Sustainability Assessment Model (IASAM) (Aiztrauta and Ginters 2013) is a new approach for evaluating technologies that combines socio-economic aspects and socio-technical characteristics of technology development and exploitation thus framing a united multi-level framework for technology sustainability assessment based on system dynamics (SD) simulation.

2. INTEGRATED ACCEPTANCE AND SUSTAINABILITY ASSESSMENT MODEL (IASAM) STEP BY STEP

The new concept of technology sustainability was proposed to evaluate the set of factors that let the technology to be developed, implemented, maintained properly (i.e. according to the needs of all stakeholders), attract long-term users and create positive output and/or outcome according to the purpose of the technology and initial intentions of its developers (financial, social, etc.) (Aiztrauta and Ginters 2013)

2.1. Methodology fundamentals

The IASAM approach is based on the opinion that technology acceptance research should not be separated from technological, economic and social evaluation.

Although the term “socio-technical systems” is loosely used to describe many complex systems, five key aspects determine a socio-technical system:

- The system has interdependent parts;
- The system adapts to and pursues goals in external environments;
- The system has an internal environment comprising separate but interdependent technical and social subsystems;
- There is a choice in the system, e.g. system goals are achievable by more than one means;
- System performance depends on jointly optimizing the technical and social aspects of the system (Badham et al 2000).

According to the IASAM methodology there are four main groups of factors that influence technology sustainability S and thus comprise the backbone.

$$S = \langle M, Q, A, D \rangle \quad (1)$$

- Management (M) – successful management of every asset;
- Quality of technology (Q) – quality of the product (simulator);
- Technology acceptance (A) – acceptance of the new product (simulator) by customers/users;
- Domain development and societal processes (D) – the development of society demands more diverse technologies, which, in turn, after development, change society and also influence the demand for new technologies.

Thus IASAM consists of four flows in SD understanding that all together constitute what we call technology sustainability (see Figure 1).

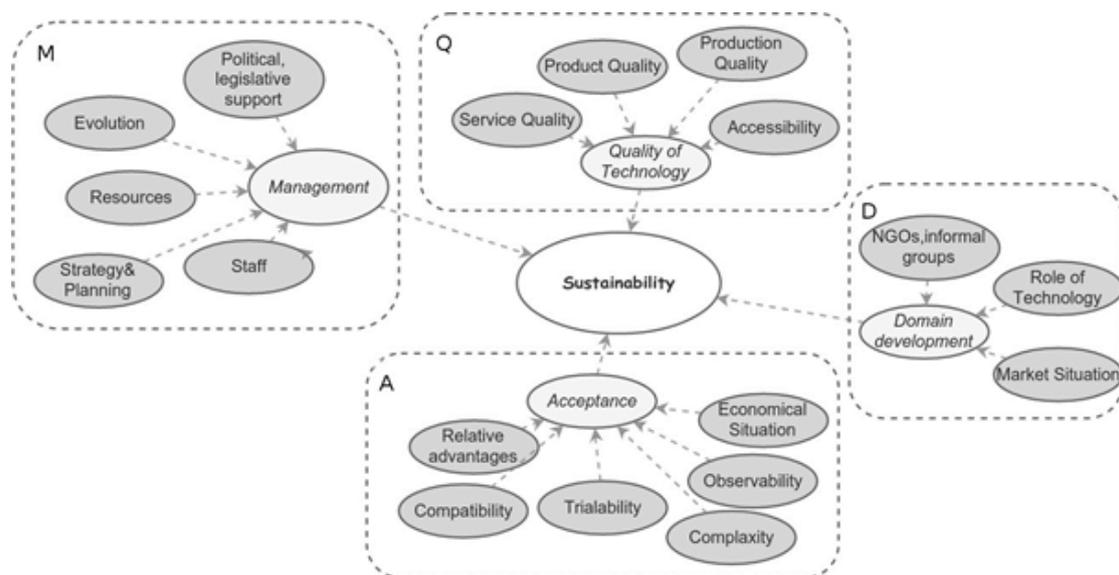


Figure 1: IASAM SD model in Insight Maker notation

Two internal flows are – Management of development and exploitation, and Quality of technology. And two external flows are – Technology acceptance and Domain development. Each includes several socio-technical factors that all together constitute the IASAM.

The IASAM approach can be viewed as the following sequence of processes that may be repeated over time based on the needs of evaluator (see Figure 2).

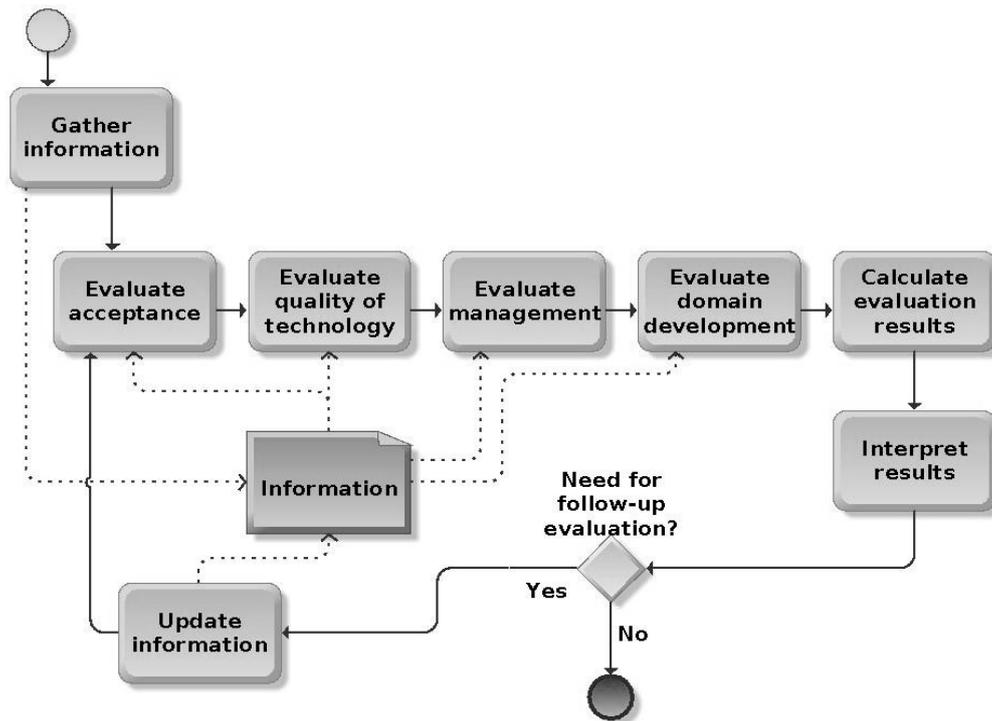


Figure 2: Assessment in conformity with IASAM rules

The evaluation of criteria is undeniably subjective, but it relies on the assumption that every evaluator, whether a technology developer or potential investor, will want to receive a reliable evaluation for decision-making. Each criterion is evaluated with the help of a specially formulated criteria description/statement. The respondent evaluates each description. The model follows these steps:

- The statement provided for each criterion is evaluated on the 7 point Likert scale. The questions are all formed in a unified manner – every statement is formulated in a positive way so that bigger score is always “better”. If it is not possible to evaluate any criteria at the time of the current evaluation (for example, the evaluator does not know the answer because the technical specification is not yet ready), the criteria should be marked with “NA”;
- The result gives a numerical value of integrated technology sustainability index (IASAM index) consistent with the assessment framework. It is calculated as the sum of all values from the questionnaire and divided by maximum possible value of questions answered:

$$IASAM 2_{index} = \frac{\sum_{n=1}^{12} F_n + \sum_{i=1}^{49} B_i}{(N - C) * 7}$$

where (F) – additional IASAM2 survey response values; (B) – initial IASAM2 survey response values, (N) – total number of questions; (C) – number of questions marked with “N/A”. The final result is then interpreted according to IASAM2 methodology;

- The IASAM methodology was validated using Skype application sustainability research (Aizstrauta and Celmina et al 2013) and the IASAM index was measured in *skypes* making the assessment more understandable and perceivable. After validation it was also decided that IASAM needs certain amendments to make it more user-friendly without losing its multi-dimensional view.

The IASAM was tested under the framework of FP7-ICT-2009-5 CHOReOS project No. 257178 (2010-2013) “Large Scale Choreographies for the Future Internet (IP)” (CHOReOS, 2013) to assess project pilots.

2.2. Development and implementation technology management

Resources include available funds (or clear perspectives on this matter), time resources, equipment and knowledge in the organization (the technology development team). IASAM criteria help to evaluate the methods used to manage these resources. The detailed

level of this and other Management flow factors and criteria used for evaluation can be seen in Table 1.

Table 1: Description of criteria “Resources”

	Criteria	Criteria description/statement
Resources	Available funds (existing or credit options)	Project management has a sound finance attraction and management strategy
	Time resources	Sufficient and reasonable time resources are planned for technology development and implementation processes
	Equipment	There is a sufficient amount of equipment or a sound plan for equipment procurement
	Knowledge in organization	Project management possesses knowledge creation, description, sharing, and other important knowledge management mechanisms

Each situation is of course different and no one can name one single project management approach that would fit all cases, but it is important that certain unified principles are applied in project management. The IASAM model being universal, evaluates only the core criteria.

The next factor is Human resources. *Human resources* have been separated from other resources on purpose as this is a very important resource in itself. Although people involved in the project can depend on other types of resources, it can play a crucial role in ensuring technology acceptance and sustainability. The evaluation criteria for this factor include knowledge and skills in technology elaboration and requirement engineering, knowledge and skills in technology distribution into the market, and motivation (see Table 2).

Table 2: Description of criteria “Human resources”

	Criteria	Criteria description
Human resources	Knowledge and experience in technology design and implementation	Involved staff has the necessary knowledge and skills or project management has defined a clear recruitment strategy

	Knowledge and experience in project management	Involved staff has the necessary knowledge and experience in project management or project management has defined a clear recruitment strategy
	Motivation	Project management has a sound and grounded motivational system for the staff

Strategic management principles include target setting, ex-ante analysis, risk assessment, market research, requirements engineering, etc. (see Table 3). Therefore the evaluation criteria are as follows – ex-ante market research, ex-ante competitor’s research, ex-ante target audience research, risk management, business plan, resource management, and project management. IASAM does not promote one specific project management approach, but it is important that the technology development team has the necessary skills and is familiar with project management tools and strategies.

Table 3: Description of criteria “Strategic management”

	Criteria	Criteria description
Strategic management	Ex-ante market research	Project management has comprehensive knowledge about the market situation in the established field
	Ex-ante competitors research	Project management has a clear notion and information about the activities of competitors as well as about competing technologies
	Ex-ante target group research	Project management has a clear notion of the target audience and its needs
	Risk assessment	Project management is aware of potential risks and possible activities to prevent them
	Resource planning	Project management uses adequate resource planning methods to ensure that the necessary resources are available in the right amount and at

		the right moment
	Project management	An adequate project management approach is chosen
	Business plan	There is a business plan (in the broadest sense)

To respect the influence from a governmental level, the model also adds the factor called **Political/Legislative support**. This factor characterizes the public policy context for the given technology. It is important because development of technology very often means eclipse in traditional realms or demands serious re-orientation, so technology developers might need to face contra arguments by lobbyists. Also legislature can have a great effect on new technology. To mention the most extreme examples – the use of a technology can be prohibited or deemed compulsory. And this undeniably impacts the acceptance and usage of a technology. The evaluation criteria for this factor are – political support, lack of political constraints, ideology (technology conforms to dominant ideology, social norms, etc.) and legislature. The criteria and their descriptions can be seen in Table 4.

Table 4: Description of criteria “Politico-legal support”

	Criteria	Criteria description
Politico-legal support	Political level support	Technology development/implementation has political support
	Absence of political obstruction	The technology development/implementation process does not have any obstructions on the political level/lobbying against it.
	Ideology	The technology is not inconsistent with the dominant ideology/social norms and processes
	Legislation	Technology use is compulsory

Another factor left is **Evolution** that is also connected with the flow Domain development and ensures feedback to project management about the necessary amendments to ensure sustainability. This concept is rather close to the concept of maintainability. Maintainability can be defined as the ability of an item to be retained in or restored to a specified state.

Maintainability has to be built into systems during design and development by realizing a maintenance concept. This built-in maintainability also means an ability to ensure changes in the provided technology as deemed necessary by situation changes in the domain environment. The evaluation criteria of this factor are – maintenance resources, evolvability of technology, feedback and maintenance processes (see the descriptions on Table 5).

Table 5: Description of criteria “Evolution”

	Criteria	Criteria description
Maintainability	Maintenance resources	Resources for maintenance are available
	Evolvability	The technology is able to evolve, its operation is not time-restricted
	Feedback	There are mechanisms to research feedback and make the necessary changes to the technology
	Maintenance processes	There are mechanisms to prevent/repair errors and flaws

2.3. Quality of technology

Product quality covers characteristics of technology outputs, the need for additional gadgets to use the output or compatibility with different platforms. The criteria for evaluation are as follows: output accuracy, output timing, output perceptibility, additional gadgets (needed or not), possibilities for technical integration. By output we mean anything that reaches the client/user and is used by him – information, services, sound, video etc. (see Table 6).

Table 6: Description of criteria “Product quality”

	Output accuracy	Output is consistent with the determined client needs
Product quality	Output perceptibility	Output is comprehensible and ready to use, no additional processing is needed
	Output timing	Output is ready when it is most convenient for the client
	Technical integration	The technology is compatible (or is customizable) with different platforms,

		formats.
	Additional gadgets/software	No additional gadgets/software are needed to use technology output

Production quality concerns the issues of the creation process – system design, coding practices and documentation management. Therefore the evaluation criteria include ICT architecture, Component and pattern re-use, algorithm complexity, coding practices, error and exception handling, system documentation as well as development process documentation (see Table 7).

Table 7: Description of criteria “Production quality”

Production quality	ICT architecture	The architecture of the technology is sound and valid
	Algorithm complexity	Algorithms are not unnecessarily complex
	Component and pattern re-use	The development of the technology involves component and pattern re-use
	Coding practices	Adequate coding methodology is chosen (waterfall, prototyping, spiral, etc.)
	Error and exception handling	An error and exception handling strategy is defined
	System documentation	The system (technology) is adequately documented
	Development process documentation	The system (technology) development process is adequately documented

Another crucial element for technology sustainability are the support services associated with the use of technology. **Service quality** looks at accessibility of support systems and the personnel working with support issues. This factor is measured using four criteria – support timing, support staff, support system (clients understand and accept it), the availability of support channels demanded by clients (see Table 8).

Table 8: Description of criteria “Service quality”

Service quality	Support channels	Support availability conforms to the clients' needs
	Support system	The support system is comprehensible and convenient for clients
	Support timing	Support is available when necessary
	Support staff	Support personnel is qualified or project management has defined a clear recruitment strategy

Besides the characteristics of quality, there is an additional factor – **Accessibility**. With the help of this factor the model evaluates overall technological accessibility (physical accessibility as well as accessibility of necessary knowledge and skills). For example, to use Internet services and for them to become successful, the Internet should be easily accessible. There are three evaluation criteria – Infrastructure accessibility, ICT usage costs and ICT skills (see Table 9).

Table 9: Description of criteria “Accessibility”

Accessibility	Infrastructure accessibility	The whole infrastructure necessary for technology usage is accessible to society (the target group)
	ICT skills	Society (the target group) possesses sufficient ICT skills
	ICT usage costs	ICT is available to society in terms of financial capabilities

2.4. Technology acceptance and moving to IASAM2

In first version of IASAM the Unified Theory of Acceptance and Use of Technology (UTAUT) model was used to evaluate user acceptance. The UTAUT combines other models and includes four key determinants for acceptance analysis: performance expectancy, effort expectancy, social influence, facilitating conditions; and four moderators: gender, age, experience, voluntariness of use (Venkatesh et al 2003). The UTAUT survey was carried out separately and the results were used in two ways. First, a separate analysis of UTAUT survey results gave a comprehensive understanding of potential user acceptance. Secondly, the integration of certain answers into IASAM results added the acceptance perspective to the model.

The use of UTAUT involved a potential user survey and that created the greatest difficulty associated with IASAM evaluation. IASAM was planned to be a model that is applicable to any stage of technology development and it was very challenging to obtain surveys of potential users. Therefore it was decided to change the methodology for user acceptance evaluation and to replace UTAUT and potential user surveys with another approach. After a thorough research it was concluded that the user acceptance evaluation can be expanded based on diffusion of innovations by Rogers (Rogers 2003). It examines innovations from many perspectives. The extensive research includes an innovation-development process, an innovation-decision process, attributes of innovations and their rate of adoption, different adopter categories, as well as topics on leadership and change agents and innovation in organizations.

The main amendments of IASAM2 are related to the inclusion of theory of diffusion. Practically, the part of IASAM where UTAUT criteria were used has been replaced in IASAM2 by these five attributes of innovations:

- Relative advantage – the innovation is technically superior (in terms of cost, functionality, “image”, etc.) than the technology it supersedes;
- Compatibility – the innovation is compatible with existing values, skills, and work practices of potential adopters;
- Complexity – the innovation is relatively difficult to understand and use;
- Trialability – the innovation can be experimented with on a trial basis without undue effort and expense; it can be implemented incrementally and still provide a net positive benefit;
- Observability – the results and benefits of the innovation’s use can be easily observed and communicated to others (Rogers 2003).

Based on the work carried out by Rogers the criteria and sub-criteria were defined as follows (see Table 10).

Table 10: Description of criteria “Accessibility”

Relative advantage	Economic profitability	Economic profitability is an advantage of using this technology
	Low initial cost	Low initial cost is an advantage of using this technology
	Decrease in discomfort	Decrease of discomfort of some kind is an advantage of using this technology
	Social prestige	The use of this technology advances the social prestige of the user

	Saving time/effort	Saving time and/or effort is an advantage of using this technology
	Immediacy of the reward	The benefits of using technology are immediate and that is an advantage of using this technology
Compatibility	Social/cultural values and beliefs	The use of technology is positioned as compatible with social/cultural values and beliefs
	Previously introduced ideas	The use of technology is positioned as compatible with previously introduced ideas
	Client needs	The use of technology is positioned as compatible with client needs
Complexity	Complexity of technology	The technology is positioned and should be perceived by potential users as easy
Trialability	Trial availability	There are mechanisms (free downloads, trial versions, prototypes), that enable the users to easily try the technology
Observability	Observability of technology	The results and benefits of technology is easily visible by potential users

One additional factor included in evaluation is the Economic situation. It affects the development of new technologies as well as existing ones, depending on their type and aims. For example, during economic hardship, people may seek ways to economize and choose cheaper technologies instead of more expensive ones. This factor is described as the lack of outer economic constraints for technology use and distribution. One criterion is evaluated here and the description can be seen in Table 11.

Table 11: Description of criteria “Economic situation”

	Criteria	Criteria description
--	----------	----------------------

Economic situation	Economic situation	There are no external economic obstacles for technology usage and acceptance
--------------------	--------------------	--

2.5. Domain development

The model also includes Domain development as a set of factors that impact technology sustainability. Despite the positive impact of technology development on society overall, looking from the perspective of technology creators, every innovation endangers its current position within the technology market.

This section is a very important part of IASAM, as it adds contextual information about technology development, implementation and usage period. This also distinguishes IASAM from other acceptance and technology success measurement methodologies. These criteria reflect other market players as well as other stakeholders outside the market (see Table 12).

Table 12: Description of criteria “Market situation”, “NGOs/informal groups” and “Role of technology”

	Criteria	Criteria description
Market situation	Similar products	There are similar technologies, this only adds new details.
	Uniqueness	The technology is unique
	Brand recognition	The technology uses (or is itself) a popular brand
NGOs/informal groups	Support	There are significant NGOs/informal groups or “opinion leaders” that (might) promote technology implementation, use and acceptance
	Lack of opposition	There are no significant NGOs/informal groups or “opinion leaders” that (might) hamper technology implementation, use and acceptance
Role of technology	Social role	The technology enables the transfer of some kind of popular idea into the electronic world
	Technical role	The technology enables the use of other technologies

Technologies among other competitors (*Market situation*) describe concurrent technologies as well as concurrent companies. It characterizes advantages among other technologies. There are three criteria to be evaluated – similar technologies, uniqueness and brand recognition (when technology uses a familiar brand).

The factor describing *Nongovernmental organizations* and activists or informal groups seek to represent different opinion-based groups that may influence the use of technology, by opposing it, or actively promoting its use. It is measured with two criteria – support and lack of opposition from such groups.

The *Role of technology* looks at the social (does it enable new ideas/paradigm shift?) and technological role (does it enable the users of other technologies?) of a technology. These criteria help to evaluate the necessity for the technology within the society at the certain point of its development. For example, a social role helps to determine whether and how the technology eases or promotes the implementation of new ideas, new social movements etc. To mention specific examples, massive open online course platform Coursera.org reflects and promotes the paradigm shift in education or stakeholders that empowers and enables a new level of public participation and local governance.

3. FROM IASAM TO IAAM2 – A SHIFT TO WIDER USABILITY

The substantial difference between IASAM and IASAM2 concerns acceptance evaluation and corresponding activities. An overview of the basic differences can be seen in Table 13.

Table 13: IASAM and IASAM2 differences

	IASAM	IASAM2
Acceptance evaluation	Potential user survey based on Unified Theory of Acceptance and Use of Technology (UTAUT) methodology (38 questions)	Five criteria (12 questions) based on attributes of innovations according to Innovation Diffusion theory
Sustainability evaluation	49 questions	49 questions

Data gathering methods	Self-assessment questionnaire + potential user survey	Self-assessment questionnaire
Data processing and analysis	2 phases – UTAUT survey analysis and then conjoint IASAM analysis	Integrated IASAM2 analysis
Result interpretation	UTAUT results are presented separately and within IASAM	IASAM2 results are presented in an integrated way

The newest IASAM2 version has several significant benefits:

- The assessment can be carried out by the interested party;
- There is no need for time and resource-consuming potential user survey;
- The model itself becomes more comprehensible, as the calculus, analysis and reporting can be done within one methodological framework;
- IASAM2 meets the initial goals of this tool – to be easy to use and universal in its applications (Aizstrauta, Ginters and Piera 2015).

4. IASAM USE FOR SIMULATOR SUSTAINABILITY ASSESSMENT

EC FP7 project FUPOL No. 287119 (2011-2015) „Future Policy Modelling” (Sonntagbauer and Ginters et al. 2012) given possibilities for different policy simulator design.

Under the framework of the FUPOL project we elaborated Zagreb Open Green Park Occupancy Analysis and Layout Design simulators, Skopje Vodno Mountain Recreational Activities Analysis tools, Skopje Inter-mobility Bicycle Routes Planning simulator and also the Yantai Economic Development Assessment tool.

IASAM2 was used for simulator sustainability assessment.

The Skopje Vodno mountain recreational activities occupancy simulator software allows definition of existing bike paths and docking stations and projects to improve the existing situation. The output is a calculation of the number of bikers using a certain resource and enabling a selection of a project that will offer the highest impact. IASAM2 index value was 0.79 and credibility value 0.91. That means that important issues regarding the management, quality of technology, acceptance and domain development are fully

performed. The technology highly satisfies the defined model criteria and is expected to be accepted among target group members (local municipality decision makers and local society members willing to participate in decision design) and sustainable in terms of existence of technology and targets set within FUPOL project.

Skopje bicycles inter-modality simulator offers the City of Skopje and its citizens the opportunity to simulate the occupancy and usage of bike stations and bike parking lots. The overall goal is to increase the number of people that use bikes as a transport means, by taking several different measures, such as establishing bike inter-modality, initiating the development of parking lots, rent-a-bike facilities, new bike paths, improving existing bike paths. Skopje Bicycles received the highest IASAM2 index value – 0.88, (credibility 0.96). Important issues regarding all four flows of IASAM2 model – the management, quality of technology, acceptance and domain development – are fully performed or deliberated. The technology is expected to be accepted among target group members and sustainable in terms of existence of technology.

The Yantai urban economics assessment simulator offers the administration of Yantai the possibility to study the current situation of these industries in order to make decisions about a company’s upgrade or closure in order to decrease resource consumption and impact on the environment. Yantai Urban Economics received IASAM2 index 0.78 (credibility – 0.95). On the whole the technology highly satisfies the defined model criteria. The lowest rating was given to criteria connected with feedback, evolvability and maintenance – these are the issues that might require further control. Nevertheless, the technology is expected to be accepted among target group members and sustainable.

Zagreb open green park occupancy analysis simulator provides the best solution for the facilities that would be included in the 2000m² of green area situated near the Autism Centre. The simulation satisfied most of the potential user’s demands and encourages interactions between autistic and non-autistic users, while avoiding possible conflicts between them. Zagreb-Open Green Park simulator had an IASAM2 index 0.80 and credibility 0.95. Also here the absolute majority of issues regarding the management, quality of technology, acceptance and domain development are fully performed or deliberated. The technology highly satisfies the defined model criteria. No criteria received an evaluation of less than five.

The high values of IASAM2 credibility indicate that the majority of questions have been evaluated and there are no issues that have not received careful deliberation. The results are not surprising taking into account the FUPOL project scale and clear management procedures.

5. CONCLUSIONS

The IASAM approach provides the opportunity to simulate a time-varying system with multiple feedback links and analyse quantitative and qualitative factors.

These application results lead to the following conclusions:

- Different factors of technology creation and distribution, and also contextual issues regarding, for example, the market and society have been taken into account in the development process of the simulators;
- According to the definition of technology sustainability these factors, when taken into consideration, allow the technology to be developed, implemented and maintained according to the needs of all stakeholders and attract long-term users;
- IASAM2 index values that are higher than 0.75 indicate that important issues regarding the management, quality of technology, acceptance and domain development are fully performed or deliberated. The technology highly satisfies the defined IASAM2 criteria and thus the evaluated technologies can be considered valuable from the acceptance and sustainability perspective;
- The model itself and the survey questions can be used as a checklist for further improvements and development planning.

Further activities of IASAM2 development will be related to methodology deployment of the Future Internet and providing SaaS access.

ACKNOWLEDGEMENTS

The article publication is initiated by FP7 FUPOL project No. 287119 “Future Policy Modeling”.

REFERENCES

- Aizstrauta D., Ginters E., Pjera M.A., 2015. Applying Theory of Diffusion of Innovations to Evaluate Technology Acceptance and Sustainability. In ICTE in Regional Development, December 2014, Valmiera, Latvia, Procedia Computer Science, Elsevier, Volume 43, 69–79.
- Aizstrauta D., Ginters E., 2013. Introducing Integrated Acceptance and Sustainability Assessment of Technologies: a Model based on System Dynamics Simulation. //In Springer LNBIP 145 Series “Modeling and Simulation in Engineering, Economics and Management”/ Ed.: Maria Angeles Fernandez-Izquierdo, Maria Jesus Munoz-Torres, Raul Leon, proceedings of the International Conference MS2013, Castellona de la Plana, Spain, June 2013, Springer Verlag Berlin Heidelberg 2013, ISSN 1865-1348, eISSN 1865-1356, 23-30.
- Aizstrauta D., Celmina A., Ginters E., Mazza R., 2013. Validation of Integrated Acceptance and Sustainability Assessment Methodology. ICTE in Regional Development, December 2013, Valmiera, Latvia. Procedia Computer Science. Bluemel, E., Ginters, E. (eds). Elsevier, ISSN: 1877-0509, Volume 26, 33-40.
- Barkane Z., Vincent H., Ginters E., 2010. Systems Dynamics Use for Technologies Assessment. In Proceedings of the 22th European Modeling & Simulation Symposium (EMSS 2010), October 13-15 2010, Fes, Morocco, ISBN 2-9524747-8-8, EAN 9782952474788.
- Bhattacharjee A., 2001. Understanding Information Systems Continuance: and Expectation-Confirmation Model,” MIS Quarterly 25(3), 351—370.
- CHOReOS, 2013. FP7-ICT-2009-5 CHOReOS project No. 257178 (2010-2013) “Large Scale Choreographies for the Future Internet (IP)” <http://www.choreos.eu/> [accessed 10 June 2015].
- Davis F.D., Bagozzi R.P., Warshaw P.R., 1989. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science, Vol. 35, No. 8, 982—1003.
- Rogers E.M., 2003. Diffusion of Innovations, 5th Edition. Simon and Schuster. ISBN 978-0-7432-5823-4.
- Silva S., Fidalgo J.N., Dalila B.M.F., 2009. A simulation based decision aid tool for setting regulation of energy grids with distributed generation. ISSN 1109-2858 (Print), 1866-1505 (Online). Springer Berlin: Heidelberg.
- Sonntagbauer S., Ginters E. et al., 2012. Deliverable 4.1 – FUPOL Simulator Software Requirements Report, 231.
- Venkatesh V., Morris M., Davis G., Davis F., 2003. User acceptance of information technology: toward unified view. MIS Quarterly. Vol. 27, 425 – 478.

AUTHORS BIOGRAPHY

Dace Aizstrauta is researcher at Socio-technical Systems Engineering Institute of the Vidzeme University of Applied Sciences. Her research interests is systems sustainability simulation. She has more than 10 scientific articles related with the research field.

Artis Aizstrauts is researcher at Socio-technical Systems Engineering Institute and lecturer in the faculty of Engineering of the Vidzeme University of Applied Sciences. His research interests are software designing and distributed simulation communication environments. He has more than 15 scientific articles related with the research fields.

DYNAMIC AND KINEMATIC SIMULATION OF KAWASAKI MANIPULATOR INDUSTRIAL ROBOT USING SOLIDWORKS AND MATLAB SIMMECHANICS

Zennir Youcef^(a), Makbouche Adel^(b), Souames Hamza^(c)

^(a,b,c) Automatic Laboratory of Skikda, Route El-Hadeaik, BP26. 21000 Skikda, Algeria

^(a)youcef.zennir@univ-skikda.dz, ^(b)adel.makbouche@univ-skikda.dz ^(c)hshhhsouames@hotmail.fr

ABSTRACT

In this paper we present a graphical Human Machine Interface (HMI) with a 3D modeling and simulation of an industrial robot manipulator Kawsaki FS03N with 6 DDF. A direct and inverse geometric model, with kinematics model of robot has been developed. A dynamic robot model is developed with SolidWorks software and Matlab SimMechanics and with a bridge between SolidWorks and Matlab has been developed. The developed models use the actual robot dimensions. The aim of our work that this Human Machine Interface will be used to test different control type before applying to the real robot. Different simulation and reference movements control were performed. Finally, and before opening persepectives on future work we present the results obtained validated the functioning of our interface both SolidWorks software and Matlab.

Keywords: Modeling, 3D simulation, Manipulator robot, Solid Works software, Matlab-Simulink.

1. INTRODUCTION

Robotic science it is a multidisciplinary field mechanical, computer science, electronics. A robot is a machine that can manipulate objects and perform various movements dictated by an easily modifiable program. Program a robot is specify the movement's sequence that will be achieve. Some robots are equipped with "sense". Is a more or less set of measuring instruments and appreciation (camera, thermometer,) to program a robot how choose the more adapted movement with the external conditions. The robots equipped with artificial intelligence devices so that they can deal were unexpected and new complex situations (the robot could gain some "experience"). Robots are mainly used in industry for performing repetitive manipulations, especially when the manufacturing process is subject to frequent changes. The advantage of a robot (robot manipulator) with a contribution-man is his consistency: It can perform the same motion thousands of times in a row without feeling any tiredness (Vibert 1987). Other thing, the robots can be constructed to withstand a basis that would be harmful or fatal to humans (harmful gases, high heat, cold body, radiation,...). A manipulator robot

(with 6 DDF) to flexibility movement and different possible trajectory's and positions (Lallemand 1994). For this raison our work consist to study and to control simulate (in 3Dimention) an industrial robot manipulator kawasaki FS03N (Kawasaki 2003), with the development of a human-machine interface.

2. MANIPULATOR ROBOT KAWASAKI FS03N

The FS03 is a compact multi-purpose robot from the Kawasaki F series (Kawasaki 2003). Weighing just 3 kg, it is designed as a portable model at the smaller end of the range. Despite its size, the FS03 is an advanced 6-axis arm and boasts the highest speed in its class. Launched in 2005, the FS03 retains the compact form of its predecessors but has improved speed and acceleration/deceleration characteristics and significantly reduced cycle times. Whether placed on the floor, suspended from the ceiling or mounted on the wall (Option), the FS03 exhibits excellent freedom of movement, tailoring its acceleration and deceleration speeds to both load weight and robot position for optimum performance in all situations. The FS03 is ideal as an industrial robot for demanding tasks such as assembly, handling and inspection of small components. Standard equipment includes AS high-level robot language and the ultra-compact D70 fully digital controller. The FS03 brings robot technology one step closer to humankind. The various robot arm elements are: The base (A), the shoulder (B), the arm (C), the elbow (D), the forearm (E) and wrist (F) (figure 1).

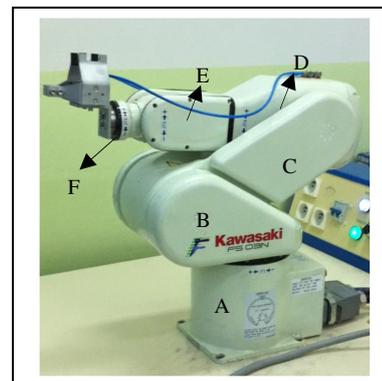


Figure 1: Kawasaki FS03N robot

Three main axes and three wrist axes deliver 6-axis performance (figure 2). The arm can be located in the desired work space and the tool location can be set, allowing greater freedom when locating peripheral equipment. The arm turning axis (JT1) and the wrist axes (JT4, 5, 6) have superior high-speed operation and feature the highest speed specifications in their class. Cycle time is very short (Cycle time: $0.4 \text{ } \Phi < 0.5 \text{ sec}^*$), and high reliability and high precision ensure that they can withstand the most stringent operating conditions of industrial robots(Ijeoma 2012)(Tuna 2001).

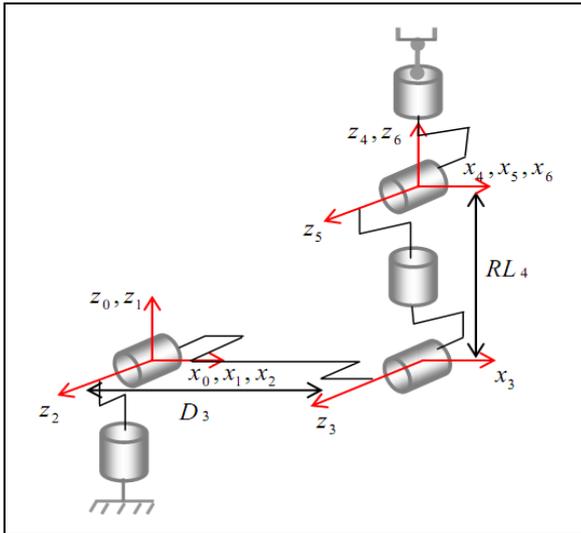


Figure 2: Axis Position.

Ideal for a variety of operating spaces: floor mount, wall mount or ceiling mount. Wiring and conduits for tool sensors are built - in inside the arm for easy operation. The utilization of absolute encoders eliminates the need to zero the unit when powering up. There is no need to worry about gravity induced interference with other equipment when the power is turned off as all six axes are broken. (Kawasaki 2003). Some specification (characteristics) of robot is illustrated in the following table:

Table 1: Kawasaki FS03N robot Specifications.

Specifications	FS03N		
Arm type	Articulated		
Degrees of Freedom	6 Axes		
Maximum Payload	3kg		
Axis Works envelope	axis	Max. Stroke	Max. Speed
	JT1	$\pm 160^\circ$	360°/S
	JT2	+150°~-60°	250°/S
	JT3	+120°~-150°	225°/S
	JT4	$\pm 360^\circ$	540°/S
	JT5	$\pm 135^\circ$	225°/S
JT6	$\pm 360^\circ$	540°/S	
Max. linear speed	6.000 mm/s		
Moment and Moment of inertia	Axis	inertia	inertia
	JT4	5.8N.m	0.12 kg.m ²
	JT5	5.8N.m	0.12 kg.m ²
	JT6	2.9N.m	0.03 kg.m ²

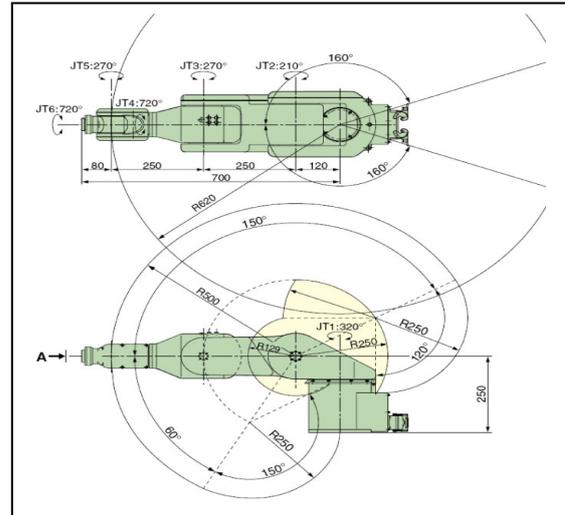


Figure 3: Motion Range & Dimensions of robot.

3. DIRECT GEOMETRIC MODEL

The direct geometric robot model is used to calculate the operational coordinates giving the position of the end-effectors based on joint coordinates. It used also to determine the configuration (position, orientation) end-effector of a robot according to the links configuration. This model is based on the determination of transformation matrix T_6^0 between R_0 and R_6 (Tuna 2000) (John 1989). The direct geometric robot parameters are illustrated in the following table:

Table 2: Direct geometric robot parameters.

J	σ_j	α_j	d_j	θ_j	r_j
1	0	0	0	θ_1	0
2	0	$-\pi/2$	0	θ_2	0
3	0	0	D3	θ_3	0
4	0	$\pi/2$	0	θ_4	RL4
5	0	$-\pi/2$	0	θ_5	0
6	0	$\pi/2$	0	θ_6	0

The homogeneous transformation matrix of the robot is:

$$T_1^0 = \begin{bmatrix} C1 & -S1 & 0 & 0 \\ C1 & S1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T_0^1 = \begin{bmatrix} C1 & -S1 & 0 & 0 \\ C1 & S1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$$T_2^1 = \begin{bmatrix} C2 & -S2 & 0 & d_2 \\ 0 & 0 & 1 & 0 \\ -S2 & -C2 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

$$T_1^2 = \begin{bmatrix} C2 & -S2 & 0 & -C2 \cdot d_2 \\ -S2 & -C2 & 1 & -S2 \cdot d_2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

$$T_3^2 = \begin{bmatrix} C3 & -S3 & 0 & d_3 \\ S3 & C3 & 1 & 0 \\ 0 & 0 & 0 & r_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

$$T_2^3 = \begin{bmatrix} C3 & S3 & 0 & -d_3 \cdot C3 \\ -S3 & C3 & 0 & S2 \cdot d_2 \\ 0 & 0 & 0 & -r_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

$$T_4^3 = \begin{bmatrix} C4 & -S4 & 0 & 0 \\ 0 & 0 & -1 & -r_4 \\ S4 & C4 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6)$$

$$T_3^4 = \begin{bmatrix} C4 & 0 & S4 & 0 \\ -S4 & 0 & C4 & 0 \\ 0 & -1 & 0 & -r_4 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

$$T_5^4 = \begin{bmatrix} C5 & -S4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -S5 & -C5 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (8)$$

$$T_4^5 = \begin{bmatrix} C5 & 0 & -S5 & 0 \\ -S5 & 0 & -C5 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (9)$$

$$T_6^5 = \begin{bmatrix} C6 & -S6 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ S6 & C6 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (10)$$

$$T_5^6 = \begin{bmatrix} C6 & 0 & S6 & 0 \\ -S6 & 0 & C6 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (11)$$

Finally:

$$T_5^6 = T_1^0 * T_2^1 * T_3^2 * T_4^3 * T_5^4 * T_6^5 \quad (12)$$

$$= \begin{bmatrix} s_x & n_x & a_x & 0 \\ s_y & n_y & a_y & 0 \\ s_z & n_z & a_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} A_6^0 & & & \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (13)$$

$$s_x = C_1 \cdot (C_{23} \cdot (C_4 \cdot C_5 \cdot C_6 - S_4 \cdot S_6) - S_{23} \cdot S_5 \cdot C_6) - S_1 \cdot (S_4 \cdot C_5 \cdot C_6 + C_4 \cdot S_6) \quad (14)$$

$$s_y = S_1 \cdot (C_{23} \cdot (C_4 \cdot C_5 \cdot C_6 - S_4 \cdot S_6) - S_{23} \cdot S_5 \cdot C_6) - C_1 \cdot (S_4 \cdot C_5 \cdot C_6 + C_4 \cdot S_6) \quad (15)$$

$$s_z = S_{23} \cdot (C_4 \cdot C_5 \cdot C_6 - S_4 \cdot S_6) + C_{23} \cdot S_5 \cdot C_6 \quad (16)$$

$$n_x = C_1 \cdot (-C_{23} \cdot (C_4 \cdot C_5 \cdot C_6 + S_4 \cdot S_6) + S_{23} \cdot S_5 \cdot C_6) + S_1 \cdot (S_4 \cdot C_5 \cdot S_6 - C_4 \cdot C_6) \quad (17)$$

$$n_y = S_1 \cdot (-C_{23} \cdot (C_4 \cdot C_5 \cdot C_6 + S_4 \cdot S_6) + S_{23} \cdot S_5 \cdot C_6) + C_1 \cdot (S_4 \cdot C_5 \cdot S_6 - C_4 \cdot S_6) \quad (18)$$

$$n_z = -S_{23} \cdot (C_4 \cdot C_5 \cdot C_6 + S_4 \cdot S_6) + S_{23} \cdot S_5 \cdot C_6 \quad (19)$$

$$a_x = -C_1 \cdot (C_{23} \cdot C_4 \cdot S_5 + S_{23} \cdot C_5) + S_1 \cdot S_4 \cdot S_5 \quad (20)$$

$$a_y = -S_1 \cdot (C_{23} \cdot C_4 \cdot S_5 + S_{23} \cdot C_5) + C_1 \cdot S_4 \cdot S_5 \quad (21)$$

$$a_z = -S_{23} \cdot C_4 \cdot S_5 + C_{23} \cdot C_5 \quad (22)$$

$$P_x = -C_1 \cdot (S_{23} \cdot RL_4 + C_2 \cdot D_3) \quad (23)$$

$$P_y = -S_1 \cdot (S_{23} \cdot RL_4 + C_2 \cdot D_3) \quad (24)$$

$$P_z = C_{23} \cdot RL_4 + S_2 \cdot D_3 \quad (25)$$

Avec $C_{23} = \cos(\theta_2 + \theta_3)$ et $S_{23} = \sin(\theta_2 + \theta_3)$

4. INVERSE GEOMETRIC MODEL

If we try to find all possible configurations for a joint position and orientation data an inverse geometric model (IGM) can meet this need. Knowing that for the serial manipulator type, the development of (IGM) is a very difficult and complex issue, where it must reverse a system of nonlinear equations which is not trivial. Nevertheless, and according to the structure of the manipulator robot, there are various methods for solving the IGM in an explicit form. In our case with the manipulator robot Kawasaki FS03N used in industry, the Paul methods can give the solutions of IGM in explicit form (Paul 1981), (Lallemand 1994). Hence we obtain the following solutions:

$$\theta_1 = \text{atan}_2(P_x, P_y); \theta_2 = \text{atan}_2(S_2, C_2) \quad (26)$$

With:

$$S_2 = \frac{Y \cdot Z - \xi \cdot X \cdot \sqrt{X^2 + Y^2 + Z^2}}{X^2 + Y^2} \quad \text{and} \quad C_2 = \frac{X \cdot Z - \xi \cdot X \cdot \sqrt{X^2 + Y^2 + Z^2}}{X^2 + Y^2},$$

$$\xi = \pm 1.$$

$$B_1 = P_x \cdot C_1 + P_z \cdot S_1; X = -2 \cdot P_y \cdot D_3;$$

$$Y = -2 \cdot B_1 \cdot D_3 \quad (27)$$

$$Z = (RL_4)^2 - (D_3)^2 + (P_z)^2 + (B_1)^2 \quad (28)$$

$$\theta_3 = \text{atan}_2\left(P_z \cdot S_2 - B_1 \cdot C_2 + \frac{D_3}{RL_4}, -B_1 \cdot S_2 + \frac{P_z \cdot C_2}{RL_4}\right) \quad (29)$$

$$\theta_4 = \text{atan}_2(a_x \cdot S_1 - a_y \cdot C_1, -C_{23} \cdot (C_1 a_x + S_1 a_y) - a_z \cdot S_{23}) \quad (30)$$

$$\theta_4 = \theta_4 + \pi; \theta_5 = \text{atan}_2(S_5, C_5) \quad (31)$$

With :

$$S_5 = -C_4 \cdot [C_{23} \cdot (C_1 a_x + S_1 a_y) + S_{23} \cdot a_z] + S_4 \cdot (S_1 a_x - C_1 a_y) \quad (33)$$

$$C_5 = -S_{23} \cdot [C_1 a_x + S_1 a_y + C_{23} \cdot a_z] \quad (34)$$

$$\theta_6 = \text{atan}_2(S_6, C_6) \quad (35)$$

With :

$$S_6 = -S_4 \cdot [C_{23} \cdot (C_1 s_x + S_1 s_y) + S_{23} \cdot s_z] - C_4 \cdot (S_1 s_x - C_1 s_y) \quad (36)$$

$$C_6 = -S_4 \cdot [C_{23} \cdot (C_1 n_x + S_1 n_y) + S_{23} \cdot n_z] - C_4 \cdot (S_1 n_x - C_1 n_y) \quad (37)$$

$$C_1 = \cos(\theta_1), S_1 = \sin(\theta_1), C_4 = \cos(\theta_4),$$

$$S_4 = \sin(\theta_4) \quad (38)$$

5. INVERSE KINEMATIC MODEL

The inverse kinematics model (IKM), it positions a joint, and generates the joints Parents configuration required to achieve the desired position. Hence an inverse kinematics problem is therefore to find a robot joints configuration in the robot skeleton for positioning a hinge according to a direction and a translation defined at the beginning. The inverse kinematics model (IKM) describes the speed of operational coordinates with joint speeds (yin 2011), (Megahed 1991).

$$\dot{X} = J(q) \cdot \dot{q} = \begin{bmatrix} V_n \\ w_n \end{bmatrix} \quad (39)$$

$J(q)$: Jacobian matrix ($m \times n$), equal $\frac{\partial X}{\partial q}$;

V_n : Translation speed of "O_n". Its equal derivative of 0P_n vector;

w_n : Rotation speed of R_n [9]

For our case (robot FS03N) we must calculating the basic Jacobian matrix.

$$\dot{X} = \begin{bmatrix} V_n \\ w_n \end{bmatrix} = J_n \cdot \dot{q} = J_n \cdot \dot{O} \quad (40)$$

We noted :

$$\begin{cases} V_{k,n} = (a_k \wedge L_{k,n}) \dot{q}_k \\ w_{k,n} = a_k \cdot \dot{q}_k \end{cases}$$

$$\Rightarrow \begin{cases} V_n = \sum_{k=1}^n V_{k,n} = \sum_{k=1}^n (a_k \wedge L_{k,n}) \dot{q}_k \\ w_n = \sum_{k=1}^n w_{k,n} = \sum_{k=1}^n a_k \cdot \dot{q}_k \end{cases} \quad (41)$$

With:

K : index k^{th} joint of the robot;

$V_{(k,n)}$ and $w_{(k,n)}$: translational and rotation speed;

$L_{(k,n)}$ denotes the original vector O_k and extremity vector O_n ;

a_k : unit vector along the Z_k axis of the articulation k .

Each column of the matrix J_6^i is expressed like following:

$$j_{6,k}^i = \begin{bmatrix} -P_{6y}^k s_k^i + P_{6x}^k n_k^i \\ a_k^i \end{bmatrix} \quad (42)$$

6. ROBOT KAWASAKI (FS03N) MODEL IN SOLIDWORKS :

Using a 3D computer aided design (CAD) software allowed us to model, simulate and make the data management and processes of system. Many 3D software has been developed by Dassault Systems like Catia, ENOVIA, DELMIA, Simula, Exalead and 3DVIA, Solidworks and other.

Various functions can be realization with each different software. We opted to use solidworks since our goal is create a link with Matlab-Simulink software and then control simulate laws develop. SolidWorks is CAD software "Computer aided design". It has been

developed in 1993 and Bought in 1997 by Dassault Systèmes. SolidWorks is design automation software and in this software, you sketch ideas and experiment with different designs to create 3D models. It's used by students, designers, engineers, and other professionals to produce simple and complex parts, assemblies, and drawings (Alejandro 2007). Our robot consists of six segments and a mass attached to the terminal member. With real demotion of each segments we obtained perfect robot reproduction. The various robot segments before and after assembly are illustrated in the following figure:

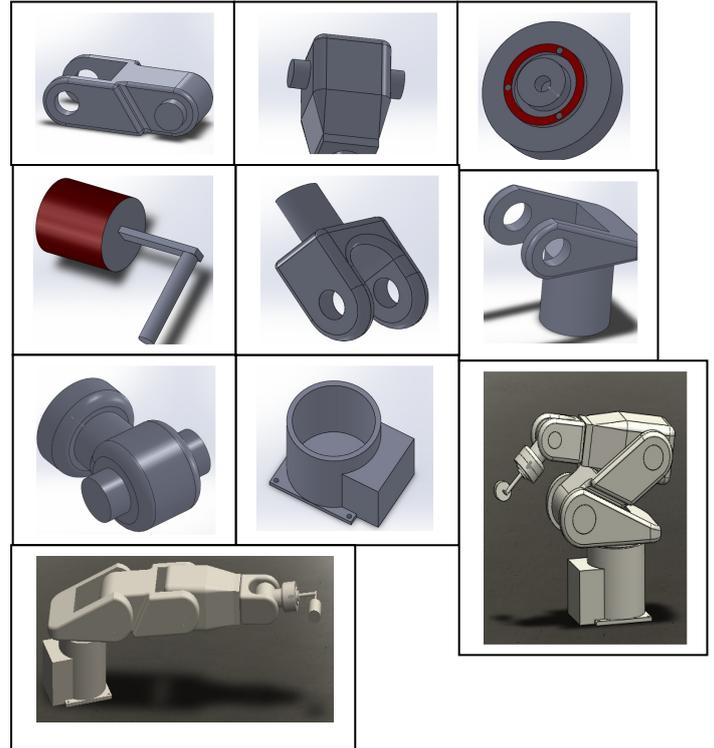


Figure 4: Various robot segments.

7. SIMULATION

3D simulation of the robot kawasakiFS03N) is constructed Malab-Simulink with simmechanics block library (Kalapyshina, 2014). The System (robot) is represented by the following blocks: the body, joints, constraints, and force. The SimMechanics block library provided us the tools to formulate and solve motion equations of complete mechanical system.

We used a bridge between solidworks_matlab with same adaptations (Simmechanics 2007), (Matlab 2010) to operate the robot model that we designed with solidworks. The Simulink modeling then appears:

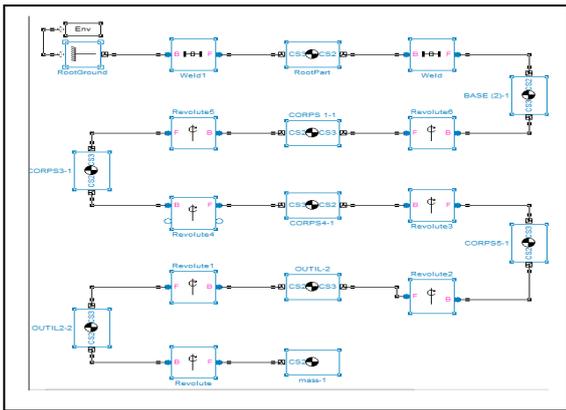


Figure 5: Block diagram of robot Kawasaki (FS03N) model.

With Ctrl+D we obtained volume robot model

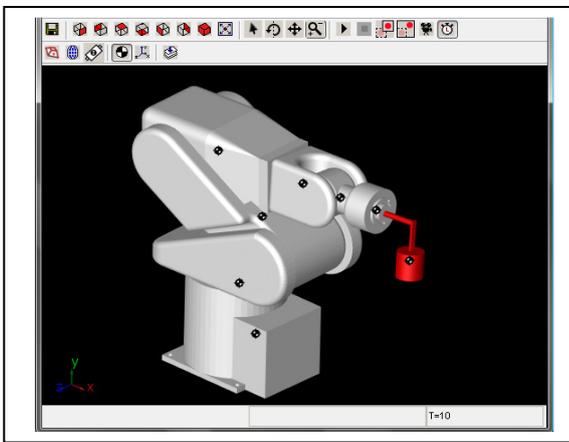


Figure 6: dynamic robot model

The results show that 3D robot model is very close to the real robot (Kawasaki FS03N).



Figure 7: Robot model in simulink and solidworks.

To simplify the simulation we have blocked all robot joints except the terminal element and after we applied a simple control signal.

A block diagram of the robot with the actuator and the sensor is illustrated in the following figure:

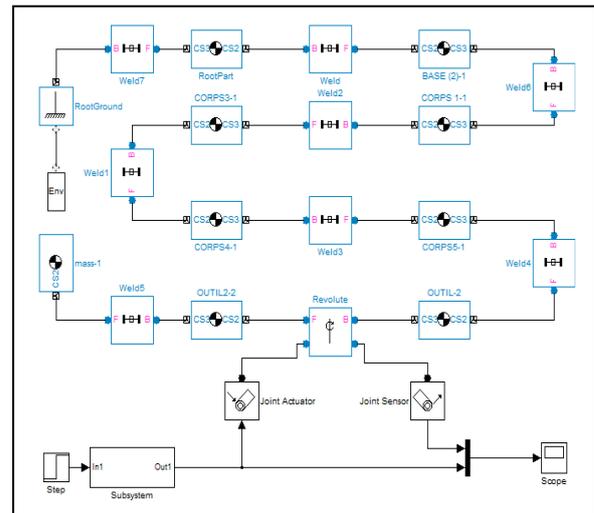


Figure 8: Control diagram block

The results obtained (Figure) illustrated the position of terminal element in four different positions.

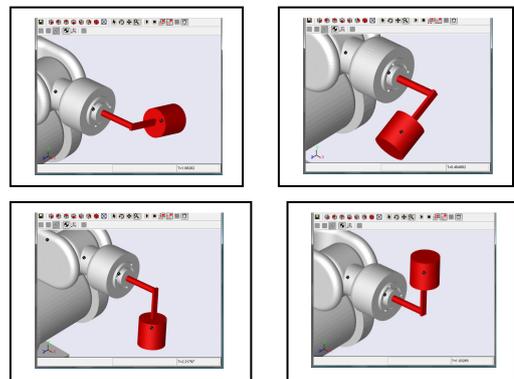


Figure 9: End effectors movement

With direct and inverse geometrical models of robot Kawasaki (FS03N) we created a 3D simulation to apply commands on the terminal element (end effectors). We chose the following simulation parameters:

- The reference mark coordinates $R_0(x_0, y_0, z_0)$.
- The operational coordinates (initial position of the end effector) (x_f, y_f, z_f) .
- The desired end effector position (x_d, y_d, z_d) .

To a reference mark:

$x_0=0$	$y_0=0$	$z_0=0$	
$x_d=100$	$x_d=200$	$x_d=100$	$x_d=100$
$y_d=100$	$y_d=200$	$y_d=100$	$y_d=100$
$z_d=0$	$z_d=0$	$z_d=0$	$z_d=100$

The results are illustrated in the following figures:

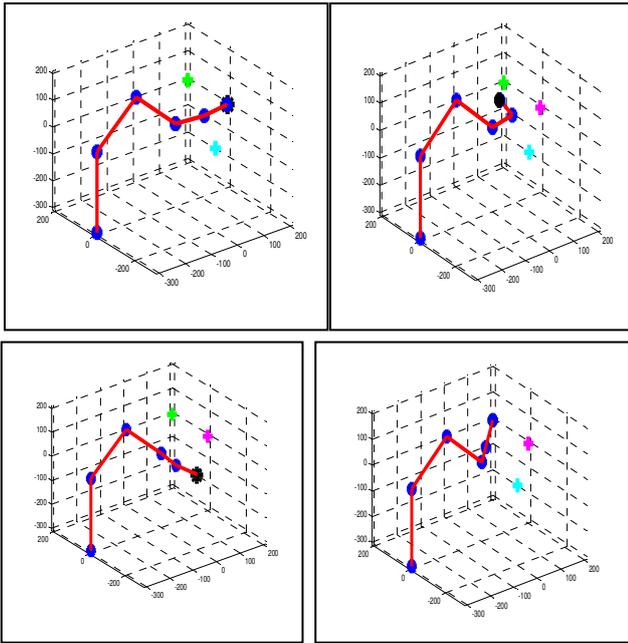


Figure 10: operational coordinates.

The following figures illustrate the comparison obtained results between Matlab model simulation and SolidWorks model software:

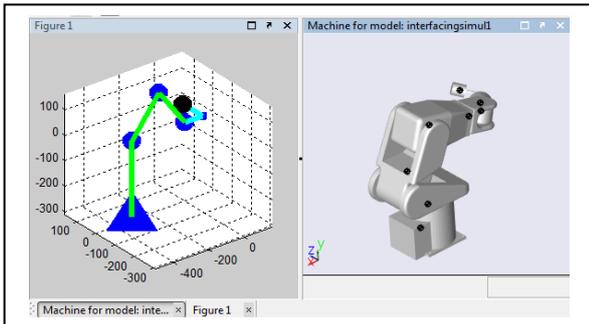


Figure 11: Matlab-Solidworks simulation with different position.

CONCLUSION

This work aimed to study and 3D simulates manipulator robot kawasaki (FS03N) and create a GUI to ease human machine interaction with control tested applied in robot. We can note that our work can be used for all simulation control work in 3D with real constraints for all industrial robots have the same type as Kawasaki FS03N. We have presented different mathematical model of the robot (forward geometric, inverse geometric, kinematics and dynamics). These models are used to simulate the robot in 3D using Matlab and SolidWorks in the same time. From the obtained results we noted that the simulations (Matlab and under solidworks) are almost identical, which will allow us to develop further control laws and improve the GUI. In

the second time compared the obtained results with experimentation results.

REFERENCES

- Yin F., Wang Y.N, Wei S.N.g, 2011., Inverse kinematic solution for robot manipulator based on electromagnetism-like and modified DFP Algorithms. Journal Acta Automatica Sinica, ScienceDirect, Vol.37, No.1, pp.74-82
- Ijeoma W., Tarig F., Al-Assadi H M A A, Mahmud I., 2012. Implementation of Industrial Robot for Painting Applications. Procedia Engineering, vol. 41, pp.1329 – 1335
- Megahed, S.M. 1991. Topological and kinematical study of tree structure robot manipulators: Symbolic computation technique. Journal Robotics and Autonomous Systems, Vol. 7, pp. 27-35.
- Tuna B., Kemal M.Ö, Sahir Arikan M.A, Murat Baykurt H., 2001. A kinematic structure-based classification and compact kinematic equations for six-dof industrial robotic manipulators. Journal Mechanism and Machine Theory, Vol.36, pp.817-832.
- Tuna B., Kemal M.Ö, Sahir Arikan M.A, Murat Baykurt H., 2000. A method of inverse kinematics solution including singular and multiple configurations for a class of robotic manipulators. Journal Mechanism and Machine Theory, Vol.35, pp.1221-1237.
- John J. Addison-Wesley C., 1989. Introduction to Robotics Mechanics and Control, 2th edition, Publishing Company, pp. 450.
- Khalil .W, Dombre .E., 2002, Modeling, Identification & Control of Robots. Hermes Penton Science, pp. 480.
- Lallemand J.-P., Zegloul, S. 1994. Robotique. Aspects fondamentaux, Masson 1994, pp. 312.
- Vibet, C. 1987. Robots. Principes et contrôle, , Ellipses, pp. 207
- Kalapyshina, I.I, Perechesova, A.A, Nuzhdin, K.A, Musalimov, V.M, Zamoruev, G.B., 2014. Modeling of Mechatronic Systems in Matlab (Simulink / Simmechanics). Proceedings of the 26th European Modeling and Simulation Symposium, 10-12 September, Bordeaux, France
- SimMechanics2, 2007. User's Guide MATLAB & SIMULIK, The MathWorks, September 2007 Online only Revised for Version 2.7 (Release 2007b)
- MATLAB®, 2010. Creating Graphical User Interfaces, September 2010 Online Only Revised for MATLAB 7.11 (Release 2010b).
- Kawasaki robot Manual, 2003. Kawasaki Heavy Industries, Ltd.90206-1034DEF. pp.371
- Alejandro R., 2007. Beginner's Guide to SolidWorks 2007. Schroff Development Corporation. pp.46
- Paul R-P., 1981. Robot Manipulators: Mathematics, Programming and Control. M.I.T Press, ISBN: 026216082X, pp. 292.

DESIGN AND DEVELOPMENT OF A MESOSCOPIC SIMULATOR SPECIALIZED IN INVESTIGATING CAPACITIES OF RAILWAY NODES

Roman Diviš^(a), Antonín Kavička^(b)

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

^(a)roman.divis@student.upce.cz, ^(b)antonin.kavicka@upce.cz

ABSTRACT

To investigate capacities of railway stations and junctions, analytic methods or methods of computer simulations can be used. Analytic methods are slowly getting outdated. Classic microscopic simulations offer very good results but the life cycle of simulation studies is relatively long. Railway companies nowadays require a lot of traffic simulations, which are supposed to be carried out in shorter periods of time than are typically related to microscopic simulations. A simulation tool *MesoRail*, now in development, works on a mesoscopic level that abstracts away from some details affected by the microscopic simulations. An appropriate choice of the level of abstraction enables to achieve still very good results. At the same time, the tool simplifies the stage of building a simulation model. *MesoRail* tool is mainly specialized in investigating the capacities of railway stations and junctions, but it can also be used for studying other traffic problems concerning railway transport.

Keywords: mesoscopic traffic simulation, railway station, capacity

1. CAPACITY OF RAILWAY NODES

Capacities of railway nodes can be, in principle, investigated using two approaches: (i) analytical methods or (ii) methods based on computer simulation. Analytical methods utilize the theory of probability and mathematical statistics in order to establish formulas thanks to which it is possible to analytically calculate the throughput of individual railway nodes and their parts. Let us mention two analytical methods for determining capacities of railway infrastructures: (i) directive *SŽDC D24 – The regulation for determining capacities of railway lines* applied by the state company *The Railway Infrastructure Administration* in the Czech republic (SŽDC 2009) and (ii) *UIC Code 406 – Capacity* developed by *UIC - International Union of Railways* (UIC 406) (ETF 2013).

1.1. Analytical methods

Analytical methods provide a result in the form of an overall indicator - the *degree of occupancy* of individual parts related to a railway infrastructure. On the other hand, simulation methods enable to observe variety of operating characteristics during the simulation - e.g.

values of delays of individual trains within particular parts of a railway station, real occupancy of individual parts of railway segments, etc. Deploying analytical methods is getting outdated nowadays, their utilization is relatively fast, but the informative value of their results is significantly lower in comparison with the outcomes from simulation methods.

1.2. Methods of computer simulation

Simulation methods enable to test thoroughly the throughput of stations under various operational conditions. Nowadays, there is a variety of high quality tools for microscopic simulations specialized in railway traffic, which can support investigations connected with capacities of railway infrastructures. As examples can be mentioned tool *Villon* (Simcon 2014), *OpenTrack* (OpenTrack Railway Technology Ltd. 2015) and *RailSys* (RMCon 2015). Applying computer simulation, it is possible to test the behaviour of the investigated railway node according to various operational scenarios. Thus, it is possible to test normal traffic without delays, or to test the sensitivity of a train schedule for various combinations of delay values from several directions etc.

2. SIMULATOR OVERVIEW

The next chapter provides some examples of simulators, which specialize in rail transport simulation, mainly the microscopic simulation tools.

2.1. Microscopic simulation

Typical representatives of microscopic simulation tools include *Villon* (Simcon 2014). Infrastructure is prepared in scale to reality and it contains very detailed data about individual rail segments. The actual simulation tool itself further requires defining characteristics of individual wagons, locomotives and trains created from them. Controlling a train in the railyard is defined by a network graph which consists of individual activities the train gradually performs. Technological activities, train paths, and other elements are easily parameterized. Thus, a wide range of train control during the simulation is provided, but the preparation of the simulation is difficult and lengthy.

A tool similar to *Villon* is the simulation tool *RailSys* (RMCon 2015). It realizes a very detailed microscopic

simulation. Although the railway infrastructure is depicted in a scheme, it still contains all the necessary underlying data. *RailSys* is well applicable for extensive simulation studies, which can automatically detect any capacity problems in individual sections of the rail network.

Another used simulation tool is *OpenTrack* (OpenTrack Railway Technology Ltd. 2015). Like *RailSys* it uses infrastructure defined in a scheme while preserving real underlying data. Train control is realized by defining a train schedule, which includes train stops, and other various parameters. The same solution is applied to train joining and train splitting, and waiting for other trains.

All the above mentioned tools support online animation of occupying the rail network by trains or the movement of trains, and they also support generating post-simulation protocols. Commonly generated outputs include tachograms of trains, occupation of individual track sections, overall statistics, and more.

2.2. Mesoscopic simulation

The disadvantage of using a computer simulation is that building a quality microscopic simulation model and conducting simulation experiments is a rather time consuming process. In addition, railway companies nowadays require that the infrastructure capacity investigations take shorter time than the current approach which uses microscopic simulations. Using the mesoscopic approach to simulation provides the ability to abstract away from several details and thus to shorten the life cycle of simulation studies and in result satisfy the needs of railway companies.

The results of mesoscopic simulations cannot provide the same level of accuracy as the outcomes of microscopic simulations, but by choosing an appropriate level of abstraction it is possible to achieve satisfactory results for practical usage.

Nowadays, models of mesoscopic transport systems are already fairly well adapted and a number of tools to simulate road traffic also exists. A summary of simulation tools and different mesoscopic techniques can be found in, for example, article by Xu, Song, Weng and Tan (2014).

In the field of railway transport systems, the usage of mesoscopic simulations is currently mainly in its investigation phase. Stallybrass in his article describes the use of mesoscopic simulations in determining the optimum interims for preparing train graphics (Stallybrass 2014). Mesoscopic simulation of freight trains using the event-oriented simulation tool *SIMUL8* is described by the authors Marinov and Viegas (2010). Due to the lack of availability of a suitable tool which would be appropriate for determining the throughput of railway stations, we decided to develop our own simulation tool called *MesoRail*.

3. THE MESORAIL TOOL

MesoRail is a mesoscopic simulator of train traffic with support of rapid prototyping of railway infrastructure and fast specifications of traffic scenarios. *MesoRail* is being developed in the programming language of Java. Mesoscopic level of simulation enables to model important aspects on the same level as in microscopic simulations, and at the same time it allows to apply higher level of abstraction to some aspects. Such approach enables to achieve reliable results from simulation experiments without any significant loss of their informative value.

The main problem is to find a suitable compromise and the degree of abstraction for individual parts of the simulator in order to achieve the desired result. To determine the appropriate abstractions, consultations with railway experts were used. The following abstractions are implemented within the *MesoRail* simulator (i) schematic depiction of the railway infrastructure, (ii) simplification of train dynamics calculations, (iii) mobile human resources (working groups) are not considered, (iv) service technological procedures provided to the trains are not considered.

The simulation tool *MesoRail* will find its application mostly in determining the throughput of railway nodes, therefore it is essential to meet the following requirements - (i) rapid infrastructure prototyping, (ii) real train ride characteristics, (iii) respecting the functionality of track signal and interlocking systems, (iv) considering station and railway line intervals, (v) simulating deterministic and stochastic train flows, (vi) animation outputs during the simulation experiment, (vii) post-simulation statistics and graphic protocols. The provided list represents only the basic requirements expected from the simulator. Based on this list it is necessary to establish the levels of abstraction for individual components.

3.1. Phases of development

Developing such a complex simulation tool is not a trivial matter and it takes quite a long time. Individual development steps are separated into stages and development takes place gradually over specified stages.

At the beginning of development, it was necessary to carefully specify the desired properties of the simulator and to determine the appropriate abstractions that move the microscopic simulation to the mesoscopic level on that basis. At this stage, consultations with railway experts with considerable experience with microscopic simulation of rail transport were found very useful.

After completing the basic specification of requirements and abstractions, it was necessary to design the architecture of the simulator – to select the method of simulation, used libraries, file formats, and to specify concrete requirements for train entities and railway infrastructure.

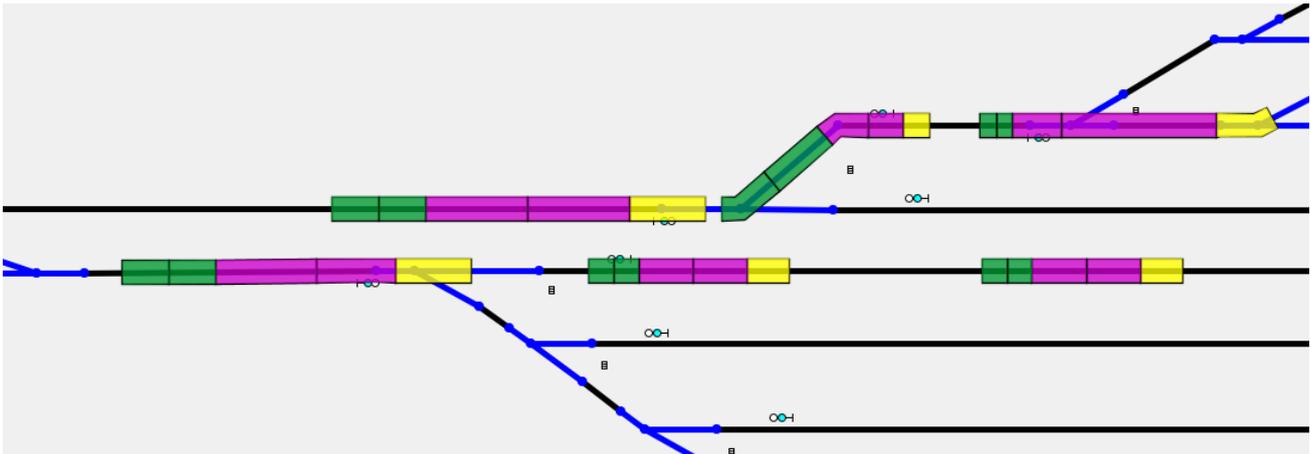


Figure 1: Train Deformation Example (single train placed and rendered in different parts of railway infrastructure)

In the following stages, specification of requirements for the *MesoRail* simulator were completed and the gradual implementation of various components of the simulator was scheduled. Individual partial results of simulations were verified and validated to ensure correct operation of the simulator and its credibility.

At the present stage of development we focus on building complex functionalities. It is already possible to perform simulations of simple railway nodes and obtain results for evaluation. However, it is necessary to further establish support for intelligent behaviour of trains in complex railway nodes.

3.2. Railway infrastructure

While simulating railway traffic, the infrastructure the trains move on plays a very important part. To prepare the railway infrastructure for a microscopic simulator is a relatively lengthy process. Creating an infrastructure and defining all parameters (speed restrictions within the infrastructure, vertical alignment, location of light signals for traffic directions, etc.) takes a considerable amount of time. Sometimes, the necessity to prepare an infrastructure in a particular scale can be a problem, as input data can be missing or containing faults.

For these reasons, a concept of schematic depiction of rail infrastructure with the support of rapid prototyping of infrastructural elements is proposed. Switches, railway crossings, and other more complex track elements are prepared as prototypes that can easily be parameterized and inserted into the infrastructure. That way it is possible to quickly build an infrastructure that can mirror real environment in detail, or it is possible to omit some parameters and leave standardized default values if needed.

The created infrastructure maintains the precision in the matter of meters. Unlike the microscopic simulations it is possible to prepare such infrastructure in substantially shorter time. For more details about

infrastructure and its editor *TrackEd*, see the article by Novotný.

3.3. Train driving dynamics

Train moving on a track infrastructure corresponds to real riding properties, i.e. train and track railway characteristics are respected (Iwnicki 2006, Diviš 2014). However, in the *MesoRail* tool it is possible to abstract away from some parameters as their influence on the train or on its speed is insignificant. Such parameters are wind resistance in a tunnel or effects of transition curve on the train ride. Defining such parameters would prolong creating a simulation model in comparison with the insignificant effect on the riding time of trains.

The train driving dynamics is described by a differential equation (Diviš 2014), which contains parameters mentioned in the paragraph above. Calculation of the equation is performed by numerical integration using the Euler method with an intermediate step of calculation for greater accuracy. Since it is a numerical integration and the characteristics of tracks are different with the changing position of the train, it is necessary to calculate dynamics repeatedly according to a defined time frame.

Driving dynamics of trains is closer to the microscopic approach in the *MesoRail* tool. Accuracy and speed of the calculation can also be affected by the choice of the time frame in which the train driving dynamics is recalculated.

3.4. Railway traffic

The simulated train flows can be either deterministic or stochastic. The trains can then follow an estimated flow diagram of train traffic or they can be burdened by various delays.

Delays are specified according to appropriate methodologies of railway companies. Simulator's

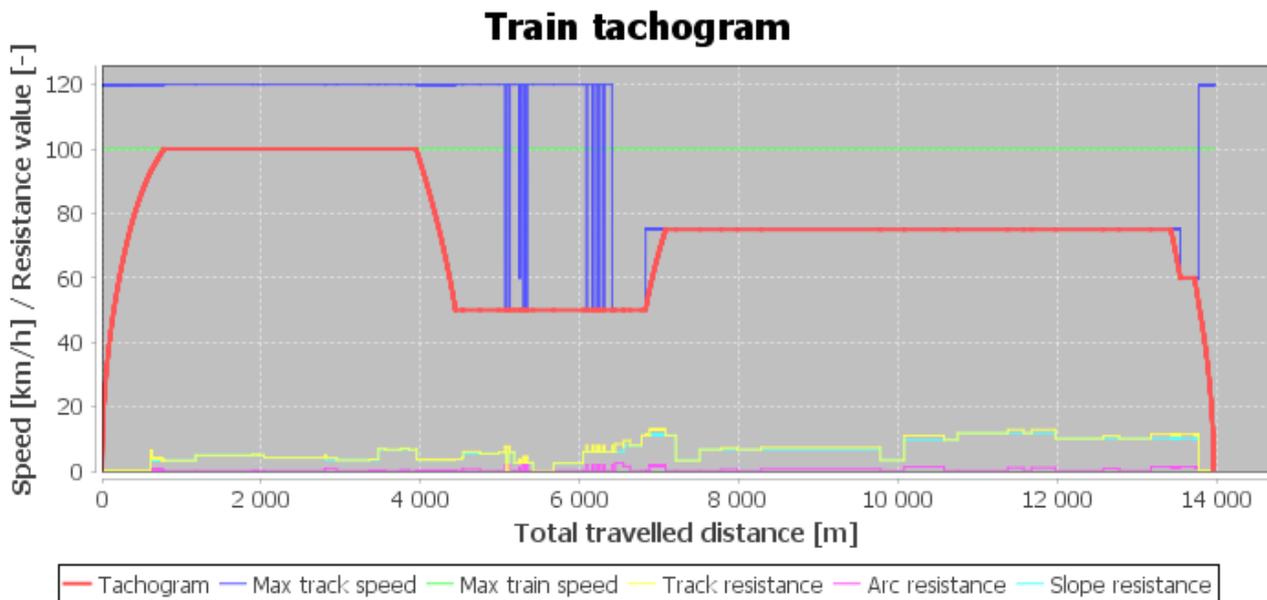


Figure 2: Train Tachogram Example

decision-making algorithms enable, for example overtaking for priority trains or choosing another alternative without user interference in case of an occupied track. All train movements are carried out in compliance with deployed signal and interlocking systems in such a way, in which the simulated scenarios mirror real railway traffic.

3.5. Animation and simulation protocols

MesoRail features online animation for easy and clear overview of the progress of a simulation. The simulator is capable of setting the speed of animation based on the user's needs or to pause the animation if a detailed inspection of the simulation status is needed.

Considering the schematic depiction of the infrastructure, deformed depiction of trains can occur. Depending on an accurate length of a track segment in scale to the depicted length, the length of a train or the speed can change, even though both values are constant. However, the simulation correctly observes the location and the speed of a relevant train and the depiction deformations are an expected consequence of the schematic infrastructure layout. Example of this behaviour could be seen at Figure 1.

After completing the simulation, post simulation protocols can be displayed. From these protocols, data can be filtered and further processed into, for example, tachograms of trains (see Figure 2).

3.6. Simulation core

The simulation engine of *MesoRail* is based on the method of discrete events planning. This method uses a calendar of events, from which events are removed in the correct chronological order and executed. In classical simulation using discrete event method, any entity, that has been included in processing the previous event, can create an event and include it in the calendar.

There is no hierarchy or delegation of competences and despite its simplicity, in complex simulators there can occur a situation in which the simulation will be messy and every entity will plan a considerable amount of its own events.

In contrast, agent-oriented architectures, such as *ABAsim*, divide competences and responsibilities between agents and allows hierarchic classification of agents (Kavička, Klima and Adamko 2005, Adamko 2009). Agent-oriented architecture has become the inspiration for the design of the *MesoRail* simulator. Although it uses a simple method of event planning, the concept maintains simulator architecture of *ABAsim* along with the Java programming language and object oriented programming features. Classes represent different types of agents. Communication within each part of the agent is performed by the class attributes (also serves to maintain the agent's state) and by method calling. Communication among various agents is also carried out by public method calling. Agents can also add an event to the calendar of events and cause a shift in simulation time.

In the actual simulation, trains represent dynamic mobile agents that move along the track. Their movement is controlled by a central dispatcher (managing agent), which controls all traffic on the railway infrastructure.

The concept of a simulation based on the use of autonomous agents allows good division of responsibilities and competences of the individual agents and thus creates a transparent environment, in which it is easier to navigate, and provides an easy possibility to expand for future development simulator.

3.7. Verification and validation

In order for the simulator to be used in practice, it is necessary for simulation results to be correct and real. Since it is a mesoscopic simulator, the results will not

reach the same level of accuracy as other simulator using a microscopic approach. Nevertheless, it is necessary to ensure correct functionality of individual parts of the simulator.

In principle, it is possible to identify several basic parts of the simulator which need to be verified, or rather validated:

- functionality of basic data structures,
- train driving dynamics,
- correctness of the decision-making algorithms governing the operation of trains (compliance with track and station intervals, obeying speed restrictions on the line, the correct function of safety devices on railways ...).

To verify the functionality and proper behaviour of data structures, it is useful to use unit testing using package *JUnit* in *Java*. The tests themselves, however, do not provide assurance of correct behaviour, tests must then be written by a developer or a tester, and it is necessary to include as many possible situations so that the behaviour of data structures is tested thoroughly. In essence, the quality of the written test to test critical points of implementation matters. Unit tests are used to test data structures such as graph implementation, basic functionality of methods for discrete event planning, and other parts in *MesoRail*.

Proper implementation of the train driving dynamics is a very important part of the simulator. To be able to validate the driving dynamics, it is necessary to first perform the simulation to obtain information about train movement in the form of a tachogram. Next, it is possible to compare it with a similar tachogram obtained from a real environment or other simulation tool. In case of the *MesoRail* simulator, validation against real data provided by *ŠZDC* (Diviš 2014) was used in the initial phase. Other, more complex experiments evaluating the effect of a gradient, curves on the track during acceleration or braking were compared with results from the simulator *Villon* (representing a validated simulation tool widely accepted by experts from railway practice).

Rail traffic control is also an essential element on which the credibility of the simulator depends. For simpler problems related to traffic management, it is possible to ask railway experts who examine the simulation results in detail and assess the accuracy of the simulation. In the case of a complex scenario, it is possible to use an alternative simulation tool, in which the same simulation will be created and then their progression can be compared. If stochastic events occur in the simulation, it is possible to compare the various statistical indicators. With ensuring a sufficient number of replications, the results should be comparable.

4. CASE STUDY

As part of the initial testing of the *MesoRail* simulator, an infrastructure, which consists of a prototype station and several adjacent sections of tracks, was used. The

total length of tracks in the station and the adjacent track sections is almost 20 km and the track includes a series of arches and various gradients. This infrastructure has been implemented in the *Villon* and *MesoRail* tools. Next, a series of simulation experiments was conducted. Types of trains used were - (i) a passenger train, (ii) an express train, (iii) and a freight train.

After the experiments were performed, statistical evaluations of the results were done and the differences in driving characteristics were compared. Sections where trains accelerated and braked were compared separately. Table 1 summarizes the average deviations of the measured distances that the trains had to travel to reach the prescribed speed in comparison with the *Villon* tool.

Table 1: Average Differences in Measured Distance During Deterministic Experiments

Type	Percent difference	Absolute difference
Acceleration	2,01 ± 0,01 %	26,33 ± 14,08 m
Braking	5,56 ± 0,03 %	33,44 ± 16,22 m

The difference of measured values between the *MesoRail* tool and the *Villon* tool is in the matter of units of percent and it is expected due to the mesoscopic simulation level. Braking, however, shows higher difference than the acceleration of trains. The reason for this difference will be the subject of further examination during the development of the simulator.

For investigating the simulation, support for time shift was implemented in the simulator. The core allows to automatically save the state space of the simulator, or it can store it under certain conditions or upon user request. During the course of the simulation and after its completion, it is possible to seamlessly move between states and to view the previous states, and thus find out what conditions led to the particular behaviour of the simulation.

Individual states of simulations are stored in memory. However, the complete status information (including all the variables at a given point in time) is not stored, but only differences from the last recorded state are created. This method saves memory but slows down the saving of state itself (it is necessary to calculate the difference), and in particular, it slows down the recovery of state to the state space of the simulator. Restoring a complete state from partial differences must proceed in an iterative way back into the past, until complete status information is not created.

Although storing states of the simulator leads to increased overhead of the simulator and also to an increase in requirements on the amount of RAM, this functionality enables detailed analysis of the simulation and identification of potential critical points, where a bad decision due to an error in the code or configuration of simulator occurred. Support for the time shift is being further developed and alternative storage options for states for comparing demand on memory, disk space

and processor utilization during saving and restoration of states are tested.

5. CONCLUSION

The need for faster examination of railway station throughput has led to the design of a new mesoscopic simulation tool, *MesoRail*. The difference between a classic microscopic simulator and the *MesoRail* tool is that *MesoRail* uses several different abstractions that enable faster execution of simulation projects at the cost of lower accuracy.

In the article, the *MesoRail* tool was introduced, and at the end a case study with comparison of results against the microscopic simulator *Villon* was demonstrated. The achieved results correspond with the estimated accuracy of the simulation, but parts that need to undergo further examination and possible repairs were also discovered.

Now, the *MesoRail* tool is able to perform simulations of simple railway nodes. Currently, the development is focused on the implementation of advanced decision-making algorithms and support for complex node modelling.

REFERENCES

- Adamko N., 2009. Enhancing discrete simulation executive with simple continuous simulation and animation support. Proceedings of 23rd European Simulation and Modelling Conference, ESM 2009, p. 35-42. October 26-28, Leicester, United Kingdom.
- Diviš R., 2014. Train movement dynamics within AnyLogic tool. Proceedings of the European Modeling and Simulation Symposium, 307-312. September 10-12, Bordeaux, France.
- ETF, 2013. Capacity leaflet (UIC 406). Available from: http://www.uic.org/etf/codex/codex-detail.php?langue_fiche=E&codeFiche=406 [accessed 6 April 2015]
- Iwnicki, S., 2006. Handbook of railway vehicle dynamics. Boca Raton: CRC/Taylor.
- Kavička A., Klima V., Adamko N., 2005. Agentovo orientovaná simulácia dopravných uzlov (Agent-based simulation of transportation nodes, in Slovak language). Žilina: Žilinská univerzita.
- Marinov M., Viegas J., 2010. A mesoscopic simulation modelling methodology for analyzing and evaluating freight train operations in a rail network. Simulation Modelling Practice And Theory Vol. 19: p. 516-539
- OpenTrack Railway Technology Ltd., 2015. Railway Simulation. Available from: <http://www.opentrack.ch/> [accessed 26 June 2015]
- RMCon, 2015. RailSys. Available from: <http://www.rmcon.de/index.php?page=railsys-classic-planning> [accessed 6 April 2015]
- Simcon, 2014. Villon. Available from: <http://www.simcon.sk/en/tools/villon> [accessed 6 April 2015]
- Stallybrass M., 2014. Mesoscopic simulation: a tool for analysing and optimising headways and capacity. Proceedings of AusRAIL 2014, 15 p. 11 - 12 November 2014, Perth.
- SŽDC, 2009. D24 – The regulation for determining capacities of railway lines. Available from: <http://www.szdc.cz/dalsi-informace/dokumenty-a-predpisy/provozne-technicke.html?page=detail&docid=1%3B%237eacfd179-1551-468f-949a-641e9600c22d> [accessed 6 April 2015]
- Xu Y., Song X., Weng Z., Tan G., 2014. An Entry Time-based Supply Framework (ETSF) for mesoscopic traffic simulations. Simulation Modelling Practice and Theory Vol. 47.: p. 182-195

ONTOLOGY AND ITS ROLE IN DOMAIN-SPECIFIC MODELING

ZHU Ning^(a), LEI Yonglin^(b), LI Xiaobo^(c), ZHU Yifan^(d)

College of Information System and Management, National University of Defense Technology, Changsha, Hunan, P.R. China, 410073

^(a)zhun87025@126.com, ^(b)samuelyonglin@gmail.com, ^(c)lixiaobo@nudt.edu.cn, ^(d)nudtzyf@163.com

ABSTRACT

Domain-Specific Modeling (DSM) aims to support domain experts building models with familiar concepts by improving the abstract level of modeling languages. And it is a trend in the Model Driven Engineering (MDE) domain. However, there are still a series of problems left to be solved well for DSM, such as how to design a domain-specific language adhering to the domain nicely and how to verify the domain model during building models. An ontology is a conceptual model of a domain with strict and explicit semantics consisting of domain concepts and their relationships. It seems to be a promising way to utilizing ontology to boost DSM up. In this paper, what's an ontology and its metamodel is discussed firstly. Then the DSM method is introduced and its problems are proposed. Finally the different ways of applying ontologies in DSM to solve its problems are summarized.

Keywords: domain-specific modeling, domain concept, ontology

1. INTRODUCTION

Domain-Specific Modeling (DSM) supports domain experts to build models rapidly with familiar domain concepts by improving the abstract level of modeling languages. And the models are more comprehensible and easier to communicate with others. What's more, DSM can enhance software production with tools to translate the domain models to target productions (such as code, executable programs) (Tolvanen and Rossi 2003).

DSM is a trend in Model Driven Engineering domain. It aggregates advantages of Model-Integrated Computing (MIC) and Model Driven Architecture (MDA), two important branches in MDE. MIC was proposed by Janos Sztipanovits and Garbor Karsai from Vanderbilt University in 1997 (Sztipanovits and Karsai 1997). It utilizes the metamodeling environment to create metamodels of a specific domain and configures a domain-specific modeling environment for the metamodel to support domain modelers to create domain models. Object Management Group put forward MDA in 2001 to solve the integration problem of software architecture and various software tools (Kleppe, Warmer, and Bast 2006). MDA decomposes

the whole development process into three models with different abstract levels, respectively Platform Independent Model (PIM), Platform Specific Model (PSM) and executable codes with transformation standards among different models. DSM absorbs ideas of MIC and utilizes technique and standards provided by MDA to support modeling domain-friendly.

However, there are still a series of problems left to be solved for DSM such as how to design a domain-specific language adhering to the domain nicely, how to reduce the learning curve for domain languages learners, how to verify and debug a domain model and so on. An ontology is a conceptual model of a domain with strict and explicit semantics consisting of domain concepts and their relationships. It seems to be a promising way to utilizing ontologies to solve the problems in DSM.

The paper is organized as the following. In section 2 what's an ontology and its metamodel is discussed and the pervasive used ontology language Web Ontology Language (OWL) and the corresponding tool Protégé OWL is introduced simply. Then the method of DSM is concluded and its problems are proposed in section 3. In section 4, the different ways of applying ontologies in DSM to solve its problems are summarized. Finally, we conclude the paper and talk about how we will use ontologies and DSM in our next research.

2. ONTOLOGY

2.1. What's an ontology

Ontology is initially a philosophical concept. It was proposed in the 17th century and accepted widely in the philosophy circle in the 18th century. And ontology became a branch of metaphysics in philosophy domain and considered the nature of existence and relationship. According to Aristotle and Peirce, it is the science of being qua being and aims to find the most general features of reality and real objects. It does not focus on specific disciplines such as physics or chemistry, but the transcendent rules over all the entities in the different domains. (Guizzardi 2005)

In 1967, ontology was introduced in Computer Science, and then it is pervasively used in the following several decades in domains of Database & Information System, Software Engineering, especially in Domain Engineering and Artificial Intelligence. When ontology

in philosophy is immigrated into Computer Science, its concept is localized according to the application domains and experience by different researchers:

- An ontology defines the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary. (Neches, Fikes, and Finin 1991)
- An ontology is an explicit specification of a conceptualization. (Gruber 1993)
- An ontology is an explicit, partial account of a conceptualization/the intended model of a logical language. (Giaretta and Guarino 1995)
- An ontology is a formal specification of a shared conceptualization (Borst 1997).
- An ontology is a set of structured terms that describing some domain or topic and provides a skeletal structure for a knowledge base. (Swartout, Patil, and Knight 1997)
- An ontology is the vocabulary related to a generic domain (like medicine, or automobiles) or a generic task or activity (like diagnosing or selling). (Guarino 1998)

Though many definition of ontology exist, there's no consensus on what's an ontology. However, examining these concepts, there are some common foundations of an ontology according to Fensel (Fensel, Harmelen, and Horrocks 2001):

- Conceptualization. An ontology is a conceptual model for reality and real object.
- Explicit. Concepts and their relationships are defined using precise semantic.
- Formal. An ontology should be machine understandable.
- Shared. Knowledge represented by an ontology is recognized commonly by the users in the same domain.

So considering about all the definitions of ontology and the research problem we are facing (Applying DSM in military modeling and simulation), we see an ontology as a conceptual model of reality which including concepts, their relationships and common entities of these concepts in some domain. And the conceptual model is described in exact and formal semantics.

2.2. Ontology metamodel

Colin Atkinson and Thomas Kuhne proposed an ontology metamodel using the biological taxonomy as an example (Atkinson and Kuhne 2003). They separated the biological ontology language into four levels, respectively O3, O2, O1 and O0. In the O2 level are the metaconcepts such as Kingdom, Phylum, Class and so on. In the O1 level are the concepts (Animal, Chordate, Mammal, et al.) instantiated from the metaconcepts while the instances are in the O0 level. However, what's in O3 level was not explained clearly and how these four levels correspond to the four-level metamodel hierarchy was not discussed, too.

OMG provided Ontology Definition Metamodel in the four-level metamodel hierarchy (OMG, 2014). In the

M3 level is the MOF, and the structure of the ontology language is defined in the M2 level. The ontology is an instance of the M2 level and dwells in the M1 level. The M2 level consists of the instances of the ontology. But The O2 level which composes of the metaconcepts in Atkinson and Kuhne's ontology language is not considered about.

Referring to Ontology Definition Metamodel provided by OMG and ontology metamodel proposed by Colin Atkinson and Thomas Kuhne, the ontology metamodel hierarchy combining them is shown in Figure 1.

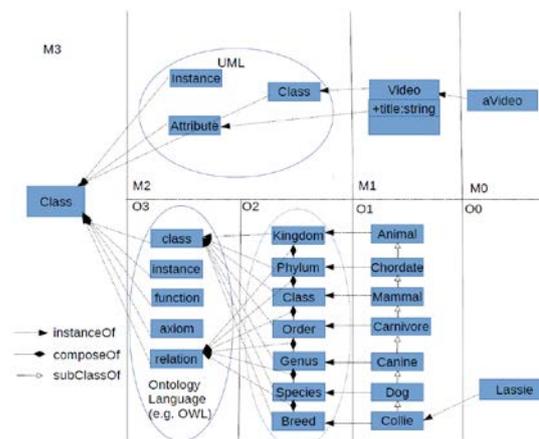


Figure 1: The ontology metamodel hierarchy

In Figure 1, the M2 level of the four-level metamodel hierarchy is divided in two levels respectively O3 and O2 levels. The O3 level is the instance of the M1 level and corresponds to the M2 level of Ontology Definition Metamodel. The O2 level is optional for the domain consists of metaconcepts. The other levels are the same as the Ontology Definition Metamodel.

2.3. OWL and Protégé

Currently, OWL is a pervasive ontology language and is adapted by World Wide Web Consortium (W3C) as a standard for ontology modeling. The foundation elements of OWL are Class, Individual and Property. OWL consists of three types of languages, respectively OWL-Lite, OWL-DL and OWL-Full. OWL-Lite is the least subset of OWL, which can be used to describe the simple class hierarchy and constraints. OWL-DL involves Description Logics and can support automated reasoning to verify consistence of the concepts and the class hierarchy. OWL-Full is the most expressive subset of OWL, but doesn't enable automated reasoning. The OWL ontology is stored based on RDF/XML. (Horridge 2011, Li 2013)

Protégé is an open-source ontology modeling tool developed by Stanford University in Java for ontology modeling and reasoning. It coheres to OWL and is composed of four main views. In the class view, the concepts and their hierarchical relationships can be modeled. Properties of a concept should be described in the Data Property view, while relationships of concepts in the Object Property view. The Individual view composes of all the instances of the concepts. Protégé

supports to be extended by plugins like Eclipse. So quite a lot of excellent customized views are developed by researchers to describe, visualizes or analyze the ontology model. Reasoners such as Racer, FaCT++ and Hermit are integrated into Protégé as plugins to support checking and reasoning of ontology model. (Li 2013)

3. DOMAIN-SPECIFIC MODELING

3.1. The DSM Method

The process in the DSM method is shown in Figure 2. It consists of five different steps, respectively domain analysis, design of domain-specific modeling language (DSML) and domain-specific modeling environment (DSME), domain model analysis and target product generation.

Domain analysis is a branch of Domain Engineering. It analyzes the domain to collect domain information and constructs domain conceptual models which describe various entities, their properties, roles, relationships and constraints in the domain (Reihartz-Berger and Sturm 2004). So domain analysis brings out the domain concepts, relationships and constraints and is the foundation of DSML design.

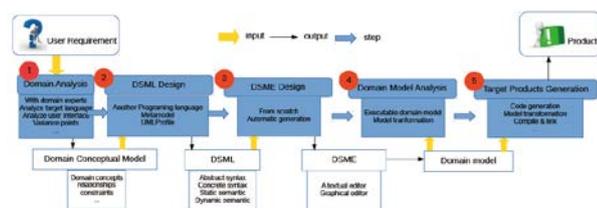


Figure 2: The process in the method of DSM and its products

DSML is the core of the DSM method. Generally, a DSML composes of abstract syntax, concrete syntax, static semantic and dynamic semantic (Cho 2011). Abstract syntax describes concepts, relationships and constraints of a language in a formal way. Concrete syntax corresponds to the representation of the models, such as a word in textual format or an icon in graphical format. Static semantic represents the static constraint relationships between concepts, while dynamic semantic shows the meaning of the components when the model executes. There is a type of DSML which is based on the common programming languages (such as Yacc, SIMULA67) and can be designed as another programming language. However, the metamodeling and UML profile are the mainstream methods to construct a visual domain-specific modeling language. DSME is related to the concrete syntaxes and static constraints of the DSML. Concrete syntaxes are correspondent to the visual icons or graphics representing the models or the highlighted text. Static constraints correspond to the limits in using the domain-specific modeling environment such as the confinement on linking one modeling element to another. DSME can be developed from scratch according to the concrete syntaxes and static semantics. But the pervasive used method to construct DSME is to generate DSME from

them automatically (Rath and Varro, Zbib, Jain and Bassu 2006). Some mainstream DSM tools such as MetaEdit+, EMP, and GME all support this method.

Domain experts build models using DSME. These models can be used to communicate with others, and also can be analyzed. Currently, there are two methods to analyze the domain model. Firstly, the corresponding executable semantics can be embedded in the DSML to make the domain model executable. Secondly, the domain models can be transformed to another modeling language or formalism to be analyzed according to the analysis ability of the target language or formalism.

The most valuable of DSM is to support domain models to generate target products, automatically transforming the design into realization.

3.2. The problems of DSM

DSM is a good way to support rapid model design, development and product implementation. But it is not perfect. There are still a series of problem left waiting to be solved well (Walter, Parreiras and Staab 2014). The problems of DSM include:

- How to design a DSML adhering to a domain well. Currently the emphasis of most DSM researches is on the middle part of the whole DSM method, while few of them consider on how to design the components of the DSML to match the domain well and if there is a better DSML for the same domain.
- How to capture formal constraints for DSML Design. The constraints of DSML in a domain are embedded in the domain knowledge, the current DSM method lacks of a mechanism to help DSML designer to extract the constraints in the domain.
- Lacking a tool for connecting domain analysis with DSML design. Though current pervasive DSM tools provide environments to build metamodels of a domain, there is not a tool to help or guide users to capture the language constructs.
- How to help domain experts to learn and use the DSML and DSME quickly. For domain modelers who are new to the DSML, suggestions should be provided to guide the use of the domain-specific modeling languages.
- How to verify and debug a domain model. Most of current DSM tools are devoid of the ability to check the domain models' correctness and completeness. What's more, few of them support domain modelers to debug the models to find the bugs.

4. ONTOLOGY'S ROLE IN DSM

Ontology is a conceptual model of reality and real objects. It consists of concepts and their relationships, constraints, some common entities in the domain, which can be used to solve the problems in the DSM proposed in the previous section.

The roles of an ontology in the DSM method are sorted according to the different steps of the DSM method, which is shown in Table 1.

In the domain analysis step, an ontology can be used as a domain conceptual model to represent the domain as the result of domain analysis. And it has been used widely in domain engineering (Tairas, Mernik, and Gray 2009).

In the DSML design phase, an ontology can guide the design of DSML. If the ontology is modeled in a formal language such as OWL, the transformation from ontology model to the metamodel in the M2 level of the four-level metamodel hierarchy can be realized to generate DSML automatically. What's more, different DSMLs for the same domain can be compared for their appropriateness based on the ontology (Guizzardi 2005).

In the DSME design step, the ontology information can be embedded to support suggestions and guidance for new modelers to build domain models. And an ontology can be utilized to verify and debug the domain model in the Domain Model Analysis (Walter, Parreiras and Staab 2014). The last step Targets Products Generation is about code generation and model transformation. In our opinion, currently ontology information is not needed there.

Table 1: Ontology's role in DSM

Ontology's role in DSM	
Domain Analysis	As a domain conceptual model As a basis for model comparison
DSML Design	Guide the DSML design Generate DSML
DSME Design	As suggestions or guidance in the modeling process
Domain Model Analysis	Verify and debug the domain model
Target Products Generation	Not yet

5. CONCLUSION

In this paper we discuss about what's an ontology, the ontology metamodel and the pervasive used ontology language OWL and its correspondent tool Protégé. And then we introduced the DSM method and separated it into five steps including domain analysis, DSML design, DSME design, domain model analysis and target products generation. The problems of DSM are proposed including how to design a DSML, how to extract the constraints from the domain and so on. Finally, we summarize the roles of ontology in the five different steps of the DSM method.

Currently, how to transform the ontology model to the M2 level of the four-level metamodel hierarchy of UML has not been studied so well for generating a DSML automatically. And it will enhance the DSM method significantly if the problem is solved. So in the future,

we will stick to the problem and research on how to transform DoDAF Meta Model (DM2, an ontology model for military architecture framework) to a DSML for domain-specific architecture modeling.

ACKNOWLEDGMENTS

The work is partly supported by the national Natural Science Foundation of China (No. 61273198)

REFERENCES

- Tolvanen J. and Rossi M., 2003. Metaedit++: defining and using domain-specific modeling languages and code generator. OOPSLA'03. October 26-32, Anaheim, California, USA.
- Sztipanovits J. and Karsai G., 1997. Model-integrated computing. *IEEE Computer*, 30(4), 110-111.
- Kleppe A., Warmer J. and Bast W., 2006. Translated by Bao Zhiyun. *MDA Explained*. Beijing: Posts & Telecom Press. (in Chinese)
- Guizzardi G., 2005. *Ontological Foundations for Structural Conceptual Models*. Thesis (PhD). Telematica Insituut Fundamental Research Series.
- Neches R., Fikes R.E., Finin T., et al., 1991. Enabling technology for knowledge sharing. *AI Magazine*, 12(3), 36-56.
- Gruber T.R., 1993. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 2(5), 199-220.
- Giaretta P. and Guarino N., 1995. Ontologies and knowledge bases: towards a terminological calrification. *Towards very Large Knowledge bases: Knowledge Building and Knowledge sharing*, 25-32.
- Borst W., 1997. *Construction of Engineering Ontologies*. Thesis (PhD), Universiteit Twente.
- Swartout B., Patil R., Knight K., et al, 1997. *Toward Distributed Use of Large-Scale Ontologies*. *Ontological Engineering Aaai Spring Symposium*.
- Guarino N., 1998. *Formal Ontology and Information Systems*. In N. Guarino(ed.), *Formal Ontlogy in Information Systems*. *Proceeding of FOIS'98*, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, 3-15.
- Fensel D., Harmelen F.V., Horrocks I., et al., 2001. OIL: an ontology infrastructure for the Semantic Web. *Intelligence Systems IEEE*, 16(2), 38 – 45.
- Colin Atkinson and Thomas Kuhne, 2003. *Model-driven development: A Metamodeling Foundation*. *IEEE Software* 20(5), 36-41.
- OMG, 2014. *Ontology Definition Metamodel*. Object Manangement Group. Available from : www.omg.org/spec/ODM/. [accessed 14 April 2015]
- Horridge M., 2011. *A Practical Guide To Building OWL ontologies Using Protégé 4 and CO-ODE Tools Edition 1.3*.
- Li X.B., 2013. *A DSM-based Multi-Paradigm Composable Modeling Method for Effectiveness*

- Simulation. Thesis (PhD). National University of Defense Technology, 44-45. (in chinese)
- Reihartz-Berger I. and Sturm A., 2004. Behavioral domain analysis – the application-based domain modeling approach. Lecture Notes in Computer Science, 410-424.
- Cho H., 2011. A Demonstration-Based Approach for Designing Domain-Specific Modeling Language. SPLASH'11 Companion, 51-53. October 22-27, Portland, Oregon, USA.
- Rath I. and Varro D. Challenges for advanced domain-specific modeling frameworks.
- Zbib R., Jain A., Bassu D., et al., 2006. Generating domain specific graphical modeling editors from meta models. Proceedings of the 30th Annual International Computer Software and Application Conference (COMPSAC'06).
- Walter T., Parreiras F.S., Staab S., 2014. An ontology-based framework for domain-specific modeling. *Softw Syst Model*, 83-108.
- Guizzardi G., 2005. An ontology-based approach for evaluating the domain appropriateness and comprehensibility appropriateness of modeling languages. *MODELS 2005*, 691-705.
- Guizzardi G. Ferreira Piress L., Sinderen V.M.J., 2002. On the role of domain ontologies in the design of domain-specific visual modeling languages. University of Alabama at Birmingham.
- Tarias R., Mernik M., Gray J., 2009. Using ontologies in the domain analysis of domain-specific languages. Lecture Notes in Computer Science, 332-342.

RESULT VALIDATION OF A RUDDER SIMULATION MODEL

Lina Meng^(a), Xiaochao Qian^(b), Haiying Wang^(c), Wei Li^(d),

^(a)Department of Mathematics, Heilongjiang Institute of Technology, Harbin, 150050, P. R. China

^{(b), (c), (d)} Control and Simulation Center, Harbin Institute of Technology, Harbin, 150080, P. R. China

^(a) fleehit@163.com, ^(b) everqxc@hotmail.com, ^(c) whydiana@163.com, ^(d) frank@hit.edu.cn

ABSTRACT

Air vehicle simulation has been extensively used with the advantages of economic and security. As an important part in guidance control system of air vehicle, the rudder plays a key role in vehicle simulation. Thus evaluating the credibility of the rudder model is of considerable significance. This paper presents an approach to the result validation of the rudder model, which uses the consistency of the rudder model outputs and the real rudder system outputs to obtain credibility of the rudder model. The procedure includes: 1) building an rudder credibility evaluation index system; 2) providing validation methods including feature consistency quantification, proximity of position quantification, similarity of shape quantification and spectral analysis; 3) achieving the validation result based on index system and validation methods.

Keywords: result validation, rudder model, data consistency, feature consistency

1. INTRODUCTION

Air vehicle simulation has been extensively used with the advantages of economic and security. As an important part in air vehicle, the rudder is not only an actuator but also a major component in the guidance and control system. Modeling and simulation of the rudder is an important aspect of air vehicle simulation, which affects the credibility of the air vehicle especially the guidance control simulation system (Sargent 2013).

The credibility of the simulation system directly determines the success or failure of the simulation application. Model validation is the primary means to research the credibility of model, and result validation is one way to achieve the validation process (Balci 2003). Result validation is the process of measurement of the similarity between the simulation data and experimental data with the same inputs.

As early as 1967, Naylor (1967) pointed out that face validation can be carried out according to the intuition, and that the similarity between the simulation model and the real object can be carried out according to the typical events, assumptions and other internal features of the two. Mckenny (1967) also considered that the validation process should be done on condition that the simulation and reference system should be with the same inputs. Simulation outputs can be divided as static outputs and time series outputs. Under this hypothesis, static validation metric can be separated into two categories: 1) parameter estimation, model parameter updating, or system identification (Oberkampf and Barone 2006); 2) hypothesis testing or significance testing (Yu 2011). Time series can be validated with Theil inequality coefficient (TIC) (Kheir and Holmes 1978), grey relational analysis (Wu J., Wu X.Y. and Chen Y.X. 2010), etc.

In this paper, we first analyze the character of the rudder system and build the index system of the validation mission; then select the validation method based on the

feature of index; finally calculate the validation result based on the simulation and experimental data.

2. INDEX SYSTEM OF THE RUDDER MODEL RESULT VALIDATION

The air vehicle rudder is a complicated electro-mechanization system. There are numerous factors which influence the credibility of the rudder model. Thus, it is important to validate the simulation model in different input condition. The output of the

rudder model we most concern is the angle of rudder reflection with diverse input condition during the result validation procedure. The typical working condition of the rudder is following the tracks of the constant value input instructions and periodic signal input. Therefore, the input conditions of the rudder are a step signal with different amplitude and a sinusoidal signal with different frequency. Thus, the index system of the rudder model result validation is established as below (see Figure 1).

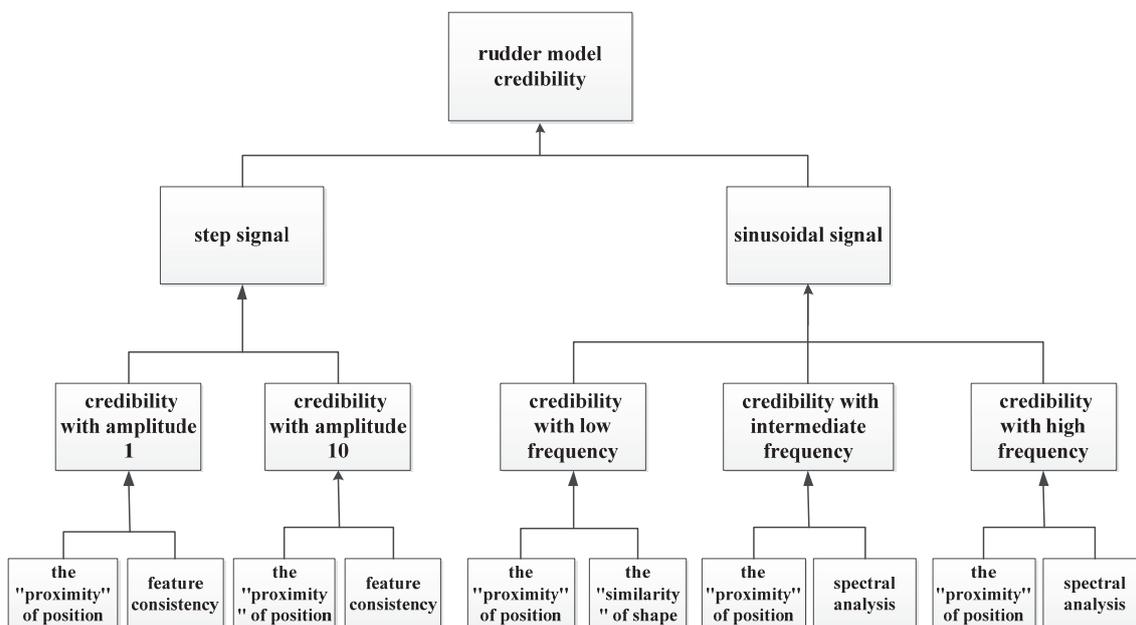


Figure 1: the Index System of the Rudder Model Result Validation

Figure 1 shows the index system. According to the characteristics of the step signal, the response characteristics (e.g. overshoot, steady state error, rise time, etc.) and the position proximity of the simulation output and the reference output under step input are concerned. According to the characteristics of the sinusoidal signal, the simulation output under sinusoidal input is divided into three frequency domain: low frequency, medium frequency, high frequency, and different indexes are used to assess consistency. Proximity of position and similarity of sharp are concerned for low frequency. For middle and high frequency, in addition to the proximity of position in time domain analysis, the consistency of frequency domain characteristics should also be concerned, so

spectral analysis is used to assess consistency. The reason for that we analyze middle and high frequency separately is that the rudder works mainly in middle frequency. Therefore, the weight of middle frequency is higher than that of high frequency when we determine the index weight. The weights of index system are determined by domain experts using the AHP method.

3. RESULT VALIDATION METHODS OF THE RUDDER MODEL

According to the index system in Section 2, we carry out result validation of the rudder model by using methods of feature consistency, and data consistency. The data consistency methods including: proximity of

position, similarity of shape, spectral analysis, etc. These methods are introduced below respectively.

3.1 Feature consistency

c_r is a feature extracted from a reference output Y_r . And c_s is a feature extracted from a simulation output Y_s . Relative error is used to represent their difference and is given by Eq. (1). Then we obtain the degree of consistency of c_r and c_s by mapping their difference into an interval $(0,1]$ according to Eq. (2).

$$\eta_c = \begin{cases} \frac{|c_s - c_r|}{|c_r|}, & c_r \neq 0 \\ |c_s - c_r|, & c_r = 0 \end{cases} \quad (1)$$

$$V(c_s, c_r) = e^{-\lambda_c \eta_c} \quad (2)$$

Where, $\lambda_c > 0$ is a model parameter of feature consistency quantification, which is given based on the specific application domain.

3.2 Proximity of Position

$y_r = \langle y_r(1), y_r(2), \dots, y_r(p) \rangle$ and $y_s = \langle y_s(1), y_s(2), \dots, y_s(p) \rangle$ represent the time series of reference output and simulation output respectively.

TIC coefficient (Ming Yang, Wei Li, etc. 2014) is often used to describe the position difference between two time series. The TIC algorithm is shown in Eq. (3).

$$T(y_r, y_s) = \frac{\sqrt{\frac{1}{p} \sum_{i=1}^p (y_r(i) - y_s(i))^2}}{\sqrt{\frac{1}{p} \sum_{i=1}^p y_r(i)^2 + \frac{1}{p} \sum_{i=1}^p y_s(i)^2}} \quad (3)$$

$$= \frac{\sqrt{\sum_{i=1}^p (y_r(i) - y_s(i))^2}}{\sqrt{\sum_{i=1}^p y_r(i)^2 + \sum_{i=1}^p y_s(i)^2}}$$

Where, $T(y_r, y_s)$ is the TIC coefficient.

According to Eq. (3), $T(y_r, y_s) \in [0,1]$. It describes a relative error, which facilitates its use and comprehension. However, if we use it in simulation model validation directly, there exist some problems as below.

In Figure 2(a), $y_r(t) \equiv c_r$, $y_s(t) \equiv c_s$, $t = 1, 2, \dots, p$. c_r and c_s are constants and $c_r \leq 0$, $c_s > 0$. Based on Eq. (3), $T(y_r, y_s) \equiv 1$. This indicates that the proximity

of position of y_r and y_s is the worst, which is unreasonable obviously. In Figure 2(b), $y_r(t) = f(t)$, $y_{s1}(t) = f(t) + c$, $y_{s2}(t) = f(t) - c$, $t = 1, 2, \dots, p$. c is a constant and $c > 0$. We can judge intuitively that $T(y_r, y_{s1}) = T(y_r, y_{s2})$, but $T(y_r, y_{s1}) > T(y_r, y_{s2})$ according to Eq. (3).

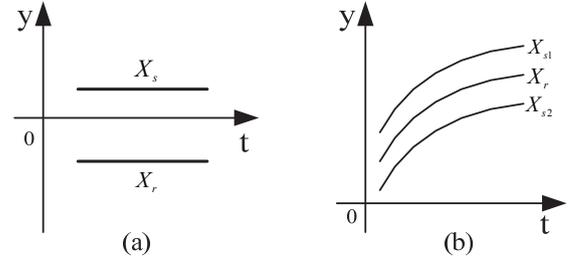


Figure 2: Reference Outputs and Simulation Outputs in Two Special Cases

The reason why the results above are not consistency is that the relative error of y_r and y_s is calculated by the benchmark of

$$\sqrt{\sum_{i=1}^p y_r(i)^2} + \sqrt{\sum_{i=1}^p y_s(i)^2} \quad (4)$$

It is not calculated by the benchmark of reference outputs. Based on the TIC coefficient, this paper proposes an improved algorithm for the proximity of position, which is shown in Eq. (5) ~ (6).

$$T(y_r, y_s) = \begin{cases} \frac{\sqrt{\sum_{i=1}^p (y_r(i) - y_s(i))^2}}{\sqrt{\sum_{i=1}^p y_r(i)^2}}, & \sqrt{\sum_{i=1}^p y_r(i)^2} \neq 0 \\ \frac{\sqrt{\sum_{i=1}^p (y_r(i) - y_s(i))^2}}{\sqrt{\sum_{i=1}^p y_r(i)^2 + \sum_{i=1}^p y_s(i)^2}}, & \sqrt{\sum_{i=1}^p y_r(i)^2} = 0 \end{cases} \quad (5)$$

$$d(y_r, y_s) = e^{-\lambda_d T(y_r, y_s)} \quad (6)$$

Where, $\lambda_d > 0$ is a model parameter for proximity of position quantification, which is given based on the specific application domain.

3.3 Similarity of Sharp

Grey incidence degree is adopted in this paper to describe the similarity of shape. The calculation formulas are listed in Eq. (7) ~ (8).

$$r(t) = \frac{\min_t \Delta(t) + \lambda_s \max_t \Delta(t)}{\Delta(t) + \lambda_s \max_t \Delta(t)}, \quad t = 1, 2, \dots, p \quad (7)$$

$$s(y_r, y_s) = \frac{1}{p} \sum_{t=1}^p r(t) \quad (8)$$

Where, $\Delta(t) = |y_r(t) - y_s(t)|$; $\lambda_s \in [0, 1]$ is a resolution coefficient and generally $0 \leq \lambda_s \leq 0.5$.

3.4 Spectral Analysis

The principle of spectral analysis is to compare the compatibility of two time series samples with same model in frequency domain. Each frequency point $\omega_i (i=0, 1, \dots, m)$ is tested respectively for consistency. If the power spectrum estimation $\hat{S}_x(\omega)$ and $\hat{S}_y(\omega)$ of each frequency point are the same, then the two time series samples are compatible. Set the spectral density estimation of two time series $\{x_i\}$ and $\{y_i\}$ are $\hat{S}_x(\omega)$ and $\hat{S}_y(\omega)$, where $\omega \in [-\pi, \pi]$.

We obtain respective spectral density through window spectrum analysis, then we describe the differences between X_{rf} and X_{sf} based on the spectral density differences. Kheir and Holmes (1978) introduced the specific principle. The spectral density difference is defined as

$$F_H(X_S, X_R) = 1 - \frac{m}{M} \quad (9)$$

Where, $F_H(X_S, X_R)$ is the spectral density difference between X_{rf} and X_{sf} ; M is the number of points that X_{rf} and X_{sf} are converted to frequency domain; m is the number of points that passed the compatibility test.

4. RESULT VALIDATION OF THE RUDDER SIMULATION MODEL

This section is to validate outputs of the rudder model based on the validation index system of the rudder model proposed in Section 2 and the validation methods proposed in Section 3. According to output styles of the rudder, the validation can be divided into step response with amplitude 1, step response with amplitude 10 and

sinusoidal response of low frequency, medium frequency and high frequency.

4.1 Result Validation under Step Input

The step input signal here has two types: step input with amplitude 1 and step input with amplitude 10. According to the step input signal characteristics, the validation under the step input can be divided into two parts: the feature consistency and the proximity of position. Figure 3 is a schematic diagram of simulation output feature extraction on condition that step with amplitude 1 is the input signal. And Figure 4 is a schematic diagram of reference output feature extraction on condition that step with amplitude 1 is the input signal. Table 1 shows the validation result when the input signal is step with amplitude 1.

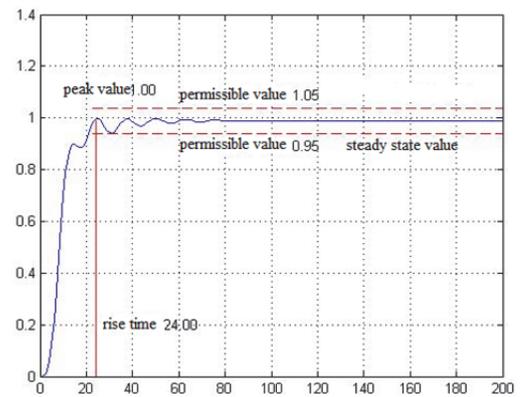


Figure 3: Experimental Output after Feature Extract under Step Single with Amplitude 1

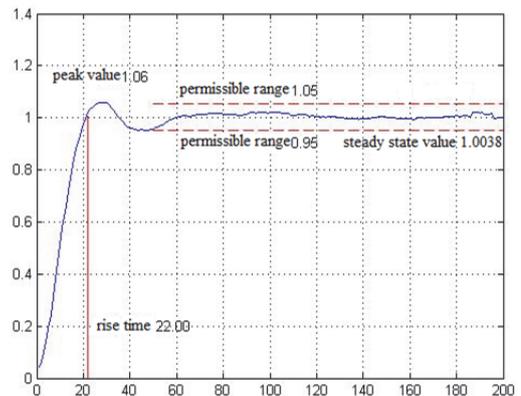


Figure 4: Simulation Output after Feature Extract under Step Single with Amplitude 1

Table 1: Date Feature and Validation Result under Step Single with Amplitude 1

Feature name	Simulation output	Experimental output	Validation result
Rise time/s	0.024	0.022	0.913
Peak value/°	1.000	1.060	0.942
Steady state value/°	0.989	1.004	0.985

4.2 Result Validation under Sinusoidal Input

Sinusoidal input is continuously variable signal from 1Hz to 50Hz. According to the characteristics of the sinusoidal signal, we divide it into three parts: low frequency, medium frequency and high frequency. Sinusoidal signal with frequency 1Hz is chosen as low frequency signal, focusing on the similarity of its position and shape. Sinusoidal signal with frequency 18Hz-22Hz is chosen as medium frequency signal, focusing on the similarity of its position and frequency spectrum. Sinusoidal signal with frequency 48Hz-52Hz is chosen as high frequency signal, focusing on the similarity of its position and frequency spectrum too.

Figure 5 is the simulation and experimental output under sinusoidal single with frequency 1 Hz. Table 2 shows the validation result under sinusoidal single with frequency 1 Hz.

Figure 6 shows the simulation and experimental output under sinusoidal single with frequency 20 Hz. The validation results under sinusoidal single with middle and high frequency are listed in Table 3.

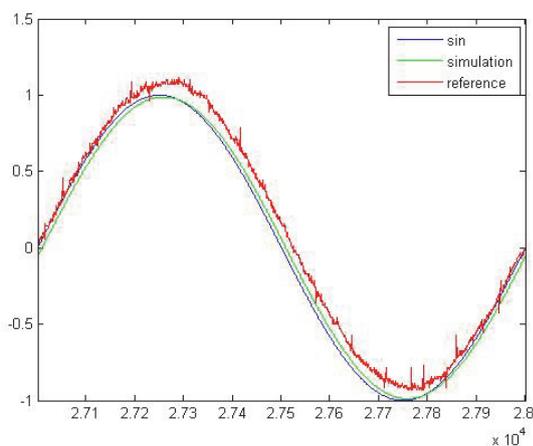


Figure 5: Simulation and Experimental Output under Sinusoidal Single with Frequency 1 Hz

Table 2: Validation Result under Sinusoidal Single With Frequency 1 Hz

Single frequency	Proximity of position	Similarity of sharp
1Hz	0.9792	0.873

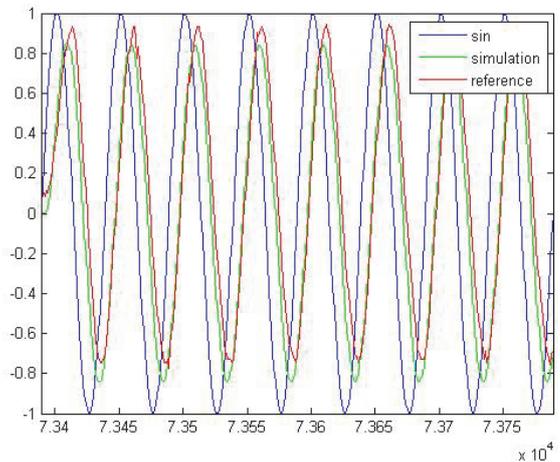


Figure 6: Simulation and Experimental Output under Sinusoidal Single with Frequency 20 Hz

Table 3: Validation Result under Sinusoidal Single With Middle and High Frequency

Single frequency	Spectral analysis	Proximity of position
18~22Hz	1.000	0.9148
48~52Hz	0.801	0.5874

4.3 Comprehensive Validation Result

The results of the rudder model validation are shown in Table 4. According to the index system given in Figure 1, comprehensive validation result is calculated using weighted average method.

According to the existing validation result, the credibility of the rudder model is 90.6 (full score of 100). Thus the rudder model is creditable. Although the entirety credibility of the rudder model is well, the credibility under high frequency sinusoidal single is poor. Therefore, the model needs some future improvements.

Table 4: Index Similarity and Model Validation Result

Output	Index		Index result	Index weight	Validation result(%)
Step single with amplitude 1	Proximity of position		0.987	0.8	98.7
	Feature consistency	Rise time	0.913	0.2	94.7
		Peak value	0.942		
		Steady-state value	0.985		
Step single with amplitude 10	Proximity of position		0.974	0.8	97.4
	Feature consistency	Rise time	0.595	0.2	86.3
		Peak value	0.996		
		Steady-state value	0.997		
Sinusoidal single with frequency 1 Hz	Proximity of position		0.979	0.7	97.9
	Similarity of sharp		0.873	0.3	87.3
Sinusoidal single with frequency 18~22 Hz	Proximity of position		0.915	0.5	91.5
	Spectral analysis		1.000	0.5	100
Sinusoidal single with frequency 48~52 Hz	Proximity of position		0.587	0.5	58.7
	Spectral analysis		0.801	0.5	80.1

5. CONCLUSIONS

As an important part in air vehicle, the rudder is not only an actuating mechanism of its guidance system, but also an important part of the guidance loop. The rudder model can influence the air vehicle system especially the credibility of the guidance control system, thus the rudder model is a major part of the air vehicle modeling and simulation. To ensure the credibility of a rudder model, this paper presents a result validation method of the rudder model, which used the consistency of simulation outputs and reference outputs to evaluate the credibility of the rudder model.

The main work of this paper includes: 1) Build the index system of the rudder model result validation; 2) Extract the rise time, peak value and the steady-state value under the step input as the validation feature; 3) TIC and feature analysis method are chosen for the result validation under the step input; 4) TIC, grey relational analysis and spectral analysis are chosen for the result validation under sinusoidal input; 5) Determine the comprehensive evaluation method of the

index system and obtain the credibility of the rudder model.

REFERENCES

- Balci, 2003. Verification, validation, and certification of modeling and simulation applications. Proceedings of the 2003 Winter Simulation Conference, pp. 150-158. December 7-10, New Orleans, LA.
- Kheir N.A., Holmes W M, 1978. On validating simulation models of missile systems. Simulation 30(4): 117-128.
- Mckenny J.L., 1967. Critique of Verification of Computer Simulation Models. Management Science 14:55-59.
- Ming Yang, Wei Li, Lingyun Lu, Song Jiao, 2014. Validation of Simulation Model Based on Combined Consistency Analysis of Data and Feature. The Third International Conference on Data Analytics, pp. 65-69. August 24-28, Rome, Italy.

- Naylor T.H., Finger J.M., 1967. Verification of computer simulation models. *Management Science* 14(2): 92-101.
- Oberkampf W.L., Barone M.F., 2006. Measures of agreement between computation and experiment: Validation metrics. *Journal of Computational Physics* 217(1): 5-36.
- Sargent R.G., 2013. An introduction to verification and validation of simulation models. *Proceeding of the 2013 Winter Simulation Conference*, pp.321-327. December 8-11, Washington, D. C., USA.
- Wu J., Wu X.Y., Chen Y.X., et al., 2010. Validation of simulation models based on improved grey relational analysis. *Systems Engineering and Electronics* 32(8): 1677-1679.
- Yu, L., Wei C., Paul A., et al., 2011. Toward a Better Understanding of Model Validation Metrics. *Journal of Mechanical Design* 133:1-13.

WEB SIMULATION AS A PLATFORM FOR TRAINING SOFTWARE APPLICATION

Stepan Kartak

Faculty of Electrical Engineering and Informatics, University of Pardubice

stepan.kartak@student.upce.cz

ABSTRACT

This paper is focused on a practical use of a web browser in a web distributed simulation. A web browser environment is used for a realization of a comprehensive solution for multiuser trainer applications. A computer with the web browser as one simulator station is usually available anywhere. Presented solution allows to create a script of simulation, to run the simulation and then to centrally monitor status of individual stations and behavior of their operation in real time. The individual stations can be in one room or they can be geographically distant. All this wherever and without a need of special software. An internet browser and PC connection to a PC network is the only requirement.

Keywords: distributed simulation, web-based simulation, WebRTC, software for training

1. INTRODUCTION

This contribution deals with the realization of the multiuser trainer applications just with an usage of web technologies. A base of the solution is based on a previous work regarding distributed web simulation (Kartak and Kavicka 2014; Hridel and Kartak 2013), that has solved only WebRTC applicability without an additional user-friendly control and software support of a remote control and central visualization. The term “trainer application” will represent a group of applications (an interactive simulation) for testing (examination, teaching) workers/dispatchers. The dispatcher (i.e. one logical process that is operated by one user at one computer/station). This user operates above specific area (then without knowledge of a global status) and responds to incoming information from nearby logical processes with which it is in direct connection. For instance let us state an operation (i.e. dispatcher) of a railway station within a region (where more dispatchers occur). Another example a management of road transport can be within an exposed territorial entity, a management of investigative, life-saving or military operations under geographical areas. Individual participants of the simulation can be in one room or they can be geographically distant (for example they need not move away from their real workplaces). PC connected to the PC network is the only requirement.

2. CHARACTERISTIC

Entire solution consists of four parts which will be described in detail in next chapters:

- administration web interface (AWI) for assembly and setting of the simulation,
- web application for central mass management of pages in browser (JSRC) – in this case it will be used for a management of logical processes,
- logical processes (LP) realized as independent web pages put above common JavaScript library,
- central visualization (CV) of individual logical processes and of entire simulation.

Due to a character but found WebRTC weak points too (see below) the entire solution focuses on distributed simulation of smaller range, in this case to 20 logical processes, however this restriction is given by current state of WebRTC only. Theoretically it is not a problem to operate a simulation of much larger range.

2.1. All in web browser

All parts are realized as web applications running in web browser (with a server support for partial additional activities – AWI, record of animation activities, etc.). Very distributed simulation then runs fully in web browsers (without using the server). A support of new HTML5 technologies (i) WebRTC (full-duplex network communication between browsers) is essential, (ii) WebSockets (full-duplex communication server-browser), (iii) canvas (medium for 2D drawing to a display area of browser) and (iv) WebGL (using of graphic card for 3D animation in the environment of web browser). These features support browsers using web core Webkit (primarily Google Chrome and many others) and Gecko (Mozilla Firefox) in both cases both desktop and mobile version. For the distributed web simulation first two above mentioned features are key which allow to communicate on network as well as standard desktop applications.

Other two features are necessary for realization of a graphic output.

2.2. Web browsers and communication on network

Due to a request for fluent animation output it is necessary to minimize network response among logical processes and it is important too to consider a fact that web browser and its programming language JavaScript do not belong among computationally efficient languages or environments. Too an option of suitable communication and synchronization of course among logical processes is essential. WebSockets communicate on basis of TCP connection and permit connection with server. This possibility would be potentially suitable for a realization of distributed simulation solved similarly as HLA (it very often selects using of centralized architecture; Kuhl, Judith, and Weatherly 2000). Here a summarizing of information on one node in the network and therefore creating of global state space of running simulation is a significant advantage of using server element. In this case we need to send a message via network twice for delivery of the message (these are four messages with a request to reply). An operation is a result that is very time-consuming and reduces possibilities of creating above mentioned type of application. WebRTC technology is a better choice since it is proposed for creating direct connection among browsers (i.e. it allows to omit “center” server element from the communication among logical processes). Network response time will be two times shorter theoretically when using WebRTC compared to WebSocket. The omitting of center element is a negative aspect as it results in a loss of simple realization of global state space (for example for a central accounting, a simulation recapitulation or for yet mentioned central visualization).

2.3. Using and problems of WebRTC

In theory connections through WebRTC are not limited in no way. It does not depend on a network topology, local networks, NAT routers (described in detail in Kartak and Kavicka 2014) not even on a number of connections. From this view it is possible to create a web alternative of a Distributed Interactive Simulation (DIS) standard in the sense of broadcast UDP communication (IEEE standards). For creating of connection in this style it is necessary to realize WebRTC peer-to-peer connection of each with each which should not be a problem in theory due to above mentioned “limitless” situation. At an extensive (connection of twenty or more logical processes) practical tests of this solution it revealed that a problem appears with frequent closing connections without an obvious reason right after connection. It is possible in the local network to create the connection of each with each among twenty computers however usually it is a need of many trials (even tens of trials with increasing number of browsers). It is a significant problem whereas after disconnection it is not possible to restore the network WebRTC connection (it depends on a type of disconnection but it is not possible in the case of above mentioned problem) without restart of application

(i.e. must page refresh in the web browser). This problem identified in local site will start to show much more at “connecting” of more browsers (usually after NAT routers) across Internet public network. Even though this status is highly unsatisfactory and largely limiting for application in practice it can be to minimize number of logical processes connections on the basis of their natural dependences among each other (see figure 1 or 3). It revealed from testing in local site that connection of 20 browsers (logical processes) over ca. 40 WebRTC connections provides stable achievable solution in the local site. As regards connection through public site problems with self-closing connections still occur however achieving all stable connections seems as possible contrary to above mentioned alternative in broadcast connection.

For an example, an attempt to connection among 6 browsers (each with one logical process, in different geographically distant net, 1 km to 150 km beeline amongst physical location of browsers) failed to repeat even after ten trials because connection between 1 or 2 browsers (usually those that were geographically furthest apart) always closed itself after connection without an apparent reason. Connection with the same topology within an urban network (geographic distance between physical location of browsers around 1 km) succeeded repeatedly among 1 and 3 trials. Even this result is not optimal. After successful connection establishment the connection is always permanent and smooth (max. 5 operation hours tested).

2.4. Synchronization of logical processes

The synchronization of logical processes is realized by a method Conservative synchronization technique of sending null messages with a lookahead (Chandy-Misra-Bryant Distributed Discrete-Event Simulation Algorithm, Fujimoto 2000). It is about standard slightly modified version of this algorithm, the version was described in detail in Kartak and Kavicka 2014.

2.5. JavaScript and computational demands

Programming language JavaScript is almost exclusively one-thread in the web browser (it generally valids for entire user source code), this means that all operations are performed gradually, by time (or order) that they have attributed by internal scheduler or in nearest possible time if at the time on which an event is planned already some calculation runs. This is a critical issue with that it is necessary to count.

The logical process performs three operation groups:

- calculations of simulation core,
- realization of the animation,
- and a whole range of minor tasks (for example sending information to the server).

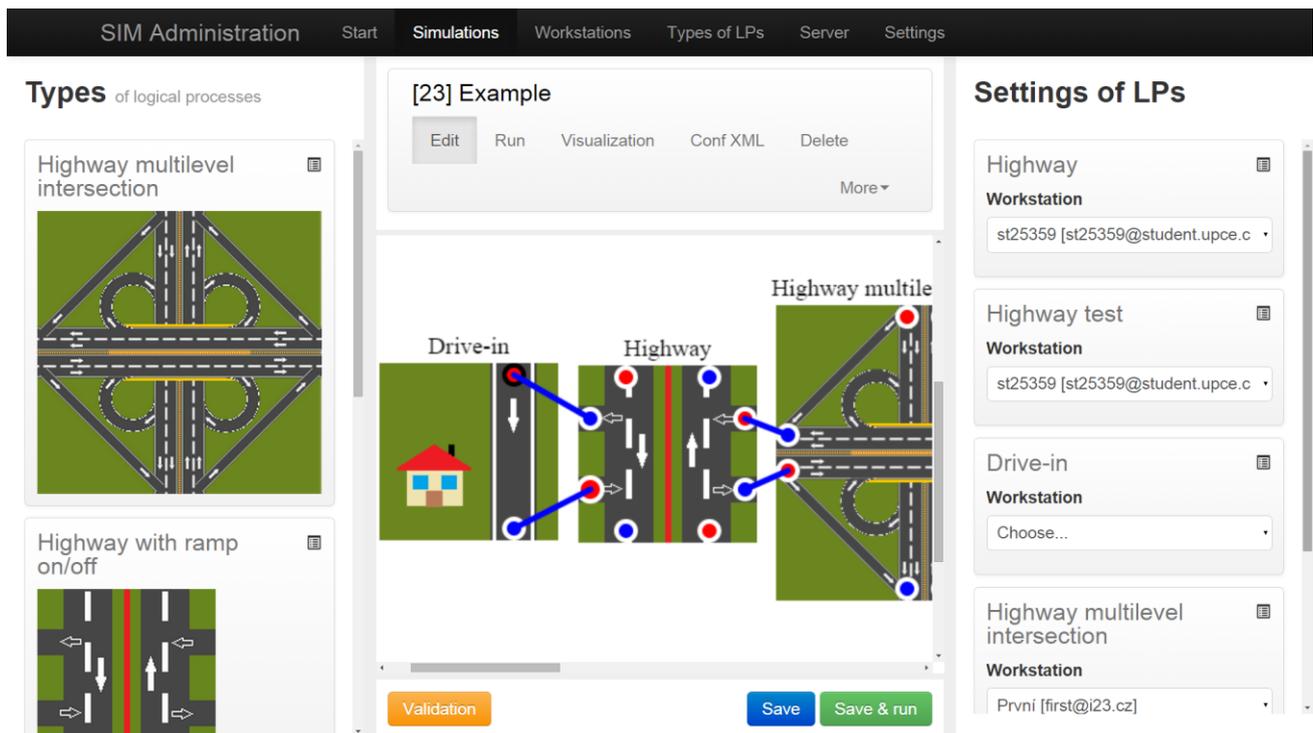


Figure 1: Administration web interface, visual editor

Just the realization of the animation currently appears as a problematic area.

Demand depends on area for drawing size (HTML element *canvas*) and too on the quantity of animation activities that it is necessary to update in each step of the animation – whereas it is about drawing here (in our case in 2D space) we will not avoid often counting of parametric curves and else not fully trivial calculations that will begin to occupy precious time in larger quantities with next operations of animation activities (time control of beginning and end of display, etc.). Generally speaking computationally intensive tasks are not an ideal type of problem for JavaScript solution and they may represent significant problem for fluency of simulation run from the perspective of service and in a combination with one-thread access (in fact it can be only about decrease of framerate of animation output). If we take into account that a man considers fluent animation it is possible at 25 FPS (generally accepted number, films are distributed with that FPS) we will get a time to a realization of calculations between redrawing of animation scene $1000 / 25 = 40$ ms. During 40 ms it must be done at worst (in brackets average measured times in ms stated, if they are average larger than 1 ms; computers in one local network, 20 logical processes, topology of connection see figure 9a):

1. Calculations of simulation core, consist of:
 - (a) calculation of single activities and planning of further activities (1 ms),
 - (b) waiting for a response of remote logical processes because of synchronization (in this solution it is about a request for a null message if we have n nearby logical

processes then a time required to realization will be $n \times 10$ ms at worst).

2. Processing planned JavaScript tasks:
 - (a) execute commands of central JSRC control,
 - (b) responses to requests for null message from nearby logical processes,
 - (c) internal planned operations, i.e. sending information to server, sending screenshots of animation scene, etc. (2 ms).
3. Calculations of animation module (10 ms):
 - (a) sorting of animation activities,
 - (b) time control of beginning and end of animation activities display.

Critical time that it is not possible to optimize in no way is 1b point (waiting for response of remote logical processes because of synchronization). On average response to this type of request lasts 10 ms when whole logical process inactively waits. The more logical processes are with that it is necessary to keep logical process as synchronized the more (in the worst case) a time can be longer between animation frames. According to presented values 3 and more nearby logical processes can mean a problem.

Standard desktop application would solve a considerable part of these operations in parallel, in JavaScript something like this is not possible so must be to pay special attention to optimizing of calculations and data structures.

3. ADMINISTRATION WEB INTERFACE

Administration web interface (AWI, figure 1) is a web application that serves primarily to a compilation of distributed simulation model and to a creation of a session of this simulation. The process of creating the distributed simulation from beginning to launch is shown in a flowchart in a figure 2.

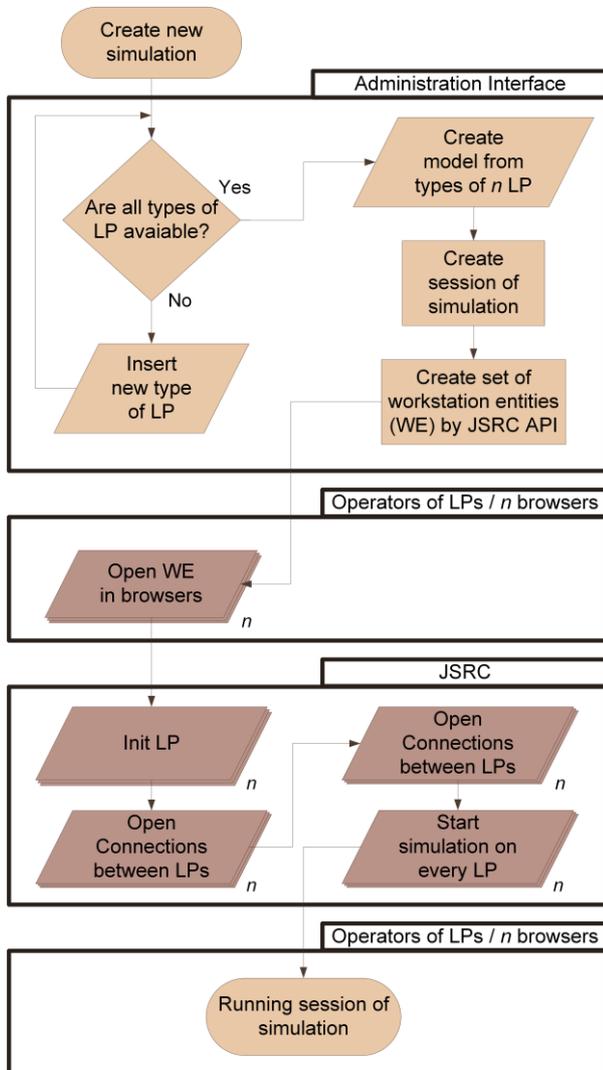


Figure 2: Flowchart of process of creating the distributed simulation from beginning to launch

3.1. Logical process, its type and definition

The distributed simulation consists of logical processes. This solution considers set web page registered in AWI as the logical process. In AWI a record about new type of logical process is created. To this record (a) name, (b) URL address of very web page (it represents very logical process) and (c) URL address with XML file that specifies possibilities of setting and network linking with other logical processes are assigned. Based on this record and XML configuration in visual editor (see next chapter) a miniature of logical process is created (figure 3) that represents logical process in the editor.

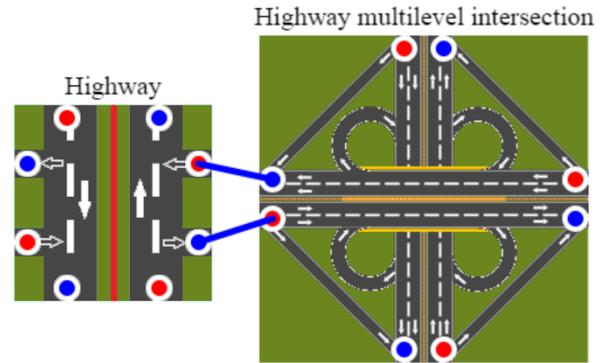


Figure 3: Display of two linked logical processes in AWI visual editor

Configuration XML file contains:

1. **Meta information** – basic information about the logical process (name, author, picture for display in AWI, etc.)
2. **Description of connection with other logical processes** – i.e. input and output entit points and their setting:
 - (a) list of entit types that are valid for the logical process,
 - (b) distinction of input and output points,
 - (c) determining if the connection is required (usually for a meaningful function of logical process some input is required) or not.
3. **Items of setting** – each requires to specify:
 - (a) name,
 - (b) description,
 - (c) default value,
 - (d) type of value (for an appropriate processing/validation of input and display of appropriate input fields - eg *string*, *integer*, *boolean*, etc.),
 - (e) and few further options (for example a key under that the logical process will expect these data).
4. **Free space** (node chosen within XML file) for any content, determined for a possibility to set logical process above editor frame options.

3.2. Visual editor for setup of distributed simulation model

The simulation model consists of a set of logical interlinked processes. Visual editor in AWI serves for creating of this set (see fig. 1).

In the left part of editor an offer of available types of logical processes is.

In the middle a design area is located that serves for setup and connecting of logical processes. The logical processes are added by dragging from a left bar. Setting and connecting of logical processes proceeds under rules said in XML configuration files of logical

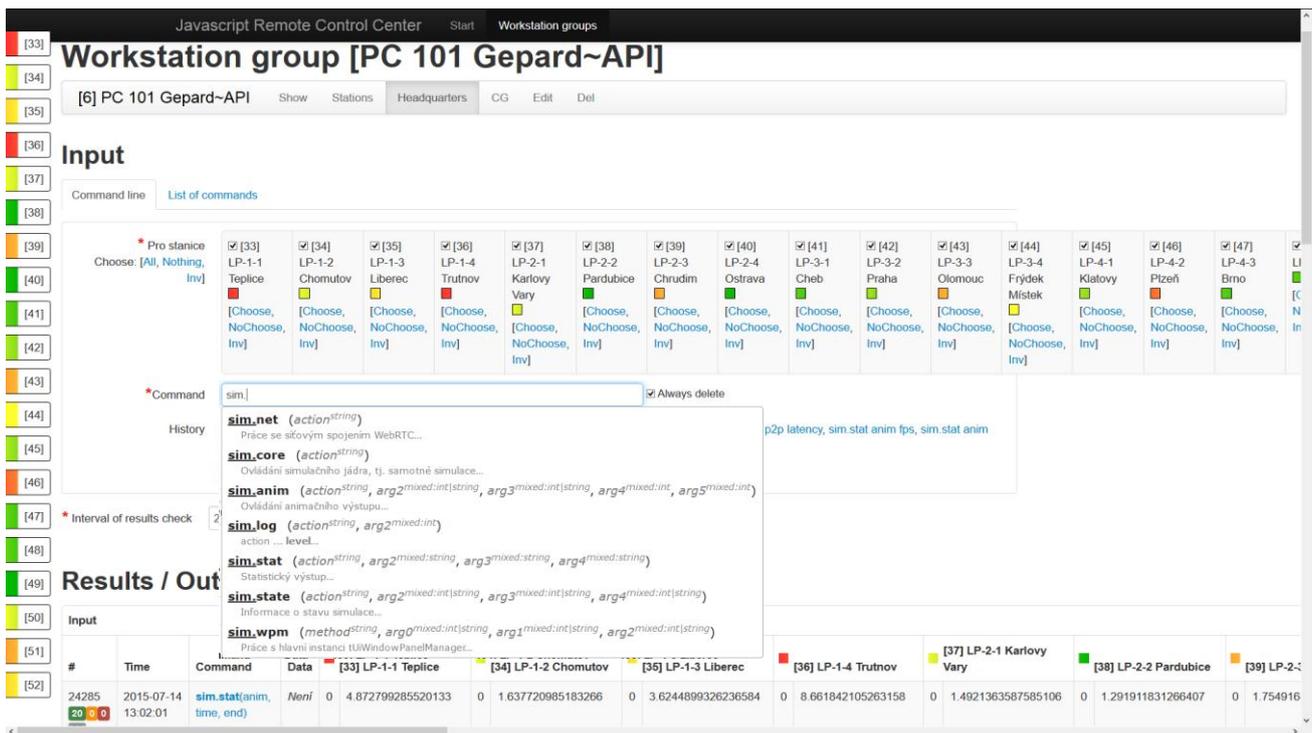


Figure 4: Remote Control web interface

processes types. In the right section a quick review of used logical processes is with a possibility of their setup.

3.3. Creating of simulation session

The simulation model created as described in part 3.2 is only an executable description of a simulation structure and setup. Usable simulation became from the model by creating of simulation session (de facto of session of the simulation model).

Such action consists of these steps:

1. Assignment of an unique identifier to each logical process in the model.
2. Creating/preparing of relevant data structures:
 - (a) preparing of data in a database,
 - (b) creating of XML configuration file describing compilation of model with added information about simulation session (especially identifiers of single logical processes).
3. Creating of a group of stations in JSRC application (see below). Creating of single stations in this group for each logical process. (All through relevant JSRC API.)

Created XML configuration file represents enter data for very logical processes, on the basis of information from this file WebRTC connection and assistant data structures is set.

4. REMOTE CONTROL

Remote control (JSRC, web interface is shown on figure 4) is a separate application intended to a multiple control of web pages. In this case it serves to a control of logical processes. Primarily to a load of logical process, creating of WebRTC connection and startup of all logical processes.

4.1. Description of working and integration to the simulation

This application as an entirely independent product serving for a mass control of web pages that is from one place (JSRC web application) any quantity of web pages is controlled. 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), moreover there is an option to create advance prepared command sets at “one click” (figure 5).

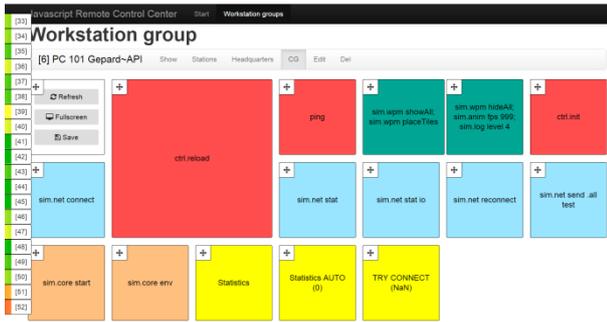


Figure 5: JSRC: Prepared command set, one square is user-defined command (or commands), prepared for touch-devices

Each work group can dispose different command set, here used for a separation of remote control of central visualization (CV, see next chapter) from very simulation.

The session of distributed simulation is realized within one work group and CV control in the second entirely independent work group (it contains only one workstation).

4.2. Description of used API

This application communicates with pages via API that allows to:

- 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).

The communication via API is displayed at picture 6.

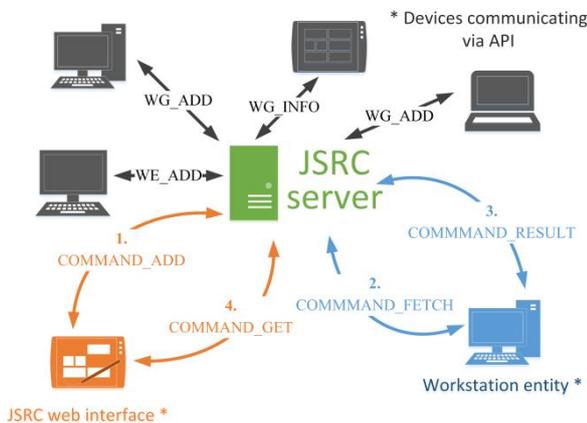


Figure 6: The communication via JSRC API with order of API calls for execute one command

An usage scheme in the simulation is displayed at figure 7.

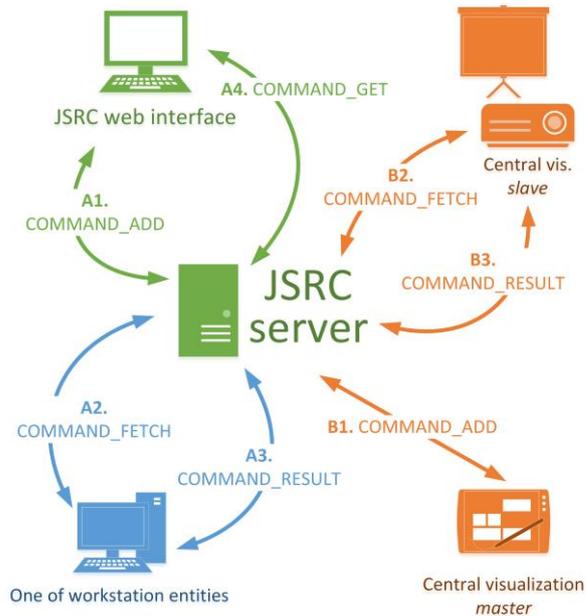


Figure 7: The scheme of usage JSRC API in the simulation

5. REALIZATION OF LOGICAL PROCESSES

The logical processes are realized as common web page in that with assistance of prepared JavaScript library (it is not a condition while respecting communication standards) is: (a) XML file for transmission of information about configuration of distributed model, (b) structure of messages transmitted among logical processes.

It can be to quickly create with this library activities for simulation cores and on the basis of several prepared prototypes of animation activities too an animation output.

While all can be created user friendly in an uniform style of control and appearance too.

So prepared web page represents type of logical process that is potentially configurable and reusable (in more models and in one model multiply too).

It is necessary to prepare a picture yet for adding new type of logical process (a miniature schematically representing the logical process) and XML configuration file specifying setup and network connections with that type of logical process disposes (described in detail in chapter 3.1).

6. CENTRAL VISUALIZATION

The central visualization (CV) is an independent web page that is created from the same components as very logical processes – especially for keeping united visual style and control. It allows to repeatedly play animation status of logical processes.

6.1. Realization of animation play of one logical process

Animation of logical processes sends new, changed and removed animation activities to the server (currently solved as a part of AWI). The sending is solved, according to server workload, in two modes: (a) sending immediately, (b) queuing of activities and their batch sending after some time. CV page in regular intervals downloads these information about animation activities that it immediately ranks to the animation output or removes them by record type. The output is due to delay in communication between logical process and server and between page with the visualization and the server burdened by delay against real in order of tens to hundreds of milliseconds (according to selected setting). Display of the communication is in a figure 7 (“b” commands).

6.2. Page with a summary of all logical processes

CV provides a possibility too to display at once status (i.e. the animation in this case) of all logical processes (within the session of distributed simulation) in bulk in the summary. Whereas fluent play of animation activities in one animation scene is computationally very difficult the realization of displaying larger number of animation scenes is difficult to achieve.

From that reason it was proceeded to that the logical processes after selected intervals (specified in XML configuration of logical process type) send to server (part of AWI) screenshots of the animation scene and these static images are then displayed in the summary – the animation is not fluent but it is enough for the summary.

Neither this solution is not ideal because creating of animation scene screenshot lasts in order of tens of milliseconds after that it is not possible to restore the animation of the logical process – so impression of choppy video generates that definitely is not an ideal status and it disturbs an user serving the logical process. Generally speaking the larger animation scene the more time creating of the screenshot will take.

6.3. Remote control of central visualization

For a comfortable CV control it is created through JSRC to control the page (primarily the animation output and selection of logical process or page with the summary) remotely from another instance/CV page.

In this case one CV is in a *master* position (i.e. control, CV_m) and the second one in a *slave* position (i.e. controlled visualization, CV_s).

A practical use CV_s display can be on a projector and its control through CV_m that can run for example on a tablet that service or trainer handles.

The animation output is always synchronized with a system time and therefore we can suppose that one second of CV_s and CV_m animation will in real time correspond to one second.

This premise is not valid relative to JavaScript “one-thread” architecture and displayed time at both

animations will not remain synchronized for a long time.

From this reason it was proceeded to running CV_m and CV_s synchronization.

Practically it works that CV_m continually sends (after 5 seconds interval, through JSRC) information about an actual time that displays animations and CV_s with respect to own actual animation time, a real measured animation speed and network communication delay adjusts speed of own animation to achieve a synchronous time.

This way of synchronization currently requires an exact time on both equipments.

7. COMPARISON WITH OTHER SOLUTION TYPES

7.1. Web vs desktop application

There are three fundamental reasons for using web browser instead of native desktop application:

- web browser is available for all ordinarily available OS without a dependency on hardware components,
- theoretically seamless network connection of logical processes in various networks,
- it is free.

The reasons why the web simulation (through WebRTC) is not an ideal way:

- offending WebRTC,
- low computing performance of JavaScript,
- a problem of a connection to yet existing architecture/product relative to an unique reports size.

7.2. Comparison with other standard solutions

As noted in chapter 2.3 relative to found problems and limitations an aim of this work is to create the environment for realization of small simulations (to 20 logical processes) not to create a competition usually by a military solution.

Only a brief look back at contemporary most used standards:

- **HLA** is a very extensive and complex solution its implementation is very demanding according to all standards. For that reason using of HLA is wholly unpractical for models with a small number of logical processes on that this work is just focused on. (Kuhl, Judith, and Weatherly 2000)
- **DIS** is in general replaced by HLA today, from this view already only a thought of network connections after this pattern (each with each) is interesting but as described in chapter 2.3 this approach is not usable.
- **TENA** is an extensive complex architecture (as HLA is) created for needs of US Army

comprising LVC distributed simulations with emphasis on live simulation thus totally exceeds scope of this work. (Powell and Noseworthy 2012)

8. TESTING A USE CASE

For testing road/highway network was chosen because it is possible to set up almost any (and still realistically looking and behaving) distributed model from several prepared types of logical processes.

These types of logical processes are used (see fig. 8):

- highway with ramp on/off (fig. 8a),
- straight stretch of highway,
- highway multilevel intersection (fig. 8c),
- road stretch with small window with refreshments (fig. 8b, primarily for testing queue realization).

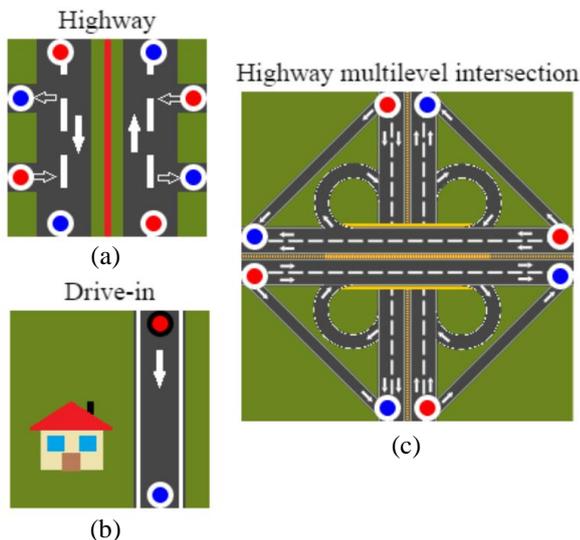


Figure 8: Examples of types of logical processes

The testing runs in local network, at 20 computers where only web browser runs with one opened panel. It is about a medium performance ordinarily available computers with Intel® Core™ i3 processor. Schemes of used models (topology of logical processes connections is primarily important) are in figure 9a-c. Model in figure 9c represents closed system that detains entities and serves to testing behavior of solution in extreme (i.e. normally absent) amount of entities. Then in a table 1 measured values for indexes are shown that normally guarantee smooth run of simulation. So these values indicate limits (especially table 1 / fig. 9c) of an applicability on above mentioned processor. A more powerful processor may enable you to smoothly animate a larger number of entities on other parameters will have no effect.

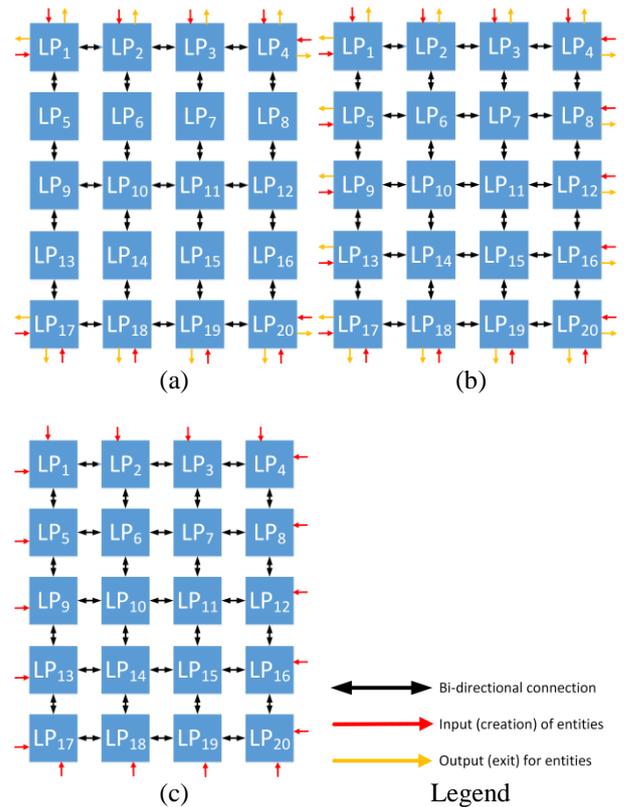


Figure 9: Schemas of tested models

Table 1: The average values of run of distributed simulations on models by fig. 9, lookahead 500 ms

Scheme	Schemes of model in fig. 9		
	Fig. 9a	Fig. 9b	Fig. 9c
Time from start [min]	30	30	15
Count of entities in	220	450	1400
Time between request for null message and response [ms]	18,7	15,4	12,2
Time between send WebRTC message and response delivery [ms]	5,39	9,5	7,7
FPS of animation output	141	125	103
Status	No problem, values continued in time unchanged		Last smooth anim.

9. FUTURE DEVELOPMENT

Still a space remains here for future development especially in terms of solving problem with more connections and stable logical processes connection among more local networks operated behind NAT routers.

The future development will be focused too to a possibility to save the status of logical processes at server and thus allow to supervisory staff to return to

required moment whether for any simulation play and for a initiation of different solution procedure.

So option will arise to later (after finishing simulation) repeat course or to change procedure from chosen moment of time.

Hereafter there is the space for the whole series of improvements starting with improving CV_m and CV_s synchronization. Creating of animation for monitoring entity transitions among logical processes, etc.

10. CONCLUSION

Presented solution creates the environment for service training or testing whenever and wherever using only web browser that today it is available practically at any computer connected to a computer network.

Just this feature is key and competitive against existing alternatives that usually cannot do without an specific, single purpose and costly software.

However JavaScript character and WebRTC problems yet close a path to extensive use and soon we cannot expect an essential progress.

A final product can find an application for smaller models (ca around 10 logical processes) or in static distributed scenes.

REFERENCES

- Fujimoto, R.M., 2000. *Parallel and distribution simulation systems*. New York: Wiley.
- Kartak, S., Kavicka, A., 2014. WebRTC Technology as a Solution for a Web-Based Distributed Simulation. *Proceedings of the European Modeling and Simulation Symposium 2014*, pp. 343–349. Genova: Università di Genova.
- Tropper, C., 2002. *Parallel and distributed discrete event simulation*. New York: Nova Science Pub Inc.
- Kuhl, Frederick, Judith Dahmann, and Richard Weatherly. *Creating computer simulation systems: an introduction to the high level architecture*. Upper Saddle River, NJ: Prentice Hall PTR, 2000.
- Bergkvist, A., Burnett, D.C., Jennings, C., Anant Narayanan, 2013. *WebRTC 1.0: Real-time Communication Between Browsers*. W3C. Available from: <http://www.w3.org/TR/webrtc/> [accessed date July 2014]
- Rosenberg, J., 2014. *Interactive Connectivity Establishment (ICE)*. IETF. Available from: <http://tools.ietf.org/html/rfc5245> [accessed date May 2014]
- Hridel, J., Kartak, S., 2013. Web-based simulation in teaching. *The European Simulation and Modelling Conference 2013*. EUROSIS-ETI.
- Powell, E.T., Noseworthy, J.R., 2012. *The Test and Training Enabling Architecture (TENA)*. Available from: <https://www.tena-sda.org/download/attachments/6750/TENA-2012-Paper-Final.pdf> [accessed date 1 July 2015]

IEEE standard for distributed interactive simulation: application protocols. New York: Institute of Electrical and Electronics Engineers, 2012.

IEEE standard for distributed interactive simulation communication services and profiles. New York: Institute of Electrical and Electronics Engineers, 1996. Print.

MODELING AND SIMULATION BASED ON FINITE ELEMENT METHOD FOR ELECTRON BEAM WELDING PROCESS

Zhengkun Feng^(a), Henri Champliaud^(b)

Department of Mechanical Engineering,
Ecole de technologie supérieure
1100 Notre-Dame Street West,
Montreal, Quebec, H3C 1K3, Canada

^(a)zhengkun.fengca@etsmtl.ca, ^(b)henri.champliaud@etstmtl.ca

ABSTRACT

Electron beam welding is a type of welding with high energy density that provides advantage for welding thick parts of high strength metal with single pass. In this paper, the development of a thermal structural model based on finite element method is presented. A conical heat source model is applied in the model. The temperature distribution and time history are reported with numerical simulation results. The residual stresses in the fusion zone and the heat affected zones have been analyzed.

Keywords: electron beam welding, conical heat source, Finite element analysis, Residual stress, ANSYS

1. INTRODUCTION

Since the environment protection gets increasing attention, high strength metals have more applications in industries. Electron beam welding has many advantages over other welding. This welding process provides deep penetration and does not require prepared welding groove and filler material and single pass can be performed for welding thick parts of high strength metals. Moreover, small heat affected zones can be obtained by this welding. During electron beam welding, the electron beam strikes the surface of the metal. As a result, melts evaporates and a hole is created. As the surface of the hole is continuously struck by the electron beam and the surface of the hole evaporates continuously, a keyhole is drilled by the electron beam metal (Andrews and Atthey 1976; Ho et al. 2007).

With the fast evolution of computing technology, modeling and simulation of complex process and system with large dimension becomes possible. Three-dimensional modeling of electron beam welding can be found in the late 1980's (Lindgren and Karlsson 1988). Rai et al. (2009) and Adamus et al. (2013) performed analyses of electron beam welding on thin plates. Lacki and Adamus (2011) performed numerical simulations of electron beam welding for joining tubes with different thicknesses. In this paper, a thermal-structural model based on finite element method is presented. The

geometry of the specimen to be welded is illustrated in Fig. 1. A conical heat source model is applied to the thermal-structural model. The temperature and stress distributions obtained from the numerical simulation results will be reported and discussed.

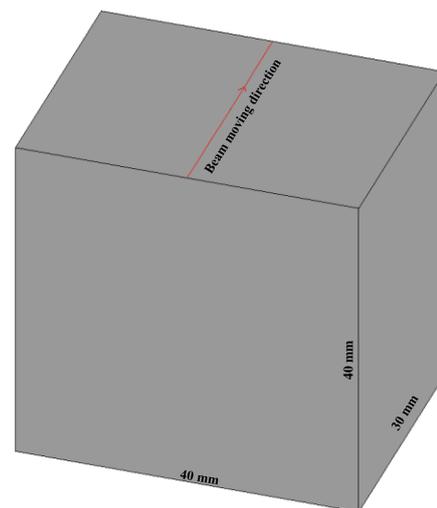


Figure 1: Geometrical set up

2. PROCESS MODELING BASED ON FINITE ELEMENT METHOD

As the geometry of the specimen was symmetric to the electron beam moving direction, when building the model, the boundary conditions were selected to make the deformation to be symmetric. Therefore, for the sake of saving CPU time when performing numerical simulations, only one half of the specimen was required for the symmetric model (Fig. 2). In the figure, yz plane was the symmetric plane and y was the electron beam moving direction. The numerical model based on finite element method with implicit time integration scheme was built on ANSYS platform. The model was divided into two parts: the thermal and structural parts. 8-node thermal and structural cubic elements were respectively selected for the mesh of the model with 38400 elements in the fine zone with 8 mm of width and 19200

elements in the coarse zone with 32 mm of width. The degrees of freedom (nodal temperature in the thermal part, nodal displacement u_x , u_y and u_z in the structural part) between the nodes of the fine and the coarse zones were coupled. Since the deformation was assumed to be small, thermal structural uncoupled sequential model was selected. The thermal part provided the temperature distributions at each instant during the welding as information to be used by the structural part to calculate the thermal expansion during the welding. For preventing the specimen from rigid body movement, u_y and u_z were zero at points (0, 0, -0.04), and u_x was zero at points (0, 0.03, -0.04). The material had the density of 7860 kg/m³ and Poisson ratio of 0.3. The other thermal and mechanical properties, such as the enthalpy H, the conductivity coefficient k, the convective coefficient h_f , the thermal expansion coefficient α , the Young's modulus E and the yield stress S_y were dependent of temperature as shown in table 1. The elastic-perfect-plastic bilinear material model with the Young's modulus E and the yield stress S_y was applied. The following conical heat flux distribution (Luo et al. 2010) was selected:

$$q = q_0 e^{\frac{-3h^2}{(z+h)^2} \left(\frac{x^2+y^2}{r_0^2} \right)} \quad (1)$$

where q denotes the heat flux in the conical zone; q_0 denotes the heat flux at the center of the cone; h denotes the depth of the heat source; r_0 denotes the radius of the cone bottom. In the model, q_0 , h, r_0 were selected as 137 W/mm³, 60 mm, 2 mm.

Table 1: Thermal and mechanical properties of material

T (°C)	H (MJ/m ³)	k (W/°Cm)	h_f (W/°Cm ²)	α (10 ⁻⁶ /°C)	E (GPa)	S_y (MPa)
20	0	51	9.5	12.18	165	345
100	296	49.5	11.8	12.18	180	
200	685	48	16.3	12.66	195	280
400	1544	42	30.8	13.47	173	150
600	2603	37	54.8	14.41	88	55
800	4840	26	90.8	12.64	75	40
1000	5859	26	141	13.37	70	15
1200	6877	28	208.1		60	
1400	7936	28	294.4		30	
1490	8412	50	340.8		17	
1570	10512	80	384.8		10	
1700	11200	120	465.4		10	
5000	28668	125	500		10	

3. SIMULATION RESULTS AND DISCUSSIONS

In the welding simulation, the simulation time was 5 seconds with the speed of the electron beam of 6 mm/s. Firstly, the thermal simulation was carried out with the thermal part of the model. Then, the structural

simulation was carried out with the structural part of the model as described in section 2.

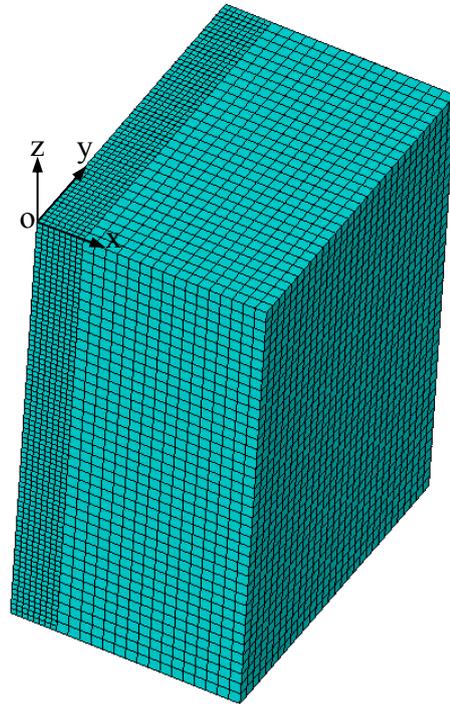


Figure 2: non-uniform mesh of the model with 8-node solid elements

3.1. Thermal Simulations

The initial temperature of the specimen was 20 °C. It was assumed that the welding was performed in vacuum chamber. The convective coefficients were selected as 100 times less than those listed in table 1. Figure 3 shows the temperature distribution on the cutting plane perpendicular to the beam moving direction at the instant of 3 seconds (For the clarity, the results are presented in complete model instead of symmetric model). It was observed that the temperature distribution was in conical shape. Figure 4 shows the temperature distribution on the symmetric plane at the instant of 3 seconds. After the welding, the specimen was cooled at free convection on all surfaces by air at room temperature of 20 °C. The convective coefficients were selected as shown in Table 2. As the temperature gradient of the specimen at the beginning was high, a small constant time step at the first 2 seconds was used to ensure the convergence. However, the temperature gradient became smaller and continued to rapidly decrease after 2 seconds. For the sake of CPU time, the current time step proportionally greater than each previous time step was selected. The total cooling time was 6 hours and the specimen temperature became 20 °C. Figure 5 shows the temperature distribution on the symmetric plane at the instant of 61 second that corresponding of the instant of 56 second after the start of the cooling. As the electron beam welding is a type of welding with high energy density and the heat affect zone is small, the temperature gradients in the center of the conical heat source and nearby are very strong. The

temperature gradient of point C near the instant of 2.5 second was the highest when the electron beam center reached to this point. This point was the geometric center of the specimen (Fig. 7). Figure 8 shows the temperature distributions on the cross line AB (Fig. 7) that passed through the specimen center and was perpendicular to the electron beam moving direction at different instants. It was observed that the temperature gradient at the position x of zero between instants 3 second and 6 second was much stronger than that between instants 6 second and 61 second. The nearby positions had the same temperature distribution tendencies.

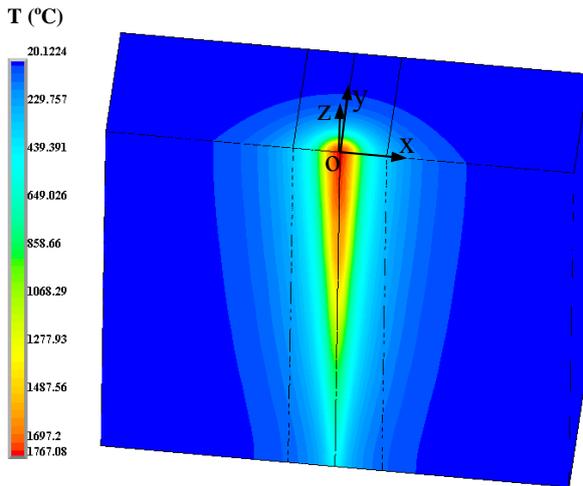


Figure 3: Temperature distribution on the cutting plane perpendicular to the beam moving direction at 3 second

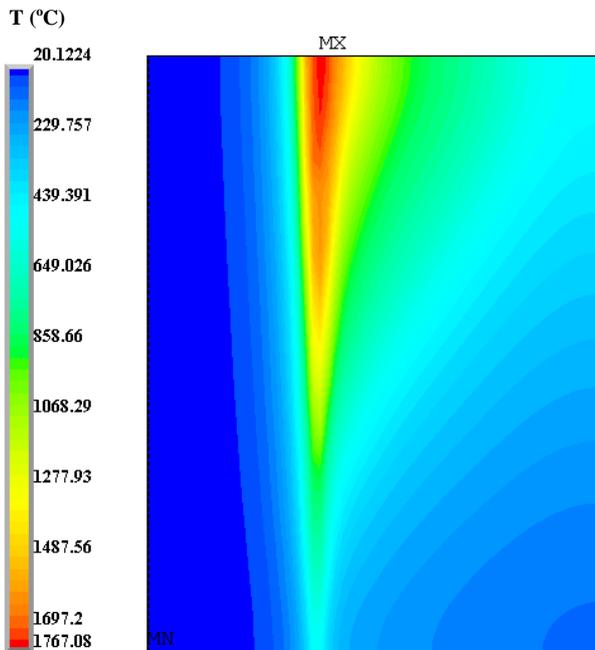


Figure 4: Temperature distribution on the symmetric plane at 3 second

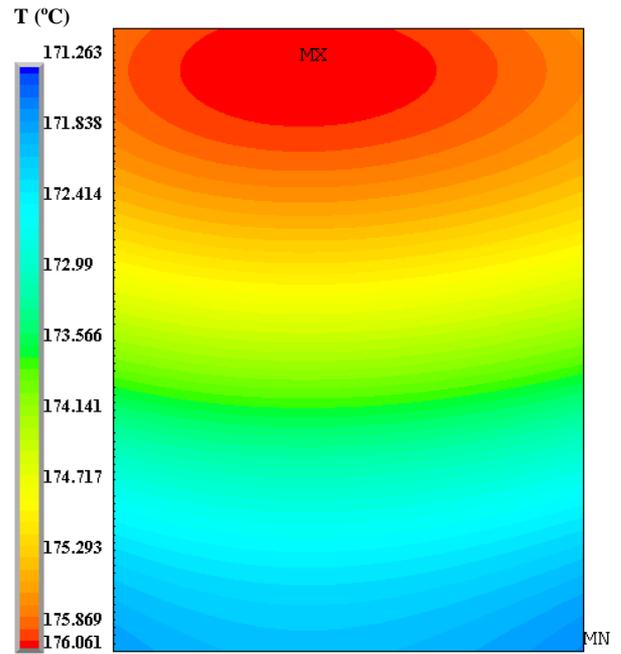


Figure 5: Temperature distribution on the symmetric plane at 61 second

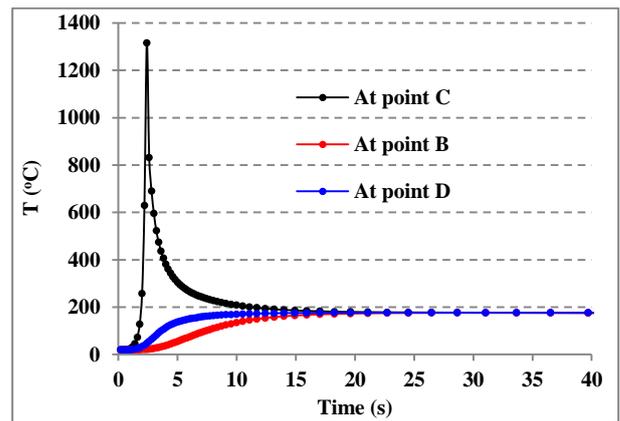


Figure 6: Temperature histories at selected positions

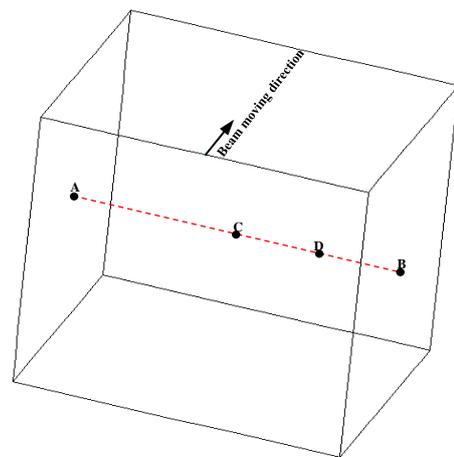


Figure 7: Position of the selected points and the cross line

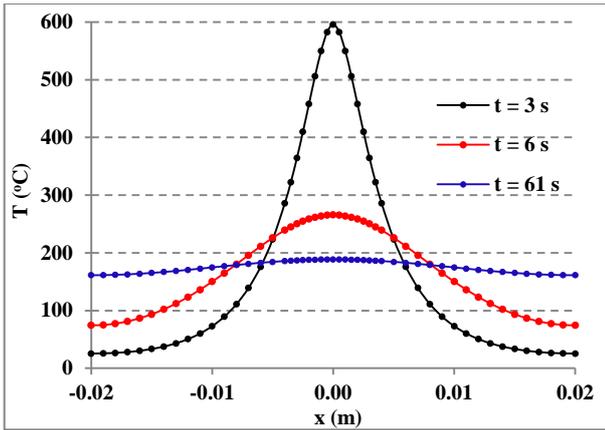


Figure 8: Temperature distributions on the cross line at different instants

3.2. Structural Simulations

After performing the thermal simulation, the temperature at each node of the model of the specimen was saved. According to this information with thermal properties, such as the thermal expansion coefficients and the mechanical properties of the specimen at different temperature, a structural simulation was performed with the structural part of the model to assess the electron beam welding process. Figure 9 shows the Von Mises stress histories of the three points in Fig. 7. It was observed that the Von Mises stress variations during the welding was very strong and got the values that approached to the final values. The Von Mises stress located at the center point C was higher than that at the point B located at the side surface and the one at the point D located between these two points. Figure 10 shows the residual stress distributions on the cross line AB of Fig. 7. The Von Mises residual stress was not at the center of the specimen but at a point in the heat affected zone with a distance of 6 mm from the center of the specimen. In addition, it was observed that there were a zone where the stress in the plane perpendicular to the thickness direction was compressive. Figure 11 shows the Von Mises stress distribution on the cross section cutting through line AB of Fig. 7 and perpendicular to the electron beam moving direction. It was observed that the maximum value was located at a point between points C and D. Figure 12 shows the Von Mises stress distributions on the cross sections perpendicular to the electron beam moving direction and cutting from the welding starting point A with distances of 2.5 mm, 7.5 mm, 12.5 mm, 17.5 mm, 22.5 mm and 27.5 mm, respectively. It was observed the Von Mises stress distribution partner was established in the range from 2.5 mm to 2.75 mm in the electron beam moving direction.

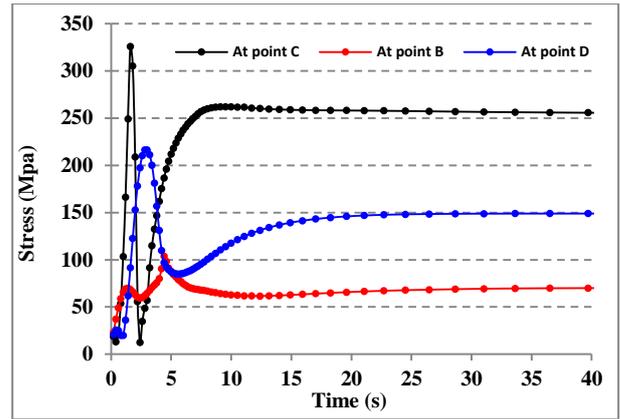


Figure 9: Von Mises stress histories at selected positions

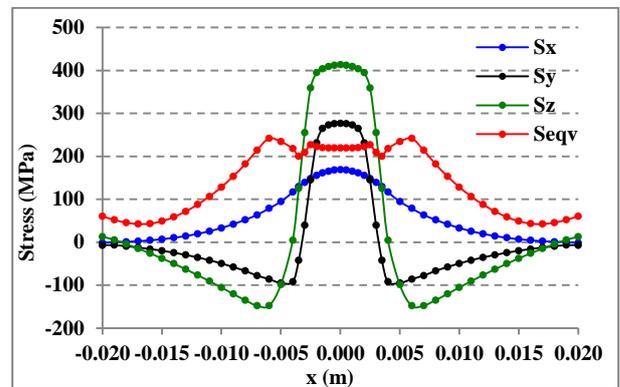


Figure 10: Residual stress distributions on the cross line AB

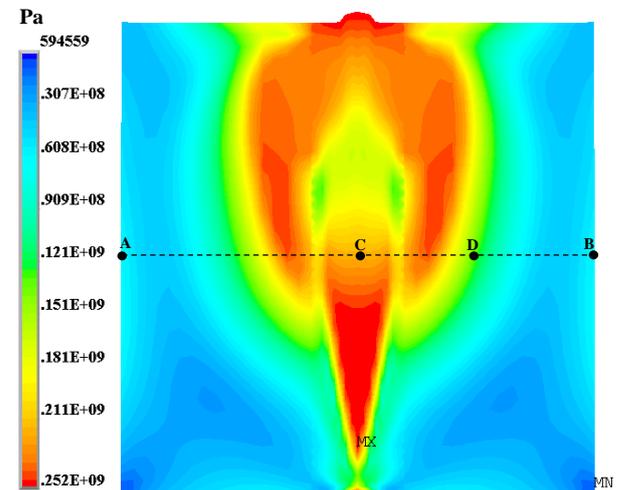


Figure 11: Residual Von Mises stress distribution on the cross section at the mid length

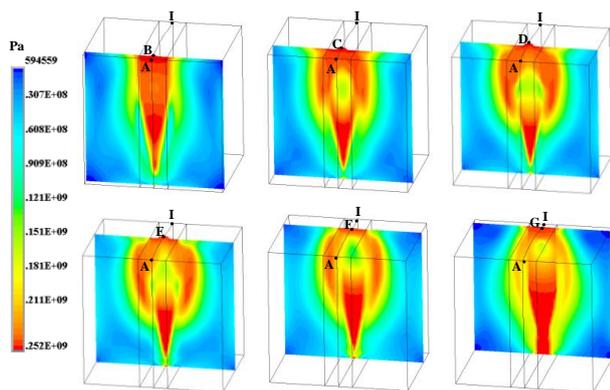


Figure12: Residual Von Mises stress distributions on different cross sections

4. CONCLUSIONS

The numerical model for electron beam welding process shows that this type of welding with high energy density provides deep penetration and is capable of welding thick parts of high strength metals. It shows also that the welding can be performed without prepared welding groove and with single pass. The heat affected zones are relatively small. However, reducing the maximum residual stress and the heat affected zones by pre-heating and/or post heating and validating the model with experimental results remain for the ongoing work.

ACKNOWLEDGMENTS

The authors thank the Natural Sciences and Engineering Research Council (NSERC), Alstom Hydro Canada Inc. and Hydro Quebec for their financial supports to this research.

REFERENCES

- Adamus, K., Kucharczyk, Z., Wojsyk, K., and Kudla, K. (2013). Numerical analysis of electron beam welding of different grade titanium sheets. *Computational Materials Science*, 77, 286-294. doi: 10.1016/j.commatsci.2013.05.001
- Andrews, J. G., and Atthey, D. R. (1976). Hydrodynamic limit to penetration of a material by a high-power beam. *Journal of Physics D: Applied Physics*, 9(15), 2181-2194. doi: 10.1088/0022-3727/9/15/009
- Ho, C. Y., Wen, M. Y., and Lee, Y. C. (2007). Analytical solution for three-dimensional model predicting temperature in the welding cavity of electron beam. *Vacuum*, 82(3), 316-320. doi: 10.1016/j.vacuum.2007.04.040
- Lacki, P., and Adamus, K. (2011). Numerical simulation of the electron beam welding process. *Computers and Structures*, 89(11-12), 977-985. doi: 10.1016/j.compstruc.2011.01.016
- Lindgren, L.-E., and Karlsson, L. (1988). Deformations and stresses in welding of shell structures. *International Journal for Numerical Methods in Engineering*, 25(2), 635-655.

Luo, Y., Liu, J., Ye, H., and Shen, B. (2010). Numerical simulation on electron beam deep penetration welding and weld appearance of magnesium alloy. *Hanjie Xuebao/Transactions of the China Welding Institution*, 31(6), 65-68.

Rai, R., Burgardt, P., Milewski, J. O., Lienert, T. J., and Debroy, T. (2009). Heat transfer and fluid flow during electron beam welding of 21Cr-6Ni-9Mn steel and Ti-6Al-4V alloy. *Journal of Physics D: Applied Physics*, 42(2). doi: 10.1088/0022-3727/42/2/025503

SEARCHING THE LONGEST COMMON SUBSEQUENCES IN DISTORTED DATA

Tomáš Kocyan^(a), Jan Martinovič^(a,b), Kateřina Slaninová^(a,b), Daniela Szturcová^(a,c)

^(a)IT4Innovations, ^(b)FEECS, Department of Computer Science, ^(c)Institute of Geoinformatics
VŠB - Technical University of Ostrava
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

tomas.kocyan@vsb.cz, jan.martinovic@vsb.cz, katerina.slavinova@vsb.cz, daniela.szturcova@vsb.cz

ABSTRACT

Many models and artificial intelligence methods work with the inputs in the form of time series. Success of many of them strongly depends on ability to quickly and precisely compare two time series or search the mutual parts. Such ability is especially crucial while recognizing characteristic patterns, indexing, prediction or compression. There are many algorithms able to handle that, however, many of them fail while processing distorted data. Unfortunately, the distortion is natural for many types of data collections, e.g. for measurements of natural phenomena such as precipitations, river discharge volume etc. This paper discusses the possibilities of searching such common subsequences in time series and presents a new approach for searching the longest common subsequences in distorted data. This approach is based on modified the dynamic time warping algorithm, which allows the effective processing distorted time series data.

Keywords: time series, dynamic time warping, longest common subsequence, distorted data

1. INTRODUCTION

Processing and analyzing time series data is very important task in many domains, especially in modeling and simulations. In this domain, time series data is often used as one of the simulation inputs, or can be produced as one of the simulation outputs. For this purpose, it is appropriate to be able to manage this type of data, e.g. describe the data nature, search in data in reasonable time, or to recognize characteristic patterns in a collection. Algorithms providing such functionality usually need a robust mechanism for comparing two time series or identifying their common parts (Kocyan, Martinovič, and Podhorányi, 2014). However, many of such mechanisms fail while processing distorted data (Muller 2007). Unfortunately, the distortion is natural for many types of data collections, e.g. for measurements of natural phenomena such as precipitations, river discharge volume etc. Moreover, methods searching the common subsequences are mostly focused only on processing categorical data. For processing the real data, they have to be modified (Esling and Agon 2012) or some kind of

categorization (Lin, Keogh, Wei, and Lonardi 2007) has to be performed.

This paper discusses the possibilities of searching such mutual subsequences in time series and presents a new approach for searching the longest common subsequences in distorted data. It is organized as follows: First, basic approaches for searching the mutual subsequences of categorical data are introduced in Section 2. Second, the DTW algorithm for comparing two distorted sequences will be described. In the Proposed solution, a new approach for searching the mutual subsequences in distorted data will be suggested. At the end, both advantages and disadvantages of the approach will be summarized and the future work will be outlined.

2. COMPARISON OF SEQUENCES

Sequence comparison is widely used in information retrieval and in molecular biology for calculating sequence alignments of proteins (Gusfiel 2008) where several algorithms and their modifications are presented. However, we can use these methods in other areas as well. There are two main basic groups of algorithms known for the comparison of two or more categorical sequences. The first group divides the algorithms by the fact of whether or not the sequences consist of ordered or unordered elements. The second group of algorithms focuses on the comparison of the sequences with different lengths and with possible error or distortion.

2.1. The longest common substring

The basic approach to the comparison of two sequences, in which the order of elements is important, is *The longest common substring method* (LCS). This is used in exact matching problems (Gusfiel 2008). It is obvious from the name of the method that its main principle is to find the length of the common longest substring. Given the two sequences $x = (x_1, x_2, \dots, x_n)$ of the length $n \in \mathbb{N}$ and $y = (y_1, y_2, \dots, y_m)$ of the length $m \in \mathbb{N}$, we can find such substring $z = (z_1, z_2, \dots, z_p)$, $\forall k \in \{1, \dots, p\} (z_k = x_{i+k-1} = y_{j+k-1})$, where $1 \leq i \leq n - p + 1$ and $1 \leq j \leq m - p + 1$.

The LCS method respects the order of elements within a sequence. However, the main disadvantage of this method is that it can only find the identical

subsequences, which meet the characteristics of substrings. Precisely speaking, this means that the elements in the subsequence must be contiguous. For some domains, in which a large amount of different sequences typically exists, this fact gives too strict limitation to solved the problems.

2.2. The Longest Common Subsequence

Unlike substrings, the objects in a subsequence might be intermingled with other the objects that are not in the sequence. *The longest common subsequence* method (LCSS) allows us to find the common subsequence (Hirschberg 1977). Given the two sequences $x = (x_1, x_2, \dots, x_n)$ of the length $n \in \mathbb{N}$ and $y = (y_1, y_2, \dots, y_m)$ of the length $m \in \mathbb{N}$, we can find the subsequence $z = (z_1, z_2, \dots, z_p)$, $\forall k \in \{1, \dots, p\} (z_k = x_{i_k} = y_{j_k})$, where $i_k < i_{k+1}$ and $j_k < j_{k+1}$. Contrary to the LCS method, the LCSS method allows (or ignores) these extra elements in the sequence, and therefore, it is immune to slight distortions.

2.3. The Time-Warped Longest Common Subsequence

When the similarity between compared sequences is defined as a function using a length of common subsequence, one characteristic of this method can be found. The length of the common subsequence is not immune to the recurrence of the identical elements that may occur only in one of the compared sequences. We can find such situations as a result of inappropriate sampling or any kind of distortion. In some applications it is suitable (or sometimes even required) to eliminate such type of distortions and to work with them like with the equivalent elements. The solution is in another method, the time-warped longest common subsequence (TWLCS) Guo and Siegelmann (2004). This method emphasises the recurrence of elements in one of the compared sequences. Due to this fact, the length of the common subsequence can be, in some cases, longer than the shorter length of the compared sequences.

2.4. Comparison of Real Time Series

Our goal is to search continuous mutual subsequences in distorted real data. However, none of the previously mentioned methods can be directly applied, because they process only a categorical data. There are usually two ways how to adapt them for processing the real data as well: On the one hand, the data can be first categorized as in (Lin, Keogh, Wei, and Lonardi 2007), and then the known methods can be used. On the other hand, some kind of tolerance for equality between two values can be defined as in (Esling and Agon 2012). If such tolerance is not overcome, the values are considered as identical, i.e. belonging into the same category.

Despite the fact there exists some ways how to adapt the methods, it usually brings an unnecessary inaccuracies. They are caused by the degradation real number fineness to an isolated category. We tried to avoid such loss of information while processing the real data, but maintain the main principles of searching

subsequences. To deal with a possible distortion, we decided to modify the dynamic time warping algorithm introduced in the following section, and supplement it by useful steps from the previously mentioned methods.

3. DYNAMIC TIME WARPING

Nowadays, searching and comparing the time series databases generated by computers, which consists of accurate time cycles and which achieves a determined finite number of value levels, is a trivial problem. A main attention is focused more likely on the optimization of searching speed. A non-trivial task occurs while comparing or searching the signals, which are not strictly defined and which have various distortions in time and amplitude. As a typical example, we can mention measurement of functionality of human body (ECG, EEG) or the elements (precipitation, flow rates in riverbeds), in which does not exist an accurate timing for signal generation. Therefore, comparison of such sequences is significantly difficult, and almost excluded while using standard functions for similarity (distance) computation. Examples of such signals are presented in Figure 1. A problem of standard functions for similarity (distance) computation consists in sequential comparison of the opposite elements in the both sequences (comparison of elements with the identical indexes).

DTW is a technique for finding the optimal matching of two warped sequences using pre-defined rules (Muller 2007). This approach was used for example for comparison of two voice patterns during an automatic recognition of voice commands (Rabiner 1993). Essentially, it is a non-linear mapping of particular elements to match them in the most appropriate way. The output of such DTW mapping of sequences from Figure 1 can be seen in Figure 2.

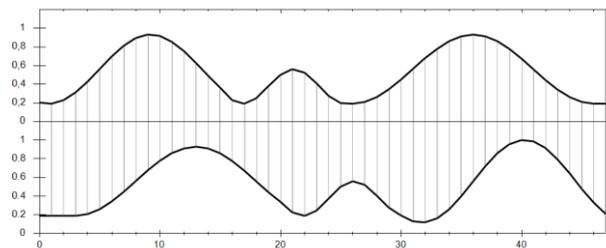


Figure 1: Standard Metrics Comparison

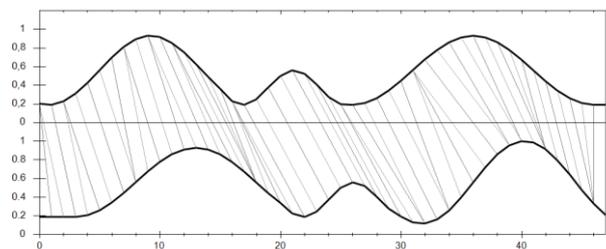


Figure 2: DTW Comparison

The main goal of DTW method is a comparison of two time dependent sequences x and y , where $x =$

(x_1, x_2, \dots, x_n) of length $n \in \mathbb{N}$ and $y = (y_1, y_2, \dots, y_m)$ of length $m \in \mathbb{N}$, and to find an optimal mapping of their elements. To compare partial elements of sequences $x_i, y_j \in \mathbb{R}$, it is necessary to define a local cost measure $c: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$, where c is small if x and y is similar to each other, and otherwise it is large.

Computation of the local cost measure for each pair of elements of sequences x and y results in a construction of the cost matrix $C \in \mathbb{R}^{n \times m}$ defined by $C(i, j) = c(x_i, y_j)$ (see Fig. 3). Then the goal is to find an alignment between x and y with a minimal overall cost. Such optimal alignment leads through the black valleys of the cost matrix C , trying to avoid the white areas with a high cost. Such alignment is demonstrated in Fig 4.

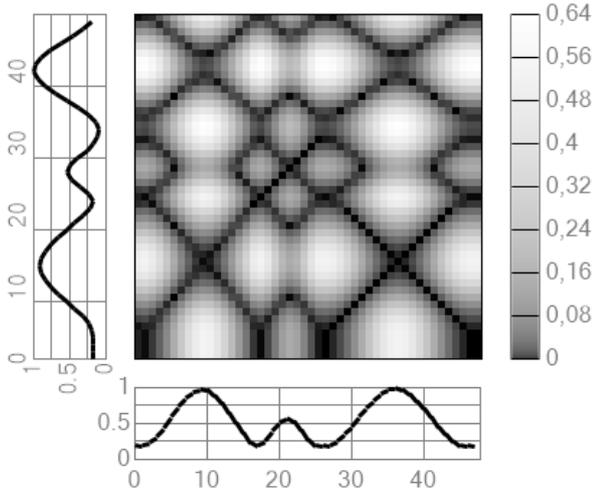


Figure 3: Cost Matrix

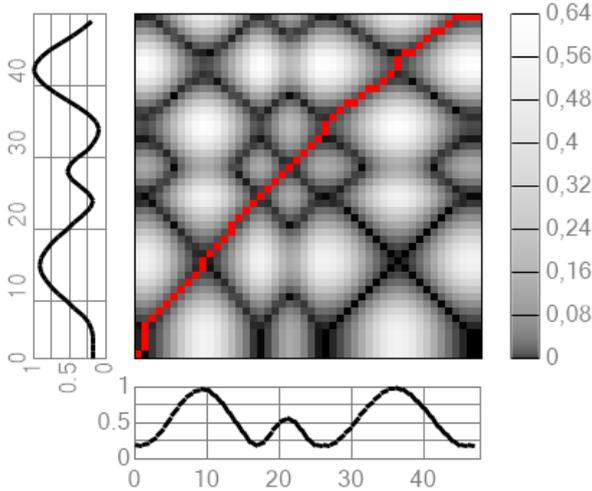


Figure 4: Cost Matrix with Found Warping

Basically, the alignment (called warping path) $p = (p_1, \dots, p_q)$ is a sequence of q pairs (warping path points) $p_k = (p_{kx}, p_{ky}) \in \{1, \dots, n\} \times \{1, \dots, m\}$. Each of such pairs (i, j) indicates an alignment between the i th element of the sequence x and j th element of the sequence y . Moreover, the found warping path has to satisfy the following three conditions:

1. Boundary condition:
 $p_1 = (1, 1)$ and $p_q = (n, m)$.
2. Monotonicity condition:
 $p_{1x} \leq p_{2x} \leq \dots \leq p_{qx}$ and
 $p_{1y} \leq p_{2y} \leq \dots \leq p_{qy}$.
3. Step size condition:
 $p_{k+1} - p_k \in \{(1, 0), (0, 1), (1, 1)\}$
for $k \in \{1, \dots, q - 1\}$.

The total cost $c_p(x, y)$ of the found warping path p between sequences x and y is then defined as a sum of partial local costs c :

$$c_p(x, y) = \sum_{k=1}^q c(x_{p_{kx}}, y_{p_{ky}})$$

As an optimal warping path p^* between x and y , the warping path having minimal total cost among the set P of all possible warping paths is selected:

$$p^* = \underset{p \in P}{\operatorname{argmin}} \{c_p(x, y)\}$$

The DTW “distance” $DTW(x, y)$ between x and y is then defined as:

$$DTW(x, y) = c_{p^*}(x, y)$$

Retrieval of optimal path p^* by evaluating all possible warping paths between sequences x and y leads to an exponential computational complexity. Fortunately, there exists a better way with a $O(n * m)$ complexity based on dynamic programming.

For this purpose, we have to define a function $subSeq: \mathcal{F}^n \times \mathbb{N} \times \mathbb{N} \rightarrow \mathcal{F}^m$, where \mathcal{F} is a feature space and $m \leq n$, that creates a new subsequence from a given sequence that is defined as $subSeq(x, a, b) = (x_a, x_{a+1}, \dots, x_b)$, where $a \leq b$. In the rest of the paper, we will use a shortened notation of this function specified as $x_{a:b} = subSeq(x, a, b)$.

For searching the definite optimal warping path, an accumulated cost matrix $D \in \mathbb{R}^{n \times m}$ can be utilized. The partial elements of this matrix are defined as:

$$D(r, s) = DTW(x_{1:r}, y_{1:s})$$

This can be easily computed in following way:

$$\begin{aligned} D(r, 1) &= \sum_{k=1}^r c(x_k, y_1) \text{ for } r \in \{1, \dots, n\}, \\ D(1, s) &= \sum_{k=1}^s c(x_1, y_k) \text{ for } s \in \{1, \dots, m\}, \\ D(r, s) &= \min \left\{ \begin{array}{l} D(n-1, m-1), \\ D(n-1, m), \\ D(n, m-1) \end{array} \right\} + c(x_r, y_s) \end{aligned}$$

for $r \in \{2, \dots, n\}$ and $s \in \{2, \dots, m\}$.

Computed accumulated cost matrix for a cost matrix from Fig. 3 can be seen in Fig. 5. It is evident that the accumulation highlights only a single black valley.

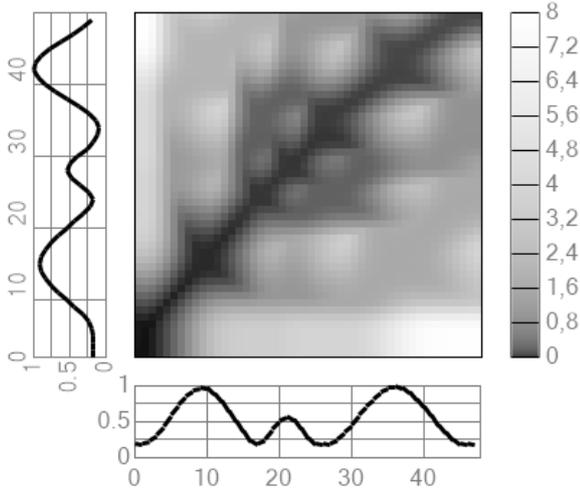


Figure 5: Accumulated Cost Matrix

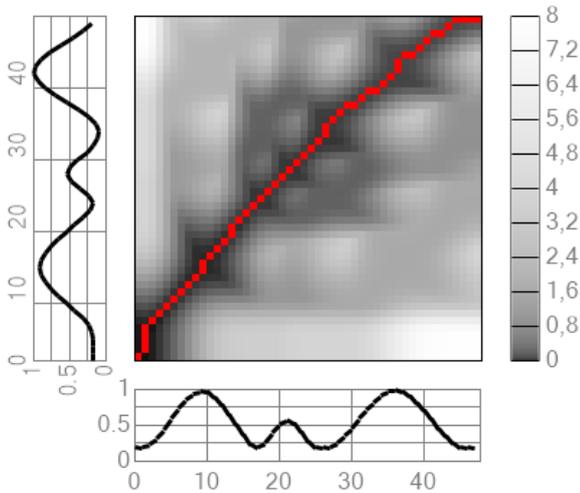


Figure 6: Accumulated Cost Matrix with a Found Warping Path

The optimal path $p^* = (p_1, \dots, p_q)$ is then computed in a reverse order starting with $p_q = (n, m)$ and finishing in a $p_1 = (1, 1)$. The rest of the points are defined recursively:

$$p_{k-1} = \begin{cases} (1, p_{ky} - 1) & \text{if } p_{kx} = 1 \\ (p_{kx} - 1, 1) & \text{if } p_{ky} = 1 \\ \argmin_{(i,j)} \left\{ D(i,j) \mid (i,j) \in \left\{ \begin{array}{l} (p_{kx} - 1, p_{ky} - 1), \\ (p_{kx} - 1, p_{ky}), \\ (p_{kx}, p_{ky} - 1) \end{array} \right\} \right\} & \text{otherwise} \end{cases}$$

3.1. Subsequence DTW

In some cases, it is not necessary to compare or align the whole sequences. A usual goal is to find an optimal alignment of a sample (a relatively short time series) within the signal database (a very long time series). It is very usual in situations, in which one disposes with a signal database a wants to find the best occurrence(s) of a sample (query). Using the slight modification, the DTW disposes with an ability to search such queries in a much longer sequence. The basic idea is not to penalize the omission in the alignment between x and y that

appears at the beginning and at the end of the sequence y . Suppose we have two sequences $x = (x_1, x_2, \dots, x_n)$ of the length $n \in \mathbb{N}$ and $y = (y_1, y_2, \dots, y_m)$ of the much larger length $m \in \mathbb{N}$. The goal is to find a subsequence $y_{a:b} = (y_a, y_{a+1}, \dots, y_b)$ where $1 \leq a \leq b \leq m$ that minimizes the DTW cost to x over the all possible subsequences of y . The modification involves the changed approach in the computation of the first row and column of accumulated cost matrix. The not penalizing the omissions at the beginning and at the end of the sequence y means not accumulating the values in the first column of the accumulated cost matrix, formally:

$$D(1, s) = c(x_1, y_s) \text{ for } s \in \{1, \dots, m\}$$

instead of

$$D(1, s) = \sum_{k=1}^s c(x_1, y_k) \text{ for } s \in \{1, \dots, m\}.$$

The remaining values of this matrix are computed in a standard manner as it was described earlier. Then, the warping path connecting the right and left side of the matrix is searched. The searching of the subsequence $y_{a:b}$ begins in a point:

$$(n, b) = \underset{i \in \{1, 2, \dots, m\}}{\operatorname{argmin}} \{ D(n, i) \}$$

and then it is computed in the same way as in classical DTW until the warping path touches the left side of a matrix in a point $(1, a)$, where $a \in \{1, \dots, b\}$. If the optimal subsequence warping path is defined as $p^* = (p_1, \dots, p_q)$, the $p_1 = (1, a)$ and the $p_q = (n, b)$. By selecting the next greatest value in the last column and searching the warping path from this point again, other paths can be found. An example of such searching the best subsequence alignment can be seen in Fig. 7. The both constructed matrices including the found warping path are then shown in Fig. 8.

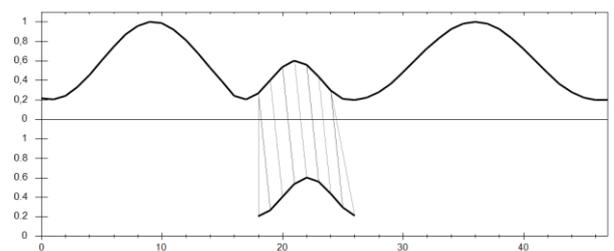


Figure 7: Cost Matrix and Accumulated Cost Matrix for Searching Subsequences

3.2. Searching the mutual subsequences

Despite the fact that the DTW has its own modification for searching subsequences, it works perfectly only in a case of searching an exact pattern in some signal database. However, in real situations, exact patterns are not available because they are surrounded by additional values (Figure 9a), or even repeated several times in the sequence (Figure 9b). Unfortunately, the basic DTW is not able to handle these situations and it fails or returns only a single occurrence of the pattern. To deal with this

type of situations, our own DTW modification was created.

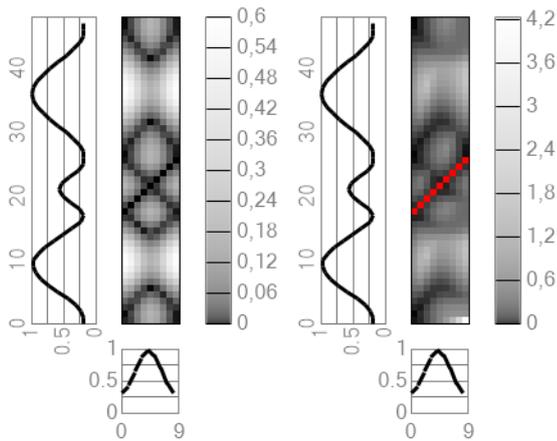


Figure 8: Cost Matrix and Accumulated Cost Matrix for Searching Subsequences

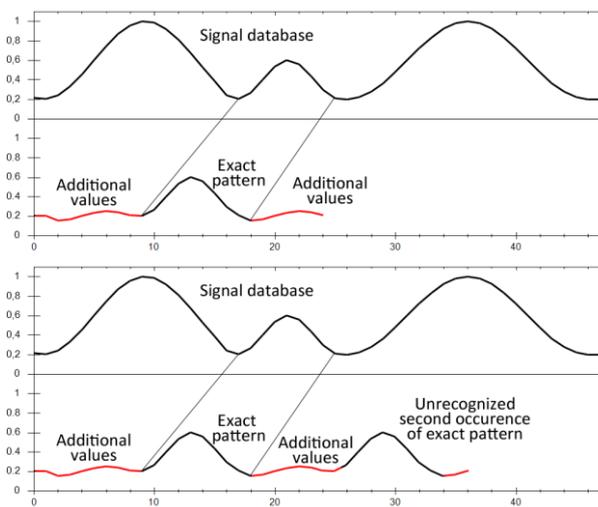


Figure 9: Basic DTW inaccuracies

4. PROPOSED SOLUTION

Proposed solution tries to eliminate weaknesses of previously mentioned approach and allows searching mutual time-warped subsequences respecting predefined conditions. The goal is to find all mutual subsequences as long as possible respecting the following three constrains (numeric thresholds):

- maximal total cost t_t ,
- maximal average cost t_a ,
- maximal single element cost t_e .

Functionality of the algorithm will be demonstrated on two sequences in the Fig. 10. Illustratively, each of these sequences consists of three same subsequences, but in the different order.

The biggest difference between the classical and subsequence DTW is in the philosophy of searching the warping path. In simple terms, the algorithm does not search the warping path from the upper right corner to the bottom left one (as in the case of classical DTW in Fig. 11a) and also it does not connect the opposite sides

of a matrix (as in the case of subsequence DTW in Fig. 11b). The main idea is to find warping paths as long as possible from any element to another one, parallel to a diagonal, as it is outlined in Fig. 11c. Moreover, the constructed warping path have to respect an adjusted constraints.

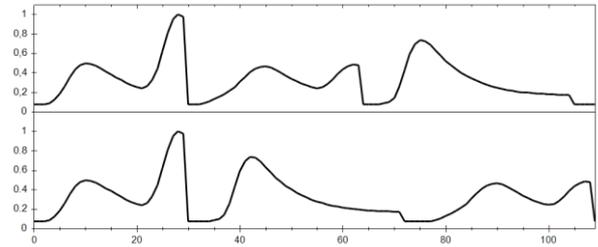


Figure 10: Sample Sequences

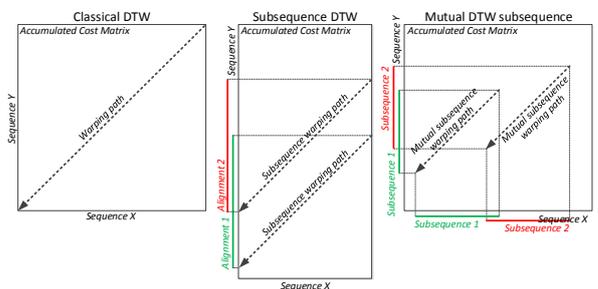


Figure 11: Philosophy of Searching the Warping Path

Visualized cost matrix in Fig. 12 suggests the optimal way for searching the warping paths. There are two important tasks to solve: Where exactly is the best to start the searching of the warping paths and how to adapt the computation of accumulated cost matrix for searching mutual subsequences.

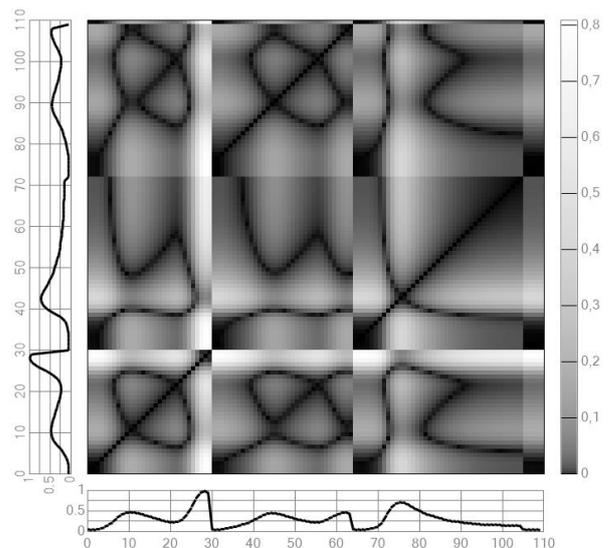


Figure 12: Visualized Cost Matrix

4.1. Determining the starting points

Unlike the DTW or Subsequence DTW, there is no clearly defined point (or set of points) where to start

searching the warping path. In this case, it is necessary to pick one or more matrix elements that are interesting in some manner. Test showed that local minima can be utilized as potential path starts. As the local minima such elements of the matrix are considered, whose values are less or equal than the value of surrounding values (except those in diagonal way). In another words, the element is considered as a local minimum if its value is lower or equal than the neighboring item above, below, right, left, below right and above left. Items below right and above left are ignored, because these elements are parallel to diagonal and lie on a potential same path. Formally, the set of minima M is defined as:

$$M = \{(i, j) \in \mathbb{N}^2 \mid i < m \wedge j < n \wedge \\ \wedge C(i, j) < C(i - 1, j + 1) \\ \wedge C(i, j) < C(i - 1, j) \\ \wedge C(i, j) < C(i, j - 1) \\ \wedge C(i, j) < C(i, j + 1) \\ \wedge C(i, j) < C(i + 1, j - 1) \\ \wedge C(i, j) < C(i + 1, j)\}$$

Fig. 13 shows that minima (red dots), found by this way for sequences from Fig. 10, exactly highlight the center of the dark areas and suggest the meaningful points.

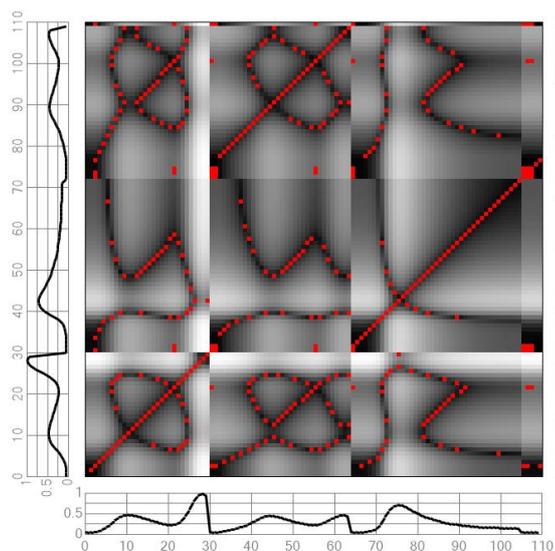


Figure 13: Visualized Cost Matrix Containing Minima

4.2. Computation of accumulated matrix

Once the potential paths' beginnings are found, searching the mutual subsequences may start. However, a standard accumulated matrix cannot be used, because the values are accumulated from the bottom left corner. Because of that, also constructed warping paths are "attracted" to the bottom left corner. It is mostly evident on the left and bottom side of the matrix in Fig. 14, where the color goes light instead of getting dark.

An original subsequence DTW allows searching the best alignment of a sample within the signal database by not penalizing omissions alignment at the beginning and end of the longer sequence by the following setting:

$$D(1, s) = c(x_1, y_s) \text{ for } s \in \{1, \dots, m\}.$$

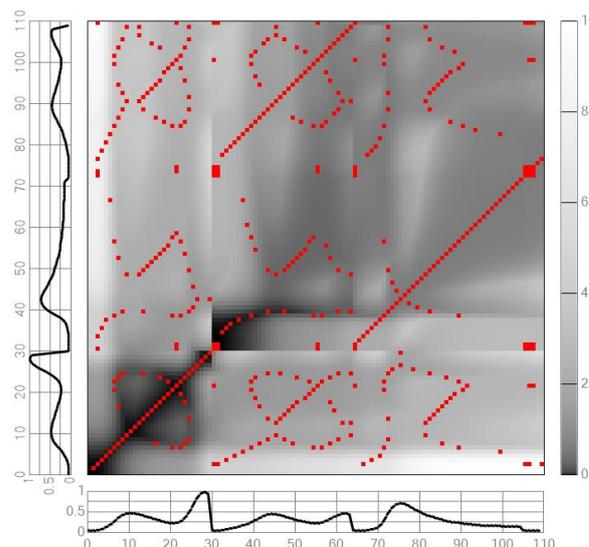


Figure 14: Accumulated Cost Matrix

Our approach additionally does not penalize even omissions of the alignment at the beginning and at the end of the shorter sequence by setting:

$$D(r, 1) = c(x_r, y_1) \text{ for } r \in \{1, \dots, n\}.$$

Accumulated cost matrix in Fig. 15 shows that subsequences located along borders are now clearly visible. However, subsequences located in the middle or at the ends of the sequences are still hidden. Moreover, the visualization suggests that warping will be "attracted" to the borders of a matrix, not diagonally as it is wanted.

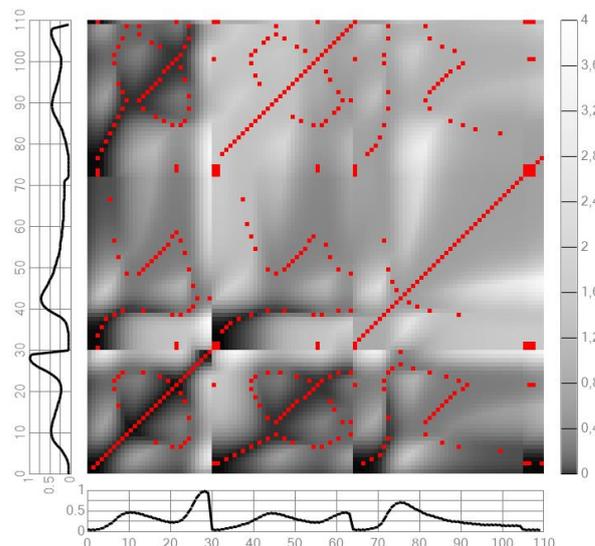


Figure 15: Modified Accumulated Cost Matrix

To solve this situation, the set of minima can be utilized again. We supplemented the computation of the accumulated cost matrix by zeroing the items located in minima.

$$D(r, s) = \begin{cases} 0 & \text{if } (r, s) \in M \\ \min \left\{ \begin{array}{l} D(n-1, m-1), \\ D(n-1, m), \\ D(n, m-1) \end{array} \right\} + c(x_r, y_s) & \text{otherwise} \end{cases}$$

for $r \in \{2, \dots, n\}$ and $s \in \{2, \dots, m\}$.

This zeroing causes resets of accumulation in minima points and changing the direction of warping path to the closest minimum. This effect is mostly evident in Fig. 16 on the right middle side.

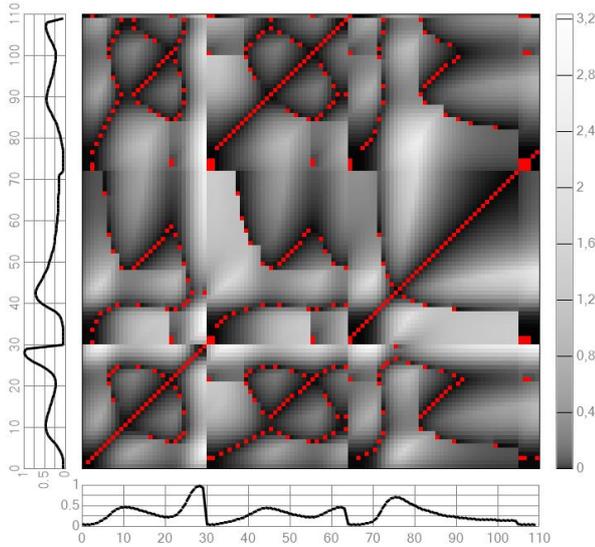


Figure 16: Modified Accumulated Cost Matrix with Zeroed Minima

4.3. Searching the subsequent warping path

Searching the mutual subsequence warping path is performed in the same way as in classical DTW. The only difference is in selecting the local minima as starting points, as it was defined in paragraph 4.1.

The local minima are taken one after another (from the top right corner of a matrix) and from the corresponding matrix element a new warping path is constructed. For a faster run of this algorithm, each of the minima does not have to be examined. A minimum contained in any of the already existing warping paths can be skipped.

Each of such warping paths is constructed with no restrictions, i.e. it is constructed until it touches the bottom or the left side of the matrix, regardless the thresholds defined in Sec. 4. It is done because it is not possible to evaluate some of the thresholds while constructing the warping path, because the overall view onto it is missing. It is evident especially for the maximal average cost t_a , where evaluating the threshold during the construction of a warping path may cause its premature end. This will be demonstrated in the following simple example.

Imagine the situation we want to find a subsequence with a maximal average cost $t_a = 0.2$ and the found warping path points' costs are:

$$c_s = \left(c(x_{p_{5x}}, y_{p_{5y}}), c(x_{p_{4x}}, y_{p_{4y}}), \dots, c(x_{p_{1x}}, y_{p_{1y}}) \right) = (0, 0, 1, 0, 0)$$

If the average cost will be computed and evaluated point by point, the construction of the path will end just after the third point, because:

$$c_{s_{1:3}} = \frac{0 + 0 + 1}{3} = 0.\bar{3} > t_a = 0.2$$

However, this ending of the construction is unnecessarily, because the average costs of the whole warping points is:

$$c_{s_{1:5}} = \frac{0 + 0 + 1 + 0 + 0}{5} = 0.2 \leq t_a = 0.2,$$

The interruption was caused by a single greater value, which does not have important meaning in the global view. To avoid this inaccuracies, the warping paths with no limitations are constructed first and then the final mutual subsequences exactly respecting the adjusted thresholds are extracted.

4.4. Extraction of mutual subsequences

In general, once the unlimited paths in 4.3 are found, the goal is to extract set of all subsequences $S = \{s | s = p_{u:v}\}$ for each warping path $p = (p_1, \dots, p_q)$, where following conditions are satisfied:

- $c_t(s) \leq t_t$,
- $c_a(s) \leq t_a$
- $c_e(s) \leq t_e$,
- found sequence $s = p_{a:b}$ is not a subsequence of any other found sequence $u = p_{c:d}$ also satisfying defined conditions, where $1 \leq c \leq a \leq b \leq d \leq q$.

The t_t , t_a , t_e are thresholds defined in Sec. 4 and the $c_t(s)$, $c_a(s)$ and $c_e(s)$ for a subsequence s are costs defined as follows:

- Total cost $c_t(s) = \sum_{k=u}^v c(x_{p_{kx}}, y_{p_{ky}})$,
- average total cost $c_a(s) = \frac{c_t(s)}{v-u+1}$,
- maximal item cost

$$c_e(s) = \max_{k \in \{u, \dots, v\}} \{c(x_{p_{kx}}, y_{p_{ky}})\}$$

To ensure extracting all the mutual subsequences and to avoid unnecessary shortening of them, the all combinations have to be examined. For this purpose, a two triangular matrices are constructed – matrix of total costs and a matrix of average costs. The partial elements of the matrices symbolize computed sums of costs (respectively averages of costs) between the i -th and j -th item of a warping path. Formally:

$C_t \in \mathbb{R}^{n \times m}$ defined by $C_{t_{ij}} = \sum_{k=i}^j c(x_{p_{kx}}, y_{p_{ky}})$.

$C_a \in \mathbb{R}^{n \times m}$ defined by $C_{a_{ij}} = \frac{\sum_{k=i}^j c(x_{p_{kx}}, y_{p_{ky}})}{j-i+1}$.

From these matrices, a sets of pairs describing the start and the end of mutual subsequences can be derived:

$$B_t = \{(i, j) \in \mathbb{N}^2 \mid C_t(i, j) \leq t_e\}$$

$$B_a = \{(i, j) \in \mathbb{N}^2 \mid C_a(i, j) \leq t_a\}$$

To receive all mutual subsequences respecting both maximal cost threshold and maximal average cost threshold, the intersection of these sets have to be defined as:

$$B_s = B_t \cap B_a$$

Unfortunately, an appropriate matrix for maximal item cost cannot be built, so we used the set of forbidden indexes which cannot be included in any mutual subsequence. At these indexes, the costs in a warping path are larger than the threshold t_e :

$$I_f = \{i \in \mathbb{N} \mid c(x_{p_{ix}}, y_{p_{iy}}) > t_e\}$$

The final set of pairs respecting all the thresholds are:

$$B_f = \{(i, j) \in B_s \mid \neg \exists k \in I_f [i \leq k \leq j]\}$$

However, this pairs include all the possible mutual subsequences including also the subsequences of subsequences. In another words, this set can contain a pair (0,10) as well as its subsequence (2,8). The (2,8) have to be ignored, because the goal is to find as long mutual subsequence as it is possible.

Omissions of such subsequences of subsequences can be easily done by ignoring all pairs having their begins and ends inside of another pair:

$$B = \{(i, j) \in B_f \mid \neg \exists (k, l) \in B_f [k \leq i \wedge l \geq j]\}$$

The final pairs $(i, j) \in B$ specify the indexes of the start and the end warping points $p_{i:j}$ enclosing the warping points of a mutual common subsequence. The values of mutual subsequences are there defined as:

$$\begin{aligned} x_s &= (x_{p_{ix}}, x_{p_{(i+1)x}}, \dots, x_{p_{jx}}) \\ y_s &= (y_{p_{iy}}, y_{p_{(i+1)y}}, \dots, y_{p_{jy}}) \end{aligned}$$

An example of such found mutual subsequences for the sequences from Fig. 10 can be found in Fig. 17. The corresponding warping paths are also visualized in the cost matrix in Fig. 18.

However, this example is presented only for a demonstration purposes, because the processed sequences are artificially constructed from the identical sequences and they are not a subject of the real distortion. Visualization of many sequences with corresponding combinations of algorithm's parameters, unfortunately, exceeds the scope of this paper.

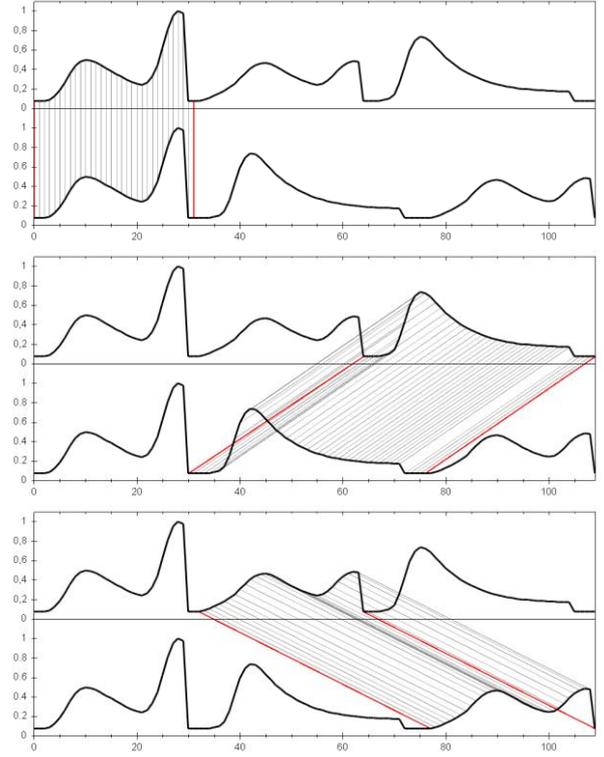


Figure 17: Found Mutual Subsequences

CONCLUSION

Tests showed that the proposed algorithm is able to find mutual subsequences in distorted time series. Since the algorithm is mostly based on DTW and mutually independent parts, it is very easy to parallelize the computations and rapidly speed it up. Future work will be focused on such optimization and automatic adjustment of algorithm's parameters.

ACKNOWLEDGEMENT

This work was supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and the national budget of the Czech Republic via the Research and Development for Innovations Operational Programme, by the project New creative teams in priorities of scientific research (reg. no. CZ.1.07/2.3.00/30.0055), supported by Operational Programme Education for Competitiveness and by the internal grant agency of VŠB - Technical University of Ostrava, Czech Republic, under the project no. SP2015/157 'HPC Usage for Transport Optimisation based on Dynamic Routing'.

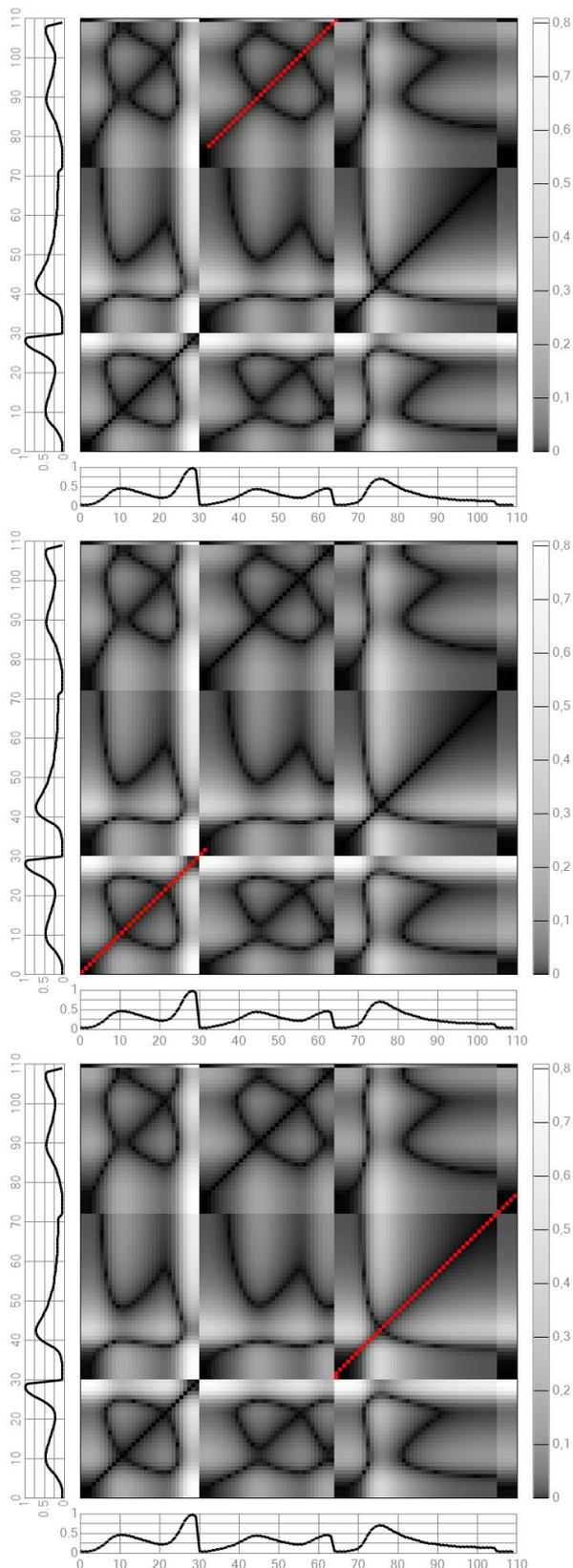


Figure 18: Found Warping Paths

REFERENCES

- Cohen, P. R., Adams, N., and Heeringa, B., 2007. Voting Experts: An Unsupervised Algorithm for Segmenting Sequences. *In Journal of Intelligent Data Analysis*.
- Chen, J., Huang, K., Wang, F., and Wang, H., 2009. E-learning Behavior Analysis Based on Fuzzy Clustering. *In Proceedings of International Conference on Genetic and Evolutionary Computing*.
- Esling, P., Agon C., 2012. Time-series data mining. *ACM Computing Surveys*, Volume 45, Issue 1, November 2012, Article number 12.
- Goldin, D., Kanellakis, P., 1995. On similarity queries for time-series data: Constraint specification and implementation. *1st International Conference on the Principles and Practice of Constraint Programming*, pages 137–153, France.
- Guo, A., Siegelmann, H., 2004. Time-Warped Longest Common Subsequence Algorithm for Music Retrieval. *In Proceedings of 5th International Conference on Music Information Retrieval ISMIR 2004*, pages 258–261. Universitat Pompeu Fabra, 2004.
- Gusfield, D., 2008. Algorithms on Strings, Trees and Sequences, *Computer Science and Computational Biology*. Cambridge University Press.
- Hirschberg, D. S., 1977. Algorithms for the Longest Common Subsequence Problem. *ACM*, 24:664–675.
- Keogh, E., Pazzani, M., 2000. Scaling up dynamic time warping for datamining applications. *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*. Pages 285-289
- Kocyan, T., Martinovič, J., Dráždilová, P., and Slaninová K., 2013. Recognizing Characteristic Patterns In Distorted Data Collections. *Proceedings of The European Modeling and Simulation 2013*.
- Kocyan, T., Martinovič, J., and Podhorányi, M., 2014. Searching and indexing distorted data collections. *Proceedings of The European Modeling and Simulation 2014*.
- Lin, J., Keogh, E., Wei, L., Lonardi, S., 2007. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, v.15 n.2, p.107-144, October 2007.
- Miller M., Wong P., and Stoytchev A., 2009. Unsupervised Segmentation of Audio Speech Using the Voting Experts Algorithm. *Proceedings of the Second Conference on Artificial General Intelligence (AGI)*.
- Muller, M., 2007. Dynamic Time Warping. *Information Retrieval for Music and Motion*, Springer, ISBN 978-3-540-74047-6, 69--84.
- Rabiner, L., Biing-Hwang Juang, 1993. *Fundamentals of Speech Recognition*, Prentice Hall, ISBN: 0130151572.

SIMULATIVE ANALYSIS OF PERFORMANCE SHAPING FACTORS IMPACT ON HUMAN RELIABILITY IN MANUFACTURING ACTIVITIES.

Valentina Di Pasquale^(a), Salvatore Miranda^(b), Raffaele Iannone^(c), Stefano Riemma^(d)

^{(a),(b),(c),(d)}Department of Industrial Engineering, University of Salerno, Via Giovanni Paolo II, 132 - 84084 – Fisciano (SA), Italy (Tel: +39 089 964106)

^(a)vdipasquale@unisa.it, ^(b)smiranda@unisa.it, ^(c)riannone@unisa.it, ^(d)riemma@unisa.it

ABSTRACT

In performing Human Reliability Analysis (HRA), the environmental and personal factors impact on human performance and enhance or degrade the human error probability (HEP). Their modelling and quantification is one of the most complex issues in the HRA field, to which many researchers recently are concentrating their efforts. This paper focuses on the eight performance shaping factors (PSFs), used in the Simulator for Human Error Probability Analysis (SHERPA), in order to understand how they affect HEP, considering especially their mutual interactions. The one-factor-at-a-time method was used to assess singularly the levels of each factor, while the analysis of variance allowed to realize if there is more or less dependence between them. The results highlight statistical relevance and interactions between factors when the factors with high multiplier value are combined, like experience, procedures and ergonomics. These outcomes can be useful in the enhancement of work context in order to reduce errors.

Keywords: Human Reliability Analysis, Performance Shaping Factors, Human error modelling and simulation

1. INTRODUCTION

The identification, assessment and reduction of risks are one of the most important elements of health and safety at work (Fera and Macchiaroli 2009). Evidence in the literature shows that human actions are a source of vulnerability for industrial systems (Griffith and Mahadevan 2011, De Felice et al. 2012, Di Pasquale et al. 2013, Iannone et al. 2004). In the workplace human errors can have serious consequences such as accidents, malfunctions and quality defects in the performed task. Human Reliability Analysis (HRA) is carried out, as part of the Probabilistic Risk Assessment (PRA), to identify and to quantify human actions and the associated impacts on structures, systems, and components of complex facilities, through the forecast of the events that can occur during the working activity. The HRA field aims to identify the causes and sources of human errors and to provide a numeric estimate of the human error probabilities (HEP). HRA typically encompasses three phases (see Figure 1), ranging from

identifying error sources, to modelling these errors as part of a systemic analysis including hardware failures, to quantifying the HEPs (Boring 2010).

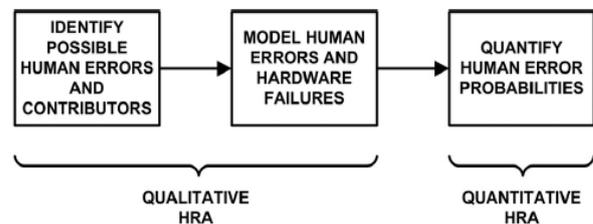


Figure 1: Three Phases of HRA (Boring 2010).

One of the undisputed assumptions in all HRA method is that the human performance depends on the conditions under which the tasks or activities are carried out (De Ambroggi and Trucco 2011). In the HRA methods, conditions that influence human performance are often referred by term performance shaping factor (PSF). They are used in qualitative approaches in order to identify contributors to human performance, while in quantitative ones, they are used to estimate a more realistic HEP. These contextual factors characterize significant facets of human error and they are environmental or personal factors, that have the potential to affect the man performance positively or negatively (Di Pasquale et al. 2013). Therefore, identifying and quantifying the PSF effects are key steps in the process of HRA (Griffith and Mahadevan 2011).

In this paper a simulative analysis of PSFs impact on human reliability in manufacturing activities was carried out thanks to the Simulator for Human Error Probability Analysis (SHERPA), recently proposed in literature. The tool is able to estimate human reliability for different workplaces and to assess the impact of context via several performance shaping factors, as described below (Di Pasquale et al. 2015).

Such factors require an evaluation to understand how they affect human reliability, above all considering their mutual interactions. The influence of every PSF level considered both in singular way and combining with the others factors, was quantified through numerous simulations in SHERPA. The one-factor-at-a-time

(OFAT) method was used to evaluate singularly the levels of each factor, while the analysis of variance (ANOVA) allowed to realize if there is more or less dependence between them.

The simulation outcomes can be useful in the enhancement of work context in order to reduce errors. The main conclusion is that the consideration of the factors influencing human performance is very important in the identification of the corrective action to reduce the risk (Farcasiu and Prisecaru 2012).

The paper is organised as follows. Section 2 summarises the state-of-the-art on performance shaping factors. Section 3 describes the SHERPA model and the PSFs using in it. Section 4 describes the methodology issues related to the simulative analysis. Section 5 shows the simulation results and analyses them. Finally the last section summarises the main findings and the conclusions.

2. STATE-OF-THE-ART PERFORMANCE SHAPING FACTORS

The PSFs are determined by the individual characteristics of the human being, the environment, the organization or the activity that enhances or decreases human performance and increases or decreases the likelihood of human error. Their goal is to provide measures to account for human performance. Performance shaping factors (PSFs) encompass those influences that enhance or degrade human performance. PSFs are used within human reliability analysis (HRA) methods to identify contributors to human errors and to provide a basis for quantifying those contributors systematically. While completing an HRA, an analyst may review a list of possible PSFs to identify possible sources of human error. The analyst may subsequently use predefined error rates associated with specific PSFs to determine a human error probability for a given task or situation (Boring et al. 2007). The PSFs are an integral part of the modelling and characterization of errors and play an important role in the HRA process.

The first-generation HRA methods are less concerned with what people are likely to do than with whether they will succeed or fail (Lee et al. 2011). None of these approaches consider explaining how the PSFs exert their effect on performance. PSFs such as managerial methods and attitudes, organizational factors, cultural differences, and irrational behaviour are not adequately treated in the first-generation. On the contrary, the second-generation considers the context in which humans make errors and derives PSFs based on these contexts. PSFs in the first-generation HRA methods were mainly derived by focusing on the environmental impacts on operators, whereas in the second one they were derived by focusing on the cognitive impacts on operators (Lee et al. 2011).

Within HRA, PSFs are often categorized as internal or external, corresponding to the individual vs. situational or environmental circumstances, respectively, that brings to bear on performance. The research literature divides the PSFs into two other categories: direct and

indirect measures of human performance (Boring et al. 2007). While some popular PSFs such as “time needed to complete a task” are directly measurable, other PSFs, such as “fitness for duty,” can primarily be measured indirectly through other measures and PSFs, for example through fatigue measures.

Their definition and classification, although complex and variable, have been carefully detailed by researchers who have proposed over time numerous taxonomies. There has been a greater emphasis recently to catalogue ways in which PSFs might also enhance performance and to develop taxonomy of performance influencing factors for HRA of emergency tasks (Lee et al. 2011, Kim and Jung 2003, Boring 2010).

On the other hand, the interrelationships between PSFs gain much attention from the HRA community. Despite continuing advances in research and applications, one of the main weaknesses of current HRA methods is their limited ability to model the mutual influence among PSFs, intended both as a dependency among the states of the PSFs dependency among PSFs influences on human performance, as shown in Figure 2 (De Ambroggi and Trucco 2011).

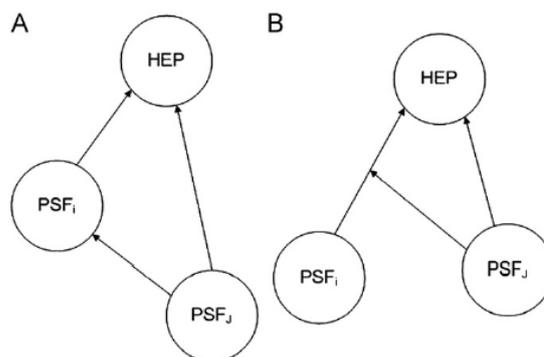


Figure 2: Possible types of dependency between PSFs: (A) dependency between the states/presence of the PSFs and (B) dependency between the state of PSF_j and the impact of PSF_i over the HEP.

Very different conceptual and analytical models are proposed for describing how these factors exert their influence on the human error probability; indeed if a PSF has an effect on human performance it is crucial to account for how this influence comes about. Several studies argued that the dependency between PSFs should be included in the quantification of HRA and suggested that the Bayesian network (BN) would be a promising technique because it can describe the casual relationship between PSFs (Groth and Swiler 2013, Groth and Mosleh 2012).

The study of De Ambroggi and Trucco (2011), instead, deals with the development of a framework for modelling the mutual influences existing among PSFs and a related method to assess the importance of each PSF in influencing performance of an operator, in a specific context, considering these interactions.

3. PERFORMANCE SHAPING FACTORS IN SHERPA

Di Pasquale et al. (2015) proposes a new use of HRA methodologies for human behaviour modelling and simulation through the implementation of a Simulator for Human Error Probability Analysis (SHERPA), which aims to predict the likelihood of operator error, for a given scenario, in every kind of industrial system or other type of working environment. SHERPA exploits the advantages of the simulation tools and the traditional method HRAs and it is able to estimate human reliability and the impact of context via PSFs.

Unlike many existing HRA methods, that are deeply qualitative and include excessive levels of detail for many assessments, SHERPA focuses on the quantitative aspect in order to obtain a significant numerical result in terms of HEP.

Furthermore SHERPA is suitable for any environment and working conditions, without limitations related to a particular sector or activity; while first and second generation HRA methods have been developed for specific contexts (e.g. aviation or nuclear power plants) and applications (e.g. single operator or crew simulation). Methods such as THERP (Technique for Human Error Rate Prediction), CREAM (Cognitive Reliability and Error Analysis Method) or SPAR-H (Standardized Plant Analysis Risk-Human Reliability Analysis method) were born and are used mainly in the nuclear field and so much effort needs to be applied in different fields, such as manual assembly or manufacturing systems (Di Pasquale et al. 2015). SHERPA can be used in the preventive phase, as an analysis of the possible situations that may occur and for the evaluation of the percentage of scraps or noncompliant processing due to human error and in post-production to understand the nature of the factors that influence human performance in order to reduce errors (Di Pasquale et al. 2015).

The working environment and the operator conditions in SHERPA are taken into account considering the performance shaping factors of SPAR-H method. While many HRA methods have often proposed a large number of PSFs, even as many as fifty, SPAR-H attempts to provide a reasonable coverage of the influence spectra of human performance in a reasonable minimum number of PSFs. The eight PSFs are:

- available time;
- stress;
- complexity;
- experience and training;
- procedures;
- cognitive ergonomics;
- fitness for duty;
- work process.

The decision to use only eight PSFs in SPAR-H is based on a review of the available HRA methods and the behavioural sciences (Boring 2010). In SHERPA this number is increased because of the sub-factors

modelled for stress, complexity and work processes (see Table 1).

Table 1. Sub-factors modelled in SHERPA.

PERFORMANCE SHAPING FACTORS			
	<i>Stress</i>	<i>Complexity</i>	<i>Work processes</i>
Sub-factors	- Circadian rhythm	- General complexity	- General work processes
	- Mental stress	- Mental efforts required	- Communication and integration in team work
	- Pressure time	- Physical effort required	
	- Workplace	- Precision level of the activity	
	- Microclimate	- Parallel tasks	
	- Lighting		
	- Noise		
	- Vibrations		
- Ionizing and non-ionizing radiations			

The effects of PSF on the HEP are summarized in the following equation (Boring 2010):

$$HEP_c = HEP_n \cdot PSF = \begin{cases} 0 < PSF < 1 & \rightarrow HEP_c < HEP_n \\ PSF = 1 & \rightarrow HEP_c = HEP_n \\ PSF > 1 & \rightarrow HEP_c > HEP_n \end{cases} \quad (1)$$

Where HEP_c and HEP_n are respectively the contextual and nominal HEP. Each PSF can have both positive and negative effects on performance. When the PSFs represent a positive effect, the different levels of effect for the PSFs correspond to a value less than one. Multiplying a nominal HEP by this fractional value associated with the PSF serves to decrease the overall HEP. When the PSFs represent a negative effect, the different levels of effect for the PSFs correspond to a value greater than one. Multiplying the nominal HEP by this positive integer serves to increase the overall HEP. When the PSFs are thought to have no effect, the PSF multiplier is set to one, thereby neither increasing nor decreasing the overall HEP.

The tool proposed quantifies the contextual HEP as following:

$$HEP_{contextual} = \frac{HEP_{nominal} \cdot PSF_{composite}}{HEP_{nominal} \cdot (PSF_{composite} - 1) + 1} \quad (2)$$

Where $PSF_{composite}$ is calculated as:

$$PSF_{composite} = PSF_1 \times \dots \times PSF_x \times \dots \times PSF_8 \quad (3)$$

Where PSF_x is the multiplier for each PSF considered in SPAR-H (Table 2). The multiplier values were attributed by analysts of the method, on the basis of several studies carried out on nuclear power plants. In order to align the evaluation of PSFs in SHERPA model, the multipliers were standardized assigning corrective coefficients for different kind of simulated task (Di Pasquale et al. 2015).

The SHERPA model was implemented in an Arena template that allows doing easily and without time consuming a lot of simulations. The tool is suitable for assessing the PSFs influencing on HEP, having regard

to all the possible combinations. Figure 3 shows the main dialogue from that the PSF levels can be clearly selected.

Table 2. Performance shaping factors in SHERPA and corresponding multipliers (Di Pasquale et al. 2015).

SPAR-H PSFs	PSF Levels	Multipliers Action	Multipliers Diagnosis
Available Time	<i>Inadequate Time</i>	HEP=1	HEP=1
	<i>Barely adequate time</i>	10	10
	<i>Nominal time</i>	1	1
	<i>Extra time</i>	0.1	0.1
	<i>Expansive time</i>	0.01	0.01
Stress/ Stressors	<i>Extreme</i>	5	5
	<i>High</i>	2	2
	<i>Nominal</i>	1	1
Complexity	<i>Highly complex</i>	5	5
	<i>Moderately complex</i>	2	2
	<i>Nominal</i>	1	1
	<i>Obvious diagnosis</i>	-	0.1
Experience/ Training	<i>Low</i>	3	10
	<i>Nominal</i>	1	1
	<i>High</i>	0.5	0.5
Procedures	<i>Not available</i>	50	50
	<i>Incomplete</i>	20	20
	<i>Available, but poor</i>	5	5
	<i>Nominal</i>	1	1
	<i>Diagnostic/ symptom oriented</i>	-	0.5
Ergonomics	<i>Missing/ Misleading</i>	50	50
	<i>Poor</i>	10	10
	<i>Nominal</i>	1	1
	<i>Good</i>	0.5	0.5
Fitness for Duty	<i>Unfit</i>	HEP=1	HEP=1
	<i>Degraded Fitness</i>	5	5
	<i>Nominal</i>	1	1
Work Processes	<i>Poor</i>	5	2
	<i>Nominal</i>	1	1
	<i>Good</i>	0.5	0.8

4. METHODOLOGY

The influencing factors play a key role in the modelling of human error and many theoretical studies have been carried out to define, to classify and to model these factors as above. The aim of this paper is the study of parameters that affect the human performance in workplace, considering how they increase or decrease the human error probability. The simulative analysis focuses the attention on factors previously described and used in SHERPA.

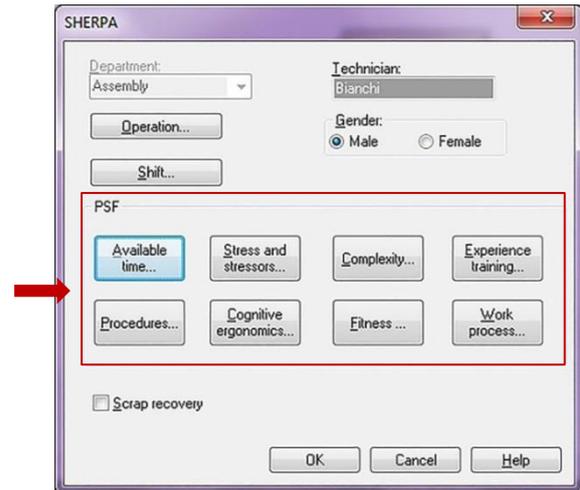


Figure 3. Main dialogue SHERPA.

The influence of every PSF level considered both in singular way and combining with the others factors, was quantified through numerous simulations with the SHERPA template. The following briefly describes the basic steps used in the simulation process:

1. Problem definition: description of the case study.
2. Experiment planning and system definition: identification of the system components to be modelled and the performance measures to be analysed.
3. Experiment execution.
4. Results analysis: list of results and discussion of study implications.

4.1. Problem definition

A manufacturing activity was simulated in an Arena model for the research purpose of this study. The construction of the simulation model takes hint from the description of different assembly stations proposed in literature (Falcone et al. 2010, Falcone et al. 2011). In particular, the Arena model reproduces the operator work station involved in manual assembly on an eight-hour shift, considering 235 working days (Figure 4).

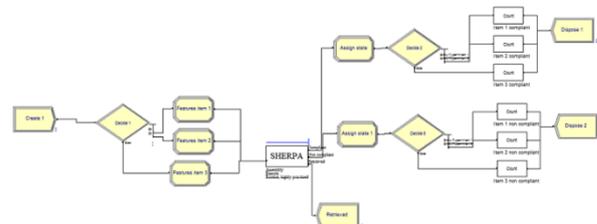


Figure 4: Simulative model used in the study.

A 30-minute break after 4 hours for the shift start is scheduled. The simulations were performed considering the assembly activities belonging to the action category: routine, highly-practiced, rapid task involving relatively low level of skill. The assembly operation was

simulated for three different items with random arrival sequences based on a production mix and with processing times characterized by a triangular distribution, with vertices corresponding to the mean $\pm 10\%$. For each item, processing times, fixed and variable costs and selling prices, as well as overall production mix, were defined and are shown in Table 3.

Table 3. Features of simulated items in the case study.

Features	Item 1	Item 2	Item 3
Mean processing time (min.)	25	36	45
Setup time (min.)	5	5	5
Price (€)	115	155	200
Fixed cost (€)	52	65	78
Variable cost (€)	18	24	32
Productive mix	25%	35%	40%

4.2. Experiment planning and system definition

The SHERPA template allows to model the context and the psycho-physical condition of the operator through 21 PSFs, both main and secondary.

Every PSF impacts in a different way on nominal HEP. SPAR-H method, in fact, uses a nonlinear levels classification and the levels classes are different for every PSF. In the case of available time, for example, there are four levels in addition to the nominal case, while stress or work processes have only two levels. Consider all these factors with a full factorial analysis, would be to make $2^{21} = 2097152$ simulations, taking into account just two levels per factor. For this reason in the experiment planning a selection of the potentially most significant PSFs was necessary.

Firstly, all factors have been classified (as reported in Table 4) compared to the experiment context in:

- Controllable: you can manage and define values in advance, as the input of the experiment itself.
- Uncontrollable: are out of hand when they appear; may change during operation of the product or process.
- Measurable: able to be measured; not subjective, perceptible or significant.
- Unmeasurable: not able to be measured objectively.

The proposed classification has been useful in the subsequent selection of the most relevant factors for the goal of simulative analysis. The choices of the factors have taken into account this classification, considering at least a factor by category.

In the second step a common method of investigating the effects of parameters on a process has been applied. The one-factor-at-a-time (OFAT) method allows to change only one factor at a time, to assess the impact of factors considered one at a time instead of all simultaneously and to notice its influence on a given

response. Although this method has the advantage of being simple, it requires a large number of trials and does not point out the possible interactions between several factors.

Table 4: Factors classification in terms of measurability and controllability.

	Measurable	Unmeasurable
Controllable	Available time Parallel tasks Microclimate Lighting	Workplace Procedures Precision level Physical effort Mental effort Cognitive ergonomics
Uncontrollable	Circadian rhythm Experience Noise Vibrations Radiations	General complexity Work processes Mental stress Fitness for duty Pressure time Communication

In our study the PSF levels have been modified one at a time keeping the others at nominal level. The contextual HEP value and the PSF composite were calculated for every simulation. Some factors have been set to scenario: microclimate, lighting, circadian rhythm and physical effort, the last one is related to the performed task). Downstream the simulations, the results were analysed and for each factor the ratio between the percentage variation of contextual HEP and PSF composite was considered (Table 5). It can be clearly seen that the increase of the PSF level determines an increase of the PSF composite and a consequent increase in the probability of error.

The negative changes, such as extra and expansive available time or good cognitive ergonomics, represent the positive effect of the factors on the performance, as seen above. A special factor is the experience. This factor, in fact, determines a very high increase of HEP ($\Delta\text{HEPc}=97.85\%$) with a modest increase in PSF composite ($\Delta\text{PSFc}=20.63\%$). Factors with similar changes in their PSF composite, for example the general complexity ($\Delta\text{PSFc}=15.25\%$) or mental stress ($\Delta\text{PSFc}=26.47\%$), respectively determine increments of HEP equal to 1.89% and 3.57%. This exception will be thorough better later.

Previous evaluations have been used to select the factors for the next step considering the factors with more impact on HEP and mainly representative of a manufacturing context.

With the aim of reducing the number of runs the parameters *available time*, *state of workplace*, *vibrations*, *radiations*, *pressure time*, *precision level*, *mental efforts* and *communication and integration in team work* were excluded. All these factors were set to the nominal level in the experimental stage and they have not had their influence on HEP.

Table 5: PSF effect on contextual HEP.

SPAR-H PSFs	PSF Levels	Δ HEPc%	Δ PSFc%	$\frac{\Delta$ HEPc%/ Δ PSFc%
Available Time	<i>Inadequate</i>	98.73	not available	not available
	<i>Barely adequate</i>	84.09	85.29	0.9859
	<i>Nominal</i>	0	0	-
	<i>Extra</i>	-1355.27	-1370.59	0.9888
	<i>Expansive</i>	-14409.55	14605.88	0.9865
Mental stress, Pressure time and Noise	<i>Extreme</i>	18.96	70.59	0.2687
	<i>High</i>	3.57	26.47	0.1348
	<i>Nominal</i>	0	0	-
Radiations and Vibrations	<i>Extreme</i>	10.39	54.54	0.1904
	<i>High</i>	1.89	15.25	0.1236
	<i>Nominal</i>	0	0	-
Workplace	<i>Extreme</i>	10.39	70.59	0.1472
	<i>High</i>	1.89	26.47	0.0713
	<i>Nominal</i>	0	0	-
General complexity, Precision level, Mental efforts, Parallel tasks	<i>Highly complex</i>	30.67	70.59	0.4346
	<i>Moderately complex</i>	6.24	26.47	0.2356
	<i>Nominal</i>	0	0	-
Experience/ Training	<i>Low</i>	97.85	20.63	4.7421
	<i>Nominal</i>	0	0	-
	<i>High</i>	-191.72	-1.94	0.9877
Procedures	<i>Not available</i>	95.72	97.06	0.9862
	<i>Incomplete</i>	91.37	92.64	0.9863
	<i>Available, but poor</i>	69.67	70.59	98.69
	<i>Nominal</i>	0	0	-
Cognitive Ergonomics	<i>Missing</i>	95.72	97.06	0.9862
	<i>Poor</i>	84.09	85.29	0.9859
	<i>Nominal</i>	0	0	-
	<i>Good</i>	-191.72	-194.12	0.9877
Fitness for Duty	<i>Unfit</i>	98.73	not available	not available
	<i>Degraded Fitness</i>	69.66	70.58	0.9869
	<i>Nominal</i>	0	0	-
Work Processes and Communication	<i>Poor</i>	53.95	77.27	0.6981
	<i>Nominal</i>	0	0	-
	<i>Good</i>	-48.71	25.37	-1.92

Based on this assessment and taking into account the classification of the factors in terms of measurability

and controllability were chosen the most significant factors:

- noise;
- mental stress;
- general complexity;
- parallel tasks;
- experience;
- procedures;
- work processes;
- fitness for duty;
- cognitive ergonomics.

For the chosen factors have been considered only two levels from those available for the analysis:

- Noise: Extreme and Nominal levels;
- Mental stress: Extreme and Nominal levels;
- General complexity: High and Nominal levels;
- Parallel tasks: High and Nominal levels;
- Experience: Low and Nominal levels;
- Procedures: Incomplete and Nominal levels;
- Work processes: Poor and Good levels;
- Fitness for Duty: Degraded and Nominal levels;
- Ergonomics: Poor and Good levels;

4.3. Experiment execution

In the system definition, nine factors were selected with two levels for each one. In this condition we can define the number of simulations to be performed to analyse the scenarios provided by all possible combinations of PSFs and to evaluate their effect on the likelihood of operator error; they are $2^9 = 512$ simulations. The experiment was conducted simultaneously changing the levels of selected factors until you cover the entire experimental plan.

5. RESULTS ANALYSIS

5.1 Discussion

Analysis of variance (ANOVA) was used to examine the effect of significant PSFs on the HEP (Gelman 2005, Scheffe 1999). This method, developed by Fisher, is at the basis of many designs of experiments and is used to compare differences of means among more than two groups. It does this by looking at variation in the data and where that variation is found. Specifically, ANOVA compares the amount of variation between groups with the amount of variation within groups. It can be used for both observational and experimental studies.

In performing ANOVA the experimental factors and the dependent variable or response are identified. The experimental factors are the source of variability whose effect is to be determined based on the results of a dependent variable or response. In the case of study, experimental factors are therefore the PSFs, while the dependent variable is the contextual HEP.

The simplest experiment suitable for ANOVA analysis is the experiment with a single factor, used in a first time to assess the impact of each factor on HEP. Table 6 lists the one-way ANOVA results. The SS stands for Sum of Squares; F-ratio is test statistic used for ANOVA, the p-value is the probability of being greater than the F-ratio. The F is a ratio of the variability between groups compared to the variability within the groups. The F-ratio will always be at least 0, meaning that it is always nonnegative. The p-values in the last column are the most important information contained in this table. The statistical significance of the effect depends on the p-value, as follows:

- If the p-value is larger than the significance level you selected, the effect is not statistically significant.
- If the p-value is less than or equal to the significance level you selected, then the effect for the term is statistically significant.

Usually, a significance level (denoted as α or alpha) of 0.05 works well. A significance level of 0.05 indicates a 5% risk of concluding that an effect exists when there is no actual effect.

Figure 5 shows the results for all the chosen factors and it underlines graphically the different impacts on error likelihood. Each graph represents the average value of

HEP when the factor is set to level one or two. The vertical bars indicate the level of confidence at 95% that is the probability that the calculated values fall in this range. It is to be noted that when the bars are large the possible values are very different from each other and fluctuate around a mean value. The most influential factors (experience, procedures and cognitive ergonomics) have a very tight confidence interval, a sign of their strong impact in the calculation of the error probability.

Table 6: One-way ANOVA results.

FACTORS	SS	F-ratio	p-value
Mental stress	0.046	0.362	0.548
Noise	0.046	0.362	0.548
General complexity	0.128	1.013	0.315
Parallel tasks	0.128	1.013	0.315
<i>Experience</i>	<i>25.196</i>	<i>327.178</i>	<i>0.00</i>
<i>Procedures</i>	<i>11.565</i>	<i>111.289</i>	<i>0.00</i>
<i>Work processes</i>	<i>2.073</i>	<i>16.942</i>	<i>0.000045</i>
<i>Fitness for duty</i>	<i>1.746</i>	<i>14.199</i>	<i>0.000184</i>
<i>Cognitive ergonomics</i>	<i>18.061</i>	<i>198.465</i>	<i>0.00</i>

Through this first analysis the greatest difference of average HEP is easily observed for those factors that have the two multipliers more distant from each other, like procedures and ergonomics. The experience is an exception, because it has a strong impact on the

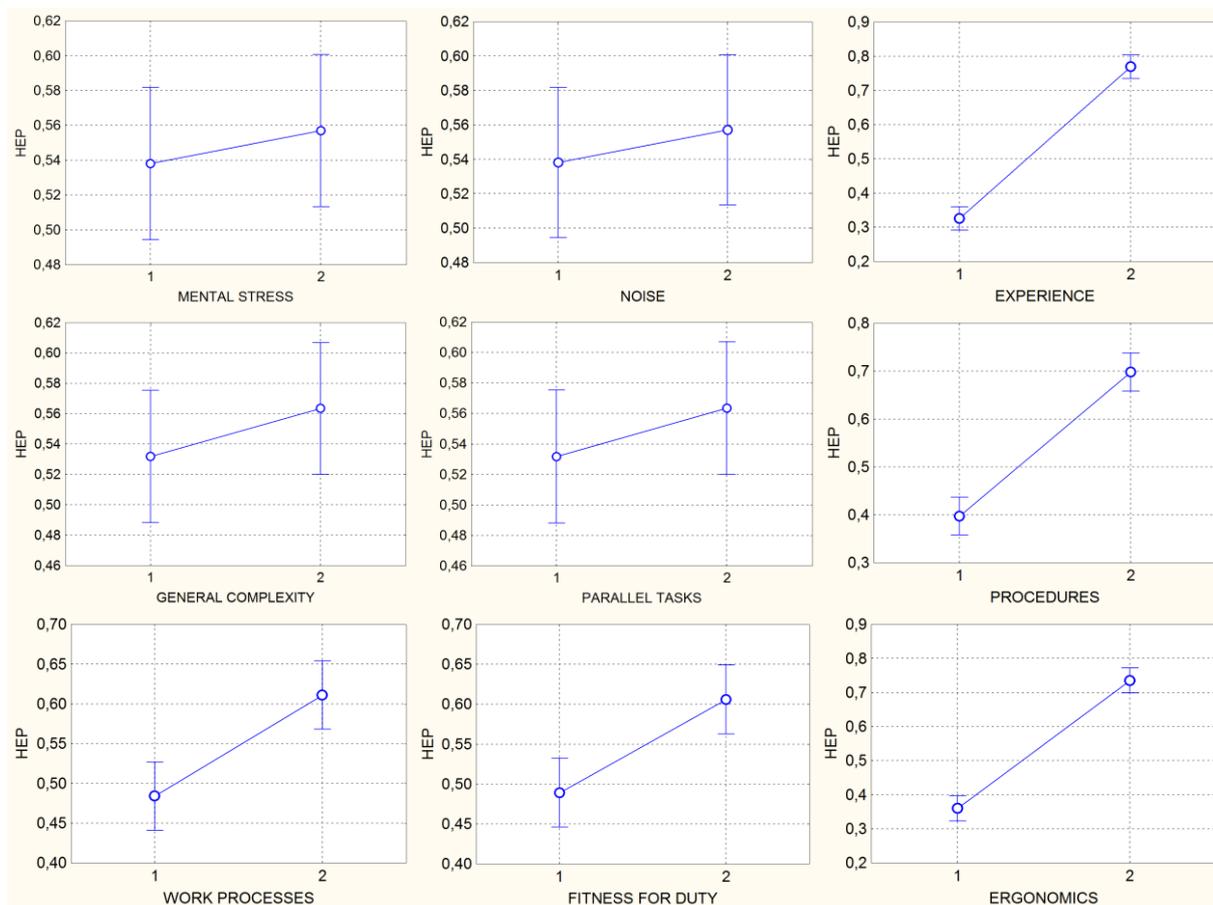


Figure 5. Single factor Analysis of Variance.

probability of error despite its multipliers are comparable to those of stress and complexity. Such behaviour is also clear in the two-way ANOVA.

The two-way analysis of variance is an extension of the previous one-way, which examines the influence of two different factors and aims at assessing if there is any interaction between factors and how the contemporary presence of two factors affects the variable result. Through this second step of analysis, the interrelationships between multiple PSFs were examined. Also in this case the p-value is used as an indicator to determine if the two factors have a significant interaction when considered simultaneously. If one factor depends strongly on the other, the F-ratio for the interaction term will have a low p-value.

The two-way ANOVA table is structured just like the one-way. Table 7, in fact, shows the SS, the F-ratio and the p-value for all factors combinations.

There is a statistically significant interaction between the effects of experience and procedures on HEP (p-value= 0.021<0.05), so the effect on the mean outcome of a change in one factor depends on the level of the other factor (Figure 6). The significant relationship between these factors depends also on the high impact on HEP of single factors. For all the other combinations there is not statistical dependence. In fact their p-values are included between 0.135 (experience x cognitive ergonomics) and 1.000 (mental stress x work processes). For example, in Figure 7 the interaction between mental stress and noise (p-value=0.976) shows clearly the statistical independence: the effects of a change in one factor on the outcome do not depend on the value or level of the other factors.

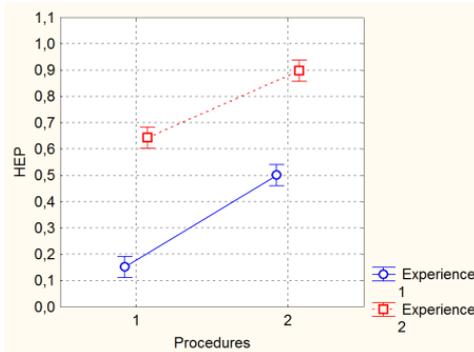


Figure 6. Procedures x Experience ANOVA results.

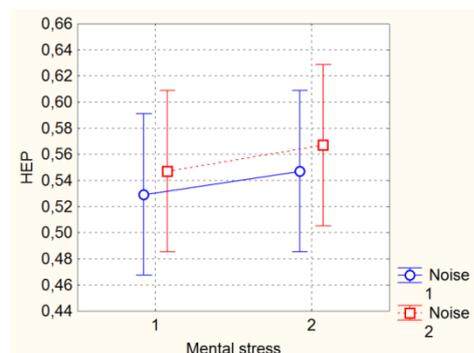


Figure 7. Mental stress x Noise ANOVA results.

Table 7: Two-way ANOVA results.

FACTORS			SS	F ratio	p-value
Mental stress	x	Noise	0.0001	0.001	0.976
Mental stress	x	General complexity	0.0006	0.004	0.946
Mental stress	x	Parallel tasks	0.0006	0.004	0.946
Mental stress	x	Experience	0.0107	0.138	0.710
Mental stress	x	Procedures	0.0007	0.007	0.934
Mental stress	x	Work processes	0.0000	0.000	1.000
Mental stress	x	Fitness for duty	0.0003	0.002	0.963
Mental stress	x	Ergonomics	0.0021	0.022	0.881
Noise	x	General complexity	0.0006	0.004	0.946
Noise	x	Parallel tasks	0.0006	0.004	0.946
Noise	x	Experience	0.0107	0.138	0.710
Noise	x	Procedures	0.0007	0.007	0.934
Noise	x	Work processes	0.0000	0.000	1.000
Noise	x	Fitness for duty	0.0003	0.002	0.963
Noise	x	Ergonomics	0.0021	0.022	0.881
General complexity	x	Parallel tasks	0.0008	0.006	0.937
General complexity	x	Experience	0.0162	0.211	0.646
General complexity	x	Procedures	0.0013	0.012	0.912
General complexity	x	Work processes	0.0001	0.001	0.976
General complexity	x	Fitness for duty	0.0002	0.002	0.965
General complexity	x	Ergonomics	0.0028	0.031	0.860
Parallel tasks	x	Experience	0.0162	0.211	0.646
Parallel tasks	x	Procedures	0.0013	0.012	0.912
Parallel tasks	x	Work processes	0.0001	0.001	0.976
Parallel tasks	x	Fitness for duty	0.0002	0.002	0.965
Parallel tasks	x	Ergonomics	0.0028	0.031	0.860
Experience	x	Procedures	0.2884	5.348	0.021
Experience	x	Work processes	0.0534	0.731	0.393
Experience	x	Fitness for duty	0.1571	2.135	0.145
Experience	x	Ergonomics	0.0926	2.237	0.135
Procedures	x	Work processes	0.0000	0.000	0.992
Procedures	x	Fitness for duty	0.0012	0.012	0.913
Procedures	x	Ergonomics	0.0075	0.109	0.742
Work processes	x	Fitness for duty	0.0001	0.001	0.973
Work processes	x	Ergonomics	0.0021	0.024	0.876
Fitness for duty	x	Ergonomics	0.0002	0.003	0.960

The experience is the most interesting factors. As already highlighted by the OFAT and one-way ANOVA analysis, the experience has one of the major impact on the error probability. This effect is further confirmed by the interactions between factors (Figure 8). The level two of experience determines a considerable decrease of human reliability and consequent increase in error probability when it is combined with every factor (i.e. mental stress, general complexity, cognitive ergonomics and procedures). The strong impact does not depend exclusively from the multiplier, but it derives also by the logic experience evaluation used by the model. The lack of knowledge of the processes, of the machines and of the procedures modifies the nominal HEP, because it impacts on the category of performed task, which can no longer be regarded as routine and highly-practiced.

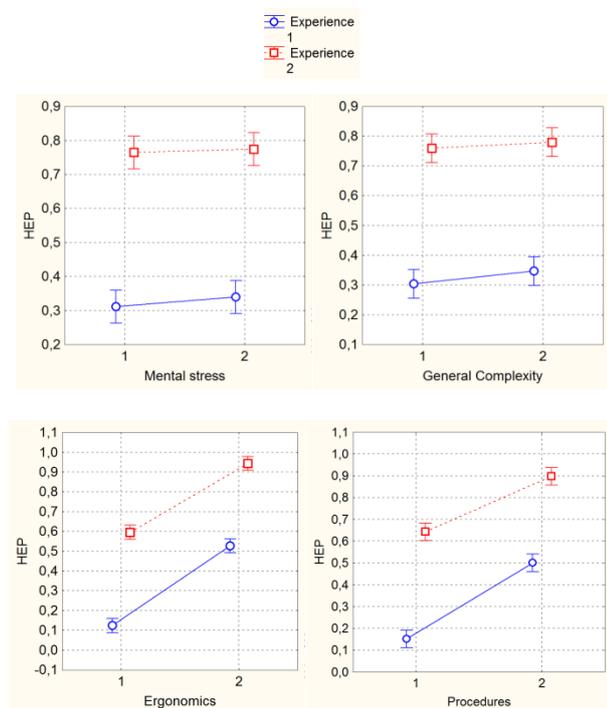


Figure 8. Comparison of the interactions with experience.

6. CONCLUSION

The performance shaping factors are an integral part of modelling and characterization of errors and they affect the productivity and the efficiency at work. Their modeling is a problem for each HRA method. Many HRA approaches introduce widespread PSF taxonomies and complex modeling of their mutual influence. Despite the efforts of HRA experts, the PSFs have not explicit role both in error identification and in probability estimation yet. The goal of this paper has been to analyse the PSFs, used in the SHERPA model, and to assess their impact on HEP in order to improve the model and to make it more responsive to working reality.

Thanks to the simulative analysis and to the results obtained from one and two way ANOVA, the influence

of every PSF level, considered both in singular way and combining with the others factors, was quantified and evaluated, allowing to realize if there is more or less dependence between them.

Several useful considerations can be made downstream of the study. First of all, through the preparatory OFAT analysis the different PSF impacts on HEP in relation to the value of its multiplier is evident. It is certainly useful as a starting point for the improvement of PSF modelling, that is currently under investigation.

The one-way ANOVA underlined the higher or lower impact on HEP of individual factors. While the results of two-way ANOVA highlight few interactions between factors. There is significance of impact only when the experience is combined with the procedures. As regards the experience, its special behavior requires further investigations and studies.

REFERENCES

- Boring R.L., 2010. How many performance shaping factors are necessary for human reliability analysis. Seattle, US.
- Boring R.L., Griffith C.D. and Joe J.C., 2007. The measure of human error: Direct and indirect performance shaping factors. Human Factors and Power Plants and HPRCT 13th Annual Meeting, pp. 170-176. 26-31 August 2007, Monterey Marriott (Canada).
- De Ambroggi M. and Trucco P., 2011. Modelling and assessment of dependent performance shaping factors through analytic network process. Reliability Engineering and System Safety, 96(7), 849-860.
- De Felice F., Petrillo A., Carlomusto A. and Romano U., 2012. A Review of Human Reliability Analysis: Management Practices and Techniques. Global Journal Of Management Science and Technology, 1(4), 131-147.
- Di Pasquale V., Iannone R., Miranda S. and Riemma S., 2013. An Overview of Human Reliability Analysis Techniques in Manufacturing Operations. Operations management. INTECH Open Access Publisher.
- Di Pasquale V., Miranda S., Iannone R. and Riemma S., 2015. A Simulator for Human Error Probability Analysis (SHERPA). Reliability engineering and System Safety, 139, 17-32.
- Falcone D., Silvestri A., Di Bona G., Forcina A. and Pacitto A., 2010. Study and modelling of very flexible lines through simulation. Proceedings of the 2010 Spring Simulation Multiconference, pp. 11-16. 11 – 15 April 2010, Orlando, Florida (USA).
- Falcone D., Silvestri A., Forcina A. and Pacitto A., 2011. Modeling and simulation of an assembly line: a new approach for assignment and optimization of activities of operators. MAS (The International Conference on Modeling and

- Applied Simulation), pp. 12-14. 12 –14 September 2011, Rome (Italy).
- Farcasiu M. and Prisecaru I., 2012. Study on the main factors influencing human performance in NPP operation. University" Politehnica" of Bucharest Scientific Bulletin, Series D: Mechanical Engineering.
- Fera M. and Macchiaroli R., 2009. Proposal of a qualitative assessment model for health and safety in small and medium enterprises. rd International Conference on Safety and Security Engineering, pp.117-126. 1-3 July 2009, Rome (Italy).
- Gelman A., 2005. Analysis of variance—why it is more important than ever. *The annals of statistics*, 33(1), 1-53.
- Griffith C.D. and Mahadevan S., 2011. Inclusion of fatigue effects in human reliability analysis.. *Reliability Engineering and System Safety*, 96(11), 1437-1447.
- Groth K. and Mosleh A., 2012. Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model. *Journal of Risk and Reliability*, 226(4), 361-379.
- Groth K. and Swiler L., 2013. Bridging the gap between HRA research and HRA practice: A Bayesian network version of SPAR-H. *Reliability Engineering and System Safety*, 115, 33-42.
- Iannone R., Miranda S. and Riemma, S., 2004. Proposta di un modello simulativo per la determinazione automatica delle pause di lavoro in attività manifatturiere a prevalente contenuto manuale. *Atti del XXXI Convegno Nazionale Animp*, pp. 46–60. 14-15 October 2004, Treviso (Italy).
- Kim J. and Jung W., 2003. A taxonomy of performance influencing factors for human reliability analysis of emergency tasks. *Journal of Loss Prevention in the Process Industries*, 16, 479–495.
- Lee S.W., Kim A.R., Ha J.S. and Seong P.H., 2011. Development of a qualitative evaluation framework for performance shaping factors (PSFs) in advanced MCR HRA. *Annals of Nuclear Energy*, 38, 1751–1759.
- Scheffe H., 1999. *The analysis of variance*. John Wiley & Sons.

COST-BASED ASSESSMENT OF PARTITIONING ALGORITHMS OF AGENT-BASED SYSTEMS ON HYBRID CLOUD ENVIRONMENTS

Chahrazed Labba^(a), Narjès Bellamine Ben Saoud^{(a) (b)}

^(a) Laboratoire RIADI-Ecole Nationale des Sciences de l'Informatique, Univ Manouba Tunisia.

^(b) Institut Supérieur d'Informatique, Univ. Tunis El Manar, Tunisia.

^(a)chahrazedlabba@gmail.com, ^(b)narjes.bellamine@ensi.rnu.tn.

ABSTRACT

Distributing agent-based simulators reveals many challenges while deploying them on a hybrid cloud infrastructure. In fact, a researcher's main motivations by running simulations on hybrid clouds, are reaching more scalable systems as well as reducing monetary costs. Indeed, hybrid cloud environment, despite providing scalability and effective control over proper data, requires an efficient deployment strategy combining both an efficient partitioning mechanism and cost savings. In this paper, we propose a cost deployment model dedicated to distributed agent-based simulation systems. This cost model, combining general performance partitioning criteria as well as monetary costs, is used to evaluate cluster and grid based partitioning algorithms on hybrid cloud environments. The first experimental results show that, for a given agent-based model, a good partitioning method used with the suitable hybrid cloud environment lead to an efficient and economic deployment.

Keywords: Hybrid Cloud, Deployment, Agent-based Simulation, Performance, Cost Model.

1. INTRODUCTION

Agent-based systems are largely used in modeling and simulating various types of complex systems in many research areas such as sociology (Macy and Willer 2002), crisis and emergency management (Bellamine Ben Saoud Ben Mena Dugdale Pavard and Ben Ahmed 2006), biology (Kuan Rui-bin Hao-ran and Jun-qing 2011), natural disasters (Mustapha Mcheick and Mellouli 2013), etc.

To cope with realistic complex situations, large-scale simulations are required. However, considering the large number of the simulated entities and their complex behaviors, scalability becomes a challenging issue. Therefore, in addition to distribution and suitable partitioning methods, using powerful computational infrastructures should ensure better performance of such large-scale simulation environments. As far as we know unlimited computational resources are not in the reach of all organizations and budget constraints are a strong

consideration in choosing a solution. Cloud computing paradigm offers new opportunities by delivering limitless resources on demand and on pay-as-you-go basis. In the current work we are interested in using the hybrid clouds. In fact, such environments consist in combining a local infrastructure and a public cloud offered by a third party. Hybrid cloud permits to overcome the challenge of scalability that can be induced by using only the private machines and minimize the costs of outsourcing the whole data and computing on a public cloud. Such environments were used with different distributed application types such as web applications (Kaviani Wohlstadter and Lea 2012), enterprise applications (Hajjat Sun Sung Maltz Rao Sripanidkulchai and Tawarmalani 2010), agent-based systems (Siddiqui Tahir Rehman Ali Rasooland Bloodsworth 2012).

However, to make an efficient use of such infrastructure, both good distribution strategy and effective cost deployment model should be taken into consideration. To the best of our knowledge, hybrid clouds have not been used to run distributed large-scale agent-based simulations. In the current work we are interested in proposing an effective cost deployment model that aims at measuring the suitability of the distribution of a large scale agent-based system on a hybrid cloud environment. Indeed an efficient deployment of a given distributed agent-based model consists of using a good partitioning method in terms of performance criteria (minimizing communication overhead, migration and global execution time) and a suitable public cloud provider (the cheapest is the best, to ensure costs savings). We used the proposed cost deployment model to evaluate and compare the performance and the effectiveness of two agent-based partitioning algorithms on hybrid clouds environments. Among others, we have chosen to evaluate: 1) A cluster-based algorithm represented by the K-means clustering approach which consists in clustering the agents based on a pre-defined criteria then assigning them to different machines. 2) A grid-based algorithm spotted by the static environment partitioning approach which consists of decomposing the space component (spatial environment) into multiple portions.

Each portion together with the set of agents residing in it are assigned to a node.

For the hybrid cloud environments, we used two public cloud providers including Amazon Elastic Compute Cloud (Amazon EC2) and Microsoft Azure, both combined to a local machine. The choice of two different hybrid cloud environments aims at comparing the value added by the public cloud provider in testing different partitioning algorithms for agent-based systems.

The rest of the paper is organized as follows: Section 2 summarizes existing work related to the partitioning and distribution of applications on a hybrid cloud environment. Section 3 defines the proposed cost deployment model. And before concluding, section 4 describes the agent-based simulator developed as a case study, the description of the technical environment as well as the performed experiments and the discussion of the obtained results.

2. RELATED WORK

A careful literature study shows that hybrid cloud environments are more and more used to deploy and execute different types of distributed applications. In order to better summarize the existing works and the cost models adopted in each hybrid cloud deployment, we defined a set of five criteria to compare them:

Added to those already defined in (Juan-Verdejo and Baars 2013) to compare the partitioning of applications on a hybrid cloud infrastructure, we consider the metrics list and the exact description of the used hybrid environment, in terms of the public cloud and local premise. Therefore, the comparison, as shown in table 1, is performed according to the (1) focus, (2) the application type, (3) the granularity, (4) the used metrics and (5) the two-side hybrid cloud environment.

In (Hajjat Sun Sung Maltz Rao Sripanidkulchai and Tawarmalani 2010), the authors present the CloudWard Bound consisting in migrating enterprise services into a hybrid cloud environment by investigating the partition and reallocation at the server level. The approach considers the application performance requirements as well as the privacy restrictions to choose the candidate components for migration. To optimize the cost models associated with the use of hybrid clouds, CloudWard Bound implements a beneficial migration strategy based on considering the application characteristics including the workload amount, the storage capacity and transaction delays.

In (Van den Bossche Vanmechelen and Broeckhove 2010), the authors address an optimization problem consisting of deciding which workloads to outsource to what cloud provider when the private datacenters are overloaded. A binary integer program for hybrid cloud was proposed to formulate the workload scheduling problem during deployment while minimizing the total costs of the execution and meeting the users' requirements. Multiple metrics and constraints are defined in the binary integer program such as the deadline constraint, the execution time, the cloud providers, the monetary costs, etc. the authors define

three scenarios including the use of public cloud only, public cloud with network costs and hybrid cloud. The solver implementing the optimization problem shows better results when it is used across public cloud provider but with hybrid cloud settings the performance of the solver decreases. The authors believe that enhancing the optimization technique with hybrid clouds presents an interesting research section.

In (Kaviani Wohlstadter and Lea 2012), the Manticore framework, which is a partitioning software service for deployment on a hybrid cloud, is presented. The authors identify two existing approaches found in the literature namely the request-based model and the static structure model. The first one consists in executing the partitioning software service request on the public cloud or on the local infrastructure. While, the second approach allows to partition a request between both of them. The existing approaches reveal some issues such as in flexibility in placement and cost-performance issues. To overcome the mentioned problems, the authors present a context sensitive model to automate the partitioning process that allows partitioning dynamically a request on both a public cloud and a local infrastructure while maintaining a distinguished transitive set of callers to each function execution. However Manticore addresses a low level of abstraction which is the code partitioning. The approach is dedicated to move code (functions) rather than components which is not ideal for enterprise applications which are developed by integrating different commercial tools and components. The authors propose a cost deployment model based on three metrics: execution time, communication overhead and monetary costs associated to the use of hybrid cloud infrastructure. The proposed model is used to compare the cost-performance of partitioning the application code on both the premise and public cloud (Amazon EC2) instead of using only one of them.

In (Fan Wang and Chang 2011), an agent-based migration framework for the migration of MapReduce jobs on a hybrid cloud is presented. Indeed, MapReduce consists into two tasks which are as follows: 1) The first phase is the map Job, which takes as input a set of data to be converted into tuples. 2) The reduce phase, takes as input the output of the previous phase and combines the received data into smaller tuples. The framework is based on agents to monitor system behavior, negotiate actions, manage resources and achieve automatic and intelligent service migration. The partitioning is based on three metrics including the job count, size of the job and estimated finish time. Unlike (Van den Bossche Vanmechelen and Broeckhove 2010) and (Hajjat Sun Sung Maltz Rao Sripanidkulchai and Tawarmalani 2010) the cost models are not taken into consideration.

In (Siddiqui Tahir Rehman Ali Rasooland Bloodsworth 2012), the Elastic-JADE (Java Agent Development Framework) platform using the Amazon EC2 resources, to scale up and down a local running agent based simulation is presented.

Table1: Research work related to the hybrid cloud partitioning approaches

Research Work	Focus	Application Type	Granularity	Metrics	Hybrid cloud composition
(Hajjat Sun Sung Maltz Rao Sripanidkulcha i and Tawarmalani 2010)	enterprise services migration	Enterprise Applications	Components	1. Applications Number, 2. Components number per application, 3. Transaction number, 4. Size of a transaction, 5. Number of servers, 6. Monetary costs	Large campus Network combined to Windows Azure Cloud Platform
(Van den Bossche Vanmechelen and Broeckhove 2010)	Workload scheduling	simulation experiments, image and video rendering codes, highly parallel analysis codes	Jobs	1.Applications Number, 2.Tasks number per application, 3.Cloud Provider, 4.Instance Type, 5.Execution time/task, 6.Deadline constraints, 7. Monetary costs.	Local infrastructure combined to multiple cloud provider (A, B,C)
(Fan Wang and Chang 2011)	Services migration	MapReduce	MapReduce jobs	1.Jobs count, 2. Size of jobs, 3. Estimated finish time	Not mentioned
(Siddiqui Tahir Rehman Ali Rasooland Bloodsworth 2012)	Agents migration	Agent-based systems applied in Image processing	Agents	1.CPU , 2. Memory	A local machine combined to Amazon EC2 instances
(Kaviani Wohlstadter and Lea 2012)	code entities partitioning	Web Applications	Code functions	1. Execution Time 2. Communication time 3. Monetary costs	A local machine combined to a large Amazon EC2 instance
(Juan-Verdejo and Baars 2013)	BI applications migration strategies using a set of predefined criteria.	BI applications	Components	1. Applications number 2. Components number per application. 3. Transaction frequency 4. Transaction delays 5. Cloud Costs	Not mentioned (no experiments)

According to the authors, JADE allows the distribution of MAS, and combining a running local JADE platform to the cloud, can enhance the scalability of such systems. In this work, thresholds are defined to determine the right moment to scale up or down the application.

The proposed approach consists in using a monitoring system on both local and cloud resources, to keep track of the available resources and exchange messages. While the proposed solution enhances scalability, it is specific to the MAS developed using the JADE platform. However, the authors did not investigate the cost of migration, or the overhead of communication between local and cloud resources that occurs during simulation. In (Juan-Verdejo and Baars 2013) a cloud migration framework for partially moving Business Intelligence (BI) applications in the hybrid cloud is described. The authors identify different criteria including functional and nonfunctional requirements related to BI

applications to define multiple alternative migration strategies. The authors consider the use of the Analytic Hierarchy Process (AHP) which assigns weights to the criteria in order to assist the migration decision. The authors don't consider the cloud offering selection such as cloud advisory service, migration service, Cloud Environment Build and Management Service (EBMS), since such services affect the cost model as well as the QoS.

As it is described in the Table.1, the deployment of distributed applications on a hybrid cloud environment requires special considerations to reach cost-performance efficiency. Therefore different metrics were defined based on the characteristics of the applications, the used hybrid infrastructure and the policies of the user. In the current work, we are interested in the deployment of distributed agent-based systems on a hybrid cloud infrastructure. An efficient partitioning approach as well

as a suitable cloud provider are required to achieve both performance in terms of running large-scale distributed application and cost savings while using hybrid clouds. Similarly to the mentioned research work in Table.1, in order to accomplish the partitioning of an agent-based simulation on a hybrid cloud, a set of metrics should be taken into consideration.

Consequently, in the current work, based on the existing research work and the characteristics of an agent-based system, we define the relevant metrics to be considered while partitioning agent-based simulation applications on a hybrid cloud infrastructure: 1) the communicated data between the public cloud and the on premise infrastructure (in our case such data can be either the messages transfer between agents residing on different machines or the agents migrations between nodes); 2) execution time; 3) monetary costs related to the use of a public cloud provider. Each metric will be detailed and discussed in the next section.

3. COST DEPLOYMENT MODELLING

Usually, in order to evaluate a partitioning mechanism used with agent-based simulations, not in hybrid cloud environments, the following metrics are considered in several research works (Wang Lees Cai Zhou and Low 2009, Wang Lees and Cai 2012, Viguera Lozano Orduña and Grimaldo 2010):

- Communication cost: denotes the number of messages that are communicated between two agents residing in different machines.
- Migration cost: presents the number of migrations of the agents during simulation time. A migration denotes the move of an agent residing on a given machine to another different one.
- Execution Time

Also, the cost of a deployment on a hybrid cloud depends on the execution time as well as the size of the communicated data associated with the monetary costs offered by a cloud provider.

Combining these two points of view and resulting literature findings, in the current work, we matched the considered metrics in the partitioning agent-based simulation with those used to evaluate the cost of a hybrid cloud deployment in order to formulate the agent-based simulation-partitioning problem on a hybrid cloud. Therefore, we introduce the notations given in Table 2. Before defining, the different costs associated with the deployment of a distributed agent-based simulation on the hybrid cloud, we made the following assumptions:

- All the communicated types of messages have the same size.
- All the agents have the same size.
- No load balancing during the simulation (static partitioning algorithms).
- The transfer of data is charged in one way (from cloud to local machine).

In the rest of this section, we define our cost model associated to the deployment of a distributed agent-based simulation on a hybrid cloud where we consider (1) The communication cost, which introduces the cost of sending one message from the public cloud to the local infrastructure; (2) The migration cost, which presents the cost of moving one agent from the public cloud to the promise; and (3) Execution Time.

Table2: Notations used to model agent-based systems deployment cost

Symbols	Definitions
a_i	agent i / ($i=1, \dots, n$)
S_i	Size of an agent a_i
S_{msg}	Size of a message
$Cost_{send-one-msg}$	Cost of sending one message from a cloud machine to a local machine
$Cost_{migrate-one- ai}$	Cost of migrating an agent from a cloud machine to a local machine
$T_{send-one-msg}$	Time of sending one message from the cloud to the on premise machine
$T_{migrate-one-ai}$	Time of migrating one agent from the cloud to the on premise machine
T_{Unit}	Time unit defined by the cloud provider to apply charges
D_{Unit}	Data unit defined by the cloud provider to apply charges
$Cost_{Tunit}$	Cost associated to T_{Unit}
$Cost_{Dunit}$	Cost associated to D_{Unit}
β	Bandwidth
λ	Communication latency for a given cloud

3.1. Communication Cost

The communication cost depends on the size of the transferred data as well as the time of communication. Thus, similarly to (Kaviani Wohlstadter and Lea 2012) we start by determining the time required to transfer data from a public cloud machine to premise machine. The first equation represents the time of transferring one message from an agent residing on the public cloud to another one located on the premise infrastructure.

$$T_{send-one-msg} = \frac{S_{msg}}{\beta} + \lambda \quad (1)$$

Where S_{msg} indicates the size of a message communicated between two agents and β is the

communication bandwidth between the cloud provider and the local infrastructure. The first part of the above equation $[\frac{S_{msg}}{\beta}]$ denotes the message transmission time

between the cloud and the premise; and λ , as it is defined in several research work (Kaviani Wohlstadter and Lea 2012, Hajjat Sun Sung Maltz Rao Sripanidkulchai and Tawarmalani 2010, Van den Bossche Vanmechelen and Broeckhove 2010) as a constant representing the communication latency associated with the use of a given public cloud provider. Then, we associate the monetary costs offered by the cloud provider as follows:

$$Cost_{send-one-msg} = \frac{S_{msg}}{D_{Unit}} * Cost_{DUnit} + \mu * \frac{T_{send-one-msg}}{T_{Unit}} * Cost_{TUnit} \quad (2)$$

$$\mu \begin{cases} 0, & \text{if the transmission delays are neglected} \\ 1, & \text{Otherwise} \end{cases}$$

Where D_{unit} is the unit of data defined by the cloud provider to apply the charges; also $Cost_{comunit}$ represents the cloud price associated to D_{unit} . The first part of the equation denotes the charges directly related to transferring one message between the public cloud and the local infrastructure. The second part introduces the cost associated to the time of sending one message. T_{unit} presents the unit of time used by the cloud provider to apply charges and $Cost_{execunit}$ is the cloud cost associated to T_{unit} . Finally, we introduce, μ , which is a binary variable that can take either zero if the time of sending one message is neglected or one in the opposite case. The second part of the equation, as it is defined, is more general and can model previous works: For (Kaviani Wohlstadter and Lea 2012), the time of communicating data is considered important and affects the overall cost of the deployment on a hybrid cloud environment, however in other research work such as in (Hajjat Sun Sung Maltz Rao Sripanidkulchai and Tawarmalani 2010, Van den Bossche Vanmechelen and Broeckhove 2010) it is neglected, only the size of the data does matter.

3.2. Migration Cost

Similarly to the communication cost, the migration cost depends on the size of the agent to be moved as well as on the time taken to process the migration in the hybrid cloud environment. Similarly, to the way used to calculate the communication cost, we determine the cost migration. Thus, we start by determining the time of moving one agent from a cloud machine to a local machine.

$$T_{migrate_one_ai} = \frac{S_i}{\beta} + \lambda \quad (3)$$

Where S_i presents the size of the agent to be moved and β is the communication bandwidth between the cloud provider and the machine on premise. The first part of

the equation $[\frac{S_i}{\beta}]$ denotes the transmission time; and λ , as it is defined above, is a constant representing the communication latency associated with the use of a given public cloud. To determine the migration cost we associate the monetary costs offered by a public cloud provider as follows:

$$Cost_{migrate-one-ai} = \frac{S_i}{D_{Unit}} * Cost_{DUnit} + \mu * \frac{T_{migrate_one_ai}}{T_{Unit}} * Cost_{TUnit} \quad (4)$$

$$\mu \begin{cases} 0, & \text{if the transmission delays are neglected} \\ 1, & \text{Otherwise} \end{cases}$$

The first part of the equation denotes the charges directly related to the migration of one agent from the public cloud to the local machines; and the second part introduces the cost that can be added due to the migration time. This factor either it is neglected when μ is equal to zero or considered in the overall migration costs. The objective from adding the second part in both the communication and migration costs is to evaluate the effect of transmission time on the overall deployment cost.

3.3. Execution Time Cost

The computational cost depends mainly on the characteristics of the used cloud virtual machines (VM), the execution time and the load running on the VM. If we assume that we are using the public cloud resources during the whole simulation (from the beginning until the end of the simulation) then the time of using the cloud machine will be equal to the overall execution time of one simulation.

$$Cost_{exec} = \frac{T_{exec}}{T_{Unit}} * Cost_{exe_Unit} \quad (5)$$

4. CASE STUDY AND EXPERIMENTATION

In this section, we present the description of the used agent-based simulator, the description of the technical environment as well as the experimental results of conducting the distribution of an implemented agent-based simulator on two different hybrid cloud providers using the defined cost model. The distribution is performed using two partitioning algorithms, which are the K-means and the static environment partition.

4.1. Description of the agent-based model

The model consists of N agents moving within a spatial environment represented as a grid and interacts between each other. According to the Vowels model (Demazeau 1995), we highlight the environment properties, the agents' characteristics as well as the interaction aspect.

- **Environment:** In the current work, we are interested in the conceptual level of MAS; therefore, we consider only the application environment where the agents and objects are embedded. In our agent model the environment represents the geographical area in which the agents move. It is characterized by its height and width.
- **Agents:** Our model contains two types of agents including rescuers and victims. The agents are initially assigned to random positions. 1) The rescuers: present the leaders who move to the nearest exit; 2) The victims: follow the nearest rescuer in term of spatial distance to reach the emergency exit. The agents can move, communicate and migrate from one node to another.
- **Interactions:** Agents in a MAS should be able to interact and understand each other in order to coordinate their actions and possibly cooperate. Therefore, similarly to (Wang Lees Cai Zhou and Low 2009, Wang Lees Cai 2012) ,we define an Area of Interest (AoI). Agents can communicate and change messages when their AOI overlaps.

To summarize, the test case considered in this current work consists of a set of agents moving within an environment (100 cells *100 cells). We consider an emergency situation where the agents are organized into two different categories: victims and rescuers. The rescuers present the leaders who move to the exits and the victims follow the nearest rescuer in terms of spatial distance. We used the Jade agent platform [20] to develop the simulator described above.

4.2. Description of the used partitioning algorithms

Based on the existing research work (Solar Suppi and Luque 2011, Solar Suppi and Luque 2012, Viguera Lozano Orduña and Grimaldo 2010), the partitioning mechanisms can be classified into two different families:

- Clusters-based approaches in which the agents are clustered based on given criteria then assigned to different nodes.
- Grid-based approaches, called also region-based approach consists of decomposing the space component (environment) into multiple portions. Each portion together with the set of agents residing in it are assigned to a node.

Both cluster and grid approaches can be either static or dynamic. Indeed, the static strategies distribute the agents without considering the load balancing. However, the dynamic mechanisms are used to load balance the partitions during simulation time. In the current work, we consider two partitioning algorithms, which are the K-means and the static environment partition.

The K-means clustering algorithm referred also as the Lloyd's algorithm (Lloyd 1982), was elaborated by Forgy (Forgy 1965) and McQueen (McQueen 1967).

Indeed, the K-means is a cluster-based algorithm consisting of grouping objects into clusters based on a given set of attributes. Indeed, K presents an initial input which denotes the number of the clusters. Each cluster has a centroid which can be determinate by calculating the central mass of all objects. The algorithm starts by initializing the centroids and assign objects to the closest centroid. The process repeats till there is no changes in the clusters.

In this current work, we used the agents' spatial locations as an attribute to cluster them into groups. Each agent is assigned to the cluster with the nearest centroid. We consider a number of clusters equal to four (k=4) including three partitions running on the local machine and a single one on the public cloud machine.

The other method is a grid-based approach represented by the static environment partition. It consists of decomposing the spatial component into regions or cells and assign each partition with the agents residing on it to a single node.

Using this method it is possible to decompose the environment in a number of ways: in a regular grid, in vertical/horizontal strips or in an irregular grid (Wang Lees Cai Zhou and Low 2009).

In order to compare its performance with the above cluster-based algorithm, we choose to decompose the application environment into four partitions using the regular grid division. Similarly to the K-means, including three partitions running on the local machine and a single one on the public cloud machine. Indeed, the grid division is described in (Wang Lees Cai Zhou and Low 2009) as the appropriate choice to minimize the communication overhead between partitions.

4.3. Description of the technical environment

We performed experiments to compare the performance of K-means as well as the static environment partition on different hybrid clouds environments. As a local machine, we used a 2.10 GHz i3-2310M CPU with 4.0 Go memory. For the public cloud machines, we used on Amazon EC2, an m1.large machine from west Europe with 4 EC2 compute units and 7.5 Go memory. For Microsoft Azure, we used the equivalent to m1.large EC2 instance, which is A3 large type with 4 CPU cores and 7 Go memory. In both cases, using either Amazon EC2 instance or Microsoft Azure, the local and cloud machines were connected with a data link of 100 Mb/sec. For the latency, we measured 124.1 milliseconds for Amazon EC2 and 163.6 milliseconds for Microsoft Azure.

For all the experiments, the Amazon EC2 instance costs 0.19\$ per hour and the A3large on Microsoft Azure is charged for 0.324\$ per hour. For the data transfer charges it was the same for both clouds 0.12\$ from the cloud machine to the on premise one and 0\$ in the opposite case.

4.4. Evaluation

We evaluate both algorithms including the K-means algorithm and the static environment partitioning method using our proposed deployment cost model.

We conducted our experiments using different configurations summarized in Table.3.

The results of our experiments are presented as the mean of 10 runs for each configuration. As it is shown in Table.3, initially, the partitions have an equal number of agents including one rescuer in each partition. However, since both of the used algorithms are static, each simulation ends up with an extremely unloaded partitions.

Table 3: Configurations' characteristics

Config.	Total Number of Agents	Number of Victims	Number of rescuers	Number of agents on local machine	Number of agents on the cloud
Config.1	200	196	4	150	50
Config.2	400	396	4	300	100
Cnfig.3	600	596	4	450	150
Config.4	800	796	4	600	200
Config.5	1000	996	4	750	250

In fact, in the case of using both the K-means and the static environment partition, a grouped movement of the agents is more efficient than considering the individual movement pattern.

Therefore in the presented results, we consider only the case where the agents are initially clustered and moving into groups to follow the nearest rescuer.

4.4.1. Costs Evaluation with neglected transmission Time (U=0)

Indeed, in the proposed cost deployment model, within both defined equations to determine the migration as well as the communication costs we define a binary variable U, which take 0 if the transmission time is neglected or "1" in the opposite case.

The mentioned variable U is defined in the order to determine the effect of the transmission time on the overall costs. In this section, we present the costs evaluations without adding the transmission delays.

The Fig.1 presents the communication costs (message transfer) associated to the execution of both partitioning algorithms (K-means and Static environment partition) on both used hybrid clouds environments. As a first result the cluster-based algorithm outperforms the grid-based method on both hybrid cloud environments in term of the communicated messages between the cloud machine and the local one. Also, the communication costs increase with the used different configurations (agents number).

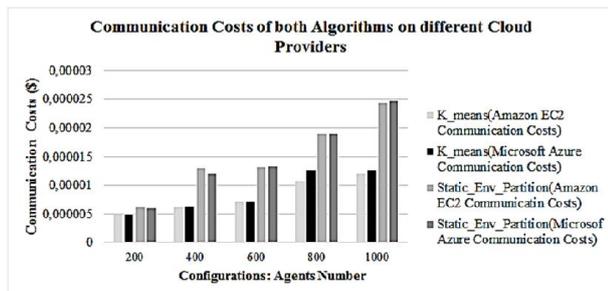


Figure 1: Communication Costs of both Algorithms on different hybrid cloud settings.

Based on previous experimental results in (Wang Lees Cai Zhou and Low 2009, Wang Lees Cai 2012, Labba Bellamine and Dugdale 2015), considering the movement pattern of agents is of great importance for the partitioning mechanism of a distributed agent-based system.

The Fig.2 represents the migration Costs of both used partitioning algorithms on different hybrid clouds settings. Similarly to the communication costs results, the K-means outperforms the grid-based method represented by the static environment partition. As it is shown in Fig.2, the overall costs of migrations increases with the used number of agents and the costs associated to static environment partition are much higher than those associated to K-means on both hybrid clouds settings either using Amazon EC2 or Microsoft Azure.

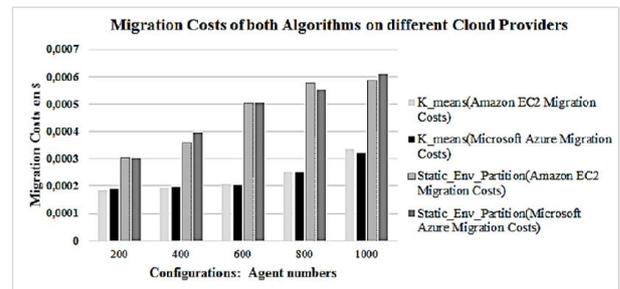


Figure 2: Migration Costs of both Algorithms on different Cloud Providers.

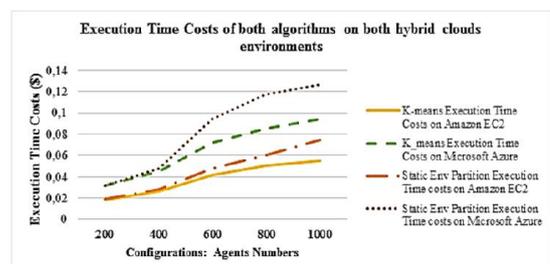


Figure 3: Execution Time Costs

The Fig.3, presents the execution time costs associated to both partitioning algorithms executed on both hybrid clouds (the local machine combined to either Amazon EC2 or Microsoft Azure).

As it is shown in Fig.3 the execution of K-means on Amazon EC2 represents better results in term of execution costs compared to the static environment partition(on both hybrid clouds environments) and even to K-means tested on Microsoft Azure. Also, with both hybrid cloud settings K-means outperforms the static environment partition in term of execution time. To summarize, based on the results of the experiments shown respectively in Fig 1, Fig 2 and Fig 3 the K-means algorithm shows better results in terms of {migration, communication and execution time} costs in both cases of using hybrid cloud environment with different settings (Amazon EC2, Microsoft Azure).

For both migration and communication aspects, the overall costs of running the K-means (same for the static environment partitioning) using both the public clouds Amazon EC2 and Microsoft Azure shows close results. Thus, we can explain these results, by the fact that both of the public clouds charges the data transfer from cloud machines to on premise ones using the same prices. However, for the execution time costs, shown in Fig.3 is more dependent to the used public cloud provider. Both of algorithms including K-means and static environment partition show better results with Amazon EC2. The increase of execution costs with Microsoft Azure can be returned to the high prices suggested by the latter one to use cloud VMs compared to Amazon EC2.

4.4.2. Costs Evaluation including transmission delays (U=1)

In the rest of this section, we present the effect of the transmission delays on the overall costs for the K-means algorithm executed on both used hybrid clouds environments.

In one hand, Fig.4 represents the communication costs including the transmission delays for K-means on both hybrid clouds environments. When the binary variable U is equal to "1", the overall costs of communication on both hybrid clouds settings are much expensive than the case where U is equal to zero.

Also, as it is shown in the Fig.4, the costs are much important with Microsoft Azure which can be returned to the high prices suggested by this latter one.

In the other hand, Fig.5 describes the migration costs including the transmission delay for K-means on both hybrid clouds environments. Similarly to the results presented with the communication costs, the transmission delay has an impact on the overall migration costs. This impact is higher with the use of Microsoft Azure compared to Amazon EC2.

To summarize, as it is shown respectively in both Fig.4 and Fig.5, the transmission delay has an important effect on the overall costs for communication as well as migration for both algorithms executed on the different hybrid clouds settings.

However, this difference is more important with Microsoft Azure due the high prices suggested by this latter one compared to Amazon EC2.

The costs depend on three aspects: 1) the size of the transferred data between the public cloud and on the local

machine. Once the overall size of data is important, the costs are important too; 2) the transmission delays are related mostly to the state of the network and the size of the transferred data; 3) the used public cloud provider. Indeed, the hybrid cloud settings is important and should be considered to minimize the overall costs of deployment.

Therefore for distributed agent-based systems, choosing the suitable partitioning method has a huge effect on the overall costs of hybrid cloud deployment. A suitable partitioning approach has to minimize the migration and communication overheads as well as the execution time. In addition, the choice of a public cloud provider is much important as the partitioning method, since each provider has its own prices and manner of charging costs.

To summarize, the deployment challenge of a distributed agent-based model on a hybrid cloud infrastructure returns to a decision making process permitting to select the best partitioning algorithm and the suitable hybrid cloud composition.

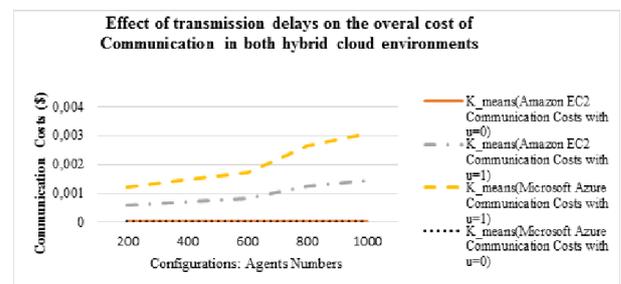


Figure 4: Communication Costs Including the delay costs for K-means on different cloud providers

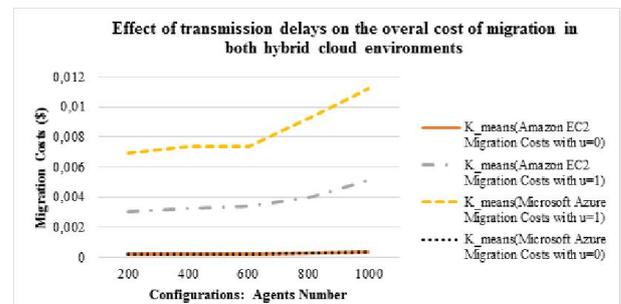


Figure 5: Migration Costs Including the delay costs for K-means on different cloud providers

5. CONCLUSION

To enhance the distribution of agent-based systems both an efficient partitioning method and a reliable/powerful infrastructure are required.

In this paper we evaluated the cost-performance of both cluster and grid based partitioning methods on two different hybrid cloud environments.

The main contributions of this paper are as follows:

- A cost deployment model dedicated to agent-based systems is proposed in order to evaluate the performance of partitioning methods and the

suitability of the hybrid cloud environment. Indeed the model takes into consideration mainly the communication overhead between the partitions, the number of migrations of agents, the execution time, the latency, the bandwidth and the costs associated to the use of the public cloud.

- Comparison of both partitioning algorithms including k-means and the static environment partition executed on two different hybrid cloud environments. The experimental results show that k-means algorithm outperform the static environment partition on both used hybrid infrastructure. Also we discovered that the transmission delays has an impact on the overall costs.

Therefore , the deployment challenge of a distributed agent-based model on a hybrid cloud infrastructure returns to a decision making process permitting to select the best partitioning algorithm and the suitable hybrid cloud composition. As a future work, we aim at automating this decision making process using a set of relevant criteria as well as the distribution and partitioning of an agent-based model on a hybrid cloud infrastructure.

REFERENCES

- Macy M.W and Willer R, 2002. From Factors to Factors: Computational Sociology and Agent-Based Modeling, Annual Review of Sociology, Vol. 28: 143 -166.
- Bellamine-ben Saoud N, Ben Mena T, Dugdale J , Pavard B and Ben Ahmed M, 2006. Assessing large scale emergency rescue plans: an agent based approach, Special Issue on Emergency Management Systems. International Journal of Intelligent Control and Systems, 260—271.
- Kuan Z, Rui-bin Q, Hao-ran Z, and Jun-qing N, 2011. An Agent-based Modeling Approach for Stochastic Molecular Events of Biochemical Networks. In Proceedings of the 2011 Fourth International Conference on Intelligent Computation Technology and Automation - Volume 01 (ICICTA '11), Vol. 1. IEEE Computer Society, Washington, DC, USA, 759-763.
- Mustapha K, Mcheick H, and Mellouli S, 2013. Modeling and simulation of agent-based complex systems and application to natural disasters. In Proceedings of the 8th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks (PM2HW2N '13). ACM, New York, NY, USA, 75-82.
- Hajjat M, Sun X, Sung Y E., Maltz D, Rao S, Sripanidkulchai K, and Tawarmalani M, 2010. Cloudward bound: planning for beneficial migration of enterprise applications to the cloud. SIGCOMM Comput. Commun. Rev. 40, 4 (August 2010), 243-254.
- Kaviani N, Wohlstader E, Lea R,2012. MANTICORE: A framework for partitioning software services for hybrid cloud, Cloud Computing Technology and Science (CloudCom), 2012 IEEE 4th International Conference on, vol., no., pp.333,340, 3-6 Dec.
- Siddiqui U., Tahir G.A., Rehman A.U., Ali Z., Rasool, R.U., Bloodsworth P, 2012. Elastic JADE: Dynamically Scalable Multi Agents Using Cloud Resources, Cloud and Green Computing (CGC), 2012 Second International Conference on , vol., no., pp.167,172, 1-3 Nov. 2012.
- Fan C-T,Wang W-J, Chang Y-S, 2011. Agent-Based Service Migration Framework in Hybrid Cloud, High Performance Computing and Communications (HPCC), 2011 IEEE 13th International Conference on , vol., no., pp.887,892, 2-4 Sept.
- Van den Bossche R, Vanmechelen K, Broeckhove J, 2010. Cost-Optimal Scheduling in Hybrid IaaS Clouds for Deadline Constrained Workloads, Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on, vol., no., pp.228,235, 5-10 July 2010.
- Juan-Verdejo A and Baars H, 2013. Decision support for partially moving applications to the cloud: the example of business intelligence. In *Proceedings of the 2013 international workshop on Hot topics in cloud services* (HotTopiCS '13). ACM, New York, NY, USA, 35-42.
- Wang Y, Lees M, Cai W, Zhou S, Low M.Y.H., 2009. Cluster based partitioning for agent-based crowd simulations, Simulation Conference (WSC), Proceedings of the 2009 Winter, vol., no., pp.1047, 1058, 13-16 Dec.
- Wang Y, Lees M, Cai W, 2012. Grid-based partitioning for large-scale distributed agent-based crowd simulation, Simulation Conference (WSC), Proceedings of the 2012 Winter , vol., no., pp.1,12, 9-12 Dec.
- Vigueras G, Lozano M, Orduña J.M and Grimaldo F, 2010. A comparative study of partitioning methods for crowd simulations, Applied Soft Computing, Volume 10, Issue 1, January 2010, Pages 225-235
- Solar R, Suppi R, .Luque E, 2011. High performance distributed cluster-based individual-oriented fish school simulation, International Conference on Computational Science, ICCS 2011.
- Solar R, Suppi R, .Luque E, 2012.Proximity load balancing for distributed cluster-based individual-oriented fish school simulations. International Conference on Computational Science, ICCS 2012.
- Demazeau Y, 1995. From Interactions to Collective Behaviour In Agent-Based Systems, In: Proceedings of the 1st. European Conference on Cognitive Science. Saint-Malo, pp. 117,132, 1995.
- Labba C, Bellamine Ben Saoud N, Dugdale J, 2015. Towards a Conceptual Framework to support Adaptative Agent-Based Systems Partitioning. 16th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking

- and Parallel/Distributed Computing (SNPD 2015),
to be appear.
- Lloyd, S.P., 1982: Least squares quantization in PCM.
IEEE Trans. Inf. Theory 28(2), 129–136.
- Forgy, E.W., 1965: Cluster analysis of multivariate data:
efficiency versus interpretability of classifications.
In: Biometric Society Meeting, Riverside,
California, 1965. Abstract in Biometrics, vol. 21, p.
768.
- MacQueen, J.B., 1967: Some methods for classification
and analysis of multivariate observations. In:
Proceedings of the 5th Berkeley Symposium on
Mathematical Statistics and Probability, vol. 1, pp.
281–297.

RAPID PROTOTYPING AND VERIFICATION OF SIMULATING SYSTEMS APPLYING ABASIM ARCHITECTURE

Emil Řezanina^(a), Antonín Kavička^(b)

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

^(a)emil.rezanina@upce.cz ^(b)Antonin.kavicka@upce.cz

ABSTRACT

When designing and developing computer simulation models it is possible to employ different architectures of simulate systems. Quite often Agent-based architectures are used. They provide a number of advantages to the needs of construction, modification and expansions of relevant models. In the field of computer simulations a variety of agent-based architectures are applied. They are available in a number of simulation tools, frameworks and programming languages.

Nowadays, when the emphasis is put on the reduction of time for software systems development, it is often required to have a graphic support for rapid prototyping of simulation models available, using the appropriate architecture.

The paper is aimed at presenting the development of a software framework (working name *ABAframe*) supporting the development of simulators using one particular type of Agent-based architecture of simulation models called *ABAsim*. Said architecture is particularly suitable for creating simulators of service, logistics and transport systems.

The main purpose of the *ABAframe* framework within an integrated development environment is to enable both rapid and partially automated graphical prototyping of simulation models based on autonomous agents and also support the implementation of simulation experiments including evaluation of their results. An important functionality the stated framework features is providing the verification of communication interfaces and communication of inter-agents within a built prototype of simulator.

Keywords:

Agent-based architecture, simulation model, rapid prototyping, prototype verification

1. INTRODUCTION

Currently, a number of simulation tools, frameworks, and programming languages (Zheng 1992) are available for the needs of production of (monolithic) simulation models, where different architectures, or rather approaches, are applied. The following architectures could be used as an example:

- events or process-oriented architecture,
- architecture based on continuous activities,
- agent-based architecture,
- combined / hybrid architecture.

Within the existing tools a support is usually available to facilitate the building of simulation models.

In this paper an attention is paid to one of the used agent-based simulation model architectures called *ABAsim* (Kavička, Klima and Adamko 2005). Using this architecture is particularly suitable for complex simulating systems reflecting traffic within the service, logistics and transport systems.

To simplify and accelerate the creation of simulation models the methodology of rapid (graphically supported) prototyping can be used. This methodology facilitates the implementation of quick configuration of a simulation model prototype, in this case composed of agents. The advantage of rapid prototyping is the possibility to bring the simulation model prototype under formal correctness control.

In order to increase the attractiveness of *ABAsim* architecture it was requested by many users to build a support for a quick and partially automated prototyping of complex simulation models, including the possibility to examine the correctness (verification) of the models. In this connection, the *ABAframe* framework and the *ABAframeIDE* tool, which is constructed above the framework, are being developed.

The concept of *ABAframe* framework, including its functionalities, is described below. Attention is paid mainly to the process of rapid prototyping of simulating systems and to the implementation of relevant prototype verification. Also a case study is presented where the

individual parts of creation and verification of a prototype are demonstrated.

2. ABASIM ARCHITECTURE

Agent-based *ABASim* architecture allows the creation of simulation models using autonomous units, the so-called agents. These agents constitute an encapsulated computer system that is able to flexibly and autonomously operate in a specific environment to achieve the stated objective (Jennings and Wooldridge 200). To "flexibly operate" means reactive, pro-active and social behaviour of the agent. The following figure shows the agent's behaviour when achieving the goal.

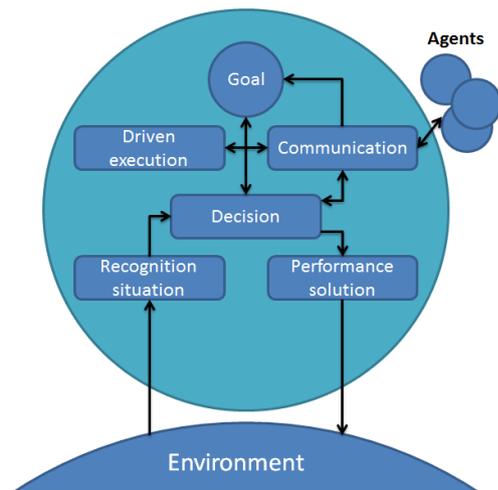


Figure 1: The structure of the agent's behaviour (Adamko 2013)

In the *ABASim* architecture the simulation model consists of cooperating agents organized in a hierarchical structure. The architecture offers two types of agents - control and dynamic agents. Control agents are static entities which exist throughout the simulation run. These agents represent a high-level system entities that cooperate together in order to achieve system goals. The second type of agents is represented by dynamic, autonomous and proactive entities. They are always current under the management of one control agent, but may be transferred under the management of another control agent. Dynamic agents mutually interact with the environment (other dynamic agents, entities) and communicate with the control agent. Dynamic agents are created and cancelled during the simulation run and receive local goals by managing agents (Adamko 2013, Kormanova, Varga and Adamko 2014).

2.1. Composition of Agent

The *ABASim* architecture supports decomposition of individual autonomous agents into specialized components that focus on performing certain activities. The main component of each agent is a *manager*. This component focuses on the control and communication activities of an agent. The remaining internal components of agents are called assistants. They are used to support the *manager*. Assistants could be classified as either *prompt* or *continuous*. The difference between them is that prompt assistant performs the action

immediately, whereas continuous assistant needs some time. Assistants can be divided into several groups according to the type of activity:

- Group of *sensors* has the task of monitoring the area of an agent. This group includes a *query* component and a *monitor* component. The query component is a *prompt* assistant that examines the surroundings at the request of the *manager* and immediately inform it of the outcome. The *monitor* component is a continuous assistant that unlike *query* component monitors the area for some time. If there is any monitored event at this time (arrival of a customer) it informs the *manager*.
- Group of *effectors* represents assistants who perform executive actions in the surroundings of an agent. For immediate performance a prompt assistant *action* is employed and for continuous operation a continuous assistant *monitor* is applied.
- Group of *advisors* is being used by the *manager* as a decision support. This group includes prompt assistant *adviser* and continual assistant *planner*.

2.2. Communication

Basic communication mechanism in *ABASim* architecture is a mechanism of messaging sending. This mechanism is used both for inter-agent communication and internal communication. Messages can be processed in several different ways. Messages are saved in a mailbox of an agent or in a central mailbox of a simulation model. Another possibility is to process the message immediately. Each message contains a timestamp that represents the time when the message should be processed.

For better organization of communication and for simplification of the simulation model design the *ABASim* architecture defines several types of messages. Message types actually enable messages to better illustrate what kind of message it is and how it should be processed. Communication via message types can be seen in the following Figure.

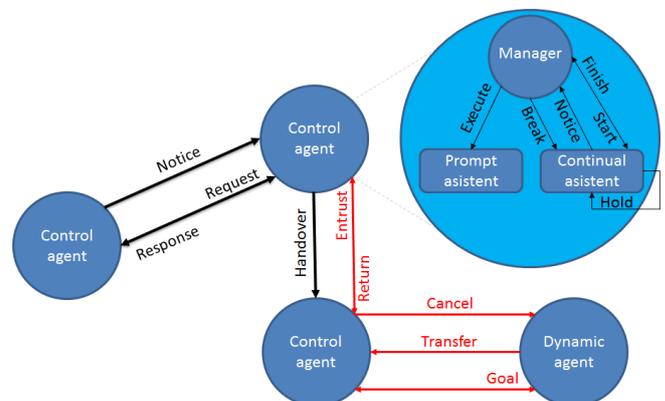


Figure 2: Inter-agent and intra-agent communication (Adamko 2013)

Message types that are used for inter-agent communication are divided into (Fikejz and Řezanina 2013, Kocifaj and Adamko 2014):

- *Notice* – a message used to inform agents about the situation that has arisen. The message only indicates given facts and doesn't wait for a response (e.g. resource was returned).
- *Request / Response* - *Request* message type is used as a request for resource. The request is sent to the addressee (e.g. Deliver mobile source). To the extent possible a response is expected in association with this type of message. Message type *Response* is used as a reply and it is sent back to the original sender.

Messages related to work with dynamic agents are also used for inter-agent communication. It is the possibility of transferring a temporary management (*Entrust / Return*) or permanent management (*Handover*) to a dynamic agent, assignment of goals (*Goal*) and cancelling of goals (*Cancel*) to a dynamic agent. In contrast, the dynamic agent sends to the managing agent that administers it a message that indicates that the goal was achieved (*Done*) or request to move the management of dynamic agent elsewhere (*Transfer*).

Intra-agent communication is open to all its components, but they must follow certain rules for this communication. The *manager* can communicate with all the other components of the agent. Other components do not communicate with each other (Adamko 2013). *Manager* and his assistants communicate through the following types of messages:

- *Start* - This type of message is sent to a continuous assistant. The addressee starts to perform its autonomous activity after receiving this message.
- *Break* - through this type of message the manager affects the autonomous run of continuous assistant. Assistant cease working after receiving this message.
- *Execute* - type of message sent to prompt assistants. Assistant responds with an instant processing of the message and returning results to this type of message.

Continuous assistants are initiated by the *manager* via message *Start* mentioned above. During their existence, continuous assistants can send the following types of messages to the *manager*:

- *Finish* - this type of message is sent when a continuous assistant completes its work.
- *Notice* - a message used by a continuous assistant to inform the *manager* about important matters that require reaction.
- *Hold* - is the only type of message that can have higher timestamp of the message than the current local virtual time of the agent. Message

type *Hold* is used in order to shift simulation time.

More specifications of inter-agent and intra-agent communication can be found in resource (Adamko 2013).

3. RAPID PROTOTYPING

Rapid prototyping is one of the ways to quickly create simulation models of complex systems. The result of prototyping is a simulation model with limited functionality, so-called prototype. Prototypes of simulation models based on Agent-based architectures can be composed of empty agents and communication links among them. In *ABAsim* architecture so-called pre-defined agents are used instead of empty agents. These agents are perceived as agents that contain external interface for receiving messages from other agents as well as a set of internal components of an agent. Mentioned components include internal communication interface, but without internal logic of incoming messages processing.

Tools that provide a graphical user interface are very often used for prototyping. These tools lead to a significant speed-up of the creation of simulation models and to convey its visualization. Visualization brings greater knowledge of the proposed model. This is very useful for complex solutions.

Another advantage of prototyping is an early control of the model. Formal correctness does not relate directly to a simulation model but to its prototype.

4. VERIFICATION

Verifying the correctness of the simulation model consists of verification and validation of the model. Model verification is a process of checking the formal correctness of the model. In contrast, model validation is a process to verify that the outputs of a simulation model correspond to the realistic characteristics of the modelled system.

When rapid prototyping simulation model in *ABAsim* architecture we deal only with the verification of the model prototype. Validation cannot be applied in this case because components of pre-defined agents do not contain internal logic and therefore the simulator is not yet built.

Verification of the prototype of agent-based simulation model determines whether all communication interface settings are correct. These interfaces can be divided in two groups, namely:

- The first group consists of the *communication interface of agents*. Every agent knows what messages it receives from other agents and what messages it sends.
- The second group is composed of *communication interfaces of internal components* of individual agents. When processing incoming messages from another agent the *manager* communicates with his

assistants. The communication interface of each assistant of the manager must be correctly configured so that the *manager* is able to communicate with the assistant.

The verification of the prototype also focuses on inter-agent communication and the involvement of agents in the simulation model.

The main advantage of prototype verification, as mentioned, is the possibility of early correctness checks. This allows to eliminate errors that might have arisen at the beginning of the draft before complete implementation of the simulation model's internal logic.

5. ABAFRAME FRAMEWORK

ABAframe Framework uses a special simulation core, which allows the simulation above the simulation models created in *ABAsim* architecture. Simulation core works with discrete events that represent moments of message processing in the *ABAsim* architecture. Simulation core performs simulation above the simulation model in two phases. Synchronization algorithm of the simulation core is based on the principles of the resource (Adamko 2013).

ABAframe Framework enables the user to create a simulation model prototype using graphical user interface (or more precisely integrated graphical development environment). The prototype consists of individual pre-prepared (partially formalized) agents and specifications of inter-agent communication.

5.1. Creating a prototype

For creating a prototype of agent's simulating system within the environment provided by the *ABAframe* Framework a tool called *ABAframeIDE* is used. This tool divides the creation of a prototype into several phases:

- (1) A draft of a model, which consists of empty agents and defined communication links among agents.
- (2) Specifications of individual internal components an agent will consist of.
- (3) Setting the communication interfaces of internal components of individual agents.
- (4) Implementation of internal model for incoming messages processing within the component manager.
- (5) Generation of frames.

In the phase of a *draft of the model at the level of agents* the user determines what agents, agent's models and inter-agent's communication links the model will consist of. Inter-agents communication link represents the sending of one message. Any agent can be a sender and the recipient of the message can be either a specific agent or also an agent model. The message must also specify the type (e.g. *Notice*), code (e.g. *Incoming customer*) and parameters (e.g. *Customer*) of the message. For message type *Response* it is necessary to set to which message type *Request* it replies.

The internal composition of every created agent must be set. Setting the composition of an agent comes under the *agent's internal components specification* phase. For each agent a tab appears that allows insertion and removal of manager assistants (i.e. *monitor*, *action*).

The next phase is the *setting of communication interfaces of internal components*. This phase is used to set what messages assistants receive and how they respond. In addition, for continuous assistants it is also necessary to define what messages they will send during their operations (to themselves or the *manager*).

The penultimate phase of developing the prototype represents an *implementation of an internal model of incoming messages processing within a manager*. At this stage, there are several ways how to implement the internal model. An imperative approach (using the resource code of the programming language) or declarative approach (based on the graphical specification of the relevant - coloured - Petri nets) may be used. If the first option is chosen, the implementation of an internal model is omitted for the time being. The programming of this model must be finished after the generation of the prototype.

The last phase is the *generation of frames*. In this phase, first of all a formal inspection of correctness of individual parts' settings from previous stages takes place. If this goes through successfully, an export of a simulation model prototype to a resource code of the relevant higher programming language (in which the model will be further developed) ensues.

5.2. Verification

It is possible to conduct a formal control of correctness within the *ABAframe* framework. It is divided in two levels.

The first level of control is performed prior to the generation of frames in the *ABAframeIDE* tool. The main objectives of this level are to check the correctness of the setting of agents' identifiers and their internal components. Furthermore, it is examined whether some communication links between two agents do not share the same type and code.

The second level of control is performed after the export of a prototype. This level is composed of the following verification:

- (i) verification of communication interfaces of agents,
- (ii) verification of communication interfaces of agents' internal components,
- (iii) verification of inter-agents communication and
- (iv) verification of the involvement of agents in a complex simulation model.

Verification of communication interfaces of agents controls which messages agents receive and which messages they send. Each message must pass the scrutiny of their requirements. These requirements include e.g. type code, sender, recipient, etc.

Verification of communication interfaces of agents' internal components controls each interface of manager's assistants. This control should check whether the interface of each manager's assistant is properly applied in the context of the internal logic of incoming messages processing. This verification does not occur if an internal model of incoming messages processing of the *manager* has not yet been specified.

Verification of inter-agents communication checks whether each message sent from agents has its existing recipient who has a relevant routine utilities of this message. Lists of agents' incoming and outgoing messages are used for this purpose. During the control it is checked whether a message from the list of agent's outgoing messages has the same message notation in the list of recipient's incoming message. Messages are the same if they have the same type, code and parameter list.

Verification of the involvement of agents in a complex simulation model aims to reveal the single alone agents. These are agents that do not communicate with any other agent, and thus are not utilized within the model.

If the model prototype fails to pass the verification successfully the framework informs the user about the formal faults of the model so that they could be removed subsequently.

The second level of control do not have to be used for generated simulation model prototype only. It can also be used for specific simulation model.

6. DEMONSTRATION APPLICATION

A procedure of creation of a simplified simulation model of queuing is illustrated here to demonstrate the work with the *ABAframe* framework and *ABAframeIDE* tools.

6.1. Description of a model of queuing

Customers enter the model from its surroundings, attend two types of service and leave the system after their completion. While the first type of service is linked to a

stable resource to which the customer must transfer, the second resource is mobile, and thus transfers itself to the customer in order to perform the second service. If the customer comes to a service that does not have a free resource, he will join the appropriate queue waiting for this resource and stay here until taking his turn (Kavička, Klima and Adamko 2005).

6.2. Creating a queuing model prototype

To create a queuing model prototype the *ABAframeIDE* tool is used. The procedure of generation of a model using this tool will be shown in this section.

In the first stage a simulation model was modelled. It includes *surroundings*, *service manager* and *resource manager* agents. Each of these agents has a specific role. In addition, it was necessary to set up communication between these agents.

Surroundings agent is responsible for the contact between the model and its environs (arrivals and departures of customers). This agent cooperates with *service manager agent*, which is in charge of organizing the service. Arriving customer will undergo two types of services here and leave the model after completing these services. To work with resources the *service manager agent* communicates with the *resource manager agent*. This agent is responsible for allocation and relocation of resources to the service. In addition it deals with queues of applicants for resources.

The design of individual inter-agents communication links consists of two steps. The first step was setting the resource and destination of the message. The second step was to set the type, code and list of parameters of the link. List of parameters contains the names of parameters the message will transmit. E.g. communications link "Incoming customer", which leads from the *surroundings agent* to the *service manager agent*, carries the Customer parameter in this list.

The result of this phase is shown in the following figure.

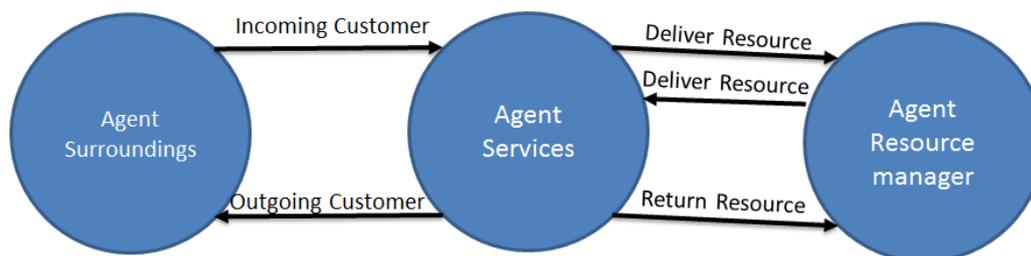


Figure 2: Design of simulation model of queuing.

The second phase dealt with the composition of individual agents. In this step components that the manager will use for his work were selected for each agent.

Surroundings agent contains only a component process *Entering Customers*. This component aims to generate

periodic messages announcing the arrival of a new customer into the system.

The manager of *Service agent* operates with four processes. The first process is the execution of service A (*Service A* process). The second process is responsible for the transfer of a customer from service A to service B (*Move customer* process). This process is

followed by the process of performing service B (*Service B* process). After completion of service B the customer leaves the service and heads to the exit of the system. The next process is employed here (*Customer outgoing*

process). The compositions of this agent in *ABAframeIDE* tool is shown in Figure 3.

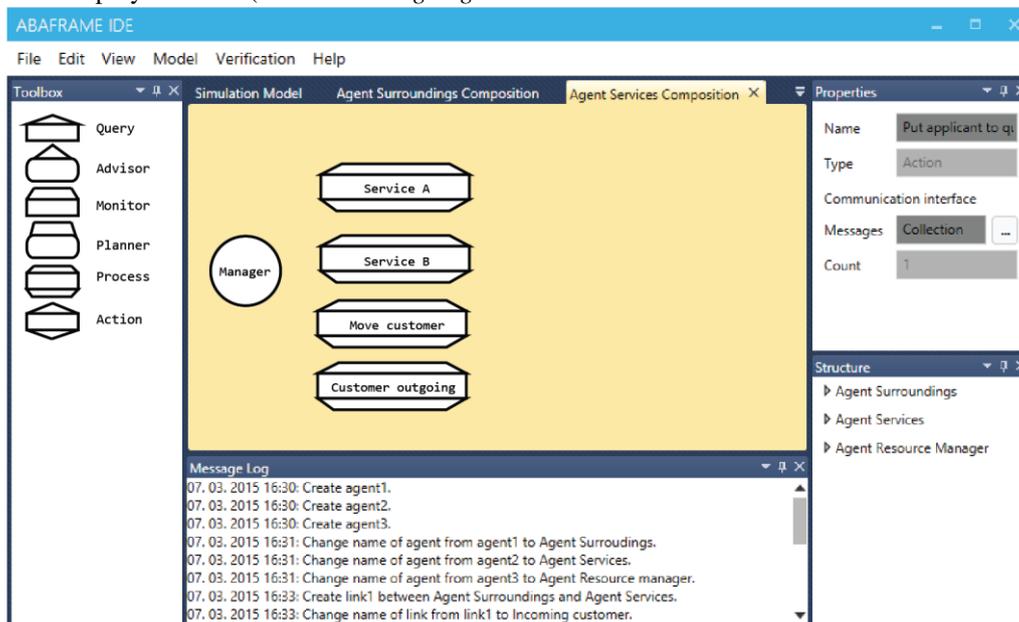


Figure 3: The compositions of service agent in *ABAframeIDE*

The last agent, the *resource manager* agent, has two groups of manager's assistants. The first group helps the *manager* with allocation and relocation of service resources. This group consists of the following components:

- action for allocation of resources (*Assign resource*)
- action for recovery of resources (*Return resource*)
- advisor for selection of free resources (*Selection of free resource*)
- request to determine whether it is necessary to relocate the resource (*Need move resource*) and process of transferring the resource (*Move resource*).

The second group is in charge of the queue of customers waiting for service. The queue depends on the occupancy of resources. The group includes an action for adding an applicant to the queue (*Put applicant to queue for resource* action) and actions for removing the applicant from the queue (*Remove applicant from queue* action).

Apart from setting the name of every assistant it was also necessary to set a communication interface. The setting of *Put applicant to queue for resource* action assistant will be used as an illustration. This component represents a prompt assistant that should add the sent applicant to the queue for the resource – as expected by the *manager*. The manager does not communicate with this assistant in any other way. Communication interface of the assistant includes only one record consisting of a *notice* message type, "*Put applicant to queue for resource*" message code and a parameter list. The parameter list contains only one

parameter called "*Applicant*". There is no need to set a return value in this case.

After setting all the agents assistants it is possible to proceed to the stage of setting the internal logic of incoming message processing. Every manager must respond to an incoming message (inter-agents and intra-agents). In our model, for example, the *Surroundings* agent must respond to three types of messages. The first type of messages comes from the *Service* agent, which sends information about an outgoing customer. The other two types of messages are from continuous component - *Entering customer* process. This assistant informs the *manager* of the arrival of a customer or of the termination of its activities.

For the implementation of internal logic a possibility of programming in the resource code was selected.

ABAframeIDE tool was used for the prototype verification within each particular phase.

The completion of each phase and successful verification was followed by the generation of the simulation model frames.

6.3. Completion of the simulation model

The result from the previous chapter can be added to the project C # programming language. It is sufficient to then add a reference to *ABAframe* framework into the project, which will allow further work with this simulation model, whether it concerns verification or launching of the simulation model.

The generated simulation model is not yet complete. The creation of an internal model of message processing by

the manager was omitted and the internal logic of generated individual assistants is missing. The missing parts must be programmed by the user.

After completion of the simulation model the *ABAframe* framework verification may be used to control the formal correctness.

Figure 4 shows the possible implementation of a complete simulation of queuing

7. CONCLUSION

This article discusses rapid prototyping and verification of simulation models in the agent oriented *ABAframe* architecture. The *ABAframe* framework was introduced here together with its tool *ABAframeIDE* that allow rapid prototyping and verification of simulation models in this architecture. A demonstration application dealing with queueing was created for better understanding of the design of simulation models using the *ABAframe* framework. The process of creating a prototype of this application is described in this article.

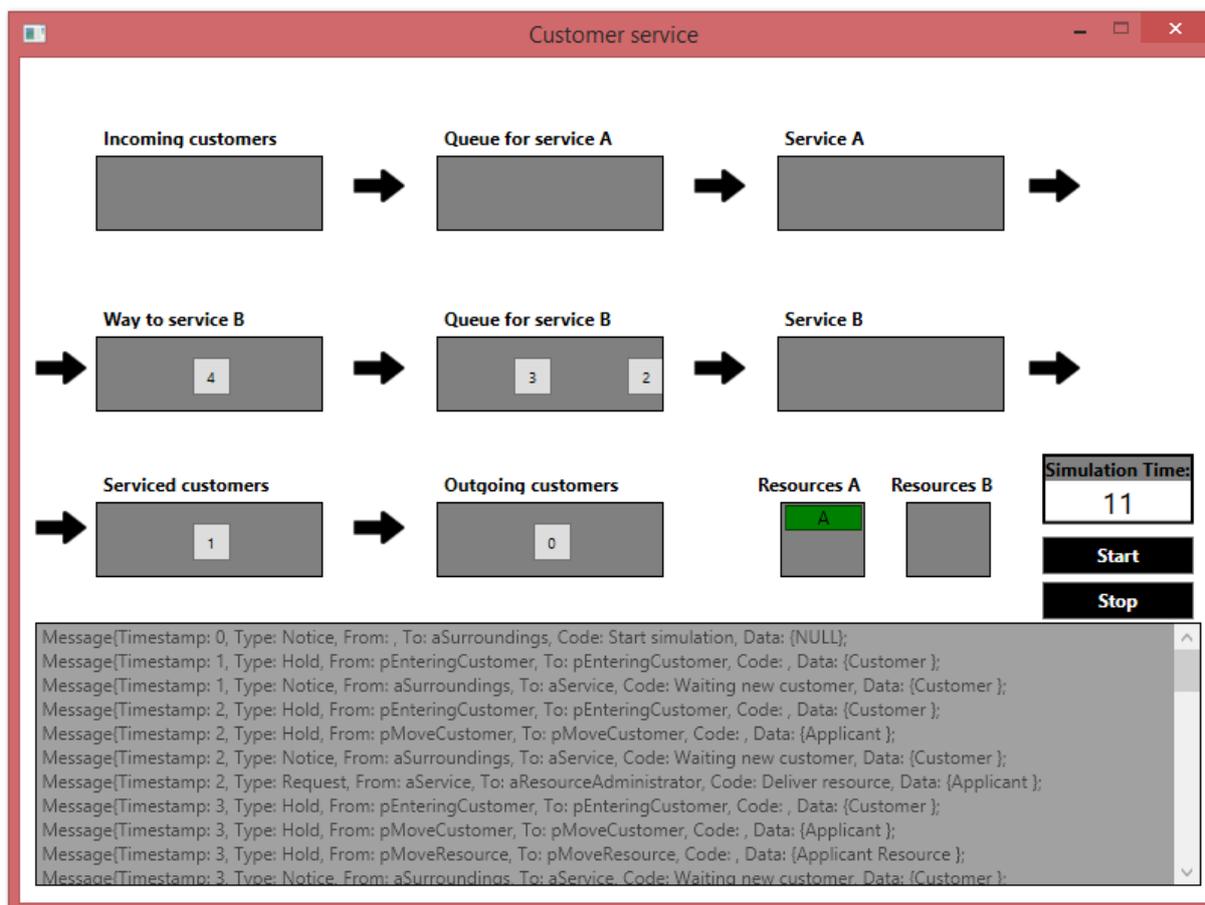


Figure 4: Resultant simulation of queuing

proto the article.

REFERENCES

Zheng, Hong, et al. A Primer for Agent-Based Simulation and Modeling in Transportation Applications. No. FHWA-HRT-13-054. 2013

KAVIČKA, A., KLIMA, V., ADAMKO, N. Agent-based simulation of transportation nodes, EDIS, University of Žilina, 2005 (in Slovak), ISBN 80-8070-477-5.

JENNINGS, N. R., WOOLDRIDGE, M. Agent-Oriented Software Engineering. Artificial Intelligence. 2000, roč. 117, s. 277–296.

ADAMKO, N. Agentovo orientovaná simulácia zložitých obslužných systémov. Habilitation thesis, Žilina, Faculty of management science and informatics, University of Žilina, 2013.

KORMANOVA, A., VARGA, M., ADAMKO, N. Hybrid model for pedestrian movement simulation. In: The 10th International Conference on Digital Technologies 2014 [online]. 2014 [cit. 2015-06-09]. DOI: 10.1109/dt.2014.6868707.

KOCIFAJ, M., ADAMKO, N. Modelling of container terminals using two-layer agent architecture. In: 2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI) [online]. 2014 [cit. 2015-06-09]. DOI: 10.1109/sami.2014.6822416.

FIKEJZ, J. - ŘEZANINA, E. Simulation of localization rolling stock within the railway network model utilizing agent-based simulation. In The European Simulation and Modeling Conference 2013. Ostend: EUROSIS-ETI, 2013. s. 290-296. ISBN 978-90-77381-79-3.

MODEL OF A RAILWAY INFRASTRUCTURE AS A PART OF A MESOSCOPIC TRAFFIC SIMULATOR

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

In order to investigate or to optimize railway traffic and infrastructure within complex railway nodes (stations and junctions), computer simulation methods are frequently utilized. Relevant simulation models can apply different degree of granularity, which defines the level of details within the computer simulating system with regard to the particular object of investigation.

For railway traffic simulators performing investigations on a mesoscopic level, it is important to support rapid prototyping of a target model in order to shorten the life cycle of simulation projects/studies. From that point of view, relevant track infrastructure submodels are supposed to be created quickly and efficiently. The contribution deals with the issue of rapid prototyping and verification of infrastructural submodels which belong to mesoscopic simulators focused on the traffic within railway stations and junctions. An appropriate editor of infrastructure provides a variety of sophisticated functionalities that support efficient and partially automated constructions of track layouts for different scenarios of simulation experiments.

Keywords: track infrastructure model, railway station, mesoscopic traffic simulation

1. INTRODUCTION

One of the main prerequisites for conducting simulations of railway traffic is to have a suitable model of railway infrastructure corresponding to the investigated segment of a railway network at disposal. The structure of the model does not have to suit the specific simulator or simulation experiment accurately, and thus even if the model is available, its modification in the professional graphic editors can be very laborious and time consuming (especially if the infrastructure model is processed on a real scale).

For these reasons, a strong motivation to develop an editing tool, which would significantly simplify, partially automate and accelerate the construction of infrastructure models, exists. The mentioned editor must meet the needs of the mesoscopic traffic simulator within railway nodes, which does not require an infrastructure model with such accuracy of metric ratios within the model of track infrastructure as microscopic simulators do, but it must contain a level of detail sufficient for investigating the throughput of railway stations or the effects of delay on railway traffic.

2. CURRENT STATE

Nowadays there are many simulation tools that use their own approaches to constructing models of a railway infrastructure. Villon, RailSys and OpenTrack can be mentioned as examples of such simulation tools (Simcon 2014, Huerlimann and Nash 2010, Radtke and Bendfeldt 2011). RailSys and Villon are simulation tools that utilize modeling of a railway infrastructure on the microscopic level and therefore the infrastructure model is constructed in a real scale. In contrast, OpenTrack uses schematic visualization, which allows it to partially accelerate the preparation of the final model. More complex railway objects, for example switches and crossings, must be created by merging individual segments.

The following two different implementation concepts of editors are used for the above mentioned simulation tools:

- implementation within the simulator (OpenTrack),
- secondary software (RailSys and Villon).

The simulation tools, mentioned above, describe the railway network using graph representation and they allow to manage properties, such as gradients, speed limits, signals, etc, and objects on the graph's edges and nodes. But only OpenTrack supports the latest version of the railML standard (RailML 2015) as a uniform prescription for storing and migrating the infrastructure model.

3. APPROACHES TO VISUALIZATION

From the viewpoint of the editor (called *TrackEd - Mesoscopic Track layout Editor*) in development by the author of this article, it is necessary to distinguish between a *visual scene* (Figure 1) presenting railway infrastructure in a graphical form and a *data scene* defining *numeric* values related to metric/topological and vertical alignments (attributes). The *data scene* is composed of appropriate data structures utilized by relevant calculations of train movement dynamics (Iwnicki 2006).

For the needs of a different approach to visualization, a higher degree of abstraction is applied. It enables significantly faster creation of the railway infrastructure model, which is based on the schematic visual depiction of the track infrastructure. Thus, from the viewpoint of visualizations there is no requirement for graphical

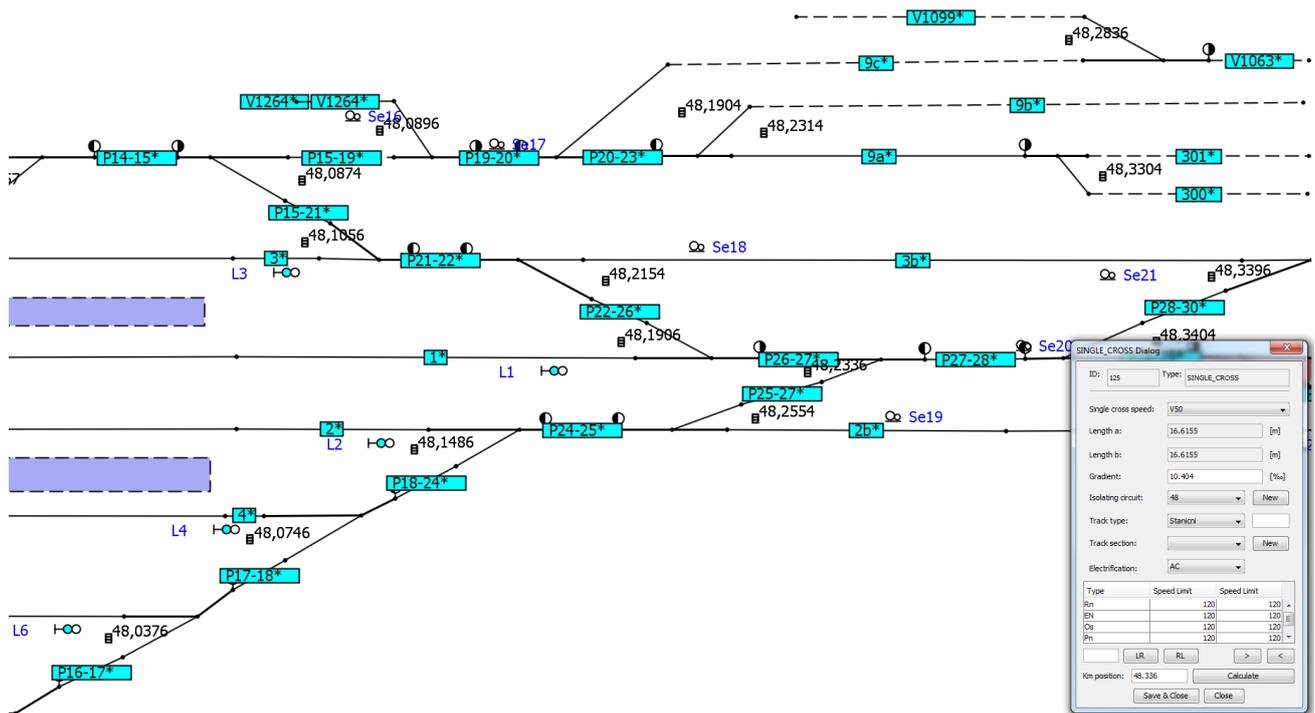


Figure 1: Schematic plan of a track infrastructure within an illustrative station

depiction of the infrastructure in a scale. On the other hand, parameterization of infrastructural objects within relevant data structures can support realistic calculations concerning train rides. The mentioned data structures store relevant entries which specify topological, metric, and vertical alignments within the track layout and mirror real situation connected with the investigated station/junction.

While creating the final model of a railway infrastructure in the *TrackEd* editor, isolated circuits (according to an installed interlocking system), locations of signaling devices, electrification of tracks and useful lengths of rails, and finally the speed limits on individual railway segments are also realistically concerned.

Schematic visual depiction of track layout is based on using polylines, which means that more complex shapes, like track arcs or track crosses, are not depicted in a highly authentic way. The concept of schematic representation of railway infrastructure with the support of visually deform individual track objects within the infrastructure without losing or changing the real topological, metric, or vertical alignments defined within the *data scene*, accelerates the constructing of the model.

4. OBSERVED OBJECTS

The typical model of railway infrastructure within the stations/junction is composed of the following basic elements (Kubat 1999, Jirsak 1979):

- tracks,
- switches,
- crossings,

- semaphores,
- limit signs for train positions on tracks,
- platforms,
- and alternatively, station buildings.

Basic atomic objects serve not only for schematic visualization, but also for describing the physical infrastructure, especially the real metric and topological ratios in the *data scene*.

Data structure of the undirected graph implemented in operation memory and composed of two types of objects, nodes and edges, type is used as a representative of a railway network. The edge represents an atomic track section, hereinafter referred to as *segment*, containing the basic characteristics of length, gradient, arc, speed limits, isolated circuits, signals, track type, electrification, etc. Individual nodes of the graph then represent a link between such segments. The *node* is an important object which describes not only the position of segment endpoints in the schematic plan, but it also contains the coordinates of their real positions in the railway infrastructure in the background (Lewis and Denenberg 1997).

5. PROTOTYPING

To accelerate the construction of the infrastructure model, it is possible to use predefined *prototypes* of selected track objects (e.g. switches, crossings, or aggregated track sections). Prototype can be seen as a part of an infrastructure comprised of elementary objects described in the previous section.

As it was said, the *atomic prototype*, used to construct the simplest track object representing one edge in the

undirected graph, is a *segment*. Decomposing the infrastructure into segments allows modelling those parts of railway infrastructure, in which its properties change (gradient, radius of the arc, speed limit, etc.). To ensure automatic calculations common for railway traffic, it is possible to aggregate a set of adjacent segments into the track. This is especially needed for tracks in stations which change the gradient in some places, but in the context of railway traffic, an example can be a calculation of useful length, they are designated as a one track.

Other typical representatives of prototypes are *switches* and *crossings*, which consist of group of segments. In the context of future development, will be implemented even more complex aggregation of prototypes mentioned above, i.e. a part of development of switches composed of atomic segments, switches and crossings.

Each selected prototype is realistically parameterized, i.e. it is related to a table of standardized dimensions (for example a switch prototype is connected with standardized maximal allowed train speed on the secondary branch) (Kubat 1999). It means in fact that lots of predefined track objects can be utilized without the need to customize their parameters and massive manual specifications are not needed.

6. OBJECTS POSITION

The *TrackEd*, the railway infrastructure editor, as already mentioned, distinguishes between the mesoscopic visual scene and real parameterization of the *data scene*. The mentioned real parameterization is important for e.g. signals, location of which is specified by real coordinates and not by the distance from the endpoint of the segment it belongs to. For this reason, each node contains an attribute of real position within the railway infrastructure. In the *TrackEd* editor, so called stationing (localization) for the input of node position in the *data scene* is supported. It is such approach, in which each node contains a kilometeric position from the start of the railway line (start of railway line corresponds to a zero kilometeric position). Stationing in the Czech Republic serves for unambiguous determination of the position of all objects in the railway infrastructure (Kubat 1999).

In aggregated prototypes forming a continuous subgraph of the railway infrastructure, which consists of nodes and edges (e.g. switch and crossing), users simply assign a location to only one node. The reason for it being the calculation of kilometeric positions of the remaining nodes based on the dimensions of the prototype listed in table of standardized dimensions (Fridrich 2008, Kubat 1999). In simple switches is it a node, which indicates the start of the switch in the setting-out diagram (Figure 2). In contrast, the location is determined only for the node which represents a point of the intersection for crossings.

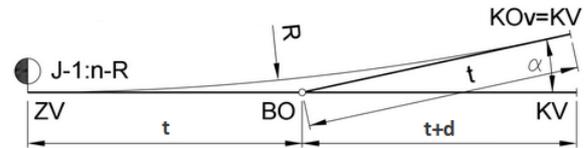


Figure 2: Setting-out diagram of a simple switch (Fridrich 2008)

In practice, there are other approaches to entering real positions of objects than stationing. An example is the use of real GPS coordinates or real coordinates of different geographical coordinate system. To use these approaches, it is necessary to have a very detailed source model to build the schematic plan of a railway infrastructure and, as already mentioned in the introduction, it does not have to be available.

7. TRACK ROUTES

Within the simulation, the train movements are closely connected with track routes. Those routes can be by either dynamically calculated (during a simulation experiment) or pre-calculated before the simulation starts. The train routes can differ for various types of trains. The *TrackEd* editor supports semi-automatic calculation of a complete set of primary and alternative train routes for all possible train transfers within the investigated track layout. In other words, each train route consists of segments forming a continuous subgraph of the railway infrastructure and each train is assigned a set of train routes with added priority. The priority is used during the simulation experiment to select a train route, which is not occupied or built for another train, for the next train movement (Hansen and Pacht 2008).

Filtering mechanisms by specified criteria, such as the maximum route length, allowed gradient, the radius of an arc, or electrification, can be applied to a selection set of train routes. In addition, for each route we can define the maximum allowed speed for a passing train in percentage to the maximum speed of each individual segment of the route. This corresponds with real train movement on track layout, which does not move by speed corresponding to the maximum design speed of each element of the railway network.

From the train movement, especially in the station, it is clear that each train can be assigned a large set of train routes. In the station, the train is assigned a route to the platform through the input development of switches and a route from the platform to the output development of switches. Because the same types of trains frequently occupy almost always the same platforms or station tracks thus use the same routes, it would be very inconvenient to assign these trains the input and output routes individually. For this reason, the train routes can be divided into groups, and before starting the simulation we can assign a group of all primary and alternate routes from the beginning of the train's movement to the end.

Such an approach enables a final control of the whole system of train routes by a professional - railway

technologist. Those pre-calculated and checked train routes are then used during traffic simulations. From the view point of determining the occupancy of train routes during simulation experiments, mainly the effect on the train movement in the station, an automatic calculation of useful lengths of station tracks is applied. The user is then informed about the maximum length of the train that can stand on the track element (Kubat 1999).

8. PLATFORM

One of the most important objects of railway stations, that should not be overlooked in the railway infrastructure editors, are platforms, because they have an influence on the movement of the train in the station. Platforms are not only important objects for visual inspection of the railway station model and an important factor for technologist that creates train routes, but they also directly affect the realization of the train arrival dynamics on the station track or in other words, they define the border place for the train to stop. In the *TrackEd* editor, each platform is always associated with a set of segments that form the platform track. For the needs of a mesoscopic railway traffic simulator, the start and the end position (border places) are parameterized for the platform as for any other objects of the infrastructure.

9. MIGRATION OF THE MODEL

An important part of every transport networks editor, in our case the railway network, must be the ability of the implemented product to store or further edit an unfinished design of an infrastructure model.

In the *TrackEd* editor, the resulting (visual) schematic plan, reflecting a given version of track infrastructure within the station, can be either further processed (i.e. another layout version can be formed), or exported into *MesoRail* - mesoscopic simulation tool currently in development. For such migration of the infrastructure between two different products, it is necessary to choose the type of migration that will be universal, easily validated, and quickly changeable in case of extended requirements for the detail of the infrastructure. For these purposes, transmission using a XML file, which can also be viewed and possibly edited in any text editor, was chosen.

During the development our own custom template, inspiration for which is the universal standard railML that prescribes uniform templates for saving the infrastructure in XML (RailML 2015), has been created. Such standard provides many XML templates, which can be used also for specifications of railway infrastructure. Some of railML-templates inspire our own template used within *TrackEd* editor (direct utilization of railML-templates was not suitable, because they are excessively complex and do not obtain any elements needed for schematic plans).

The custom template is designed in a way, in which any change of object properties of the infrastructure has little influence on the template structure and the mechanisms of its processing in the editor. The template is divided into separate blocks that describe different parts of the infrastructure model, i.e. nodes, segments, track routes, and platforms. For the needs of the editor, the template structure enables to save prototypes, all of its elementary objects (nodes and segments), in such a way, in which the editor is able to identify and correctly put them together during their import into the editor. This is especially important in situations, in which we need to save an unfinished model, which we want to modify in the future.

Nowadays, requirements for interoperability of software applications are more and more prominent. Such requirements are met by the railML standard, which can be implemented in a mesoscopic editor in the future and therefore use this editor with other tools than *MesoRail*, for which it has been primarily developed.

10. THE CONSEQUENCES OF A SCHEMATIC PLAN

The consequence of using the concept of a schematic layout plan for the needs of mesoscopic traffic simulations is realization of the train movement on a deformed infrastructure. That means that the graphic representation of a train, or its parts, always has to dynamically conform to the track element (with a given degree of deformation) on which the train is currently moving. The cause is, as already mentioned, the difference between the visual presentation of the track layout and metric and topologic ratios set in the *data scene*. Such consequence can be partially eliminated by the user when creating the infrastructure model.

11. MODEL VALIDATION

From the perspective of the simulator every inaccuracy in the infrastructure model can be the cause of distorted or bad results of simulation studies. The most common reason is an insufficiently checked model of the railway infrastructure in the editor before its final importing into the simulator. Inaccuracies, such as incorrectly input gradient, radius of the arc, or speed limit of every track elements, are not undebuggable in the simulator. Therefore, a strong emphasis to implement sophisticated functionalities in the editor for detecting coarse and even minor faults of this type exists.

Functionality for validating the infrastructure model can be divided into the following groups:

- visual support,
- automated control.

11.1. Visual support

For visual inspection of the constructed infrastructure model, we can use *visual views*, which emphasize user-defined groups of attributes for the relevant track objects. In the *TrackEd* editor, two types of visual views are supported. They are visual views that highlight

nodes with incident edges (segments), for which the gradient, radius of the arc, and the maximum allowed speed for the selected train for both directions of the move change. The second type is visual views that color code individual segments according to their values set in the *data scene*.

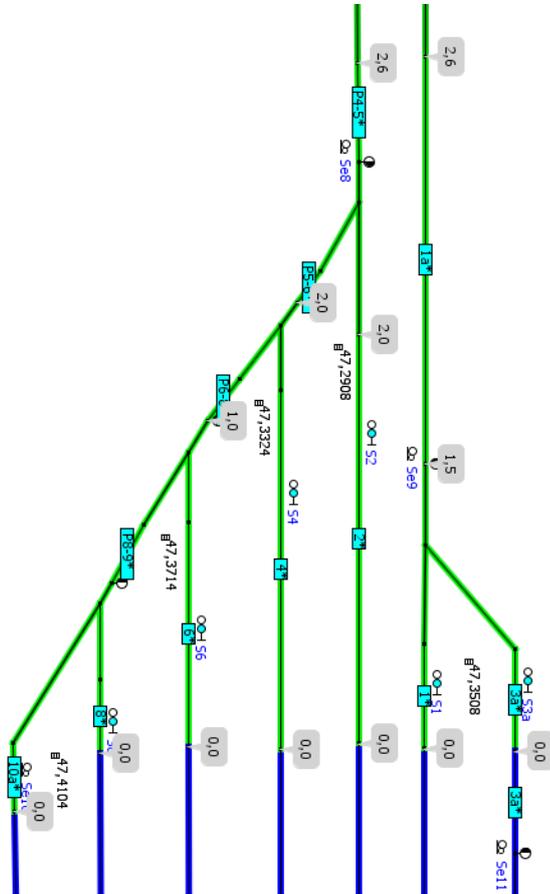


Figure 3: Visual view of the gradient changes

With this support, the user has an overview of changes across the infrastructure using one selected visual view (Figure 3).

11.2. Automated control

In the *TrackEd* editor an automated control of the result infrastructure model, in addition to the visual views, is supported. It represents a group of functionalities which serve to detect gross and informative errors in the model. A gross error is such an error which has a major effect on the simulation study. A list of gross errors is as follows:

- localization of all nodes,
- localization of neighboring nodes,
- localization corresponding to the segment length,
- isolated circuit in all segments and their adjacency,
- continuity of the train routes.

On the other hand, an informative error is the absence of a warning signal or inappropriately chosen type of a signal device on the track element and their incorrect placement from the beginning of the isolated circuit according to the standards or missing limit of switches (Kubat 1999).

Results of the automated control are always shown by the respective track element (Figure 4) on which they are discovered. It gives the user the possibility of a quick change without having to search for it.

Verification of the concept of creating a railway infrastructure model was conducted above testing schematic plans of selected railway stations in the Czech Republic. The resulting models were validated with the official schematic plans used by *The Railway Infrastructure Administration* (SZDC), management of the railway infrastructure in the Czech Republic, in mind.

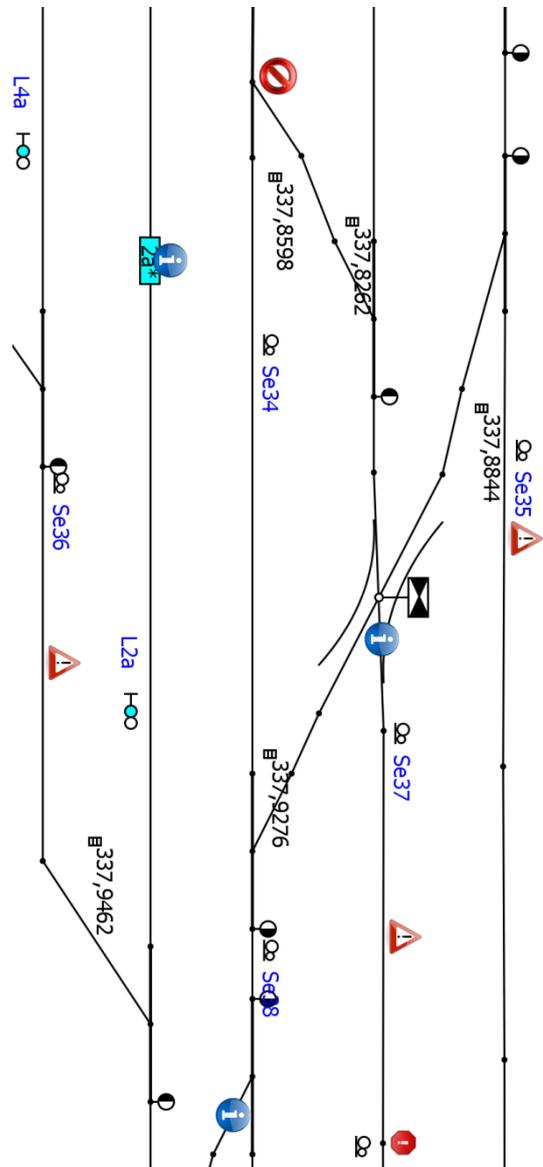


Figure 4: The result of automated control

12. PERSPECTIVE OF DEVELOPMENT

The perspective of further development is related to certain extensions of *TrackEd* editor. Within the frame of a new editor mode the special constructions of scalable track layout models are supposed to be supported. It means that the infrastructural model can combine submodels built on different levels of details (mesoscopic and macroscopic) (Hansen and Pachtl 2008).

Another perspective option is to extend the functionality of the editor by automatic transformation of the designed model, or rather its visual schematic plan, into a plan in a real scale by using metric ratios from the *data scene*. The benefit of such transformed model is brought especially for users performing simulation experiments. User will have a better idea about the actual metric properties behind the scene (eliminates the effect of deformed infrastructure to animate the movement of the train).

13. CONCLUSION

The developed *TrackEd* editor is used to construct railway infrastructure models by using a schematic concept of visualization and thus distinguishes between *visual scene* and *data scene*.

It applies accelerating and innovative practices while constructing the infrastructure model, especially *rapid prototyping*, mainly if we do not have a digitized plan of tracks layout in real scale of the investigated segment of a railway network (for example in AutoCAD).

The editor also supports many sophisticated features which help automate some phases of the methodological approach focused on creating track layout models. An example is an automated principle of calculating train routes for different types of trains. In the editor, support functionalities helping to validate the model before it is imported into the railway traffic simulator are also kept in mind.

REFERENCES

- Iwnicki Simon, 2006. Handbook of railway vehicle dynamics. BocaRaton: CRC/Taylor.
- RailML, 2015. RailML: The XML interface for railway applications. RailML.org. Available from: <http://www.railml.org/index.php/home.html> [Accessed 11 July 2015].
- Simcon, 2014. Villon. Simcon. Available from: <http://www.simcon.sk/en/tools/villon> [Accessed 11 July 2015].
- 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 11 July 2015].
- Huerlimann Daniel, Nash Andrew, 2010. OpenTrack - Simulation of Railway Networks: User's Manual Version 1.6. ETH Zurich: Institute for Transportation Planning and Systems.
- Fridrich Karel, 2008. SZDC S3 rule – Switches and railway station. Prague: CVUT. Available from: http://kzs.fsv.cvut.cz/4/zst3/zst3_vyhybky.pdf. [Accessed 11 July 2015]. [in Czech]
- Lewis Harry, Denenberg Larry, 1997. Data Structures and Their Algorithms. Boston: Longman.
- Kubat Bohumil, Fliegel Tomas, 1999. Railway buildings 30. Prague: CVUT. [in Czech]
- Jirsak Zbynek, 1979. Railway stations and junctions. Bratislava: Alfa. [in Czech]
- Hansen Ingo, Pachtl Jorn, 2008. Railway timetable and traffic. Hamburg: Eurailpress.

URBAN TRANSPORT INFRASTRUCTURE: A SURVEY

Idalia Flores^(a), Miguel Mújica Mota^(b)Salvador Hernández^(c)

^(a) Faculty of Engineering, UNAM

^(b)Aviation Academy

Amsterdam University of Applied Sciences

^(c)Technological Institute of Celaya

^(a)idalia@unam.mx, ^(b)m.mujica.mota@hva.nl

^(c)salvador.hernandez@itcelaya.edu.mx

ABSTRACT

The urban transport infrastructure is one of the most important problems for the cities, and involves many aspects that concern to citizens, governments and the economical growth of the countries. The objective of this paper is to show how this issue has been studied in recent years, with emphasis in the use of simulation and optimization at the whole planning process. We also consider the important key topics as sustainability, costs and risks, mobility and environment impact. Some study cases are shown in order to clarify the concepts presented.

Key words: Urban transport, planning, simulation, optimization.

1. Introduction

Nowadays one of the bigger problems in cities is the transportation system and its infrastructure. There have been lots of studies and research in recent decades trying to find solutions. In general there is an economic impact when countries make an investment in this sector. Most of the studies on transport infrastructure, in particular, focus on its impact on growth. In the past two decades the analytical literature has grown enormously, with studies carried out using different approaches, such as a production function (or cost) and growth regressions, as well as different variants of these models (using different data, methods and methodologies), the majority of these studies found that transportation infrastructure has a positive effect on output, productivity or growth rate Calderon & Serven (2008). One of the pioneers was Aschauer (1991) who, in his empirical study, provided substantial evidence that transport is an important determinant of economic performance. Another example is the study of Alminas, Vasiliauskas and Jakubauskas (2009), who found that transport has contributed to growth in the Baltic region. Another study on

the Spanish plan to extend roads and railways that connect Spain with other countries concludes that these have a positive impact in terms of Gross Domestic Product (GDP) Alvarez-Herranz & Martínez-Ruiz (2012). In a study of the railroad in the United States, it is mentioned that many economists believe that the project costs exceed the benefits Balaker (2006). However, the traditional model of cost-benefit assessment does not include the impact of development projects De Rus (2008). In these studies focused on growth, we see there is a bias towards economic rather than social goals. That is why it is important to emphasize the impact of transport infrastructure on development and not just growth.

Some papers that deal with economical and social impact are shown here.

In order to show the subject clearly, we will use a systems approach, dividing urban transport infrastructure according to Calgary (2009)

Urban Transport infrastructure:

- Transportation Planning
- Transportation Optimization
- Transportation Simulation

According to this system paradigm, this paper is focused on the description of the research made in the last five years, mainly considering optimization, and simulation and the intelligent systems. The structure of this paper is as follows. Section 2 shows the state of the art for the general transportation planning issue. Section 3 is devoted to transportation optimization and simulation techniques that have been used in different ways in accordance with the problems they are meant to solve. Section 4 is about intelligent transportation systems and how the development of new technologies interacts with the whole system and

where they are being used. Short conclusions are given at the end.

The impact that transport infrastructure has in increase of the quality of life can be seen in the next figure:

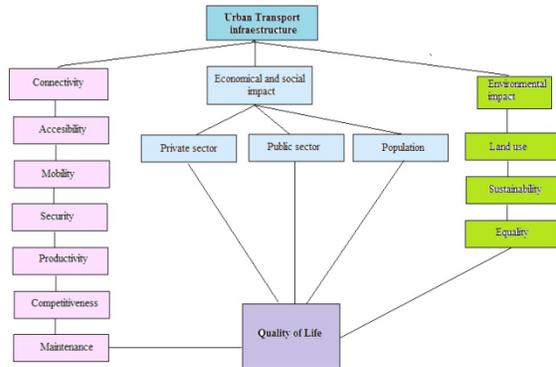


Figure 1 Impact of the transport infrastructure.

Three tables show some of the papers published about mobility and sustainability because they are close related. Later the other themes are shown along the paper.

2. Transportation Planning

Transportation planning covers a lot of different aspects and is an essential part of the system. According with Levy (2011), “Most regional transport planners employ what is called the rational model of planning. The model views planning as a logical and technical process that uses the analysis of quantitative data to decide how to best invest resources in new and existing transport infrastructure.”

Phases for transportation planning

There are three phases: The first, preanalysis, considers what problems and issues the region faces and what goals and objectives it can set to help address those issues. The second phase is technical analysis. The process basically involves the development of the models that are going to be used later. The post-analysis phase involves plan evaluation, program, implementation and monitoring of the results, Johnston (2004).

Transportation planning involves the following steps:

- Monitoring existing conditions;
- Forecasting future population and employment growth, including assessing projected land uses in the region and identifying major growth corridors;

- Identifying current and projected future transportation problems and needs and analyzing, through detailed planning studies, various transportation improvement strategies to address those needs;
- Developing long-range plans and short-range programs of alternative capital improvement and operational strategies for moving people and goods;
- Estimating the impact of recommended future improvements to the transportation system on environmental issues, including air quality; and
- Developing a financial plan for securing sufficient revenues to cover the costs of implementing strategies.

Transportation planning process in Figure 2

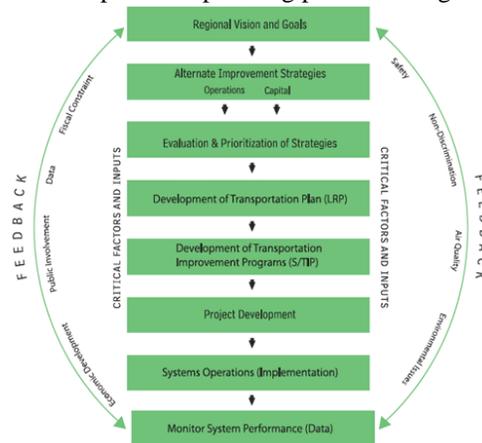


Figure 2 Transportation planning process. Source (FHWA, 2007)

Urban Infrastructure

Urban infrastructure, a human creation, is designed and directed by architects, civil engineers, urban planners among others. These professionals design, develop and implement projects (involved with the structural organization of cities and companies) for the proper operation of important sectors of society. When governments are responsible for construction, maintenance, operation and costs, the term “urban infrastructure” is a synonym for public works. Road infrastructure is the set of facilities and equipment used for roads, including road networks, parking spaces, traffic lights, stop signs laybys, drainage systems, bridges and sidewalks.

Table 1 Sustainability in Urban Transport Infrastructure

Title/Year	Author(s)	Description
A new transportation system for efficient and sustainable cities: Development of a next generation variable speed moving walkway (2015)	David Rockwood, David Garmire	Authors address how to safely control and operate such walkways. In comparison with existing forms of public transportation, this mode of transportation allows lower latencies with higher throughput by operating continuously and allowing acceleration and deceleration of passengers
Sustainable urban transport (Book- 2011)	Carlos Felipe Pardo	Shanghai Manual – A Guide for Sustainable Urban Development in the 21st Century
Technology and the notion of sustainability (2010)	Sven Ove Hansson	Instead of choosing between weak and strong sustainability, the two notions should be included in the same analysis, since they are needed to account for different kinds of assets.
A bilevel multi-objective road pricing model for economic, environmental and health sustainability (2014)	Judith Y. T. Wanga, Matthias Ehrgott, Kim N. Dirks, Abhishek Gupta	Authors consider three objectives at the upper level: minimizing system travel time; total vehicle emissions; and negative health impacts, modeled as the level of pollutant uptake.
Strategies for Safer and Sustainable Urban Transport in Bangladesh (2002)	Md. Mazharul Hoque & Jobair Bin Alam	Emphasis has been placed on the mobility needs of the urban poor. Key transport issues and possible solution strategies for enhancing mobility, safety and the environment by means of better traffic management measures are further discussed in the light of major themes of the conference.
Sustainable Urban Transport Management and Its Strategies (2012)	Touba Amirazodi	Sustainable transport management considers the effects of transportation development on economic efficiency, environmental issues, resources consumption, land use and social justice and helps reduction of environmental effects.
Advancing Sustainable Safety (BOOK) National Road Safety Outlook for 2005-2020. SWOV Institute for Road Safety Research	Ed. Fred Wegman Letty Aarts	The Sustainable Safety vision of road safety is based on five principles. These principles are the functionality of roads, the homogeneity of mass and/or speed and direction, physical and social forgiveness, recognition and predictability of roads and behavior, and state awareness. This fact sheet describes the theories and scientific background on which these principles are founded.
Indicators for sustainable urban mobility –Norwegian relationships and comparisons (2012)	Vibeke Nenseth, Petter Christiansen and May Hald	Transport data for calculated greenhouse gas emissions at city level and for local emissions still need improvement to be more easily available. In the quantification of environmental goals in transport policy, increasing emphasis is put on the indicator relationships and distance-to target indicators.
Encouraging good practice in the development of Sustainable Urban Mobility Plans (In Press) (2014).	Anthony. D. May	The purpose of this paper is to review experience with the provision of such guidance, at a European and national level, assess the underpinning research and identify areas in which further research is needed. The paper reviews the background to the preparation of guidance at a European level and also at a national level in Belgium, France, Germany, Italy, Poland, Scandinavia, Spain and the UK.

Table 2 Mobility in Urban Transport Infrastructure

Title/Year	Author(s)	Description
Políticas de movilidad urbana e infraestructura urbana del transporte Documento de proyecto CEPAL (2009) IRAP-FUPOl	Jorge A. Lupano, Ricardo J. Sánchez.	The availability of adequate urban transportation infrastructure that allows goods and people to mobilize worthy, timely, reliable and economic way that integrates certainly common core of basic needs. Specially relevant for developing countries. http://movilidadadable.org/programas
Defining urban mobility indicators (2009)	Fabio Nussio POLIS Conference 2009	http://www.airqualitynow.eu/comparing_smi.php . To obtain a time-dependent indicator of traffic and mobility that is consistent and comprehensive requires in theory the continuous monitoring of the entire transport network, which implies measuring every few minutes the flow state (speed and density) of at least each major link. Today, through an advanced methodology based on the simulation of transport demand, of driver's route choice and of the traffic phenomenon.
Estimation of Performance Indices for the Planning of Sustainable Transportation Systems (2013)	Alexander Paz, Pankaj Maheshwari, Kachroo, and Sajjad Ahmad	This study proposes a system of systems (SOS) and a fuzzy logic modeling approach. The SOS includes the Transportation, Activity, and Environment systems. The fuzzy logic modeling approach enables the treatment of the vagueness associated with some of the relevant data. Performance Indices (PIs) are computed for each system using a number of performance measures.
Sustainable Urban Transportation: Performance Indicators and Some Analytical Approaches (2002)	John Black, Antonio Páez, Putu Alit Suthanaya	The analytical framework for sustainable urban transportation analysis includes descriptive statistics—exploratory and graphical methods, spatial mapping, spatial statistics _to identify geographical patterns and to identify outliers in the data_, regression analysis, travel preference functions based on Stouffer's intervening opportunity.
Measuring Transportation Traffic, Mobility and Accessibility (2011)	Todd Litman	Traffic-based measurements (such as vehicle trips, traffic speed and roadway level of service) evaluate motor vehicle movement. Mobility-based measurements (such as person-miles, door-to-door traffic times and ton-miles) evaluate person and freight movement. Accessibility-based measurements (such as person-trips and generalized travel costs) evaluate the ability of people and businesses to reach desired goods, services and activities.
Mobility indicators and accessibility of transport (2007)	Peter L'os	There are various indicators commonly used to measure and evaluate levels of mobility and accessibility. To what extent are these indicators suitable for the purpose of comparison between various cities and investigation of influence of accessibility on mobility, is the subject of discussion in this paper.
Can national survey data be used to select a core set of indicators for monitoring the sustainability of urban mobility policies? Working papers 2009-2011	Francesca Marnelli, Gerardo Marletto	A national survey has been carried out to gather citizens' perceptions over a basic conceptual framework of dimensions and objectives of urban mobility policies in order to provide a ranking of the associated performance indicators. The results showed that different sets of performance indicators may be chosen according to city size and transport modes mostly used by citizens.

Table 3 Mobility in Urban Transport Infrastructure Part 2.

Title/Year	Author(s)	Description
La Movilidad urbana. El reto de la sostenibilidad (2013)	Angel L. Cayuela Prieto	The problems generated by urban transport are result of the increase in the number of vehicle-km in circulation, and an efficient mobility is very important for this.
Hacia una estrategia nacional integral de movilidad urbana	Bernado Baranda Sepúlveda, Jorge Cañez Fernández et al.	This document intends to translate the proposal from the civil society becomes to design a national strategy for sustainable urban mobility in Mexico.
Planes Integrales de Movilidad Lineamientos para una movilidad urbana sustentable (2012)	Salvador Medina Ramírez, Jimena Veloz Rosas, Alfonso Iracheta Cenecorta y Jimena Iracheta Carroll	To give key elements to Mexican cities to plan and manage urban mobility through strategies focused on reducing car use, improve public transport and bicycle and pedestrian infrastructure elements.
Multicriteria decision making for sustainability evaluation of urban mobility projects (2013)	Anjali AWASTHI, Hichem OMRANI, Philippe GERBER	In this paper, authors investigate four multicriteria decision making (MCDM) techniques namely TOPSIS, VIKOR, SAW and GRA, fuzzy numbers for sustainability evaluation of urban mobility projects under qualitative data and demonstrate their application through a numerical example.
Strategic Directions and Ecosystems to Address China’s Urban Mobility Challenges (2014)	François-Joseph Van Audenhove, Oleksii Kornitichuk, Jeff Hou, Antoine Doyon	The report summarizes some of the key insights from the “Future of Urban Mobility 2.0” study and puts them in perspective by looking into specific challenges and opportunities within Greater China.
A GIS as a Decision Support System for Planning Sustainable Mobility in a Case-Study (2012)	Pietro D’Amico Ferdinando Di Martino and Salvatore Sessa	Authors support the strategic environmental assessment process of the Coordination Plan of the District of Naples through the selection of indicators set to monitor and evaluate the planned actions towards a sustainable mobility in the District. This process is implemented inside a GIS and its reliability is tested by linguistic hedges modelled from fuzzy sets.
SUMO – Simulation of Urban Mobility An Overview (2011)	Michael Behrisch, Laura Bieker, Jakob Erdmann, Daniel Krajzewicz	SUMO is an open source traffic simulation package including net import and demand modeling components. Authors describe the current state of the package as well as future developments and extensions.
The Resilience of Road Transport Networks Redundancy, Vulnerability and Mobility characteristics. Ph. D thesis (2014)	Rawia Ahmed Hassan El Rashidy	This thesis is concerned with the development of a composite resilience index for road transport networks. The index employs three characteristics, namely redundancy, vulnerability and mobility, measuring resilience at network junction, link and origin-destination levels, respectively. The mobility indicator for road transport networks is formulated from two mobility attributes reflecting the physical connectivity and level of service. The combination of the two mobility attributes into a single mobility indicator is achieved by a fuzzy logic approach.

Urban infrastructure includes transportation infrastructure, which can, in turn, be divided into three categories: land, sea, and air, they can be found in the following modalities:

- Street networks, high or low-speed railway lines, together with such as bus stops, road signs, traffic lights, parking bays, and so forth. This applies to all the cases cited below:
- Infrastructure for mass transit or bike paths and footpaths
- Canals, bridges
- Ports, airports and lighthouses, etc.

Some papers published concerning transport planning are as follows:

Introduction to Multi-Modal Transportation Planning Principles and Practices (2012), Todd Litman

This paper summarizes basic principles for transportation planning. It describes conventional transport planning, which tends to focus on motor vehicle traffic conditions, and newer methods for more multi-modal planning and evaluation.

The Research on Practical Approach for Urban Transport Planning. This study developed by JICA(The Japan International Cooperation Agency)is the result of a huge experience in more than 60 countries, and its purpose is to contribute to the formulation of medium-to long-term development strategies for urban transport. A schema is shown next:

Some other papers concerning to planning are case studies in different countries and will be mentioned in section 5.

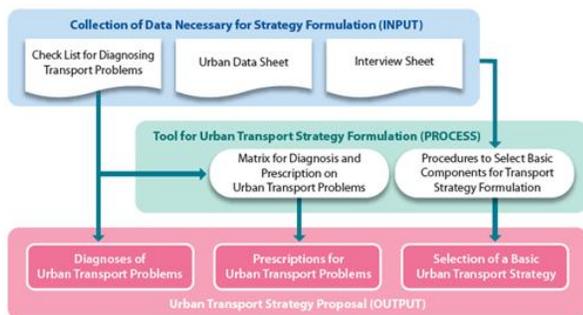


Figure3. Guidelines for formulating an urban transport strategy (Source JICA)

3. Transportation Optimization and Simulation

The goal of transportation optimization is to identify, evaluate and plan enhancements that optimize the operation of a transportation system. With this in mind many countries have specific policies for this and a lot of research has been developed over recent years to this end. Optimization deals mainly with networks operation, traffic, costs, maintenance, sustainability, mobility and management infrastructure. These aspects require balance between the performance of the structure and the total cost accrued over the entire life-cycle. For instance in Worm and Harten (1996), the authors developed a Decision Support Model. “This model is a tool for the road manager to assist in generating an optimal

maintenance plan for a road. Optimal means: minimizing the Net Present Value of maintenance costs, while the plan is acceptable in terms of technical admissibility, resulting quality, etc. Global restrictions such as budget restrictions can also be imposed.”

There is a site that considers some principles for transportation optimization (<http://www.jda.com/>) “With the transportation management system market expanding, the time has come to clarify misperceptions and reveal the best practices in optimization. The Transportation Management System Investment must optimize beyond carrier selection and rates to address myriad constraints and granular functions that apply these three key principles:

1. One size does not fit all
2. The details matter
3. Concurrency is king”.

Considering multi-objective optimization, there is the paper of Wang et al. (2014) and proposed a bilevel multi-objective approach to optimize tolls in a road network. “Consider three objectives at the upper level: minimizing system travel time; total vehicle emissions; and negative health impacts, modeled as the level of pollutant uptake”.

Another paper related to multi-objective is the one of Guerreiro et al. (2014). “In the Dial-a-Ride public transportation systems, each customer requirement is specified in terms of a pickup (origin), of a delivery (destination) and of a time window within it has to be satisfied. The aim is to find a set of routes, each assigned to a vehicle, in order to satisfy the set of requests, under capacity, time windows, precedence and pairing conditions. This paper addresses the problem with the aim of optimizing, at the same time, the maximum total ride time and the total waiting time. Then, a bi-objective PDPTW with a constraint on the maximum duration of each route is proposed and solved by a two-step approach.”

From the network optimization point of view, it can be considered the next papers:

Fanello et al. (2014) developed an integrated performance indicator in urban road infrastructure for evaluating network functionality and the impact of transport system interventions. The complex indicator has been elaborated using a multicriteria algorithm, based on concordance analysis.

An interesting application for the optimization users is the one presented by Spiliopoulou et al. (2014) authors validate a macroscopic traffic flow model. “The reported study tests, validates and compares two well-known macroscopic traffic flow models in the special, but quite frequently occurring case, where congestion is created due to saturated freeway off-ramps. In particular, the comparison includes the first-order model CTM (Cell Transmission Model) and the second-order model METANET.”

Travel time in a network is different in public and private transport, for this reason Salonen and Toivonen (2013). Analyzing the accessibility disparity between different travel modes is recognized as an efficient way to assess the environmental and social sustainability of transport and land use arrangements. “This paper aims to assess the comparability of different methods for calculating travel time by different travel modes. First, we briefly review the methods used in previous studies and identify different typical approaches, which we then compare. We use three computational models respectively for car and public transport (PT), implemented in our case study area, the capital region of Finland.”

The transport issue can be seen as a transportation network, and in this way the relationships among the main elements to analyze and study, figure 4 shows these ideas.

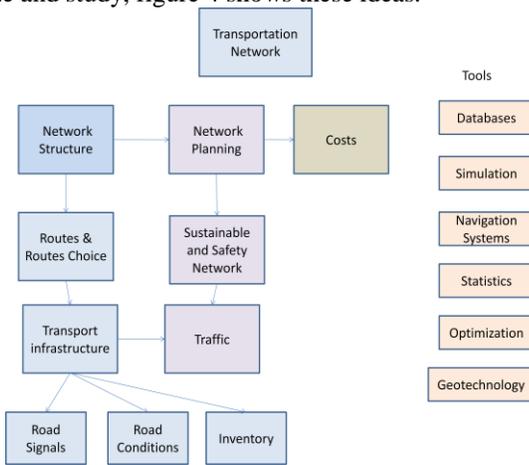


Figure 4 Transportation network

Transportation networks have as it can be seen in Figure 4 a traffic component, in this way simulation has been used with success. One paper is the one of Solecka and Žak (2014). “The paper presents the overall methodology of designing and assessment of transportation solutions that result in the integration of an urban public transportation system. The proposed approach is based on the application of heuristic – expert design of an integrated urban public transportation system, its simulation with the use of a computer-based macro-simulation system VISUM and its multidimensional – multiple parameter comparison with the existing solution.”

However, there is also the need to optimize an entire transportation network. There are two main methods for this, the exact and heuristics, though a hybrid that that combines them both can also be used. In most cases, though, some heuristic algorithms are used because of the size of the problems involved. This paper will show the new developments in this area.

In the case of networks, some routing studies have been developed using either genetic algorithms hybridized with Dijkstra algorithms to find the shortest routes, or just some

advanced label algorithms as the one shown in Klunder, (2006).

As has already been mentioned, metaheuristics are used because they provide very good solutions in a short time, like the neural networks that are used by Yu et al (2011)

The exact methods we are referring to include branch and bound, branch and cut, dynamic programming. The location and routing problem presented by Belenguer et al (2011) uses branch and cut for the design of logistic networks. In this case the overall distribution cost may be excessive if routing decisions are ignored when locating depots. In order to overcome this problem they propose a branch and cut algorithm for solving it. The proposed method is based on a zero-one linear model reinforced by valid inequalities.

Berman et al (2011) gives us an example of search paths for a service problem that is stated as follows: “A customer residing at a node of a network needs to obtain service from one of the facilities; facility locations are known and fixed. Facilities may become non operational with certain probability; the state of the facility only becomes known when the facility is visited. Customer travel stops when the first operational facility is found. The objective is to minimize the expected total travel distance”. This problem is NP-hard and a forward dynamic programming procedure is developed.

A micro-simulation related with trips consider single data and census data, Lovelace et al. (2014), studied the statistics on mode and distance of travel, which vary depending on a range of factors that operate at different scales. With this problem in mind they proposed “a spatial micro-simulation approach, which combines individual-level survey data with geographically aggregated census results to tackle the problem.”

Some transport simulators can be found in the following sites:
<http://www.simulatorcentral.com/index.php/simulators/transpo rt.html>
<http://www.simutrans.com/en/>
<http://www.aimsun.com/wp/>
<http://www.caliper.com/transmodeler/>

A transport simulation table¹:

¹

https://upload.wikimedia.org/wikipedia/commons/7/71/Simulation_T able.JPG

Table 4 Types Of Simulation In Transportation

Time	State	Space		
		Continuous	Discrete	N/A
Continuous	Disc.	Real Transportation Systems * Traffic flow, pedestrians Dynamic traffic assignment		Discrete Event Systems * queueing inventory manufacturing
	Cont.	PDE Traffic flow models Pedestrian models		ODE vehicle motion car suspension queueing (fluid approx)
Discrete	Disc.		Cellular Automata * Traffic, pedestrians Land use Urban sprawl Random Number Generation	Discrete Event Simulation * queueing inventory manufacturing
	Cont.	Car-following models * Microscopic traffic flow models *	Numerical PDE methods Godunov, Variational	Numerical ODE methods Euler, Runge-Kutta time-series * ARIMA
N/A	Disc. or Cont.	Monte Carlo method * : use of pseudo-random number Simulation of static probabilistic problems Integration, Optimization		Econometric models trip generation, distribution, modal split Optimization static traffic assignment

For traffic some of them are:

Paid: traffic flow simulators:

- Quadstone paramics (microscopic)
- VISSUM (macroscopic)
- VISSIM (microscopic)
- AIMSUN

Free:

- Matsim
- SUMO (microscopic)
- Repast
- MAINSIM

There are also general-purpose discrete simulation software packages such as SIMIO, Promodel or Flexim, as well as other more specific packages, such as S Paramics, or Simleader.

The maintenance of urban infrastructure consists of a series of actions that require knowledge and experience about the needs of different types of infrastructure (bus stops, signage, benches etc.) to be done in the optimum manner. To achieve this, infrastructure can be changed, expanded and/or replaced in an efficient manner in order to meet the needs of the users of a city.

Some papers about maintenance are:

Hana, and Thakur (2014). In this paper authors explain the use of new material for maintenance since concrete, asphalt pavements, and ballast are removed during the re-construction of existing roads and have been increasingly recycled as aggregates for the construction of roadways. As they explain “This paper reviews recent research work on the use of geosynthetics to stabilize recycled aggregates in roadway construction and summarizes the main findings on permanent deformation, creep deformation, degradation, stress distribution, and/or crack propagation.”

Another paper in the same way is the one of Corinthias et al (2014). “This study investigates the eco-burden effect of road construction in its maintenance phase along with the influence of above factors in shortening road life-cycle. By looking at previous studies which tend to focus on eco-burden in ideal

condition of road life-cycle, this study throws new light on the effects of abnormal phenomena on road usage which shorten pavement life-performance as well as increase eco-burden impacts.”

Urban transport infrastructure has a direct impact on people’s daily lives, which can, in a positive or negative way (depending on its condition), affect the competitiveness of people in general and the country at large (depending on the competitiveness of its urban infrastructure on a global level).

There are many factors that lead to the growth of urban transportation, but we must not forget other important factors such as rural development, use of the countryside or urban development.

That is why forecasting and transportation data are two other important topics in this section, considering forecasting as an important tool for designing, building, operating, and maintaining models for forecasting the demand for transport. These models are built using optimization algorithms as well as simulation software.

Forecasting is used in Durango-Cohen (2007) in order to address performance prediction and maintenance optimization for transportation infrastructure facilities.

4. Technology for the Transportation Systems

The issue of the proper and efficient administration of urban transport infrastructure contains many technological, political, and social aspects. So it is necessary to use an interdisciplinary approach such as geotechnology, whereby digital technologies can be integrated for a spatial analysis of reality.

Geotechnology, in other words, is presented as a new vision of geographic space that enhances the field of computer systems using cybernetic human and electronic systems for the analysis of physical and social Buzai (2012) and its scope is ever expanding geoBlog (2007).

Some geotechnological tools are:

- Geographic Information systems (GIS).
- Global Positioning System (GPS).

GIS integrates hardware, software and data for capturing, managing, analyzing and displaying geographically referenced information. GIS allows us to view, understand, question, interpret, and visualize data in many ways that reveal relationships, patterns and trends in the form of maps, globes, reports, and charts. A GIS helps us to answer questions and solve problems by looking at all the available data in a way that is quickly understood and easily shared. Some of the top 5 benefits that GIS has to offer are the following:

- Cost saving and increased efficiency
- Better decision making
- Improved communication
- Better record keeping
- Geographical management

Nowadays we are living in an era characterized by technological advances, mobile devices are much stronger, more efficient and capable than they used to be and for this reason a new type of commerce has been created, called Mobile Commerce, where people can make transactions through their mobile device. A subcategory of Mobile Commerce is Location Based Commerce whereby a mobile device can inform its user through a GPS system certain information that can make the user's life easier; for example a user can be informed whether he is near a gas station, hospital or restaurant. Thus we can see that geography and the technologies associated with it are connected with humans to such a degree that they can help us in our daily round. Location based m-commerce, according to Turban et al. (2008), can be divided into the following 5 categories:

- i. **Location:** the service that can determine the place of a person.
- ii. **Mapping:** the service relating to the creation of maps for specific locations.
- iii. **Tracking:** the surveillance of a person through his/her route.
- iv. **Navigation:** the creation of the ideal route between two locations.
- v. **Timing:** the calculation of the time that a vehicle needs to cover a specific route.

The need for efficient transport networks means that their operation has become a primary focus, so intelligent transport systems have been used as tools to make this efficiency possible. ITS systems apply transport systems technology to solve problems and achieve optimum performance.

Some example of ITS can be found in these papers:

Astarita et al. (2014) propose an automated sensing system for monitoring of road surface quality by mobile devices.

South Korea continues to lead the way in digital opportunity with its recent, innovative and ubiquitous city projects. Author Dong-Hee Shin (2009) with his paper: Ubiquitous city: Urban technologies, urban infrastructure and urban informatics, analyze and evaluate the process of design and development of u-city.

Planning local public transport is done by Pensa et al. (2014), "This paper describes a visual tool for data analysis applied to a case in public transport. Our tool is based on geo-referenced dynamic maps, created with free Web GIS applications, and allows users to visualize data and interact readily with a large database of public transport service information"

Vaiana et al. (2014) use GPS and other technologies for the road safety, by developing a road safety performance index.

Simonyi et al. (2014) developed a smartphone application (i.e., an app) for facilitating the assessment of urban public transport services is presented and its applicability and versatility. Samples of the data collected and evaluated are

presented in conjunction with public transport services in Budapest, Hungary.

Campos Ferreira et al.(2014) present an innovative ticketing solution based on customers' mobile devices, which intends to minimize the investment cost from the Public Transport Operators and customers' perspective, maximizing consumer's acceptance. Project was developed in Porto city.

Vitale et al. (2014) in their words "this paper presents a Decision Support System that relies on a logical network architecture characterized by the communication paradigm REST and powered by the use, on Client side, of smartphones that today have an enormous social relevance."

Blaschke et al. (2014). The amount of scientific literature on (Geographic) "Object-based Image Analysis – GEOBIA has been and still is sharply increasing. This article investigates these development and its implications and asks whether or not this is a new paradigm in remote sensing and Geographic Information Science (GIScience)".

Colomina, and Molina (2014) discuss the evolution and state-of-the-art of the use of Unmanned Aerial Systems (UAS) in the field of Photogrammetry and Remote Sensing (PaRS).

Belgiu et al. (2014) in their article Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery, analyze "the increasing availability of high resolution imagery has triggered the need for automated image analysis techniques, with reduced human intervention and reproducible analysis procedures".

The iRAP project uses technologies such as GIS and GPS with cameras in vans for the purpose of preventing road accidents and, when combined with ITS, can prove to be a really useful tool in the hands of the agencies responsible for the urban infrastructure, helping them to obtain in-depth knowledge of the state of the urban infrastructure in order to be able to meet future needs through a well-defined planning strategy. The International Road Assessment Programme (iRAP) is a registered charity dedicated to preventing the more than 3,500 road deaths that occur every day worldwide.

A Comment

There was more work but for lack of space was not possible to write down here, as environmental and social impact, connectivity, accessibility, security, more planning, and cases in many countries.

References

- Alminas, M. Vasiliauskas, A. V. Jakubauskas, G. 2009. The Impact of transport on the competitiveness of national economy. Department of Transport Management, 24(2): 93-99.
- Álvarez-Herranz, A. Martínez-Ruiz, M.P. 2012. Evaluating the economic and regional impact on national transport

- and infrastructure policies with accessibility variables. *Transport*, 27 (4), 414-427.
- Amirzodi, T. 2012. Sustainable Urban Transport Management and Its Strategies. *World Academy of Science, Engineering and Technology Vol:6*.
- Aschauer DA. 1991. Transportation spending and economic growth: the effects of transit and highway expenditures. (Report). Washington, D.C: American Transit Association.
- Astarita V., A., Vaiana, R., Iuele T., Caruso, M.A, Giofrè Vincenzo P., De Masi F. 2014. Automated sensing system for monitoring of road surface quality by mobile devices. *Procedia - Social and Behavioral Sciences* 111, 242 – 251.
- Awasthi, A., Omrani, H., Gerber, P. 2013. Multicriteria decision making for sustainability evaluation of urban mobility projects. Working papers No 2013-01 January 2013 CEPS.
- Balaker, T. 2006. Do economists reach a conclusion on rail transit? *Econ Journal Watch*, 3(3): 551.
- Baranda Sepúlveda, B. Hacia una estrategia nacional integral de movilidad urbana. *Movilidad Urbana Sustentable.ITDP México*
http://mexico.itdp.org/wp-content/uploads/Movilidad-Urbana-Sustentable-MUS_.pdf [Accessed July 2015].
- Behrisch, M., Bieker, L., Erdmann, J., Krajzewicz, D. 2011. SUMO – Simulation of Urban MObility. SIMUL 2011 : The Third International Conference on Advances in System Simulation
- Belenguer, J.M, Benavent, E., Prins,C., Prodhon, C., WolflerCalvo, R. 2011, A Branch-and-Cut method for the Capacitated Location-Routing Problem. *Computers & Operations Research* 38, 931–941
- Belgiu, M., Draĝu, L., Strobl, J. 2014. Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 205–215.
- Berman, O., Ianovsky,E., Krass, D. 2011, Optimal search path for service in the presence of disruptions, *Computers & Operations Research* 38, 1562–1571, Elsevier.
- Black, J. Páez, A., Suthanaya, P.A. 2002. Sustainable Urban Transportation: Performance Indicators and Some Analytical Approaches, *Journal of Urban Planning and Development*.
- Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff,H., van Coillie, F., Tiede, D. Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, 180–191.
- Buzai, G. 2012. El ciberespacio desde la Geografía. Nuevos espacios de vigilancia y control global, *Meridiano*, Revista de Geografía No.1. Available from <http://www.gesig-proeg.com.ar/documentos/articulos/2012-Buzai-Meridiano1.pdf>. [Accessed 10 November 2012]
- Calderón, C. Servén, L. 2008. Infrastructure and economic development in Sub-Saharan Africa. *The World Bank Policy Research Working Paper*, 4712.
- Campos Ferreira, M., Nóvoa, H., Galvão Dias, T., Falcão e Cunha, J. 2014. A proposal for a public transport ticketing solution based on customers’ mobile devices. *Procedia - Social and Behavioral Sciences* 111, 232 – 241.
- Cayuela Prieto, A.L. 2013. La Movilidad urbana. El reto de la sostenibilidad. Instituto del Transporte y Territorio ETS. Ingenieros de Caminos UPV
<http://www.upv.es/upl/U0462089.pdf>.
- Colomina, I., Molina, P. 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing* 92, 79–97.
- Corinthias P., Sianipara, M., Dowakia, K. 2014. Eco-burden in pavement maintenance: Effects from excess traffic growth and overload. *Sustainable Cities and Society* 12. 31–45.
- D’Amico, P., Di Martino, F., Sessa, S. 2012. A GIS as a Decision Support System for Planning Sustainable Mobility in a Case-Study. *Contemporary Engineering Sciences*, Vol. 5, no. 1, 9 – 32.
- De Rus, G. 2008. *The Economic effects of high speed rail investment*. University of Las Palmas, Spain, Discussion Paper, 2008-16.
- Dijkstra, A. 2011. *EN ROUTE TO SAFER ROADS. How road structure and road classification can affect road safety*. SWOV.Available from
http://www.swov.nl/rapport/Proefschriften/Atze_Dijkstra_.pdf. [Accessed 3 March 2013].
- Dong-Hee Shin. 2009. Ubiquitous city: Urban technologies, urban infrastructure and urban informatics. *Journal of Information Science*, 35 (5), pp. 515–526.
- Fancello G., Carta M. and Fadda P. 2014. A modeling tool for measuring the performance of urban road networks. *Procedia - Social and Behavioral Sciences* 111. 559–566.
- Flores, I., Chatziioannou, I., Segura, E., Hernández, S. 2013. Urban Transport Infrastructure: A state of the art. *Proceedings of the European Modeling and Simulation Symposium*.pp.83-92.
- Guerrero, F., Pezzella, F., Pisacane, O., Trollini, L. 2014. Multi-objective optimization in dial-a-ride public transportation. *Transportation Research Procedia* 3. 299–308.
- Hana, J., Thakur, J.K. 2014. Sustainable roadway construction using recycled aggregates with geosynthetics. *Sustainable Cities and Society* 14.342–350.
<http://www.irap.net/en/> [Accessed July 2015].
<http://www.jda.com/> JICA [Accessed July 2015].
- Hassan El Rashidy, R.A. 2014. The Resilience of Road Transport Networks. Redundancy, Vulnerability and Mobility characteristics. Ph. D thesis, Sep. 2014. The University of Leeds, Institute of Transport Studies, Faculty of Environment
- Johnston, R. A. 2004. The Urban Transportation Planning Process. In S. Hansen, & G. Giuliano (Eds.), *The*

- Geography of Urban Transportation (pp. 115-138). The Guilford Press. Multi-Modal Transportation Planning.
- Klunder, G.A. Post, H.N. 2006, The Shortest Path Problem on Large-Scale Real-Road Networks, *Networks* 182-194.
- KPMG 2010. International, Success and failure in urban transport infrastructure projects. A study by Glaister, Allport, Brown and Travers KPMG's Infrastructure Spotlight Report.
- Levy, J. M. 2011. Contemporary Urban Planning. Boston: Longman.
- Litman, T. 2012. Introduction to Multi-Modal Transportation Planning Principles and Practices. Victoria Transport Policy Institute.
- Litman, T. 2011. Measuring Transportation Traffic, Mobility and Accessibility. Victoria Transport Policy Institute.
- Los, P. 2007. Mobility indicators and accessibility of transport. *Slovak Journal of Civil Engineering*. March.
- Lovelace, R., Ballas, D., Watson, M. 2014. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. *Journal of Transport Geography* 34, 282–296.
- Lupano, J. A., Sánchez, R.J. 2009. Políticas de movilidad urbana e infraestructura urbana del transporte. Documento de proyecto Cepal y Gobierno de Francia.
- Mameli, F., Marletto, G. 2011. Can national survey data be used to select a core set of indicators for monitoring the sustainability of urban mobility policies? CENTRO RICERCA ECONOMICA NORD SUD (CRENOS). Working papers 2009-2011.
- May, A.D. 2014. Encouraging good practice in the development of Sustainable Urban Mobility Plans (In Press). Institute for Transport Studies, University of Leeds, LS2 9JT England, United Kingdom
- Mazharul Hoque, Md, Bin Alam, J. 2002. Strategies for Safer and Sustainable Urban Transport in Bangladesh. Urban mobility for all. Proceedings of the 10th international Codatu conference, pp.559-565.
- Nenseth, V., Christiansen, P., Hald, M. 2012. Indicators for sustainable urban mobility –Norwegian relationships and comparisons. Institute of Transport Economics. TØI Report 1210/2012.
- Nussio, F. 2009. Defining urban mobility indicators. http://www.airqualitynow.eu/comparing_smi.php.
- Ove Hansson, S. 2010. Technology and the notion of sustainability. *Technology in Society* 32, 274–279.
- Pardo, C.F. 2011. SUSTAINABLE URBAN TRANSPORT Chapter 4. Shanghai Manual – A Guide for Sustainable Urban Development in the 21st Century
- Paz, A., Maheshwari, P., Kachroo, P., Ahmad, S. 2013. Estimation of Performance Indices for the Planning of Sustainable Transportation Systems. Hindawi Publishing Corporation Advances in Fuzzy Systems Article ID 601468, 13 pages.
- Pensa, S., Masala, E., Arnone, M., Rosa, A. 2014. Planning local public transport: a visual support to decision making. *Procedia - Social and Behavioral Sciences* 111, 596 – 603.
- Rockwood, D., Garmire, D. 2015. A new transportation system for efficient and sustainable cities: Development of a next generation variable speed moving walkway. *Sustainable Cities and Society* 14, 209–214.
- Simonyi, E., Fazekas, Z., P. Gáspár. 2014. Smartphone application for assessing various aspects of urban public transport. *Transportation Research Procedia* 3, 185 – 194
- Solecka, K., and Žak, J., 2014. Integration of the urban public transportation system with the application of traffic simulation. *Transportation Research Procedia* 3, 259–268
- Spiliopoulou, A., Kontorinaki, M., Papageorgiou, M., Kopelias, P., 2014. Macroscopic traffic flow model validation at congested freeway off-ramp areas. *Transportation Research Part C* 41, 18–29.
- SWOV Institute for Road Safety Research, Advancing Sustainable Safety. National Road Safety Outlook for 2005-2020. Ed. Fred Wegman and Letty Aarts.
- The City of Calgary, Transportation Department. Available from <http://www.calgary.ca/Transportation/Pages/Transportation-Department.aspx>. [Accessed 8 march 2013].
- The International Road Assessment Programme (IRAP). Available from <http://www.irap.net>. [Accessed 7 January 2013].
- Turban, E.; Aronson, J. E.; Liang, T. & Sharda, R. 2008. Decision Support and Business Intelligence Systems (eight edition). Pearson Education, ISBN 0-13-158017-5.
- Vaiana, R., Teresa, I., Astarita, V., Carmine, V., Tassitani, A., Rogano, D., Zaffino C. 2014. Road safety performance assessment: a new road network Risk Index for info mobility. *Procedia - Social and Behavioral Sciences* 111, 624 – 633.
- Van Audenhove, F.J., Korniiichuk, O., Hou, J., Doyon, A. 2014. Strategic Directions and Ecosystems to Address China's Urban Mobility Challenges. The report summarizes some of the key insights from the "Future of Urban Mobility 2.0" study and puts them in perspective by looking into specific challenges and opportunities within Greater China.
- Vitale, A., Carmine Festa, D., Guido, G., Rogano, D. 2014. A Decision Support System based on smartphone probes as a tool to promote public transport. *Procedia - Social and Behavioral Sciences* 111, 224 – 231.
- Wanga, J. Y. T., Ehgott, M., Dirks, K., Gupta, A. 2014. A bilevel multi-objective road pricing model for economic, environmental and health sustainability. *Transportation Research Procedia* 3, 393 – 402.
- Worm, J.M., van Harten, A. 1996. Model based decision support for planning of road maintenance. *Reliability Engineering and System Safety* 51, 305-316.
- Yu, B, Lam W, Lam Tam, M. 2011, Bus arrival time prediction at bus stop with multiple routes, *Transportation Research Part C*. Elsevier.

Ambulance location using linear programming: The case of the National Autonomous University of Mexico (UNAM)

Esther Segura Pérez^(a), Idalia Flores^(b), Alexander Vindel^(c)

^{(a),(b)}Engineering Faculty, UNAM

^(c)Department of Postgraduate Studies in Engineering, UNAM

^(a)esegurap@ingen.unam.mx, ^(b)idalia@unam.mx, ^(c)alexander_vindel@hotmail.com

ABSTRACT

Every year the Universidad Nacional Autónoma de Mexico (UNAM) holds four entry exams for higher education and high school, for which the candidates do a written exam in any one of 25 schools in the Metropolitan Area of the Valley of Mexico (ZMVM) that the UNAM uses to hold its entry exams. A medical services module manned by a doctor and nurse with an equipped first-aid kit is installed in each school. Moreover, five properly equipped ambulances are available and distributed around the Metropolitan Area to reinforce this medical attention as regards the treatment and transfer of patients. At the present time the five ambulances are located empirically based on the experience of the Emergency Medical Technicians (TUMs) so that the distance between the current location of the ambulance and the modules or schools is what they believe to be the shortest. This study proposes a three-step methodology for optimizing the location of five ambulances based on the Location Set Covering Problem (LSCP) model and the p -median model. The solution compares these two models, allowing us to shorten the response time of the Pre-Hospital Care (PHC) service by locating the five ambulances better, considering a variety of speeds and coverage times. This enables us to lower the rate of patient mortality and morbidity.

Keywords: Coverage models, p -median, Pre-Hospital Care, Ambulance location, Heuristic algorithms.

1 INTRODUCTION

The UNAM, in conjunction with eight other institutions, forms part of the admissions processing organization known as the High School and Higher Education Entry Examination Board for the Metropolitan Area of the Valley of Mexico (COMIPEMS), that gives people who are interested in continuing their high school or university studies the opportunity to enter a public school in the open or regular system. The UNAM uses 25 colleges throughout the Metropolitan Area for every entrance exam. The UNAM Medical Services Bureau (DGSM) is in charge of the pre-hospital care service (PHC), which consists of installing a medical service module in every college and deploying five ambulances to provide emergency services. The module consists of a doctor, a nurse and a first-aid kit. The ambulances are properly equipped and capable of providing service to the 25 schools where the entrance exam is being applied. The ambulances are staffed by two emergency medical technicians.

Pre-hospital care is a service that is based on timely medical attention and the immediate transfer of patients to a hospital. It is offered to the sick or injured who are not inside an institution where they could be given the medical service they require, its primary purpose is to take care of and transfer the and sick or injured person to the respective hospital and to treat them in the shortest time possible, thus contributing to lowering the rate of mortality and lessening any sequels that patients could have as a result of organ failure.

2 DESCRIPTION OF THE PROBLEM

The demand for places has grown year by year. Just in the last entrance exam there were 120 thousand candidates (Olivares, 2014) representing a large number of applicants for very few places, as the UNAM can only offer seven thousand places (Olivares, 2014). Thus candidates have to have very good exam results to get a place in this institution. This means that examinees and their relatives are faced with a stressful situation that can give rise to medical emergencies. According to the PHC Coordinator of the Medical Services Bureau, the main ones are: crises in diabetics, road accidents, sunstroke, anxiety attacks, hypoglycemia, convulsions and heart problems. The first three happen more often to the relatives while the rest are problems suffered by the candidates. Based on information provided by the UNAM Entrance Examination Registration and Application Office (SRAEA), we were able to obtain an approximate number of the incidents that occurred in each school during the 2009 admission process. Unfortunately the SRAEA refused to give us any further information, so the study only focuses on that year. However, the information has proven very useful by giving us a parameter of approximate demand in each school.

The Emergency Medical Technicians experimentally establish five regions in the Metropolitan Area, locating one ambulance in each region with the corresponding schools to be served, thus shortening the response time for any incident. The Emergency Medical Technicians makes sure that there is at least one public hospital, belonging to the Institute of Social Services for State Employees (ISSSTE), the Mexican Institute of Social Security (IMSS) or the Government of the Federal District, in each region. There are two shifts for each entrance exam: the first shift from 7.00 a.m. to 2.00 p.m. and the second from 2.00 p.m. to 9.00 p.m.

2.1 Objective of the study

This study proposes the minimization of the response time of the pre-hospital care (PHC) service by locating the ambulances based on the model of the location set covering problem (LSCP) and the p-median model.

3 LOCATION MODELS

Location problems arise from the need to find the most convenient place to locate facilities such as: distribution centers, production plants, garbage dumps, fire, police and ambulance stations, among many others. In general terms, the problem can, according to (Daskin 1995), be expressed as: Given the location of each user, demand and costs (time, distance, etc.) of transport in the region in question, the number of services, the geographical location and capacity of each must be determined in order to optimize the costs of transport, operation, etc.

Location models have been studied since the 1970s. The best known basic problems are cited by Current (2002). Daskin (1995) classifies location models as: continuous, network and discrete models. The models based on coverage and the models based on the p-median are to be found within the discrete models. In this project

we are focusing on the LSCM and p -median models

3.1 Review of the literature on ambulance location models

The optimization of the PHC system or Emergency Medical Service (EMS) is a subject that has been studied since the middle of the 1960s. It is a very attractive area for applied mathematics and operational research. Some of the reasons for the attraction are that it is a very important issue for society. On the one hand, owing to the high costs involved in maintaining the equipment and having the highly qualified personnel that is required, it is vitally important to guarantee that the best use possible is made of the available resources (Restrepo, 2008). On the other hand, these are deep, interesting problems from a mathematical point of view, requiring one to keep up to date with their inherent subtleties and complexities while finding approaches that can be implemented in practice, given the constraints on the available data and computer resources.

There are hundreds of articles that approach the PHC from the perspective of developing models to support significant decisions, such as (Restrepo, 2008):

1. Location, skills and bases of personnel;
2. Crew schedules;
3. The number and type of vehicles to be located at each base;
4. Choosing the type of vehicle that will attend to the emergency; and
5. The redistribution of vehicles as a function of the state of the system.

According to (Henderson and Mason, 2004) these decisions can be classified as operational, strategic and tactical decisions. This set of decisions directly influences the time the system

takes to arrive at the place where the patient is to be found (Sasaki, 2010).

There are two basic documents, Brotones et al. (2003) and Parra (2011) that show the state of the art in PHC, where the models are classified as deterministic, stochastic and dynamic. The first location models were explained by Hakimi (1964) and Toregas et al. (1971) with the proposal of the p -median model and the Location Set Covering Model (LSCM) respectively. Church and Re Velle (1974) improve the LSCMP (Location Set Covering Models P Median) problem by restricting coverage of the population and maximizing the number of available resources to make better use of them. This model is known as a Maximal Covering Location Problem (MCLP). In this project, the p -median models and LSCP (Location Set Covering Problem) model are used to optimize the location of the PHC system's ambulances during the UNAM's entrance exams.

4. LOCATION METHODOLOGY

In this study we propose a three-phase methodology for improving the current location of ambulances. Phase 1 is for gathering information (distances between schools, number of incidents per school and speeds that the ambulances can achieve). Phase 2 is when an initial solution is obtained for the p -median model using the myopic heuristic algorithm (Daskin, 1995), while the schools are also assigned to the ambulance location that is found. In other words, five regions are determined with an ambulance located in each one and the set of schools assigned that are to be served by each ambulance. At the same time a set of initial solutions is obtained by solving the LSCP model and they are established as coverage parameters: ambulance speeds 20, 40, 60 and 90 km/h, and response times 5, 10, 15 y 20 min. Therefore a total of 17 scenarios are generated, 16 for the

LSCP model and one for the p -median model. In each scenario the 20 schools are assigned to the five ambulances located in schools. In phase 3, the 17 initial solutions are improved by neighborhood search algorithms and this solution can, in some cases, be improved by applying a second improving algorithm called an exchange algorithm (Daskin, 1995). It is worth mentioning that the models employed (p -median and LSCP) to optimize ambulance location in the specific case of the UNAM satisfy the particular constraints detected in the University's admissions procedure. A set of scenarios were generated to represent the real situation. For example, the speed variations represent situations where the city streets are not congested so the ambulances can travel at high speed (90Km/h) or else situations where the ambulances cannot travel at high speeds (20 Km/h). The purpose of the solutions to the scenarios is to give the emergency medical technicians a broader vision of the location of the ambulances under different circumstances. This is why we use simulation in order to be able to assess the various scenarios.

4.1 First step: information gathering

The information was provided by the Medical Services Bureau of the UNAM through interviews with the coordinator of the PHC service department, who knows the type and number of incidents that normally occur at each entrance exam. It is important to point out that the emergency medical technicians do not keep a record of the incidents, so, in parallel, we went to the SRAEA of the UNAM where we were given the number of students per school who presented the exam, as well as the number of recorded incidents. This case study had the limitation that we were only given information for the year 2009. Finally we used Google maps to calculate the distances between the 25 schools.

4.2 Second step of the methodology: initial location of ambulances

In this step 17 initial solutions are obtained, the first is obtained by solving the p -median model using the myopic algorithm and a set of 16 initial solutions is obtained by solving the LSCP model. We programmed, in this study, the myopic algorithm in Excel and the LSCP model in the LINGO optimizer (Lindo Systems Inc.). That is to say that in each solution or scenario we obtained the location of the five ambulances as well as the set of schools to be served by each one of them.

4.2.1 Initial location based on the p -median model

The p -median model considers the distance between the schools and the demand in each one of them, i.e., the number of incidents in each school, as well as the five available ambulances. Bearing in mind the above, the model in this case study is posed as follows: Equation (1) minimizes the weighted distance between 25 schools, where h_i is the demand in the module or school i and d_{ij} is the distance between the schools or modules i y j . Equation (2) assigns each school to only one ambulance. Equation (3) assigns school i to point j only if there is an ambulance in that module or school. Equation (4) determines that the number of ambulances to be located is five. Equation (5) indicates that the binary variable x_{ij} is 1 if school i is assigned to ambulance j , 0 if not. Equation (6) indicates that the binary variable w_j is 1 if the ambulance is located at j , 0 if not.

$$\text{Min } Z = \sum_{i=1}^{25} \sum_{j=1}^{25} h_i d_{ij} x_{ij} \quad (1)$$

$$\sum_{j=1}^{25} x_{ij} = 1 \quad i = 1, 2, \dots, 25 \quad (2)$$

$$x_{ij} \leq w_j \quad i = 1, 2, \dots, 25 \quad j = 1, 2, \dots, 25 \quad (3)$$

$$\sum_{j=1}^{25} w_j = 5 \quad (4)$$

$$x_{ij} = (0, 1) \quad i = 1, 2, \dots, 25 \quad j = 1, 2, \dots, 25 \quad (5)$$

$$w_j = (0, 1) \quad j = 1, 2, \dots, 25 \quad (6)$$

Myopic algorithm (Daskin, 1995) is applied to solve the p -median model (Equations 1 to 5). This algorithm is designed to find a potential location point for each ambulance in each iteration. This is achieved by minimizing the weighted distance (Equation 1) in each iteration, the prior location is kept fixed and it stops when there are five located ambulances.

4.2.2 Initial location of ambulances based on the covering model (LSCP)

For the application of the LSCP model variations are carried out in desired response times and speeds of the ambulances considering that time [h] is equal to distance [km] divided by the speed [km/h], $t = d / v$. This provides a set of scenarios that determine locations based on distances and speeds to cover patient demand. In this case, it was necessary to add the constraint (equation 9) on the number of ambulances there are to obtain results that fit the available resources.

The LSCP model is applied to this case study for determining the location of the five ambulances. Equation (7) expresses the objective function which is to minimize the number of ambulances required to cover all the points of demand. Equation (8) contemplates those schools or modules that fulfill the various time scenarios established for each one of the proposed speeds. Equation (9) determines that there only are five ambulances. Equation (10) indicates that the binary variable y_i is 1 if the ambulance is located at i , 0 if not.

$$\min z = \sum_{i=1}^{25} y_i \quad (7)$$

$$\sum_{i=1}^{25} y_i \geq 1 \quad (8)$$

$$\sum_{i=1}^{25} y_i = 5 \quad (9)$$

$$y_i = 0, 1 \quad i = 1, \dots, 25 \quad (10)$$

4.2.2.1 Getting scenarios that involve variations in speeds and service times

For the solution of the LSCP model, we consider the distance between schools and several speeds that the ambulance can reach. This is achieved owing to the fact that $v = d / t$. Assuming that the ambulances can maintain the following speeds 20, 40, 60 and 90 km/h, we get four tables (with size of 25*25 elements), one per speed. The tables contain the times between schools in minutes. The coverage constraints are generated when considering a desired response time. Response time refers to the time in which we want the ambulance to reach the point where

the patient is to be found, which can range from 5, 10, 15 to 20 minutes. Thus 16 total scenarios are generated. For example, if we consider an ambulance speed of 20 Km/h and a desired response time of 5 minutes, a set of constraints (equations 8) will be generated that comply with the desired 5-minute response time between each pair of schools, which is known as a 5-minute radius of coverage. The model for each and every one of the 16 scenarios is developed and we get the location of the 5 ambulances with the respective schools to be served by each one of them.

4.3 Improvement of initial location

In this step the 17 initial scenarios are improved (the one proposed by the p -median model and the 16 proposed for the solution of the LSCP model) by applying the neighborhood search algorithm and, once the solution is improved, the exchange algorithm is applied so that we can get an even better solution. Every scenario consists of the location of five ambulances and the assignment of the schools they need to serve. If the solution is based on the minimum distance, this is obtained by using the p -median model, whereas if the solution is based on a desired response time, the solution is obtained by using the LSCP model.

4.3.1 Improvement of initial solution using the neighborhood search algorithm

The neighborhood search algorithm, which is responsible for finding the closest modules to each ambulance, is applied to each and every one of the 17 scenarios obtained. For further reading about the neighborhood search model, we recommend Daskin's book (1995).

4.3.2 Improvement of initial solution using the exchange algorithm

This method is based on swapping around every one of the regions, in this case, the 5 regions. In other words, it considers each one of the schools as a possible ambulance location point by analyzing where the value of the weighted distance is smaller.

5. Analysis of results

After applying the proposed methodology we get the following results. Table 1 gives the 16 possible initial scenarios found by solving the LSCP model. It is important to point out that three scenarios were found not to be feasible. In other words, for an ambulance speed of 20 and 40 km/h and a desired response time (radius of coverage) of between 5 and 10 min, we find pairs of schools that comply with both parameters at the same time.

Table 1: Scenarios proposed for the coverage model

	5 min.	10 min.	15 min.	20 min.
20 km/hr.				
40 km/hr.				
60 km/hr.				
90 km/hr.				

Table 2 shows the results obtained after applying the methodology. We observe that the current

situation is drastically improved going from an average ambulance location distance of 3570.27 meters to a distance of 1238.98 meters. This

solution was achieved by solving the LSCP model considering a speed of 40 Km/h and of 20-minute radius of coverage. The first ambulance must be located in the UVM school and serves three other schools, the second ambulance must be located at the CSB school and serves three other schools, the third ambulance must be located in the CMA school and serves two other schools, the fourth ambulance must be located in the UL school and

serves three other schools and finally the fifth ambulance must be located in the ISEC school and serves 9 other schools. Furthermore, the solution we got, based on the p-median model, also represents a good solution, achieving an average distance between the location of the ambulance and the assigned schools of 1451.27 meters. In Table 2 scenarios in grey are used for the simulation analysis.

Table 2: Summary of the results obtained in each case

Case	Ambulance location (number of modules to be served)	Average distance (in meters)
Current situation	CM (5), CUM (4), ULV (6), UVMC (6)	3570.27
P-median	PSAPE (6), UL (2), CUM (3), CH (5), ULV (4)	1451.27
20 km/hr. - 15 min.	PSAPE (6), UL (3), CUM (6), CH (5), CCP (0)	1548.71
20 km/hr. - 20 min.	CMA (2), UVM (5), ULV (7), CH (6), CCP (0)	1740.01
40 km/hr. - 10 min.	CMA (2), UVM (5), ULV (6), CH (7), CCP (0)	1956.42
40 km/hr. - 15 min.	PSAPE (4), USJ (0), IN (0), UL (3), ISEC (13)	1398.72
40 km/hr. - 20 min.	UVM (3), CSB (3), CMA(2), UL (3), ISEC (9)	1238.98
60 km/hr. - 5 min.	PSAPE (6), ILM (2), ISEC (7), CH (5), CCP (0)	1557.94
60 km/hr. - 10 min.	PSAPE (5), CCP (0), IN (0), UL (2), ISEC (13)	1872.05
60 km/hr. - 15 min.	UVM (2), IN (0), CMA (2), UL (3), ISEC (13)	1924.11
60 km/hr. - 20 min.	PSAPE (5), CCP (0), IN (0), UL (2), ISEC (13)	1872.05
90 km/hr. - 5 min.	IN (0), PSAPE (5), UL (2), CH (7), ULV (6)	1685.13
90 km/hr. - 10 min.	UVM (4), IN (0), CMA (3), UVMC (2), ISEC (11)	2310.26
90 km/hr. - 15 min.	CMA (2), PSAPE (0), UVM (5), ISEC (13), CCP (0)	2275.90
90 km/hr. - 20 min.	CMA (2), PSAPE (0), UVM (5), ISEC (13), CCP (0)	2275.90

5.1 Validation using simulation

To validate the results obtained using the location models, simulation is used considering

the more favorable scenarios, that are the ones with less distance, as it can see in next table 3

Table 3 Analyzed scenarios with simulation

Case	Ambulance location (number of modules to be served)	Average distance (in meters)
20 km/hr. - 15 min.	PSAPE (6), UL (3), CUM (6), CH (5), CCP (0)	1548.71
40 km/hr. - 15 min.	PSAPE (4), USJ (0), IN (0), UL (3), ISEC (13)	1398.72
40 km/hr. - 20 min.	UVM (3), CSB (3), CMA(2), UL (3), ISEC (9)	1238.98
60 km/hr. - 5 min.	PSAPE (6), ILM (2), ISEC (7), CH (5), CCP (0)	1557.94

The model programmed in Simio is a representation of the proposal for locating ambulances during the UNAM entrance exam. This exam is sat in different centers, which constitute the demand nodes, the sites of the incidents that are served by the various ambulances that the UNAM puts into service when this exam is being applied. For that reason they are treated as source nodes where each node has a probability distribution that was adjusted based on some historical records of the services provided by the ambulances.

The general structure of the model contemplates 5 areas, in accordance with the p-median location model, each area has an ambulance to cover it and said ambulance has an assigned speed of 20 km / hr. For the transfer of a patient,

we consider that when the ambulance is inactive, it returns to the base node. Base nodes are the nodes where the ambulances are located in the location model. We also, in addition to this node, consider a sink node, while the arcs that join the nodes are two-way.

After analyzing the data, as location, type of service, etc. obtained by the pre hospital care reports that record the events attended during the day of evaluation; These data are adjusted to a Poisson distribution, which describes the time between arrivals at a server, in this case it describes the time between occurrences of events that require pre-hospital care provided by ambulance. Once the probability distribution is defined, scenarios are constructed as shown in Figure 1

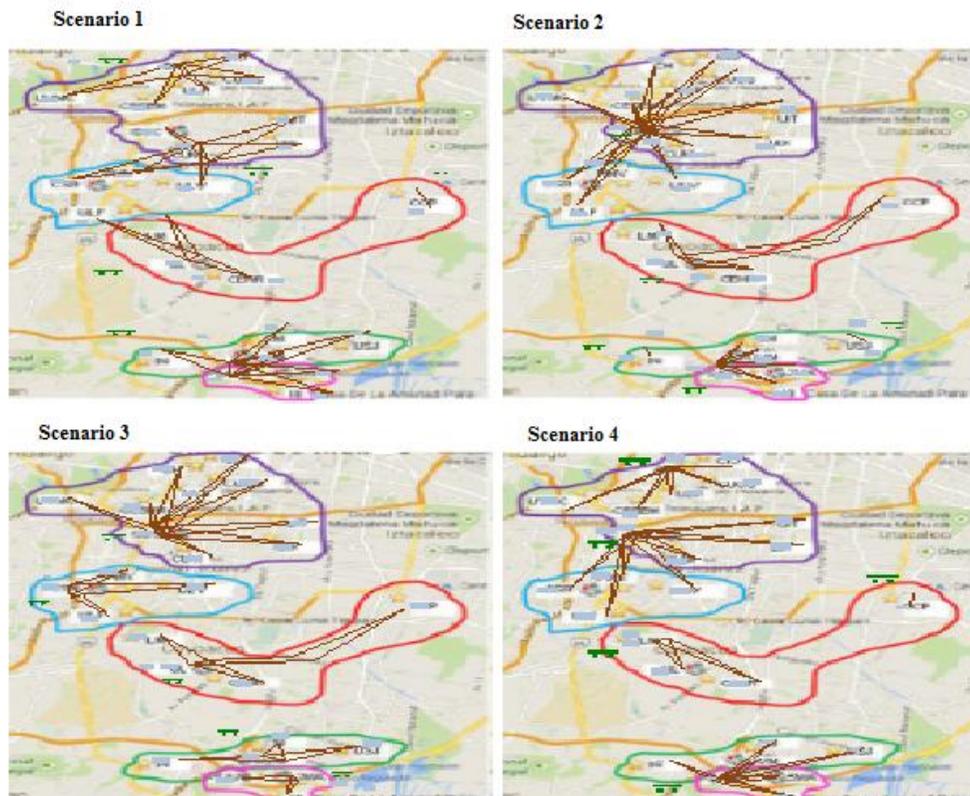


Figure 1. SIMIO scenarios

The model's scale tries to be as close as possible to a ratio 1:1 in order to contemplate the distances of the arcs that join the nodes. An experiment consisting of 100 replicates with a

confidence level of 95% was done for the four scenarios. Results obtained are shown in Figures 2 and 3.

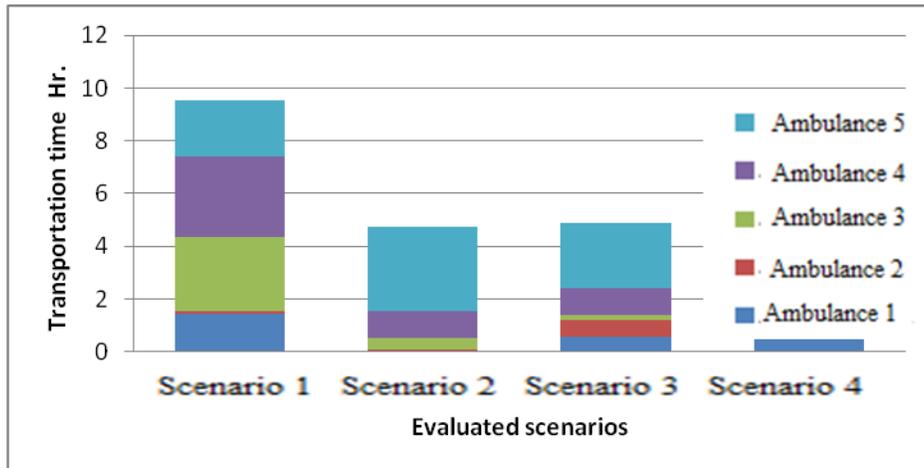


Figure 2 Distribution time per ambulance and scenario

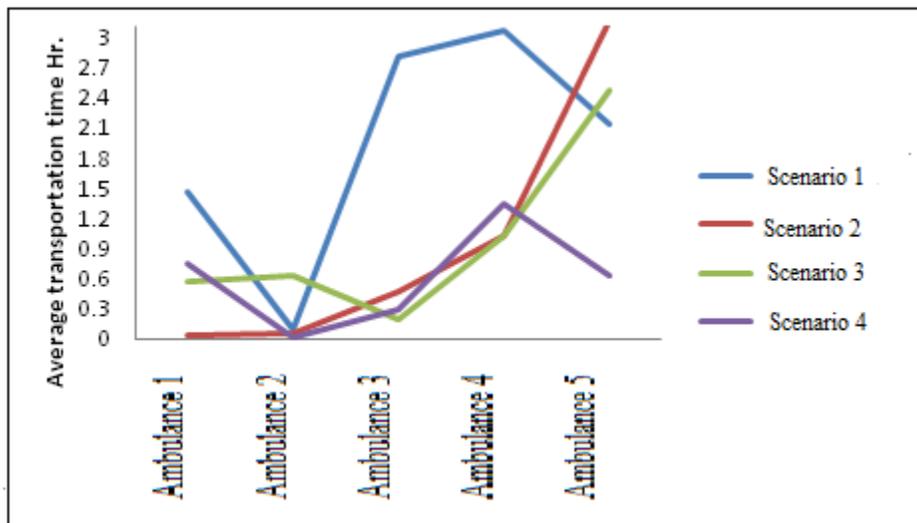


Figure 3 Average time vs. scenarios and ambulances

Based on figures 2 and 3 it can be seen that scenario 4 is better and in section 6 these results will be part of our conclusions.

6. Conclusions

We conclude that the linear programming models based on p -medians and covering (LSCP) for ambulance location during the UNAM entrance examinations are an excellent tool, as they considerably shorten the average

distances between the location of the ambulance and the schools to be served. A set of scenarios is given that the Emergency Medical Technicians can use for better decision-making in the reduction of the patient response time and a considerable reduction in the probability of patient mortality and morbidity.

Thanks to Simio we can easily expand the detail of the model, with some observations that arise after the model is built such as:

- Defining the routes there and back that can be marked out on a sketch to represent the roads taken by the ambulances.
- Improving the sketch
- Considering the possible sink nodes

And finally it was possible to validate the scenarios built with optimization and to conclude that as it was said in the previous section, from Figures 2 and 3 is noticeable that scenario 4 shows a shorter transportation time, however, one of the conditions for this scenario is a restriction of 5 min as radio coverage, and given the conditions in Mexico City, where the test is applied, it is almost impossible for this reason is better to choose scenario 3, considering that in Figure 3 the line the average time per ambulance remains under scenario 2.

References

1. L. Brotcorne, G. Laporte, F. Semet. Ambulance location and relocation models. *European Journal of Operational Research*, 147, 451-463. 2003.
2. R.L. Church y C.S. Re Velle. The maximal covering location problem. *Papers of the regional science association*, 32, 101 – 18. 1974.
3. J. R. Current and H. A. Eiselt. Location Analysis. *Computers & Operations Research* 29, v-vii (Invited Editorial). 2002.
4. M. S. Daskin. *Network and Discrete Location. Models, Algorithms, and Applications*. Ed. John Wiley & Sons, Inc., New York, 1995. ISBN: 0-471-01897-X.
5. Google Earth:<http://www.google.es/intl/es/earth/index.html>
6. S. Hakimi. Optimum Location of switching centers in a communication network and some related graphs theoretic problems. *Operation Research* 13, 462-475. 1965.
7. S.G. Henderson y A.J. Mason. Ambulance Service Planning Simulation and Data Visualisation. Operations Research and Health Care International Series in *Operations Research & Management Science*. 70, 70-102, 2004.
8. LINGO: www.lingo.com
9. E. Olivares. Buscan 120 mil aspirantes uno de 7 mil espacios en la UNAM. Periódico la Jornada, México D.F., 10/Marzo/2014. <http://www.jornada.unam.mx/2014/03/10/sociedad/036n1soc>
10. O.J. Parra. Revisión del estado del arte en modelos de localización y relocalización de vehículos para atención de emergencias. *Revista Elementos* 1, 2011.
11. M. Restrepo. Computational methods for static allocation and real-time redeployment of ambulances. Dissertation Faculty of the Graduate School of Cornell University. 2008.
12. S. Sasaki., A. Comber, H. Suzuki, C. Brunson. Using genetic algorithms to optimise current and future health planning - the example of ambulance locations. *International Journal of Health Geographics*, 9 (4), 1-10. (2010)
13. C.R. Toregas, R. Swain, C.S. Re Velle, L. Bergman. The location of emergency service facilities. *Operations Research*, 19, pp. 1363 – 1373. 1971.

A CLOUD SERVICE BASED MODELING & SIMULATION FRAMEWORK

Xiao Song^(a), Feng Li^(a), Lin Zhang^(a), Cheng'en Wang^(b)

^(a) School of Automation Science and Electrical Engineering, Beihang University

^(b) Northeastern University

^(a) johnlin9999@163.com

ABSTRACT

Modeling & simulation takes a crucial role in the development process of practical industry and cloud based service is becoming more and more widespread popularity. So it is essential to research the combination between them. In this paper, we give a conceptual outlook on behalf of cloud service based modeling & simulation framework and state a business model ensuring the execution of cloud service based modeling and simulation. At the same time, two factors aroused from this new framework, namely security and cost, are highlighted in our work.

Keywords: cloud service; modeling and simulation; business model; security; cost

1. INTRODUCTION

In the field of modeling and simulation, the problem of fusing module, model, application to form System of Systems(SoS) has been addressed for the last two decades. Up to now, a large number of simulation systems have realized interoperability and communication based on high level architecture (HLA) since it was proposed by United States Department of Defense (DoD) in 1990s. However, it is coarse-grained because HLA regards federation as the basic integration unit, and lacks the ability of combination among components and models when develop and run the simulation system.

This issue has been resolved to some extent in Discrete Event Simulation Specification (DEVS) proposed by Zeigler(Zeigler 2013). DEVS provides a modular, hierarchical systems modeling and simulation execution framework, in which each subsystem is regarded as a model with independent internal structure, behavior and explicit input and output interface. Various models can be composed into a coupled model, which can also be used as an element model for other couplings. In this way the hierarchical and modular description of model is formed.

However, large-scale distributed simulation systems are faced with problems of low model reusability, short of networked component specifications and lack of low-cost integration platforms etc.

Meanwhile, expanding simulation scale and system structure require simulation system to be realized in a more easy-to-use and cheaper platform. Actually, with

the rapid development of advanced mode like cloud computing and its related technologies (such as cloud build platform of open source), most enterprises and scientific research institution are building their own private cloud platform or renting public cloud platform (such as amazon's EC2) and use IaaS (Infrastructure as a Service), PaaS (Platform as a service), SaaS (Software as a service) provided by cloud platform (Cayirci 2013a; Cayirci 2013b; Hwang, Fox and Dongarra 2011; Laborde, Barrere and Benzekri 2013; Rosetti and Chen 2012). Based on this, Cayirci(Cayirci 2013) proposed the concept of modeling and simulation as a service(MSaaS) and noted that the simulation services based on cloud computing framework can also have the advantages in the aspects of allocating resource on-demand, high-speed Internet, fast response featured by cloud. But the thorough process of cloud based modeling and simulation was not elaborated and the problems aroused from cloud service were not put into consideration.

The reminder of this paper is organized as follows. Section 2 presents the cloud service based modeling & simulation framework and the business model of using it, the security issues are discussed in Section 3. Section 4 highlights the cost concerns. A case is presented in Section 5 whereas the conclusion is summarized in Section 6.

2. CLOUD SERVICE BASED MODELING & SIMULATION FRAMEWORK AND ITS BUSINESS MODEL

Cloud based service has become a popular and universal paradigm for service oriented computing where computing infrastructure and solutions are delivered as a service(Foster, Zhao, Raicu and Lu 2008). In order to allow the efficient sharing of modeling and simulation resources, it is essentially to encapsulate modeling and simulation resources into cloud services. When the task of modeling and simulation is encapsulated into Cloud based service, the advantages of cloud computing such as on-demand use, pay-per-use, broad network access, scalability can be reflected in cloud service based modeling and simulation.

In order to attain a better understanding and a common conceptualization, we have developed a cloud service based modeling & simulation framework as shown in figure 1, which learned from literature(Tsai, Fan and

Chen 2006). In this framework, all processes are encapsulated into service.

First, users should transfer their tasks into model and scenario by using modeling service which can provides unified model and specification language; then model checking service checks if the model can meet the requirements of completeness and consistency; finally, code generation service can generates code for simulation according to the right model. In most instances, users' requirements cannot be satisfied by only one model, so it is necessary to combine several models into a complex function. In this condition, simulation engines or bus is needed to provide communication between different models.

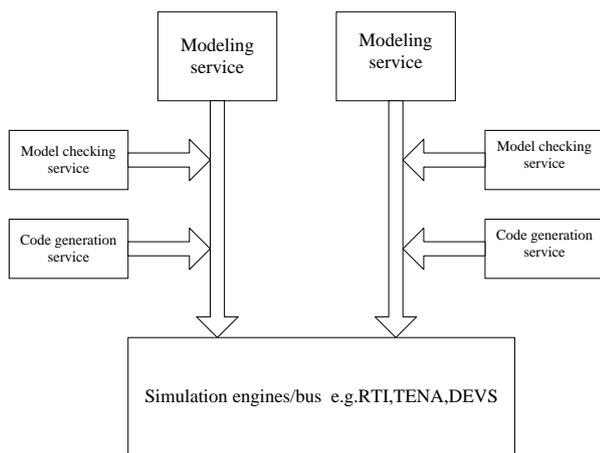


Figure 1: Cloud Service Based Modeling & Simulation Framework

All the aforementioned components can be registered into cloud-based resources pool since they are already encapsulated into services, in which model can be leased and released by modeling users. Customers can use modeling resources provided by cloud pool and pay according to use. On the other hand, the resources of cloud pool can be re-used as soon as they are released by other users, which can improve the utilization of model resources. Meanwhile, for a particular user, they do not have to possess extraordinary expertise pertaining to cloud technologies and domain knowledge about modeling and simulation, so it lowers the threshold of modeling and simulation.

After the process components of modeling & simulation are already encapsulated into cloud based services, there are two forms using the services in cloud pool, which we call it business models, namely download services and running services. Download services refer to the services that can be downloaded into local station and users can run their function locally. The other way is that users run their function in sever side without downloading services, which is called running services. Different model can induce different issues.

3. SECURITY ISSUES IN CLOUD SERVICE BASED MODELING & SIMULATION

Despite the tremendous advantages of Cloud based service, the obstacles and risks, such as user privacy,

authentication, data security, interfacing with internal systems, system availability, business continuity, and ownership of content and other legal requirements (Raval 2010), should not be neglected. In this chapter, security is the most concerned areas since it is the most important factor that users care.

Security is one of the top concerns and is also one of the most important barriers to widespread use for cloud service based modeling & simulation. Some research efforts have been injected to counter this issue.

Cloud security not only encompasses the common computing security, which includes the design of security architectures, minimization of attack surfaces, protection from malware, and enforcement of access control, but also entails some special aspects, such as the cloud is typically a shared resource and easy to be attacked; Cloud-based data is usually intentionally widely accessible by potentially insecure protocols and APIs across public networks; Data in the cloud is vulnerable to being lost, etc.(Ryan 2013). So if cloud service based modeling & simulation realizes to be widely used, the problem of security must be resolved.

In (Sun, Chang and Sun 2011), the author summaries security issues in cloud computing environments from six sub-categories, which include: (a) how to provide safety mechanisms, (b) how to keep data confidentiality for all the individual and sensitive information, (c) how to avoid malicious insiders illegal operation under the general lack of transparency into provider process and procedure environments, (d) how to avoid service hijacking, where phishing, fraud and exploitation are well known issues in IT, (e) how to management multi-instance in multi-tenancy virtual environments, (f) how to develop appropriate law and implement legal jurisdiction. Since cloud based modeling & simulation service is a kind of cloud computing services, so the security issues of cloud computing are also considered in cloud based modeling & simulation service, especially when users choose to realize their function on the cloud server side. In (Subashini and Kavitha 2011), the author points out that the best security solution is to develop a development framework that has tough security architecture. As described in this paper, though there are many advantages in using a cloud-based service, there are still many practical and potential issues that have to be solved. In (Zissis, and Lekkas 2012), the author proposes using a trusted third party within a cloud environment by enabling trust and using cryptography to ensure the confidentiality, integrity and authenticity of data and communications, while attempting to address specific security vulnerabilities. In order to handle the security in cloud service based modeling & simulation, the common or similar methods should be employed since cloud service based modeling & simulation is an extension of cloud computing, but some specific features should be emphasized as there are different parts between them.

4. COST ISSUES IN CLOUD SERVICE BASED MODELING & SIMULATION

The increase of simulation scale and the refinement of granularity require a larger scale computing resource. Cloud computing system has reduced the cost. However, cost calculation model has not been unified in cloud platform service and existing large-scale simulation framework less concerns cost factor. The cost elements include: electricity, hardware, software, human resources, business prospects, cloud services quality, and deployment, etc. (Altmann and Kashef 2014)

Altmann and Kashef (Altmann and Kashef 2014) propose a comprehensive cost model deploying simulation service for cloud platform. The total cost function is divided into four subfunctions: data transfer cost function, service deployment cost function, fixed cost function and change cost function. Data transfer cost is mainly computing in/out cloud data and transfer data; fixed cost is the fees of purchasing hardware, software and network facility; the changing cost is mainly the use-cost of resource including electricity cost, labor maintenance costs and the cost of network connections. And the cost model, when applying federated hybrid clouds, was discussed in (Altmann and Kashef 2014).

For the system of cloud service based modeling and simulation, when the components in it are encapsulated into service and are putted into cloud pool, the migrate cost, as one factor of cost model, should be considered. The costs of server usage, data transfer and storage usage should also be concerned. In addition to these factors, the cloud service can be used in two forms as mentioned above: download services and run service, so different cost and business model should be shown which is appropriate for different business model.

5. USE CASE

In order to realize service based modeling & simulation framework proposed in section 2, we developed a Complex system Modeling and Simulation Language (CMSL) (Song, Chai and Zhang 2010, Song, Zhang, He, and Ren 2012) based modeling service and a CMSL compiler service which act as code generation service. The detailed process can be shown in the following figure.

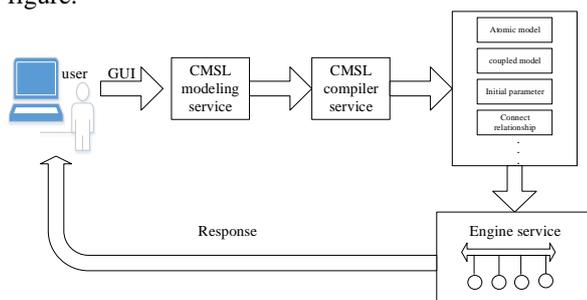


Figure 2: Use Case About Service Based Modeling & Simulation Framework

First, users submit their model or scenario in the form of CMSL by GUI. Second, service compiler transfers CMSL-based model into several component services coded by executable language, namely C++. These components and models can be atomic model, coupled model, initial parameters, connect relationship, etc. Finally, engines service can be combine these model components into a simulation federation that can fulfill user's demand, and the simulation result is returned back to users. The aforementioned services are all encapsulated in a cloud service pool. When a certain service is used, it can be invoked from cloud service pool.

6. CONCLUSION

In this paper, we summaries the framework of modeling and simulation and encapsulate them into cloud based services as well as point to two business models. Despite cloud service based modeling and simulation has offered several advantages, the security concerns and cost issues fostered by it should not be neglected in that they hamper the widespread popularity of cloud based service. Some more efficient counter measures should be carried out by cloud organization and users.

7. ACKNOWLEDGEMENT

The research in this paper was supported by grants 61374199 and 61473013 from the Natural Science Foundation of China and grant 4142031 from Beijing Natural Science Foundation, and grant 2013AA041302 from the National High-tech R&D Program. The authors thank the reviewers for their comments.

REFERENCES

- Altmann J, Kashef M M. , 2014. Cost model based service placement in federated hybrid clouds[J]. *Future Generation Computer Systems*, 41: 79-90.
- Bernard P. Zeigler, Hessam S. Sarjoughian., 2013. *Guide to Modeling and Simulation of Systems of Systems*, Springer-Verlag London
- Cayirci E., 2013. A joint trust and risk model for MSaaS mashups. In: *proceedings of the 2013 winter simulation conference*, Washington DC, December
- Cayirci E., 2013. Modelling and simulation as a cloud service: asurvey. In: *proceedings of the 2013 winter simulation conference*, Washington DC, December
- Dimitrios Zissis, Dimitrios Lekkas., 2012. Addressing cloud computing security issues [J]. *Future Generation Computer Systems*.583–592.
- Erdal Cayirci., 2013. Configuration schemes for modeling and simulation as a service federation, *Simulation: Transactions of the Society for Modeling and Simulation International*, 89(11): 1388–1399.
- Foster, I., Zhao, Y., Raicu, I., Lu, S. 2008. Cloud computing and grid computing 360-degree

- compared. In: Grid Computing Environments Workshop, GCE '08. 2008, p. 1–10. doi:10.1109/GCE.2008.4738445.
- Hwang K, Fox G and Dongarra J., 2011. Distributed and cloud computing. San Francisco, CA; Morgan Kauffmann Publishers.
- Jörn Altmann, Mohammad Mahdi Kashef., 2014. Cost model based service placement in federated hybrid clouds, *Future Generation Computer Systems*, 41,79–90.
- Laborde R, Barrere F and Benzekri A., 2013. Toward authorization as a service: a study of the XACML standard. In: proceedings of spring simulation multi-conference, April.
- Rosetti M and Chen Y., 2012. Cloud computing architecture for supply chain network simulation. In: proceedings of winter simulation conference, December.
- Ryan M D., 2013. Cloud computing security: The scientific challenge, and a survey of solutions[J]. *Journal of Systems and Software*, 86(9): 2263-2268.
- S. Subashini n, V.Kavitha., 2011. A survey on security issues in service delivery models of cloud computing[J]. *Journal of Network and Computer Applications*. 1–11.
- Song X., Chai X., Zhang L., 2010. Modeling Framework for Product Lifecycle Information, *Simulation Modeling Practice and Theory*, 18(8): 1080-1091.
- Song, X, Zhang, L, He, DJ, Ren, ZY., 2012. DEVS Based Modeling and Methodology-COSIM [J]. *Applied Mathematics & Information Sciences*,6(2):417-423.
- Sun D, Chang G, Sun L, et al. , 2011. Surveying and analyzing security, privacy and trust issues in cloud computing environments[J]. *Procedure Engineering*, 15: 2852-2856.
- Tsai W T, Fan C, Chen Y, et al. , 2006. A service-oriented modeling and simulation framework for rapid development of distributed applications[J]. *Simulation Modeling Practice and Theory*, 14(6): 725-739.
- Raval V., 2010. Risk landscape of cloud computing, *ISACA J*. 1: 26

TOWARDS HUMANS ISSUES IN CLOUD MANUFACTURING PARADIGM

Yuqiuge Hao ^(a), Rayko Toshev ^(b), Petri Helo ^(c)

^{(a),(b),(c)} Department of Production, University of Vaasa, Wolffintie 34, Vaasa, Finland

^(a) yuqiuge.hao@uva.fi, ^(b) rayko.toshev@uva.fi, ^(c) petri.helo@uva.fi

ABSTRACT

Cloud manufacturing is a service oriented, customer centric and demand-driven manufacturing model that enable manufacturers to face a dynamics business environment. Due to the products and the corresponding manufacturing processes get more and more complex, operators' daily working life are also getting more difficult. Enhanced human-machine interaction is one of the core tasks for the success of next-generation manufacturing. Most of current research only focuses on the automation and flexibility features of cloud manufacturing, the interactions between human-machine and the value co-creation among operators are missing. Therefore, attentions are needed to the operators, and also their works. One promising approach is future factory concept, with the objective to reduce time and costs of machine control and maintenance. This paper describes the architecture of a system that uses the technologies of augmented reality, internet of things, and wearable technologies to support operators' working and communication in discrete factories.

Keywords: Cloud Manufacturing, Internet of Users, Augmented Reality, Smart glasses

1. INTRODUCTION

Manufacturing is the backbone of the current industrial environment. It plays an indispensable role in global economy. The traditional manufacturing industry was gradually superseded by a global chain of resources and various stakeholders. Therefore, a new dynamic and global business form is needed in the current manufacturing environment. The agility and quick reaction to market changes are essential characteristics, and the high availability and capacity to adequately fulfill customers' requirements is one of the main sustainability criteria. Moreover, centralized operation and management of distributed manufacturing equipment and resources are required (Xu, 2012).

Recently, cloud manufacturing emerges as a new solution to address all the challenges. It is a manufacturing version of cloud computing (Xu, 2012; Putnik, 2012). In this context, manufacturing resources and capabilities are virtualized and considered as services (Li, Zhang and Chai, 2010; Putnik, 2012) and organized in a resource pool (Li, Zhang and Chai,

2010). Virtualization refers to the abstraction of logical resources from their underlying physical resources (Xu, 2012). Cloud manufacturing is oriented towards service provision. Manufacturing services include machines, processing centers, and computing equipment used during the whole manufacturing lifecycle from pre-manufacturing (argumentation, design, production and sale), manufacturing (product usage, management and maintenance), and post-manufacturing (dismantling, scrap, and recycling) (Li, Zhang and Chai, 2010).

Many of existent infrastructures are already cloud-based or changing towards this virtual architecture. To fully and efficiently use those infrastructures, the services should have the ability to dynamically adapt to changes with minimal human intervention (Xu, 2012; Wu et al., 2013). However, it's difficult to avoid human actions in some services. Cloud manufacturing covers the entire manufacturing lifecycle. Therefore, some cloud manufacturing activities are requiring human involvement. Mostly they are after-sales services, which are supporting services of the main products, such as field services, maintenance, diagnosis, user assistance and training, and so on (Putnik, 2012; Camarinha-Matos, Afsarmanesh and Koelmel, 2011).

Xu (2012) considers participators as one of the manufacturing capability, which refer to human resources needed for the job. Besides the human resources, knowledge is also an important factor that represents all the knowledge required to do a specific job. It requires a sophisticated training and advanced knowledge to participate and enable the cloud manufacturing activities.

Integration in cloud manufacturing should be a systemic paradigm to organize humans and machines as a whole system. The integration should be not only at the field level, but also at the management and corporate levels. Wu et al. (2012) analyzed a series of research issues in cloud-based manufacturing paradigm, and cloud-enabled human-computer interaction and human-human collaboration were issues to be considered. They pointed out that more and more computers and other mobile devices get connected to the cloud, it's important to improve the human-computer interaction and user experiences. Mezagar (2011) also pointed out the outstanding importance of the collaboration and cooperation among users in networked enterprises.

However, in previous studies related to cloud manufacturing, the human resources and the knowledge management are rarely discussed. Major problems remain with respect to the interface between the enterprise corporate level and the manufacturing shop floor level (Panetto and Molina, 2008). Integration of humans with software and hardware is one of the fundamental requirements to satisfy the new development in the industry (Panetto and Molina, 2008). Based on previous study, we discovered some weaknesses of the current research in cloud manufacturing:

- Insufficient methods to support the field services.
- Lack of collaboration among operators.
- Lack of communication between the field services and the back office.
- Inefficiently knowledge management in field

Our research question is how to connect “humans” through updated networking technology in a contextually aware manner. Based on our recently discovering, a new solution is needed to facilitate a closed loop in the cloud manufacturing. We will employ different state-of-the-art technologies to support the collaborative manufacturing training, maintenance and management among different participants and departments. While other research more focuses on Internet of Things, our standpoint is the human factor in cloud manufacturing, namely Internet of Users. We discuss the solution within a practical case study. The company is a laser machine builder, and they sell machines to customer companies. They also provides centralized helpdesk and remote assistance, training, etc.. Our solution can increase the value of cloud manufacturing, avoid waste and increase sustainability, and in the meanwhile it can help to reduce risk and prevent human error on the shop floor. We will perform a pilot study to evaluate the use of wearable technologies (smart glasses) in improving the communication channel.

2. STATE-OF-THE-ART REVIEW

2.1. Cloud Manufacturing

Cloud manufacturing is a new multidisciplinary domain that encompasses state-of-the-art technologies such as networked manufacturing, manufacturing grid, virtual manufacturing, agile manufacturing, Internet of Things, and cloud computing (Xu, 2012; Li, Zhang and Chai, 2010). It is a value creation manufacturing process across globally networked operations. It involves global supply chain management, product service connection and management of distributed manufacturing units (Mezgar, 2011).

Under the umbrella of cloud manufacturing, manufacturing resources and capabilities can be intelligently sensed and connected to the wider internet,

and automatically managed and controlled using Internet of Things (IoT) technologies (e.g. Radio-frequency identification (RFID), wired and wireless sensor network, embedded system) (Tao et al., 2011a). These resources and capabilities can be fully shared and circulated based on users’ demand. There are three category users, namely providers (who own and provides the manufacturing resources and capabilities involved in the whole lifecycle of manufacturing process), operators (who operate the cloud manufacturing as a mediator, mostly deal with sales and business management) and consumers (who subscribe these resources and capabilities according to their needs and pay for they used) (Tao et al., 2011a). In some other research, the users are also defined as consumers, producers and managers (Wu et al., 2013). Dissecting and analyze deeply, the managers can be divided into cloud broker (who manages the use, performance, and delivery of cloud services, and negotiates relationships between cloud providers and cloud consumers) and cloud carrier (who provides connectivity and transport of cloud services from cloud providers to cloud consumers) (Wu et al., 2012).

In the introduction to cloud manufacturing, Xu (2012) uses two important concepts to describe cloud manufacturing: “integration of distributed resources” and “distribution of integrated resources”. Providers publish their resources into this cloud manufacturing platform, and then the distributed resources are encapsulated into cloud services and managed in a centralized way. Consumers can use the cloud services according to their requirements. They can request services ranging from product design, manufacturing, testing, management and all other stages of the product lifecycle (Xu, 2012).

Although the concept of cloud manufacturing is mirroring from the definition of cloud computing, most of the resources in cloud manufacturing need to be operated manually by a human (Hu et al., 2012). This is unlike the virtual resources in cloud computing. In addition to the performance of manufacturing equipment, the human activities also have a great impact on the quality of the product. It’s very difficult to maintain high-performance operation because the system consists of a group of devices, such as robots, numerical control machines, sensors, and so on (Park and Jeong, 2013). To improve the quality of manufacturing processes, products, the focus can move to improve people’s performance. So the human activity must be considered as an important factor (Hu et al., 2012; Chituc and Restivo, 2009). Ford et al. (2012) suggest that enhanced human-machine interfaces and collaboration software are drivers of distributed manufacturing.

The scopes of human resources in cloud manufacturing formation vary depending on different definitions. Put it in a simple way, human resources are the personnel who engaged in the manufacturing process, i.e., designers, operators, managers, technicians, project teams, customer service, etc (Wang and Xu, 2013). Lv (2012)

defines human resources including technology person and management person needed in the product developing process. They are the most precious resources in the activities of manufacturing business, plenty of knowledge and experience storage in the brains of these persons, is the inheritance and crystal of human knowledge and mental (Lv, 2012). Jiang et al. (2013) describe human resources in two parts: the personnel in the traditional manufacturing and the personnel in cloud computing development. Human resources include product design and development personnel in the traditional manufacturing, product management and maintenance personnel, product planning and marketing personnel and the personnel of cloud computing development and maintenance and so on.

2.2. Human beings Activities

Wu et al. (2013) emphasis on their research that a high level of automation will be required to ensure the efficiency and effectiveness of machines and manufacturing process on the shop floor with minimal effort, but it does not imply the absence of human beings.

Dual Resource Constraints (DRCs) are known as when capacity constraints stem from both machines and human operators (Xu, Xu and Xie, 2011). Hu et al. (2012) created a classification of the cloud manufacturing services by the degree of human involvement. The first type of services can run automatically without human resources (such as computing resources). The second type of services involves human pure manual activity that has nothing to do with proficiency (such as driver's driving skills has few effects on the quality of logistics services resource) while the third type of services is ranked based on the skill level of a worker (for instance lathe operator's operation skills can seriously affect the quality of workpiece).

Based on the intensive research review (Xu, Xu and Xie, 2011), they summarized several issues should be addressed when considering both humans and machines factors in a system: job release, job dispatching, worker flexibility, worker assignment, and worker transfer costs. In cloud manufacturing concept, these humans-related activities are not neglected. This finding implies the importance and priority to face these challenges in cloud manufacturing context.

Human operators, as a manufacturing resource, include both knowledge and skill, the capability of learning and acquiring new skills, and also related performance of operation (Luo et al., 2013; Xu, Xu and Xie, 2011). The knowledge refers to all kinds of knowledge possessed by the resources elements and experiences accumulated in manufacturing process (Luo et al., 2013). It is build up and accumulated based on the repeated interactions with machines, which is a learning process. More precisely, human knowledge includes experiences in performing a particular manufacturing task and manufacturing method, i.e., engineering knowledge,

product models, standards, evaluation procedures and results, customer feedback, etc. (Luo et al., 2013; Wang and Xu, 2013). Knowledge management (KM) is the core and most essential part of cloud manufacturing reference models (Zhang and Jin, 2012; Zhang et al., 2014; Meier et al., 2010; Luo et al., 2013; Tao et al., 2011; Wu et al., 2012). Effective and efficient knowledge sharing and collaboration among operators are the keys to implement successfully KM and improve the operator's activities. Thus it's crucial to facilitate a well-designed information communication among all operators.

2.3. Technologies Supporting Human Beings

When use Information Communication Technologies (ICTs) to support business activities, there always some problems to be considered. The term ICT includes and focuses on not only technological improvement, but also improvement of aspects such as organization, human interaction, psychology, and man-and-machine interaction (Karim and Söderholm, 2009).

2.3.1. Internet of Things

Internet of Things (IoT) is one of the most critical technologies in cloud manufacturing. Based on IoT and virtualization, the distributed manufacturing resources are identified and connected to the cloud manufacturing and the resource sharing platform via network (Zhang et al., 2014; Wang and Xu, 2013; Huang et al., 2013), and then providing the service dispersedly with these centralized resources (Wei et al., 2013 2). It provides a promising strategy for large-scale distributed manufacturing system cooperation and integration (Liu, Li and Wang, 2011) by real-time monitoring (Liu et al., 2011). IoT makes computing truly ubiquitous, in which the internet extends into the real world and provides a standardized way for the cyber-space to understand the conditions, events and material movements in the real world (Liu et al., 2011; Michel, 2014). IoT is an essential enabler of cloud manufacturing.

With IoT, various data required (e.g., the information of sound, light, heat, electricity, mechanics, chemistry, biology, and location) can be acquired in real time by taking advantage of sensor networks (Tao et al., 2014; Chen, Chen and Hsu, 2014). It has been used in various applications; especially, it has been fundamentally changing the practical production and logistic operation with the aim of intelligent manufacturing (Tao et al., 2014; Chen, Chen and Hsu, 2014).

Tao et al. (2014) highlight three different applications of IoT in manufacturing from three perspectives, respectively in the workshop, in the enterprise and among enterprises. The general application in the workshop is using the data identified and acquired from IoT enabled manufacturing layers to achieve the automatic control. IoT technologies catalyze influential innovations in the business model and industries such as manufacturing (Maguire and Chan, 2014). However, Michel (2014) pointed out that the IoT is still not mature especially on plant floors. The challenges for

broader adoption include security concerns for plant-floor networks and a fuzzy understanding of how the IoT translates into practical use.

2.3.2. Internet of Users

Under the support of rapid development of RFID, sensor technologies and also modern wireless communication, IoT promote the stable interconnection of anything in the manufacturing field (Li, Zhang and Chai, 2010). However, most of current research (Xu, 2012; Zhang et al., 2014) focuses on the connection between products, physical devices or enterprise systems. The “things” refers to machines, automation controllers and other physical devices being connected to the internet (Michel, 2014). Since human resource is also one of the significant manufacturing resources and capabilities, it’s critical to consider a human as a “thing”. Michel (2014) provides an example to describe the importance of considering interconnections with people, processes and data in this IoT concept.

Tao et al. (2014) mention that current rapid development of embedded systems and technologies provide enabling technologies for realizing the intelligent embedding of physical terminal manufacturing equipment and the interconnection of M2M (including man-to-man, man-to-machine, and machine-to-machine) in manufacturing.

The idea of Internet of Users (IoU) as a deeper level of IoT application in cloud manufacturing was proposed (Tao et al., 2014). The users (including service provider, consumer, and operator) of cloud manufacturing are primarily internet users, while IoT is one of the enabling technologies to realize the connection and communication among these users, and to form the IoU. Many smart devices these days can be used for different applications. Wearable technology as one of the favor is inextricably linked with the consumer Internet of Things as a new category of connected devices evolves. A wearable device is giving easy access to a range of services able to connect via a range of technologies including GPS, wireless, Bluetooth, and also via direct connection to a range of fixed infrastructure terminals. Wearable technologies can provide user augmented view of the world and real-time data, beyond this, they can also form an integral part of the “Internet of Things”. The idea is to enable sensor-equipped “things” to communicate with one another in meaningful, actionable ways.

Thierer (2014) emphasis that wearable technologies are networked devices to collect data, track activities, and customize experiences to users’ needs and desires. It can be considered as a subset of IoT. Wearable technologies are supported by augmented reality (AR), which refers to the ability to superimpose virtual, registered information over a user’s view of the real world such as a real object or a space, so people can get information from the augmented data about the real object or space (Paul and Park, 2013; Furht, 2011).

Wearable AR employs a head-mounted display (HMD), a wearable computer, and a set of sensors to determine

the position and orientation of the user’s head. As users move their heads, the HMD updates their virtual world view accordingly. Thomas and Sandor (2009) discuss about the benefits of wearable AR, and one of the benefits is that users can receive information while performing other simple tasks. Also, a wearable AR can give users navigation guidance to a particular location in real time.

There is an enormous opportunity for wearable computing in manufacturing, especially smart glass solutions such as Google glass. Plex promotes its cloud ERP as a comprehensive solution for manufacturing, in which the “IN-YOUR-EYE ERP” is integrated with Google glass and offer people working without using their hands. The user will be able to interact with the real-time, transactional materials, manufacturing and financial information captured by ERP (Gould, 2014).

Thomas and Sandor (2009) also demonstrate a scenario in the manufacturing industry that wearable AR can help assembly line worker to improve their industrial working processes.

BMW Company announced the usage of Augmented Reality like a visual guideline in real time for its workers. The application consists of glasses with headphones, thanks to this mechanic sees and hears the exact instructions about how to repair a car, while at the same time he can ask for the information which tool is right for next step of assembly or repair (Kubac et al., 2013). Similar tutorials are easily transferable to other sectors of human activity. They can serve as a training and educational applications for students in schools, in preparing manuals purchased goods (e.g. furniture), or the rapid training of employees for any device. However, most of the applications and implementations of AR are still in theory and only very few of them are presented as practical demonstration.

Overall, the technologies and problems mentioned above highlight the possibilities and needs for development of a comprehensive solution to support the human and machines connected in term of cloud manufacturing.

3. METHODOLOGY AND CORE CONCEPTS

The cloud manufacturing system is defined as a centralized management of distributed factories with their distributed manufacturing resource, such as machines, tools, materials, people and information, to produce a value-added physical product based on specific customer’s requirement. Its ultimate goal is covering the entire manufacturing lifecycle. Various studies have been carried out to understand the manufacturing resources virtualization and sharing process. Human resources are also important elements to fully realize the cloud manufacturing formation. In this research, we focus on the human resources and human activities in the real factories and propose an integrated solution to manage the interconnection and collaboration among humans with different roles, and our solution can support the effective human participation in term of cloud manufacturing.

In order to implement such architecture, a communication channel needs to be established where factory information are available in real-time to support the humans (users) on the shop floor. Figure 1 presents the reference model of the humans and machines both connected and embedded in the cloud manufacturing with specific information flows.

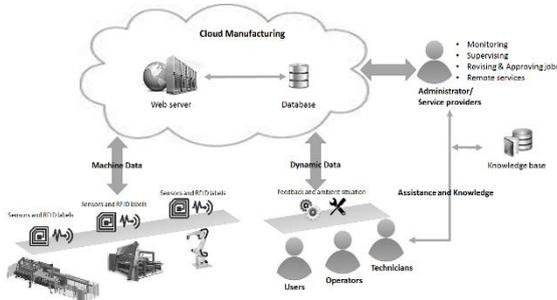


Figure 1: Reference Model for Humans and Machines in the cloud manufacturing.

There are two core components to enable this communication channel: “Internet of Machines (IoM)” and “Internet of Users (IoU)”

- **IoM:** All manufacturing resources, i.e. machines, pieces of equipment, robotics, and devices are connected with the internet. Their manufacturing capabilities are perceived by an adoption of the connecting sensors, embedded systems and RFIDs, and all related data are collected to achieve the automatic control and monitor. The adoption of IT promises the cloud manufacturing system having a real-time understanding of conditions, events and material movements in the physical world. This IoT does not only enable the manufacturing resources virtualization but also support the intelligent operation.
- **IoU:** The users on the shop floor including operators that control and manipulate the machines daily-based and technicians who can perform supervision, maintenance and repairing services. IoU is to the term to describe the connection and communication among the users. Besides the users working on the shop floor, administrators and remote assistants are also part of this IoU. Administrators’ main tasks are monitoring the shop floor in real time and making adaptation when risk or error happens. Remote assistants mainly work in the help center and respond to the urgent events and hotlines.

Because the data can be captured from multiple sources in the distributed local factories, a data aggregator is needed to integrate the data. It provides important functions, such as domain service, location service, server load balancing, and more.

Knowledge-based is also a critical component. It is a document library that retrieves knowledge from users’ activities. It includes how to control the machines, how to operate the equipment, how to use the devices, and so on.

The aim of this research paper is to provide a new concept of cloud manufacturing and technically provide industrial operators with a better approach in their working life and making the right decisions optimally.

4. PROJECT APPROACH

This section presents a case study of a real factory. It is a global company A which provide sheet metal processing machinery. Its customers are the sheet metal fabrication factories. They order customized metal sheet fabrication lines combining punching or laser cutter machines with automated bending and sorting capabilities from our case company A.

Based on the observation and analysis of the machine operator’s actions and factory’s activities on the shop floor, several processes were found to be suitable for improvement, such as the operator’s machine control capabilities, maintenance and repair operations, and also the factory helpdesk processes. The main requirements from the case company are: to use a better remote desktop to control the machines, video sharing to monitor the factory production lines and data access thru cloud services. These requirements were discussed with the company personnel teams who are responsible for the machine operating, maintenance, and helpdesk services. Respectively, three primary stakeholders were identified: machine operator, maintenance personnel and helpdesk unit.

The solution provided in this research is that: Company A manages all the production lines in the form of cloud manufacturing, and its customers are cloud services consumers while company A is hosting this cloud manufacturing platform as a cloud services provider. Company A provides cloud services such as monitoring, supervising, revising & approving jobs, remote services. Figure 2 is a cloud manufacturing architecture that represents all the required functionalities and services integrating for:

- **Real-time data capture service:** for real-time data acquisition from the machines/pieces of equipment/devices/robotics through the embedded intelligent sensors. All data is maintained in the cloud-based repository.
- **Process monitoring service:** The helpdesk unit can monitor the production process in real-time. The users can access all the relevant actual and historical data based on their roles and the data is available to different users in different views. It provides data analysis and management suggestions on multiple screens communicational interface for creating a virtual presence environment.
- **Field work assistance service:** The users are mainly machine operators and maintenance

personnel. They can receive data through mixed-reality technologies to enhance their capabilities and skills. This collaborative and value co-creation environment is an extension of the knowledge management.



Figure 2: Overall System Architecture for Implementation

The ‘cloud’ infrastructure is for integration of representation, mixed-reality representation, real-time management model, and communication for collaborative management. It consists three layers: the data, business and presentation layer.

The presentation layer supports all interfaces, views, presentations and communications for users. The traditional single screen desktop environment needs to expand to large screens for control, monitoring and communication features. Therefore, this presentation layer provides two different kinds of display for back office and for the field working.

The business layer defines applications and functionalities such as Collaboration among users, Sharing Telemetry from machines to users, Real-time information from the shop floor to back office, etc.

The data layer represents cloud-based data repository and data management, also including knowledge bases. The technologies used to implement this data layer are querying, selection, and refinement.

4.1. Augmented Reality and Smart Glasses

In fact, at the back office of the helpdesk, a browser-based dashboard into production status is displayed on several oversized flat screens (see Figure 2). This specific solution includes four environments: 1) Production management feeding from ERP/MES 2) Operator’s live view from smart glasses 3) Conference call for collaborative environment 4) Real-time video streaming from local production line.

The capabilities for video calling and augmented reality applications of Smart Glasses were appreciated as highly useful for both operators and maintain personnel. The smart glasses include sensors, accelerometers, cameras, microphones, and other capabilities that can be used to collect and transmit various types of user information. The real-time connection between them and helpdesk allows for significant time saving and

improved field services. When data captured by smart glasses is fed into cloud-based data repository in the data layer, the presentation layer will push the refined and analyzed information out to the browser-based dashboard and show on helpdesk unit’s screen. By using smart glasses, the real-time insight and instance feedback can be provided to helpdesk unit, this will enable remote assistances and services. Figure 3 shows smart glasses can support services for different stakeholders, namely machine operator, maintenance personnel and helpdesk unit.



Figure 3: Services can be supported by smart glasses

4.2. Use Case of Smart Glasses

This purpose of this pilot study was to measure the usefulness of smart glasses at the application level and to set the stage for future works to fulfill the setup of cloud manufacturing. In this paper, only the onsite maintenance scenario is described from the communication channels, roles and relevant tasks at the factory floor. This use case presents how company A provides field maintenance services to their customer factories.

When a customer, identifies the need for a maintenance service, he can difficultly describe the current fault or issue. The company A needs more information about the status of machine/equipment/device and potential error codes to give the customer a hint for a potential solution. If it is a trivial case, the maintenance personnel are required to solve the problem.

The activities performed by the maintenance personnel are depicted using an UML use case diagram. We observed prolong time for fixing faulty machines. Wearable devices can provide considerable improvement in this scenario above. We focused on implementing new functionalities as showing telemetry on the smart glasses customized menu for calibration and settings. Remote connection with helpdesk personnel, display machine usage history etc. Figure 4 presents these features. It shows the use case map for maintenance and remote assistance.

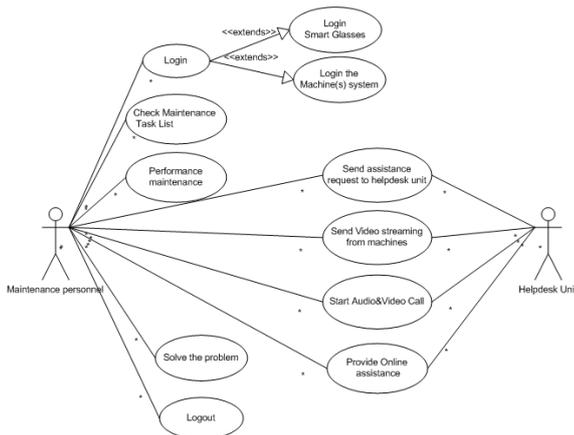


Figure 4: Use case for maintenance personnel hands free concept

In this solution we explore the possibilities for use of state-of-the-art technologies to map value streams and increase value, to avoid waste and increase sustainability, and to reduce risk and prevent a human error on the shop floor. The collaboration environment will be aided by the smart glasses and functionalities such as instant communication and file sharing. More research can be further investigated for the technology to be widely implemented and adopted on the shop floor in reality. Figure 5 provides a vision of using smart glasses to identify machines and display machine datasheet on the right corner of the screen.



Figure 5: Vision of Smart Glasses

5. CONCLUSION

Cloud manufacturing is to encompass the whole product manufacturing lifecycle: from market analysis to design and production, testing, training, usage, and maintenance; and finally dismantlement (Li, Lin and Chai, 2010). Cloud manufacturing business model can help SMEs to innovate their process, product and production side, but the support for the purchasing step, maintenance and training sides always be neglected. Moreover, in order to implement the cloud manufacturing successfully, it's very critical to share a huge amount of data and unstructured manufacturing related information across upstream design and downstream manufacturing in the cloud manufacturing environment (Wu et al., 2013). However, it lacks of a framework for seamless information sharing

mechanisms to facilitate communication and collaboration in distributed and collaborative settings.

Our goal is to connect people into this cloud manufacturing as Internet of Things, and design a communication channel between the shop floor and the back-office. This communication channel ensures while workers serve on-site the administrators or remote helpdesk can collaborate and provide remote assistance or services.

Smart glasses represent a new approach for business communication. There are several applications available, such as navigation guidance, however, not so many innovations in the manufacturing industry. This solution can be easily applied to SMEs for monitoring and assist their field service purposes, because they are flexible, portable and cost limited. Workers can wear the smart glasses to receive specific guidance and instruction when they are performing on-site tasks, such as repair, maintenance and training. Having this wearable system enables hands-free viewing of in situ information that is location based and registered to physical objects and also workers can still receive information while performing other simple tasks (Thomas and Sandor, 2009). The advanced technologies in personal wearable devices combined with cloud manufacturing concepts provide an opportunity to co-create value and streamline human-machine interactions. This will help manufacturers create customized information and assess the workers' performance.

Although this smart glasses solution totally changes the manufacturing and it can bring even more possibilities for manufacturing, it's a controversial solution. First of all, the smart glasses continually raise various concerns regarding to privacy and safety from many international government regulations. The privacy refers to both critical information and operation records. Secondly, using the small display over the eye-field will cause potential health risks (headache) (Kubac et al., 2013). Thirdly, due to the supporting of information visualization, it might cause an issue of information overload (Ford et al., 2012). It's very important to filter the information and make sure the information are effective and efficient (Kubac et al., 2013). Another big barrier hinder this smart glasses usage in factories is the disputes over the access to adequate wireless spectrum to facilitate ubiquitous networking capabilities (Thierer, 2014).

Besides, the manufacturing systems should be designed considering the socio-economic context and other factors, such as product cost, level of required flexibility, desired product quality, therefore, it's important to balance the automation and human activities (Chituc and Restivo, 2009).

REFERENCES

Chituc, C. M., & Restivo, F. J., 2009, June. Challenges and trends in distributed manufacturing systems: are wise engineering systems the ultimate answer. In Second International Symposium on

- Engineering Systems MIT, Cambridge, Massachusetts.
- Camarinha-Matos, L. M., Afsarmanesh, H., & Koelmel, B., 2011. Collaborative networks in support of service-enhanced products. In *Adaptation and Value Creating Collaborative Networks* (pp. 95-104). Springer Berlin Heidelberg.
- Chen, S. L., Chen, Y. Y., & Hsu, C., 2014. A New Approach to Integrate Internet-of-Things and Software-as-a-Service Model for Logistic Systems: A Case Study. *Sensors*, 14(4), 6144-6164.
- Hu, C. S., Xu, C. D., Cao, X. B., & Fu, J. C., 2012. Study of classification and modeling of virtual resources in Cloud manufacturing. *Applied Mechanics and Materials*, 121, 2274-2280.
- Ford, S. J., Rauschecker, U., & Athanassopoulou, N. A., 2012. System-of-system Approaches and Challenges for Multi-Site Manufacturing. *Proc. Of the 2012 7th International Conference on System of Systems Engineering*, Genoa, Italy-16-19 July 2012.
- Huang, B., Li, C., Yin, C., & Zhao, X., 2013. Cloud manufacturing service platform for small-and medium-sized enterprises. *The International Journal of Advanced Manufacturing Technology*, 1-12.
- Jiang, Y. X., Zhang, Y., & Ma, J. W., 2013, August. Research on virtualization of manufacturing resources. In *Measurement, Information and Control (ICMIC), 2013 International Conference on* (Vol. 2, pp. 1060-1064). IEEE.
- Kubac, L., Benes, F., Kebo, V. and Stasa, P., 2013. RFID and Augmented Reality. *Carpathian Control Conference (ICCC), 2013 14th International* pp. 186-191
- Lv, B., 2012, July. A multi-view model study for the architecture of cloud manufacturing. In *Digital Manufacturing and Automation (ICDMA), 2012 Third International Conference on* (pp. 93-97). IEEE.
- Li, B., Zhang, L., & Chai, X., 2010. Introduction to cloud manufacturing. ZTE Corporation. National Basic Research Program of China, (2007CB310900).
- Luo, Y., Zhang, L., Tao, F., Ren, L., Liu, Y., & Zhang, Z., 2013. A modeling and description method of multidimensional information for manufacturing capability in cloud manufacturing system. *The International Journal of Advanced Manufacturing Technology*, 1-15.
- Maguire, E. & Chan, C., 2014. *Deep Field Discovering the Internet of Things*. CLSA Americas LLC. Speical Report
- Meier, M., Seidelmann, J., & Mezgár, I., 2010. *ManuCloud: The Next-Generation Manufacturing as a Service Environment*. ERCIM News, 2010(83), 33-34.
- Mezgár, I., 2011, August. Cloud Computing Technology for Networked Enterprises. In *Preprints of the 18th IFAC World Congress*. Milano (pp. 11949-11954).
- Michel, R., 2014. 4 ways the Internet of things will reshape manufacturing. *Modern Materials Handling*, July, 2014. pp.46-52.
- Park, J. H., & Jeong, H. Y., 2013. Cloud computing-based jam management for a manufacturing system in a Green IT environment. *The Journal of Supercomputing*, 1-14.
- Panetto, H., & Molina, A., 2008. Enterprise integration and interoperability in manufacturing systems: Trends and issues. *Computers in industry*, 59(7), 641-646.
- Putnik, G., 2012. Advanced manufacturing systems and enterprises: Cloud and ubiquitous manufacturing and an architecture. *Journal of Applied Engineering Science*, 10(3), 127-134.
- Phillips, J., 2014. These SMARTGLASSES Have a JOB TO DO. *PC World Jun2014*, Vol. 32 Issue 6, p95-102.
- Tao, F., Zhang, L., Venkatesh, V. C., Luo, Y., & Cheng, Y., 2011. Cloud manufacturing: a computing and service-oriented manufacturing model. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 225(10), 1969-1976.
- Tao, F., Cheng, Y., Xu, L. D., Zhang, L., & Li, B. H., 2014. CCIoT-CMfg: cloud computing and Internet of Things based cloud manufacturing service system.
- Thomas, B. H., & Sandor, C., 2009. What wearable augmented reality can do for you. *Pervasive Computing, IEEE*, 8(2), 8-11.
- Thierer, A., 2014. *The Internet of Things and Wearable Technology: Addressing Privacy and Security Concerns without Derailing Innovation*. Mercatus Working Paper. Mercatus Center.
- Wang, X., & Xu, X. W., 2013. An interoperable solution for Cloud manufacturing. *Robotics and Computer-Integrated Manufacturing*, 29(4), 232-247.
- Wei, C. M., Zhang, C. L., Song, T. X., & Huang, B. Q., 2013, April. A cloud manufacturing service management model and its implementation. In *Service Sciences (ICSS), 2013 International Conference on* (pp. 60-63). IEEE.
- Wu, D., Greer, M. J., Rosen, D. W., & Schaefer, D., 2013. Cloud manufacturing: Strategic vision and state-of-the-art. *J. Manuf. Syst.* Available at: <http://www.sciencedirect.com/science/article/pii/S>.
- Wu, D., Thames, J. L., Rosen, D. W., & Schaefer, D., 2012. Towards a cloud-based design and manufacturing paradigm: looking backward, looking forward. *innovation*, 17, 18.
- Xu, X., 2012. From cloud computing to cloud manufacturing. *Robotics and computer-integrated manufacturing*, 28(1), 75-86.

CYBERNETICS APPLIED TO A SIX-LEGGED ROBOT

Michael Bogner^(a), Thomas Haspl^(b), Franz Wiesinger^(c)

^{(a),(b),(c)}University of Applied Sciences Upper Austria – Embedded Systems Design
Softwarepark 11, A-4232 Hagenberg, AUSTRIA

^(a)michael.bogner@fh-hagenberg.at, ^(b)thomas.haspl@fh-hagenberg.at, ^(c)franz.wiesinger@fh-hagenberg.at

ABSTRACT

Robots are used for so many different reasons such as industrial fabrication, medical assistance or military tasks. But what is it, that makes machines doing the right moves in order to accomplish a given task? And what is it, that makes them doing this autonomously and accurately timed? Essential for explaining this, is cybernetics. To understand the cybernetic principles, it is fundamental to be aware of the fact that it does not arise in robotics. It appears in many other areas such as economy, project management and even politics. From animals we can learn a lot about motion sequences for the development of walking robots, because of evolution animals are acting with a perfect coordination of their movements. On the long term the development of robots should lead to autonomous acting, because finally the independence from leading controlling processes makes the robot a real useful additive. The aim of this paper is to depict the requirements to robotic systems in order to enable cybernetic operations, which is finally indispensable for autonomous acting.

Keywords: cybernetics, robotics, hexapod, system learning

1. INTRODUCTION

The original idea of a robot was an autonomous operating machine, which was supposed to support humans in their everyday life. First models had very little abilities only for simple tasks. But with the amending technology the complexity of robots increased and they were able to accomplish more complex tasks in less time. But on the other hand, it became more difficult to control these robots. When a robots complexity achieves a certain range, it becomes extremely difficult for a human to control the robots processes just by his natural senses. So the next logical step has got to be to teach the machines controlling their movements and their tasks by their own.

A great support for achieving this is the science of cybernetics. It is a natural science, which describes the functioning of things in general. Cybernetics combines the basic theories information and communication as well as regulation and control. By regarding these major theories and by using the mechanisms of self-control, self-regulation and self-organization, cybernetics can

serve to keep a process under control and even to make it viable (Malik 2013, Malik 2008).

In order to operate without the influence of humans, it is necessary for a robot to use sensor technology. The process of measuring data with a sensor and reacting to it in an appropriate way is described by cybernetics. So, the main purpose of this work is to view the huge field of robotics as an application area of cybernetics. It will be investigated, how big the impacts of cybernetics in robotics are and how much certain processes and structures in robotics make use of the cybernetic principles. This processes can include a robots basic movement patterns or even its visual appearance and kind of construction. In order to make cybernetic processes even better, bionics can be very helpful. Bionics describes the art of learning from nature. For example, the functional principle of an aircrafts wing is copied from a real animals wing.

Controlling the basic movements of robots is just one kind of a cybernetic process. But it is much more suitable when correct reactions to unexpected influences are needed. Imagine an obstacle appearing in front of a moving robot. In which direction is it supposed to go? And how can the decision be made before the robot crashes into the obstacle? That are the two most difficult aspects cybernetics has to deal with. The best solution to a problem has to be proceeded accurately timed. All the impacts, that could affect to the correct operating of a robot annoyingly, have to be detected. Unfortunately, there could be an infinite number of different kinds of influences. So theoretically, the implementation of a system, which is able to react to all different kinds of influences, is needed.

This work views a six-legged walking robot as research platform and investigates several concrete scenarios, which are likely to occur in the robots later operating environment. These scenarios will be verified concerning the effective usage of cybernetics. The aim of this should be to obtain useful insights about the requirements to a hexapod in order to work in a cybernetic and viable kind of way and not to get out of control with rising complexity.

The paper has the following structure: In section 2 the cybernetic principles are described with their main aspects. Section 3 depicts a brief state-of-the-art of a hexapod robot and what parts play a major role for

enabling cybernetics. The fourth section describes the main elements a hexapod needs to enable cybernetics in general. Afterwards, in section 5 the cybernetic rules are discussed again regarding the hexapod with the help of a number of real-life scenarios. The needed actions to certain occurring influences will also be presented in this section. Finally, section 6 concludes these findings and declares the main requirements to a hexapod robot in order to be able to operate in a cybernetic kind of way.

2. CYBERNETIC PRINCIPLES

In order to apply cybernetics to the huge area of robotics and especially to a six-legged walking robot it is required to understand its basic working principles. First of all, it is necessary to declare why systems, which do not consider the cybernetic rules, can never be viable.

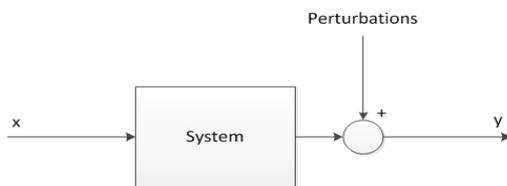


Figure 1: System with straight forward oriented Control

Figure 1 shows a model of a system with its input x and its output y . Possible perturbations are added to the systems output and distort it. This kind of systems are simple, but exactly this simplification unavoidably leads to the main problem cybernetics has to face. When the systems complexity rises, this kind of controlling becomes impossible because of the missing information if a desired target has been achieved or not. The input can not be adapted to the output dynamically if the outer perturbations to the system change.

2.1. Complexity

In fact, it is not the system itself that needs to be controlled, but rather its complexity. But when trying to control any kind of system like shown in figure 1, it is very important to differ from simple and complex systems (Malik 1998).

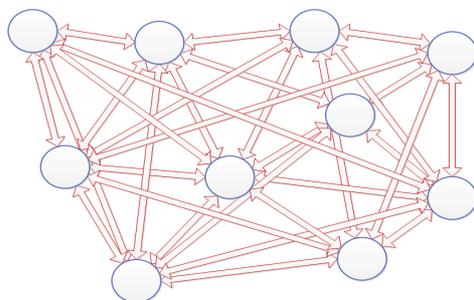


Figure 2: System with a high Complexity

When systems become more complex as shown in figure 2, it gets impossible to keep it under control or grasping it with our senses, because the degree of cross-

linking between the elements is simply too high. In this way, the system becomes unstable and this will lead to an undefined output. That means that a little change of the input can cause an enormous change of the output. Systems with networked connections and mostly acausal behavior do not appear in our everyday life very often, so we have a natural averseness to it. The complexity of a system rises with the increasing number of elements and also with the range of networking between its elements (Vester 1999). It is a matter of fact, that such complex systems with nearly no structure are not viable for very long. So from this point of view, it would be the best to keep a system simple and reduce its complexity as much as possible. But it is essential to clarify that this is not always the best way, because complexity means functionality, it makes a system powerful. The worst case when reducing a systems complexity would be to destroy some important or even its main functions.

It seems that this matter must inevitably lead to another strategies of organization. With the increasing number of elements and higher degree of cross-linking a structural transformation of these organization strategies needs to be done. In other words, it is possible to handle more complexity with a higher developed strategy of organization, or even more general, only complexity can cope with complexity (Vester 1999). This can be achieved by grouping several elements of a system to a subsystem and view it as self-organized. This brings the advantage that a major control process only have to manage the communication between the subsystems instead of all the single systems. Such a structural organization is shown in figure 3.

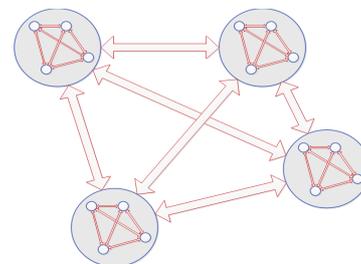


Figure 3: Organization of Organizations

2.2. Control with the Help of Cybernetics

As complexity is now known as the main problem cybernetics has to cope with, some core issues can be defined. How can you control a systems complexity? How can a system be controlled and regulated when it is complex? What characteristics does the systems structure need to have in order to cope with its complexity (Malik 1998)?

To answer these questions you have to make aware the three core values of nature:

- matter
- energy
- information

If you only consider the matter of a system and the energy you implement to it, a typical straight forward oriented control like in figure 1 is the result. If the perturbations suddenly change, the system would react differently to the given input, which can lead to an undesirable output. Consequently, you have to regulate the input information manually according to the influences to the system. But in order to do this efficiently, you have to know the systems characteristics in detail. You always have to know what output will have to be expected when the input data changes. This procedure could be very difficult under certain circumstances, especially when the influences change rapidly. It seems obviously that only considering matter and energy is too less for an efficient regulation.

So apparently the key value for success is information. In this context, information can also be seen as organization. In order to achieve the desired result, a so called feedback loop is built.

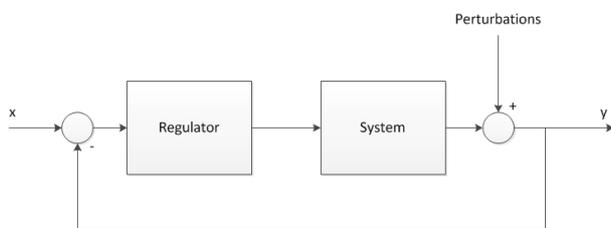


Figure 4: Basic Control Cycle

In figure 4 you can see a classic control cycle with this feedback loop. The actual output of the system is fed back to its input side. There it is compared to the input and the resulting difference, or the regulation deviation, is the input signal for the so called regulator. This unit has the task of calculating the new input for the system, according to this deviation. This process lasts until the wanted output is achieved, which means it must be equal to the input. Such a procedure is called adaptive readjustment and is the basic enhancement, which cybernetics brings. The main requirement for such control cycles is, that the result is achieved in finite time. Norbert Wiener, an American mathematician and founder of cybernetics expressed this matter as follows (Wiener 1961):

”When you desire a specific output to follow a given input pattern, the difference between this pattern and the actually performed output value is used as a new input to cause the part regulated to change in such a way as to bring its output value closer to that given by the input pattern.”

This exactly describes the big improvement, which organization in form of information should bring. The control cycle as in figure 4 describes the technical requirement of cybernetics to all cybernetic processes. With the help of the feedback loop it will be prevented that a systems gets out of control.

3. THE HEXAPOD WALKING ROBOT

A hexapod robot is a six-legged walking robot, which is supposed to imitate a real animal as much as possible. Its intents are to move in a fluent kind of way and react appropriately to unexpected situations. It should be able to recognize its environment and especially obstacles in order to move around them successfully. Another purpose of the hexapod is to receive an aim in the form of coordinates and then walk there autonomously and on the shortest path. There are many different implementations of such robots, but the main tasks are always the same.

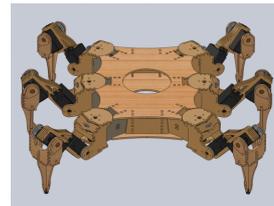


Figure 5: A possible Chassis of a Hexapod Robot

A concrete implementation of the University of Applied Sciences in Upper Austria is shown in figure 5. It is powered by 18 servo motors, three for each leg. That also results in three degrees of freedom for each leg. The chassis of the robot itself consists of ply wood. That is lightweight and stable simultaneously. The brain of the hexapod is a self developed hardware platform, which is called *Sandbox*. It combines a field-programmable gate array (FPGA) with a micro controller. On the FPGA the so-called *Robotics Chip* is implemented, which is responsible for generating the pulse-width modulation (PWM) signals, which serve as input for the servo motors. The micro controller takes care of the communication and the signal routing between the *Robotics Chip* and the servo motors. This implementation of a hexapod is our departments main platform for research in the field of mobile robotics. It is able to receive commands in terms of mathematical vectors, which are related to the hexapods local coordinate system. From these input vectors it calculates the direction, in which it should go. So, all the collected information and results in this paper base on this platform.

4. REQUIREMENTS TO THE HEXAPOD

There are certain parts and elements a robot needs, which are indispensable for enabling cybernetic processes. The availability and accuracy of these elements determine the reachable effectiveness and power of the desired cybernetic sequences.

4.1. Sensors and Actuators

Sensors and actuators might be the most important parts of a robot to allow the implementation of a cybernetic process.

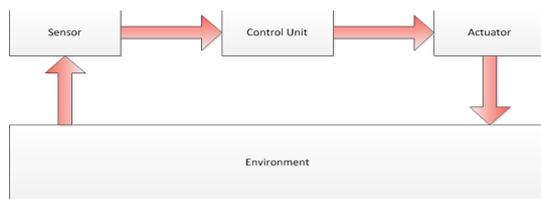


Figure 6: Structure of a simple Sensor Cycle

This technical appliances scilicet represent the beginning and the ending of a cybernetic control cycle as shown in figure 6. The needed data for the feedback mechanism is grasped with sensing elements and after passing through all the parts of the control cycle, the calculated controlling signals affect the actuating elements, which take the desired actions. To enable a completely autonomous operating machine in order to observe the environment as accurate as possible, a high number of such elements is needed.

4.2. Bionics

Similar to cybernetics, bionics is interdisciplinary and can therefore be applied to many different industries. In order to make use of bionics means learning from the construction, process and development principles of nature in order to cross-link humans, environment and technics in a positive kind of way (Nachtigall 2002).

So what can we learn from nature to develop a hexapod? The answer to this question is obvious - everything we need for the effective operating of the hexapod. Ranging from the basic movement pattern to get the ability of walking in general, to the best kinds of leg coordinations in order to get a fluent and energy saving locomotion. You can also copy the geometric structures of an insects corpus, its legs and especially the particular limbs to achieve the best results. It is essential to understand, why an insects leg is designed as it is. Once you have done this, you can try to rebuild it and apply all the gained knowledge.

5. CYBERNETICS AND THE HEXAPOD

As we already know that cybernetics can be applied to control complex systems like the hexapod and after depicting the main abilities a robot has to provide in order to enable cybernetics, we can now model a number of specific scenarios, which are typical for the hexapod, and analyze them regarding their cybernetic aspects. By simulating the behavior of the hexapod within these specific processes, the appearance of cybernetics will be investigated and the criteria to the hexapod in order to apply cybernetics successfully will be clarified. The processes we want to observe are as follows:

- basic locomotion
- changing environmental impacts
- overcoming of obstacles
- error handling
- system learning

These chosen processes are tasks, which are typical and crucial for a hexapod walking robot. Nevertheless, it is just a selection of some major operations and do therefore not describe the whole behavior of the hexapod.

5.1. Movements

With the help of bionics we can learn how the hexapod is supposed to walk in order to use its resources in the best kind of way. But what about the motion pattern itself? The different gaits just provide information about the deflection of the particular limbs and the order the legs are accessed. But this is not enough to achieve autonomy. It is too less just to tell a leg, that it has to move its tip to a fixed point in the coordinate system, because this would lead to a stiff looking movement without any information about the achievement of the desired position.

When you remember the classic control cycle in figure 4, the input for the control cycle in this case is the set of desired angles of the particular limbs. Thus, the limbs, which form the system of the control cycle, apply these angles and the feedback can be done with an angle measuring sensor or by calculating them with the help of inverse kinematics. As long as the desired angles are not achieved yet, a regulator has to calculate the ruling difference permanently and provide the according input to the legs engines. With all these parts the control cycle can be closed and a cybernetic process can be performed. When the control deviation of the control cycle is zero, the desired position is reached and the leg keeps this position until a new input is valid. A visualization of this case you can see in figure 7.

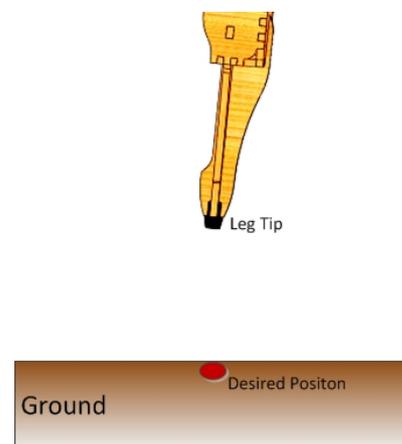


Figure 7: Basic Movement within a Control Cycle

One possible problem with this control cycle could now be, that the calculation of the particular angles with the help of inverse kinematics lasts longer than the interval, within the feedback data is sent to the comparator of the control cycle. Thus, it is the duty of the developer to choose the correct feedback interval in order to avoid this problem.

5.2. Changing environmental Influences

5.2.1. Acclivity

For an effective walking over all kinds of grounds it is required to adapt the kind of walking to the actual ruling ground conditions. To achieve this ability, the environment needs to be observed permanently.

To regard to the cybernetic control cycle from figure 4 again, an inclination sensor could serve as the feedback delivering item and give information about the actual tilt of the ground. When this tilt exceeds a certain limit value, the motion controlling algorithm should switch to another gait. But just switching to another gait does not form a control cycle.

Firstly, it is important to figure out, if ascending or descending acclivity has to be handled. This information you get from the inclination sensor. Then the required action would be to shift the hexapod's weight into the direction, which is closer to the slope. This action should not only prevent from falling over, it also correlates with a basic cybernetic rule, which declares the usage of existing surrounding forces, which would be gravity in this case. When the hexapod should be pushed to the ground by the gravity, a balanced position is required. In order to achieve this, a new angle set for all the legs, which describes the desired position to keep the hexapod stable, is calculated according to the particular value of acclivity. This information is the input for the control cycle. The feedback data can be determined with the help of inverse kinematics again. When all the limbs' angles are calculated, you can deduce the whole position of the hexapod from these angles. This calculated position is sent to the control cycles' comparator and the control deviation defines the new input of the system, which are all the limbs in this case. When the required position is occupied, the hexapod has its new initial position, to which the actual gait has to be adapted now.

5.2.2. Ground Unevenness

It is not very innovative to assume that the hexapod will exclusively walk on even underground. Exactly this ability to walk through uneven terrain should be the big advantage of walking robots over wheel robots.

Imagine the situation, a stone lies exactly on that point, to which the leg tip is supposed to move according to the algorithm. Such a trivial scenario makes it clear, that even for the movement of one single leg already a cybernetic control cycle is needed.

First of all, the hexapod needs sensing elements, which are able to give information about the tip of a leg. This could be a force sensor in the leg tip, which tells the controlling unit, that the contact with the ground or another object is given, or a kind of visual sensor like a camera, which observes the legs.

The classic control cycle like in figure 4 is not really correct in this case, because here we have a so-called interleaved control cycle. You can consider this as a control cycle within a control cycle and the inner cycle, which regulates the forward locomotion itself, forms the

system of the outer cycle. When you now consider this outer control cycle, the earlier mentioned sensor gives feedback about the connection from the leg tip and any object. If an object lies between the leg tip and the desired position on the ground, or even a hole is at the position the leg tip is supposed to move to, the inner control cycle must be aborted or delayed accordingly. This means, that the outer control cycle must have a higher priority than the inner one, because it is wrong to move to the given coordinates obstinately, when it is not possible.

5.2.3. Aerodynamic Resistance

In order to cope with changes in aerodynamic resistance, a velocity sensor could be used to measure the wind force, because the velocity of the wind is proportional to its force. When the wind force increases now from a certain direction, the hexapod has to shift its weight into this direction and also move closer to the ground in order to counteract the wind. If the hexapod does not have the ability to react in that kind of way, the wind has a maximum contact surface and the hexapod could fall over in the worst case. The measured wind force is compared to the limit values, which have to be defined by the developer, and then the appropriate position of the hexapod is the input of the control cycle. Now the actual position needs to be observed in order to close the cybernetic loop. This can be done with the help of inverse kinematics again. The calculated position is the feedback data within the control cycle and when it is equal to the input position, the desired position is achieved. This target position has to be dependent from the wind power. It has to be found out experimentally in order to counteract the actual level of wind.

5.3. Obstacles

To achieve the ability of passing obstacles successfully, the hexapod generally needs the skill to recognize the obstacle. Possible sensing elements for this could be a distance sensor or a visual kind of sensor like a camera. Or the even better solution would be the accurate knowledge of the area before the hexapod starts acting. This would be possible with a cartographic recording of the area. Once the hexapod recognized the obstacle, it has to find out the best way around it, if there is anyone at all.

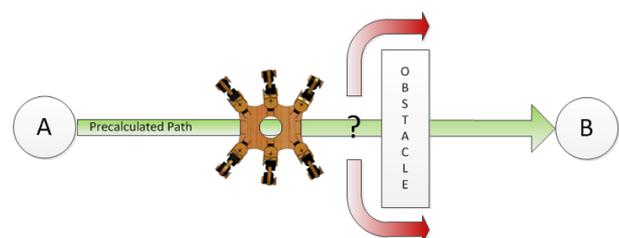


Figure 8: The Problem of choosing the correct Direction when an Obstacle appears

In figure 8 it does not matter, if the hexapod passes the obstacle on the left or the right side, because the

distances are the same. Potential factors impacting such decisions could be the grounds actual inclination, the accessibility of both sides or even the luminance on each side. It is not very likely that the hexapod always reaches the obstacle exactly in its center, so another influence to this decision could be, that into one direction the distance to the obstacles edge is shorter than into the other. However this decision is made, once a direction is chosen, the actual controlling process begins and cybernetics comes into play.

The hexapods position represents the system controlled. The input of the control cycle are all points of the precalculated path from the position in front of the obstacle to point B. Now the hexapod is supposed to walk along the obstacles edge until it reaches the precalculated path again. In order to manage this, a distance sensor can be used to control the distance between the hexapod and the obstacles wall. Due to the set of inputs and the permanent comparison of them to the actual position, the adaptive readjustment is successful when one point out of this input vector is reached. If this is the case, the obstacle has been bypassed successfully. From this achieved point, which is nothing else than a position related to the inertial coordinate system, the hexapod continues its way on the precalculated path.

But what is the hexapod supposed to do, if the obstacles edge never leads back to the precalculated path? This situation could cause a big trouble and an appropriate error handling strategy needs to be accomplished.

5.4. Error Cases

Even the best controlling algorithms can cause errors. A perfectly autonomous operating hexapod would need to have an error handling procedure for as many problems as possible. The biggest problem of all is probably to recognize a problem as such. Within every control cycle and especially within their controlling algorithms there have to be defined several limit values in order to recognize problems. These limit values can have all different kinds, because every particular task has its own special limitations. Examples could be extreme values of distances, temperatures or even time. Basically, there must be defined one or more limit values for every quantity the robot can measure. Every set of limit values has to be matched to its particular controlling algorithm. In case one of these values is exceeded, an error has occurred and the error handling algorithm takes control over the system. There can not be an universal process for handling errors, because every robot has different functions and requirements. For the hexapod it is essential to identify occurring problems early enough in order to have the chance to continue its regular work.

5.5. A learning System

If an autonomous working hexapod is supposed to operate in the same region for a longer period of time, it would be highly innovative to make him learn certain details of this region like the shortest path and the

combination of the most effective gaits to an already known destination.

A prerequisite for this is the availability of memory. When finding out the best solution to a problem, a so called evolutionary algorithm can be applied. This is a controlling algorithm for a certain problem, which is developed by itself. This procedure is called genetic programming and can lead to quite impressive results (Banzhaf and Ziegler 2005).



Figure 9: The Hexapod has to escape the Maze

The functionality of genetic programming can be best described by an example. Let us have a look at figure 9. A motion sequence with the quickest way out of the maze is wanted and the hexapod has to memorize this way in order to get out much faster the next time. For this reason, the hexapod needs some basic commands, which are processed in a recursive kind of way. These are moving forward, left, backward and right. Now the hexapod applies these commands always in the same order starting at a defined position, which is the right upper corner in this example. After every successful execution of one command, this step is buffered and the hexapod starts again with the first command from the new position. If the execution of one command is not possible, because there is a wall in front of the robot, the next command is tried out. If this is successfully, this step is buffered, and the procedure begins with moving forward again from the new position. If it comes to the problem, that no direction is possible from a certain position, the hexapod moves a step back and removes this step from the buffer. And now the next command from this position is tried out. In this way, the hexapod walks through the maze until it reaches the exit in the right bottom corner. After the exit is recognized, the buffered steps are memorized permanently as the way out of this particular maze.

This example is supposed to show the principle how to make a robot learn. Many other problems can be solved with genetic programming. It would also be possible to find the best way of walking in general with this principle. The requirements to enable genetic programming are to provide a set of basic commands or processes, which are needed to build the higher desired process.

6. CONCLUSION

The prior observed use cases of a hexapod walking robot should now allow significant statements about the requirements to robotics in order to make cybernetic processes possible and to work in an autonomous kind of way.

You always have to be aware of the most important components a cybernetic control cycle consists of. Firstly, some kinds of sensing elements to grasp the required information. Secondly, the feedback loop in order to compare the measured data with the input. And thirdly, a regulator, which is responsible for calculating the appropriate control signal for the system in order to adjust the regulation deviation to zero. These are just the basic prerequisites to enable a cybernetic process at all. But it is much more important to understand the cybernetic principles and be able to think in an interconnected kind of way. You need to be able to recognize subsystems within a bigger system and convert them into self-controlled elements consisting of one or more cybernetic control cycles. With this strategy you can reach a high flexibility and a good stability of the system as well. It is also essential to identify the best way of cross-linking between these subsystems. In this way, the whole system loses complexity and the robots supervising control algorithm can be able to handle the information flow between the subsystems.

As you can see in the shown examples it can be possible to implement an autonomous operating robot, but it also means a huge amount of effort. No matter how far developed a robot is, there will always be certain limitations in its autonomy. A completely autonomous acting robot would need the ability to react to every problem appropriately and accurately timed. But unfortunately, this is only restricted possible nowadays. The rising number of sensing elements and control units are the reasons for this limitation. So the main interest should be in developing robots, which are able to react to as many different kinds of problems as possible or to make them best suitable for specific applications.

Basically, you can say that cybernetic processes of all kinds have to follow specific rules in order to keep a system under control. By modeling concrete use cases of a specific platform in this work, it has been shown that cybernetics is an interdisciplinary science, which can be applied to many different areas. Furthermore, by simulating a robots behavior within these scenarios the needed effort for operating in a cybernetic kind of way could have been estimated.

As we can see, it is not constructive to concentrate only on the technical matters when developing robots. It is much more important to have correct and powerful organization structures, locomotion sequences, materials, and so on. You can obtain these things from many other sciences and then apply them within the robot developing process.

REFERENCES

- Nachtigall W., 2002. Bionik – Grundlagen und Beispiele für Ingenieure und Naturwissenschaftler. Heidelberg: Springer-Verlag.
- Wiener N., 1961. Cybernetics: or the Control and Communication in the Animal and the Machine. Cambridge, Massachusetts: The MIT Press.
- Vester F., 1999. Die Kunst vernetzt zu denken – Ideen und Werkzeuge für einen neuen Umgang mit Komplexität. Stuttgart: Deutsche Verlags-Anstalt GmbH.
- Malik F., 2013. Strategy: Navigating the Complexity of the New World. Frankfurt am Main: Campus Verlag GmbH.
- Malik F., 2008. Strategie des Managements komplexer Systeme: ein Beitrag zur Management-Kybernetik evolutionärer Systeme (Management Strategy of Complex Systems: A Contribution of Management Cybernetics of evolutionary Systems). Haupt Berne.
- Malik F., 1998. Komplexität – was ist das?. Cwarel Isaf Institute. Available from: <http://www.managementkybernetik.com> [accessed 8 April 2013].
- Banzhaf W., Ziegler J., 2005. Vom Kriechen zum Laufen – Evolution des Laufens mit Genetischer Programmierung auf beliebigen Morphologien. Heidelberg: Springer-Verlag.

CONDUCTING EXPERIMENTAL DESIGN AND OPTIMIZATION ON AN INNOVATIVE CAR RENTAL BUSINESS

Natalia P. Basán^(a), Mariana E. Cóccola^(b), Carlos A. Méndez^(c)

INTEC (UNL-CONICET)

^(a)nbasan@intec.unl.edu.ar, ^(b)marcoccola@santafe-conicet.gov.ar, ^(c)cmendez@intec.unl.edu.ar

ABSTRACT

This paper introduces an advanced 3D object-oriented discrete event simulation model, developed with SIMIO software, to optimize the design and operation of an innovative car rental service. The proposed simulation model is used to perform an exhaustive sensitivity analysis over the main process variables. The main goal is to determine the best system configuration, which allows maximizing the company profit while satisfying a suitable service level to the customers.

Keywords: Discrete-Event Simulation, Experimental Design, Car Rentals, Optimization.

1. INTRODUCTION

In order to represent and optimize the design and operation of modern real-world service processes, two solution strategies, among others, are typically used: (i) analytical methods based on linear / non-linear mathematical formulations (MILP / MINLP models) or (ii) discrete event simulation-based methods. On the one hand, exact mathematical optimization approaches may be able to rigorously find optimal or near optimal solutions to several interesting problems. However, for realistic complex decision-making processes, the problem representation may be usually highly simplified and the computational effort may be really significant. On the other hand, modern discrete event simulation packages provide a set of tools to easily generate a better representation of more complex systems. However, one of simulation's greatest disadvantages is that, on its own, it does not serve to estimate the optimality gap (Law et al. 2002). To mitigate this inherent drawback, a deep sensitivity analysis can be performed. This technique is used to easily evaluate how different values of a set of independent variables will impact a set of dependent variables under a given set of assumptions.

In the last years, the commercial simulation software has incorporated efficient optimization codes to automatically search for near optimal solutions. Simulation optimization can be defined as the process of testing different combinations of values for controllable values, aiming to find the combination of values which offers the most desirable output results for simulation models (Harrell and McComas 1996). Unfortunately, the performance of the method is highly affected by the number of variables being manipulated. This paper proposes a simulation optimization strategy

to determine the best configuration of an innovative car rental business. The simulation model, developed with the SIMIO software, was used to analyze and evaluate the dynamic behavior of the process, considering different operative schemes and possible alternatives of investments. During this process, several types of experimental designs, e. g. multifactorial and fractional, were executed in order to select the most important variables of the model which directly impact on the response variables. The selected variables were then properly studied in order to find the best configuration for the car rental service.

2. DEVELOPING THE SIMULATION MODEL

This work is motivated by the challenging problem presented in the 2014 SIMIO Student Competition. The problem addressed is the design and operation of an innovative car rental system for travelers in an airport. In the next subsections, the system characteristics, model assumptions, and simulation model development will be discussed.

2.1. System Description

The system to investigate is based on a new airport business that allows passengers to drop-off their car at a parking facility at the airport, and in addition to free parking they may receive rental income for their cars while they are away. The company offers to its customers free transfer between parking facility and airport.

2.2. Model Assumptions

The following assumptions are considered for developing the simulation model:

- The company must operate a uniform fleet of vehicles. There are three size options: Mini, Small and Large.
- Each shuttle follows a fixed route (Figure 1) between the airport and the rental company to load and unload passengers. In addition, they are uniformly distributed over time in order to visit the collection points (airport and rental office) periodically.
- For both shuttle drivers and check-in clerks, there are three work shifts: Early, Midday and Late.
- A FIFO (First In First Out) queue policy is adopted.

- Arriving passengers can use a mobile application to select from the available cars.
- Arriving passengers, who could not find an appropriate car in the rental office, are returned to the airport.
- Both the parking arrivals and renting arrivals have a party size. These groups of passengers are not separated during the simulation.
- Rental cars are assigned in a specific category based on the model type.
- The owner of the car receives 50 % of the rental income.
- Cars entering the parking must be cleaned.
- Both the parking duration and rental duration are in integer days.
- Cars must be parked between a minimum and a maximum number of days.
- When a car is not accepted for parking, the customer is lost to competition.
- Parking passengers are expected to reach the airport in 15 minutes at most. Rented passengers expect to spend no more than 20 minutes to travel to rental facility and complete the check-in-process.

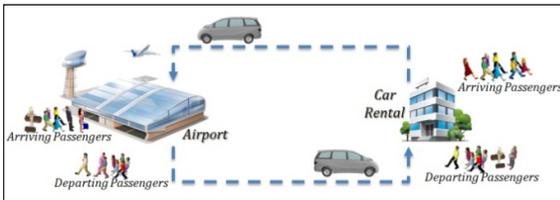


Figure 1: Shuttle routes

2.3. Simulation Model

The real system is represented as a discrete event simulation model by using the SIMIO software (Kelton et al. 2011, Thiesing et al. 1990). SIMIO is a modern simulation modeling framework based on intelligent objects. Actually, this software is used to represent a wide-range of production and service systems (Basán et al. 2013). 3D views of the simulation model for the car rental are given in Figure 2 and 3.



Figure 2: 3D graphical view of Airport



Figure 3: 3D graphical view of Car Rental

Each element and its behavior in the real system are modeled by specific objects and steps of the simulated software. The major characteristics of the computer model are presented below.

2.3.1. Model Entities

Two instances of Entity Object are defined into the model. They are:

- Car
- Group (parking and rental passengers)

A Batch Element is defined for combining both types of entities.

2.3.2. Parking and Rental Passenger

When a *Parking Passenger* arrives to the parking facility, he leaves his car at the cleaning station to then be transported from the parking facility to the airport by using the provided shuttle fleet. When returning from their trips, passengers are transported by means of the vans from the airport to the parking facility to find their cars.

We assumed that of the overall of *Rental Passengers* arriving at the airport: 50 % use the mobile app to reserve a car; 50 % go to check-in counter to reserve a car personally. For rental passengers using the mobile app:

- If the desired type of car is not available, the customer is not moved to the parking facility and is lost to competition.
- If the desired type of car is available, the customer reserves the car and then is moved from the airport to the parking facility by the shuttle fleet.

For rental passengers that do not use the mobile app:

- They are transported from the airport to the parking facility by the shuttle fleet.
- Then, they go to the check-in counter to reserve a car.
- If the preferred type of car is not available, they are lost to competition. Passengers use the van fleet to return to the airport.

2.3.3. Cars

New cars are accepted for parking only if:

- There are parking slots available.
- The number of cars that have been received by the company (parked + rented) is lower than the value of $(AcceptanceRelation) * (ParkingCapacity)$.
- When a car is not accepted for parking is sent to *ParkingCompetition* (Sink Object).

2.3.4. Parking Slots

Parking slots are modeled by using a Station Element. The number of parking slots available is determined by variable *ParkingCapacity*.

2.3.5. Check-in

The Check-in Counter is modeled as a Server Object with the following properties:

- Capacity is supposed to be Infinity.
- Processing Time sets to a triangular distribution between the largest and smallest possible values, with a peak at the most likely value: $Random.Triangular(1, 2, 3)$ minutes. Three resources are included into the model to represent the check-in staff (early, midday and late clerks).

2.3.6. Shuttle Fleet

Vans are modeled as instances of the Van Object (subclass of Vehicle Object). The Vehicle Object has an additional state variable named *RideQueueContents* which allows determining the quantity of travelers that is being transported on the shuttle. The additional variable is needed because of each arrival have a party size that varies between 1 and 4 travelers.

Vans follow a fixed route defined by the Sequence Table *VanRoute*. At each visited node, the vehicles for passenger pickups wait for a specified “dwell” time which is dependent of van type.

When the vans are off-shift, they park at Home Node named *VanParking*.

When the vans go on-shift, an add-on process is triggered to establish a suitable arrival frequency of the vehicles to pickup / drop off nodes. The above strategy allows a better utilization and distribution of the shuttle fleet and reduces the waiting time.

3. EXPERIMENTAL DESIGNS AND RESULTS

After carrying out an exhaustive verification and validation of the proposed simulation model, a comprehensive sensitivity analysis was conducted following next stages:

1. Factorial Design,
2. Full Factorial Design,
3. Optimization.

3.1. Experimental Scheme

In the experimental design, the input parameters and structural assumptions composing the simulation model are called factors. In this case, the identified input factors are:

1. Staffing level for the rental check-in counters for each of the three shifts: early clerks (EC), midday clerks (MC) and late clerks (LC),
2. Staffing level for the shuttle drivers for each of the three shifts: early drivers (ED), midday drivers (MD) and late drivers (LD),
3. Number of parking slots (PS),
4. Type of shuttle (TS),
5. Dwell Time (DT),
6. Acceptance Relation (AR).

These control variables can affect output performance measures of the model, which are called responses variables. In the simulation model, several different responses or performance measures of interest are used:

1. Company Profit,
2. Shuttle Fleet Utilization,
3. Check-In Counter Utilization,
4. Parking Slots Utilization,
5. Service Times,
6. Customers that are lost to competition.

The optimization objective is to maximize the company profit having into account the following constraints:

- Services times for parking passengers should be less than 15 minutes.
- Services times for rental passengers should be less than 20 minutes.
- Services times for returning passengers should be less than 15 minutes.
- The number of times that a car is returned by a rental passenger and there is not parking slot available should be zero.

Simulation is a useful tool to determine the sensitivity of a model to changes in input parameters. In this way, the factor identified above were combine to define and simulate alternative system configurations. After conducting a sensitivity analysis of controlled process variables, the range of input values for the control variables was determined as follows:

- Check-in clerks for each shift: 1 to 2.
- Shuttle drivers for each shift: 1 to 3.
- Number of parking slots: 600 to 800.
- Type of shuttle: Mini (4), Small (8) or Large (12).
- Dwell time: 1 to 4 minutes (depending of the type of shuttle).
- Acceptance relation variable: 2 to 3.

3.2. Results

In the simulation model 10 replicates for each scenario were run and their results were compared. The same test was replicated multiple times in order to create a variation in each response variable which then was utilized to evaluate experimental error. The proposed system configurations were tested on the simulation model during 105 days with a warming-up period of 15 days. Table 1 shows the average results for the top six alternatives.

Table 1: The best alternative system configurations

	Scenario					
	1	2	3	4	5	6
EC	1	1	2	1	2	1
MC	1	1	1	1	1	1
LC	1	1	2	1	2	1
PS	600	700	700	800	800	800
TS	8	8	8	8	8	12
ED	2	2	2	2	2	3
MD	1	2	2	2	2	1
LD	2	2	2	2	2	3
DT*	2	2	2	2	2	3
AR	2.9	2.6	2.6	2.5	2.5	2.5
CP**	2.21	2.17	2.13	2.12	2.10	2.09
PT*	11.6	10.6	9.9	10.7	9.9	9.4
RT*	16	14.6	12.7	14.8	12.7	13.2

* Dwell time (DT), parking service time (PT) and renting service time (RT) were measured in minutes.

** Company profit (CP) variable was measured in millions of dollars.

From Table 1, it follows that the best system configuration, which allows maximizing the company profit, is determined by the input values shown in Table 2.

Table 2: Proposed configuration system

Number of Parking Slots	600		
Maximum Number of Accepted cars	1740*		
Type of Shuttle	Small Van (8 seats)		
Van Quantity	3		
Dwell Time	120 seconds**		
Drivers Quantity	Early	Midday	Late
	2	1	2
Rental Check-In Counters Quantity	Early	Midday	Late
	1	1	1

* A new car is accepted whether $(2.6 * PS \geq \text{current parked cars} + \text{current rented cars})$. This is due to that the company is able to offer parking to more customers than the available parking spaces.

** In order to improve the use of the van fleet, a minimum dwell time is required in each pickup point visited by the vehicles.

3.2.1. Company Profit

Table 3 shows the benefits and costs of the resources for the best found scenario. The confidence level used to calculate confidence interval half-width statistics for result averages across replications is 95 %.

Table 3: Company profit

+Gross Profit (\$)	2,759,774
-Parking Spot Cost(\$)	385,714
- Van Operation Cost(\$)	13,500
-Shuttle Fleet Drivers Cost(\$)	104,400
-Rental Check-in Counters Cost(\$)	45,360
Average Net Profit(\$)	2,210,800
Confidence Interval of the Mean	$2,146,650 \leq u \leq 2,186,097$

3.2.2. Resources and Parking Utilization

The average utilizations of each resource are given in Table 4. For parking facility, the maximum quantity of cars parked in simultaneous was 580 lots while the average utilization was 72 %.

Table 4: Resources utilization

	Early	Midday	Late
Vans	45.0%	45.6%	40.6%
Check-in Counters	29.0%	19.4%	24.7%
Cleaning	75.7%		

3.2.3. Passenger Arrivals

On one hand, if parking customers arrive when the lot is full are lost to competition. On the other hand, the renting passengers are lost to competition when the desired type of car is not available. Figure 5 shows on the total the customers arrivals, the percent of arrivals lost.

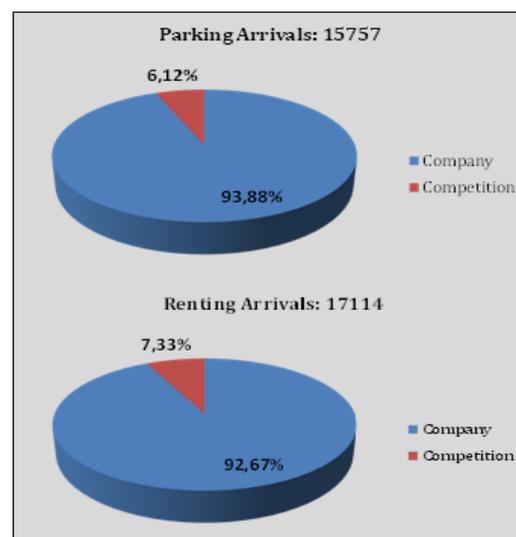


Figure 5: Customers lost

3.2.4. Service Times

Service times required for each passenger transported between parking and airport are detailed in Figures 6 and 7. These graphs show the service times computed in one replication of the proposed configuration. The mean and confidence interval were calculated considering 10 replications. The confidence level used was 95 %. Based on confidence intervals, the proposed configuration is able to satisfy the required level of service to customers.

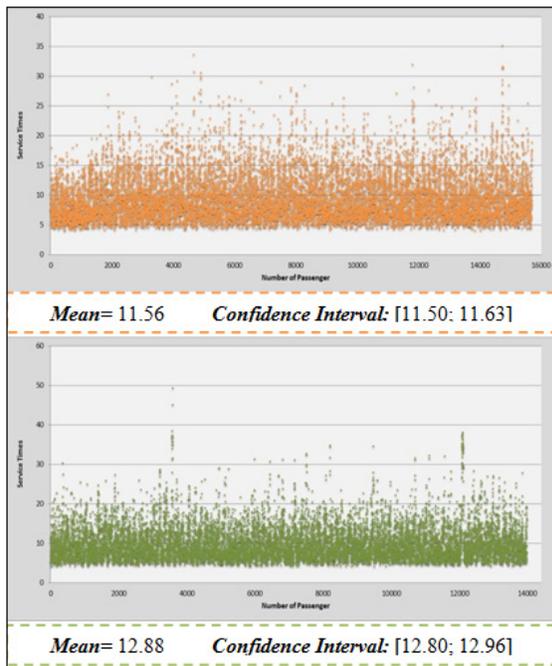


Figure 6: Services times for transporting parking passengers from parking facility to airport and vice versa

4. CONCLUSIONS

In this paper, a novel discrete event simulation modeling tool has been used to evaluate different configurations and operating modes for an innovative car rental business. The computer model was developed together with an attractive user friendly 3D interface to expedite the validation and illustration of simulation results. The proposed simulation model was used to evaluate the performance of alternative system designs and operation modes and, at the same time, to estimate the expected service quality and profitability of a realistic complex business. In this stage of the simulation project, an exhaustive sensitivity analysis was carried out to evaluate how different values of a set of critical independent variables will impact on a particular set of performance variables under a given set of assumptions. The sensitive analysis process was very useful to find the configuration system that allows maximizing the company profit while satisfying a suitable service level to the customers.

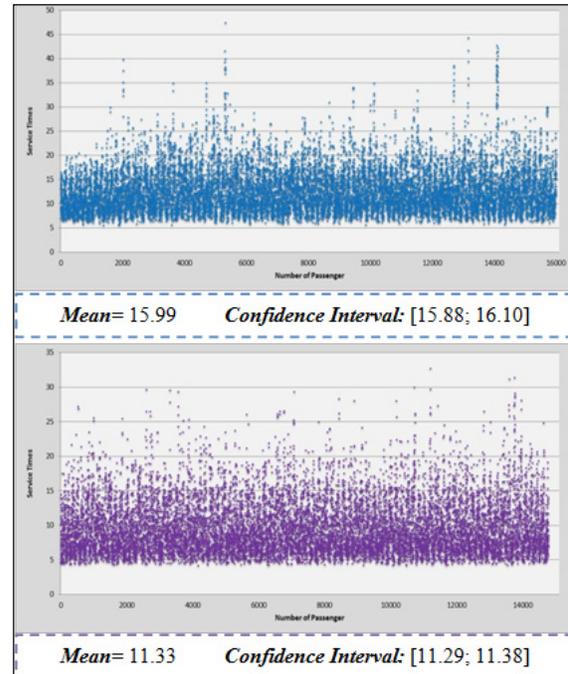


Figure 7: Services times for transporting renting passengers from airport to parking facility and vice versa

ACKNOWLEDGMENTS

This paper was partially funded by CONICET under Grant PIP-2221 and from UNL under Grant PI-81-481.

REFERENCES

- Basán N., Ramos L., Cóccola M., and Méndez C. A., 2013. Modeling, simulation and optimization of the main packaging line of a brewing company. Proceedings of the 25rd European Modeling & Simulation Symposium, 551-560. September 23-25, Athens, Greece.
- Harrell C., Bateman R., Gogg T., Mott, J., 1996. System Improvement Using Simulation. Orem, Utah, United States. Promodel Corporation.
- Kelton D., Smith J. S. and Sturrock D. T., 2011. Simio & Simulation: Modeling, Analysis, Applications, 2nd Ed. Boston, MA, McGraw-Hill Learning Solutions.
- Law A. M. and McComas M. G., 2002. Simulation-based optimization. Proceedings of the 2002 Winter Simulation Conference, 41-44. December 8-11, San Diego, CA; United States.
- Thiesing, R., Watson C., Kirby J., and Sturrock D., 2009. SIMIO Reference Guide, Version 6. SIMIO LLC.

AUTHORS BIOGRAPHY

NATALIA P. BASAN is an Industrial Engineer and PhD student at National Scientific and Technical Research Council (CONICET). She is lecturer of Simulation at Universidad Nacional del Litoral (UNL). Her research interests include hybrid optimization &

simulation tools for production planning and scheduling of automated production systems. Her email address is nbasan@intec.unl.edu.ar.

MARIANA E. COCCOLA is an Information Systems Engineer and a PhD student conducting research in hybrid optimization & simulation tools for production and logistics systems. Her email address is mcoccola@intec.unl.edu.ar.

CARLOS A. MENDEZ is a Titular Professor of Industrial Engineering at Universidad Nacional del Litoral (UNL) in Argentina as well as a Researcher of the National Scientific and Technical Research Council (CONICET) in the area of Process Systems Engineering. He has published over 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/>. His email address is cmendez@intec.unl.edu.ar.

MODULAR APPROACH FOR MODELLING AN AIRPORT SYSTEM

Paolo Scala^(a), Miguel Mujica Mota^(b), Nico de Bock^(c)

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

^(a)p.m.scala@hva.nl, ^(b)m.mujica.mota@hva.nl, ^(c)n.de.bock@hva.nl

ABSTRACT

Airport capacity, expressed as the maximum number of air traffic movements that can be accommodated during a given period of time under given conditions, has become a strict constraint to the air transportation, due to the scarce amount of resources on the ground and restrictions in the airspace. Usually the problem of capacity at airports is studied separating airspace operations from ground operations, but it is evident that the two areas are tied each other. This work aims at developing a simulation model that takes into account both airspace and ground operations. The approach used is a modular approach, in other words, the model developed is the result of the combination of four different models which in turn make a more powerful one. The four models refer to the airside (runway and taxiway and turn-around), and airspace operations (Terminal Maneuvering Area). This integrated approach allows to assess the system as a whole, identify possible bottlenecks and elements that affect the real system.

Keywords: simulation model, airport ground operations, airspace operations, modular approach, data driven decisions

1. INTRODUCTION

As many authors stated, airports are the main bottleneck of the global air transportation. Looking at the trends for the next coming years, it is expected that the demand of traffic will increase, therefore, the study of an efficient use of the resources involved in air transport operations is a key aspect. The actors involved in the air transportation industry come up with different solutions to cope with the problem of lack of capacity. For instance, a solution could be the building of new facilities such as new runways or new gates, other solutions regard the optimization of the existing resources in order to improve the efficiency of the various processes involved in airport operations. The former solution seems to be the more logical, but it is the worst in terms of time and investments needed, on the other hand, the latter seems to be a challenging one in which techniques from operations research gain importance due to the type of problems to be solved. Another solution could be the development of the so called multi-airport system (De Neufville and Odoni 2003), in which capacity is spread on a network of

airports constituted by a main airport (hub) and some secondary airports (non-hub). The purpose of the multi-airport system is to relieve capacity at the main airport (hub) and increase the capacity of the overall system in order to satisfy the demand of their catchment area. Airports included in a multi-airport system need to be close to each other and well connected to each other in order to make the choice of the airport from/to which depart/arrive irrelevant.

Many studies have been made about the improvement of airport capacity, some refer to the optimization of ground operations and others to airspace operations and terminal capacity. Regarding ground operations, most of the studies were related to gate allocation, optimization of taxi-in and taxi-out routes, scheduling of departing aircraft and operations related to the turnaround process. From literature we can see that lots of methodologies were used for solving these kinds of problems, most of them employ analytical and heuristic solutions. For instance, in the work of Dorndorf et al. (2007) is provided a large survey about various models and solutions techniques utilized for the gate scheduling problem, Bolat (2000) solved the gate assignment problem using a Branch&Bound algorithm, combined with the use of two heuristics. In the work of Narciso and Piera (2015), a colored Petri net formalism was implemented with the objective of calculating the number of stands necessary in order to absorb the arrival/departure traffic. Other authors proposed models to avoid congestion on the ground using pushback control strategy (Pujet et al. 1999, Simaiakis and Balakrishnan 2014, Khadilkar and Balakrishnan 2014). For the terminal operations, the work of Mujica (2015a) presents the allocation of desks using simulation and optimization techniques.

Concerning the airspace, in the specific Terminal Maneuvering Area (TMA), the main findings are related to the sequencing and merging of aircraft flow problems, while other works face the problem of scheduling arrival aircraft. The main objectives are to ensure separation minima and avoid conflicts between aircraft (Michelin et al. 2009, Zuniga et al. 2011-2013), and optimizing the sequence of aircraft in order to decrease delays (Beasley et al. 2000-2001-2004, Balakrishnan et al. 2010, Murca et al. 2015). Recently we have found some works that employ discrete-event simulation approaches regarding airport capacity issues.

For example Mujica et al. (2014-2015b) put focus on the ground operations and also Scala et al. (2015) on airspace operations aiming at evaluating the performance of the systems in terms of capacity.

The links of airspace with ground operations are more evident when the ground is congested, then the airspace should to be capable of absorb this congestion accommodating as much aircraft as possible in order to avoid disruptions of the service. If the airspace has not the capability of absorbing the congestion occurred on the ground, a chain effect is triggered, with delays imposed by air traffic control to the aircraft in the airspace (e.g holding pattern procedures, alternative routes, speed control). On the contrary, delays occurred to aircraft in the airspace (e.g. due to weather conditions) are translated on the ground as well. This fact drove the study presented in this paper taking a holistic perspective, therefore, in this work an integrated model of airspace (Terminal Manoeuvring Area TMA) and ground operations was developed. The scope of the work was to evaluate the performance of the entire system and identify the main factors that affect it. To do that, it was developed a simulation model employing a modular approach in which four different models were integrated to obtain a representation of the entire system (Airspace + Ground side). It was used a Discrete Event Simulation (DES) (Banks et al. 2010) which suits for modeling the operations of a wide range of systems.

The paper is organized as follows, in section 2 is presented the modular approach employed in order to obtain an integrated simulation model, describing the characteristics of the sub-models and the way in which they were combined. In section 3 the different scenarios and the results obtained from the model were presented, and in section 4 some conclusions were made.

2. DEVELOPMENT OF THE SIMULATION MODEL

In this work, the simulation model of Lelystad airport was developed. Lelystad airport is a regional airport of the Netherlands, currently, it accommodates only general aviation traffic, but in the future it will accommodate commercial traffic (Alders 2014). This work takes into account not only ground operations but also airspace (TMA) operations, obtaining an integrated perspective of the airport system.

2.1. Integrated model

This section explain how a modular approach was applied in order to develop an integrated model, including four different sub-models. This approach allows to analyze the airport system as a whole. In line with a modular approach, starting from the sub-models, each one representing a part of the airport system, the integration of them has made possible to come up with a unique and integrated model that keeps the characteristics of the sub-models and gives as an output a more accurate and realistic view of the system. If we want to analyze the entire system, an integrated approach results better than analyzing separately the

sub-models. In real cases, the operation represented in the sub-models, interact each other affecting the performance of the entire system. As mentioned in the introduction of this paper, all the studies related to the improvement of efficiency of the airport as a system were done separating airspace from ground operations, thus, this work results to be a novel approach to address that issue. Due to the modular approach, it was possible to have a larger view of the system, and also to be able to carry on a more accurate analysis obtaining more significant results. In order to develop the integrated model was used a discrete-event simulation software called SIMIO, it is a general purpose discrete-event simulation software that allows modelling a wide range of systems. The strength of this simulation software is that it allows to load different models as libraries, and in turn, run them separately or run them together if necessary. When the different model are combined together, some entities are shared, and then these entities undergo the logic of each of the sub-models. Figure 1 represents the interaction between the different sub-models.

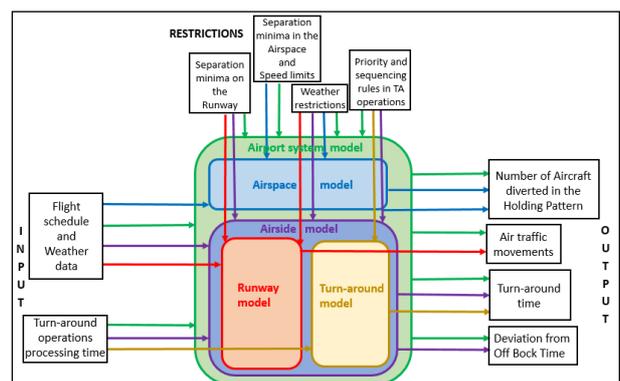


Figure 1: Representation of the interactions between the different sub-models in the integrated model

In the integrated model, the entity “aircraft” starts from the Airspace sub-model, where the logic of that specific model is applied such as separation minima between aircraft, speed limits, effect of the wind and the holding pattern procedure. Then it continues through the Airside sub-model where the “aircraft” entity lands, complete the taxiing-in, turnaround, taxiing-out and take off operations following the logic implemented in that model. In this case the airside model is already an integration of two models, Runway and Turnaround models, therefore, here the entity implicitly follows the logic implemented in these two models. Finally, after the “aircraft” crosses the runway for taking off, it enters again in the airspace model following the restrictions given by the logic implemented into the Airspace model until it reaches the last point of the model. Tables 1 and 2 show the characteristics of the system and the main assumptions taken into account.

Table 1: Characteristics of the Airport system

Number of Runways	1
Number of exitways	1
Taxiway type	Parallel
Number of Stands	16
Holding Pattern characteristics (number, capacity and speed limit)	1 (for each route), 2 Aircraft, 200kt
Mix of Aircraft	Code C (B737 – A320)

Table 2: Main Assumptions

Flight Schedule	Built taking into account traffic at Amsterdam Schiphol Airport in the peak hours
Airspace Routes	One route for each of the Runway configuration
Noise	No noise issues were taken into account

The main outcomes of the integrated model include the outcomes of the different sub-models, the difference is that from the integrated model can be obtained more accurate values since all the components of the system are taken into account simultaneously. At first glance, the main improvement obtained by the development of an integrated system, is the analysis of the delay, since both airspace and ground delays are taken into account. The evaluation of ground operations performance can be done with more accuracy when the airspace is also taken into account, for instance, when aircraft are diverted into the holding pattern, they will land later than the expected causing new scenarios on the ground side.

2.2. Sub-models

Each of the sub-models included in the integrated model, represent the following operations:

- Runway operations (first sub-model):

Operations related to runway, therefore, runway occupancy, mix of movements (landings and takeoffs handled) (Mujica et al. 2014).

- Turnaround operations (second sub-model):

Operations related to the turnaround process, therefore, boarding/deboarding of passengers, refueling, water service, catering service and baggage in/out. All these operations are made when the aircraft is parked at the stand (apron).

- Airside operations (third sub-model):

Operations related to the airport airside, therefore, landing process, taxiing processes, turnaround process and departing process. (Mujica et al. 2015b)

- Airspace operations (fourth sub-model):

Operations related to the airspace, in the specific the Terminal Manoeuvring Area (TMA). The main logic applied to the TMA are related to the operations made by aircraft during their approach and departure to and from the runway, they are: separation between aircraft; speed limit; holding pattern procedure; change of runway configuration; effect of wind direction and crosswind (Scala et al. 2015).

In the following paragraphs all the sub-models that were used to build the integrated model are presented, providing a description of their main characteristics and their main outcomes.

2.2.1. Runway operations

In this model all the operations related to the runway system of the airport (Lelystad airport) were represented. The infrastructure taken into account, includes the runway and the exitways. The main logic implemented were the separation between arrival-arrival, arrival-departure, departure-departure, departure-arrival aircraft on the runway; priority between arrivals and departures; weather conditions (crosswind and visibility); operational time. Different scenarios were tested based on different flight schedules (different volume of traffic) and different configuration of exitways (normal exitways or high-speed exitways) (see figure 2-3), the most relevant results obtained were about number of air traffic movements and runway usage among others. The model represented an early stage of the development of the entire airport system, from this model it was possible to obtain realistic values about runway occupancy time and runway usage, accordingly, instead values about number of air traffic movements needed a more accurate model with more restrictions to take into account such as the influence of the taxiway network, turnaround operations and airspace sectors. Results from the model pointed out the difference between the scenarios in terms of runway occupancy time, depending on the two different runway configurations, leading to have a more efficient use of the runway.

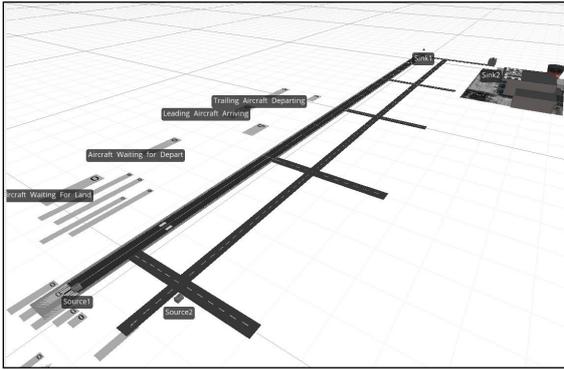


Figure 2: Runway Configuration with normal exitways

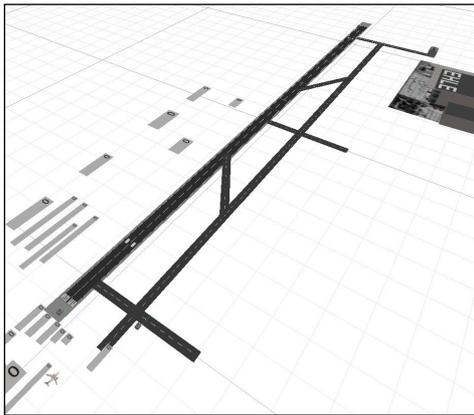


Figure 3: Runway Configuration with high-speed exitways

2.2.2. Turnaround operations

In this model all the operations made while the aircraft is parked at the stand were represented. The main operations are boarding and de-boarding of passengers, loading and unloading baggage, re-fueling, water service, cleanings and catering. Every operation needs a vehicle to be used, therefore we have these different types of vehicles in use: fuel truck, bus for passengers, trucks for baggage loading/unloading, catering service, cleaning service and water service. The purpose of the model was to simulate the total turnaround time, assuming that every operations spend a certain amount of time, and exist some sequencing rules between the operations to be made (e.g. boarding operations should be done after deboarding operations and fuel operations). Values about time were represented by random variables and data was gathered from the main manufacturers of aircraft for specific types of aircraft (Boeing 737, Airbus A-320) (see table 3). Depending on constraints about priority and sequencing rules within the turnaround operations, it was possible to identify the operations that most affect turnaround time. The logic implemented in this model make it general enough to be adapted for every airport type and for every aircraft type. (see figure 4)

Table 3: Times for Turn around operations

Operation	Distribution	Time
Positioning stairs	Random triangular	90,120,150 (sec)
De-boarding	Random triangular	3,4,5 (min)
Luggage out	Random triangular	5,7,11 (min)
positioning	Random triangular	40,60,80.9 (sec)
Luggage in	Random triangular	5,7,9 (min)
positioning	Random triangular	40,60,80.9 (sec)
Fueling	Random triangular	7,8,9 (min)
positioning	Random triangular	4,5,9 (min)
Cleaning	Random triangular	8,13,16 (min)
positioning	Random triangular	1,2,3 (min)
Water service	Random triangular	4,5,6 (min)
positioning	Random triangular	1,2,3 (min)
Boarding	Random triangular	4,5,6 (min)
Headcount	Random triangular	90,120,130 (sec)

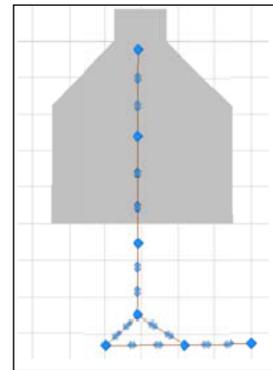


Figure 4: Turnaround operations model

2.2.3. Airside operations

In this model, the previous two sub-models (Runway operations, Turnaround operations) were combined. The result is a model that represent the airport airside, including runway operations (landings and take offs); taxiing operations; and turnaround operations. The Runway model was linked to the stands by the taxiway network, in this case we have 16 stands, in other words, the turnaround model was replicated 16 times. All the logic concerning the runway model and turnaround model were kept. In the taxiway network, the logic about speed regulation and path choice was implemented in order to let the aircraft cross the taxiway at the right speed and to go to and from the assigned stand. Once aircraft are landed, they occupy the first stand available from left to the right. The logic implemented make sure that aircraft do not have conflicts during taxiing operations, therefore, it is likely to have aircraft waiting in queues in some parts of the taxiway network, especially when there is high volume of traffic.

Different scenarios were based on incoming volume of traffic, different layout of taxiways geometry (see figures 5 to 7) and different number of vehicles involved in the turnaround operations. These scenarios were based on the characteristics inherited from each

one of the two models included in this model, in the specific, different layout of taxiways for what concern the first sub-model (runway and taxiways), and number of vehicles for what concern the second sub-model (turn-around).



Figure 5: Configuration L-shaped stand with partial parallel taxiway



Figure 6: Configuration linear stand with parallel taxiway



Figure 7: Configuration taxi-in taxi-out concept

The main results from this model are related to taxiing times, turnaround time and delay at the stand due to conflict between vehicles on the taxiway (deviation from off block time-DOBT). The integration between these two sub-models shows that the system is affected by the interactions between them, for instance, traffic congestion could be seen on the taxiway due to unavailability of stands, and this could be a consequence of high turnaround times, that is, in turn, a consequence of the unavailability of vehicle for turnaround operations. This make us able to conduct a more accurate analysis of the system, identifying possible interactions between components of the system and having a larger view of the entire system.

2.2.4. Airspace operations

In this model, the focus was put on the airspace of Lelystad airport, in the specific the TMA. The TMA is the airspace that surrounds an airport, it could be a very congested zone because of the volume of incoming and outgoing aircraft converging on the runway(s). In the TMA, aircraft should fly under some restrictions such as:

- speed
- altitude limit
- separation minima between aircraft
- weather (crosswind and wind direction)

In the model representing the TMA the aircraft approaching and departing phase to and from the runway was modeled. Another component that characterize this model is the holding pattern procedure, the holding pattern is an area in the airspace used to divert temporarily aircraft that needs to gain a delay due to congestion on the ground or congestion along the route in the airspace or due to disruptions (crosswind occurrences). The holding pattern has its own capacity, depending on the airspace sector in which it is placed. Moreover, aircraft into the holding pattern should fly at a certain speed, a turn in the holding pattern is assumed to take around four minutes. The airport taken into account as case of study has one runway that can be used in both directions, having, in turn, Runway 23 and Runway 05, therefore, we have two different routes for approach and take off to and from runway 23 and runway 05, respectively. The use of these specific runway configurations depends on the wind direction. The routes in the TMA, including the holding pattern, were modeled as a network, in figure 8 it can be seen where the main nodes of the network were placed (entry points, initial approach fix-IAF, final approach fix-FAF, exit points) and the topology of the routes. In this network, separation minima between the aircraft and speed limit was checked in each of the nodes, in tables 4 and 5 values for separation minima and speed limits are shown. Separation minima between aircraft (time based) was fulfilled controlling aircraft speed, accelerating and decelerating in turn. Moreover, the effect of the wind was taken into account (direction and crosswind), it was modeled referring to data of wind direction and crosswind collected in the 2014 and then translated into the model as probabilities of occurrences.

Table 4: Separation minima (NM) ICAO

		Leading Aircraft		
		Heavy	Medium	Light
Trailing Aircraft	Heavy	4	3	3
	Medium	5	3	3
	Light	6	4	3

Table 5: Aircraft speed range in the TMA

	Upper bound	Lower Bound
Entry Point	250 Kt	160 Kt
Initial Approach Fix	160 Kt	130 Kt
Final Approach Fix	-	130 Kt

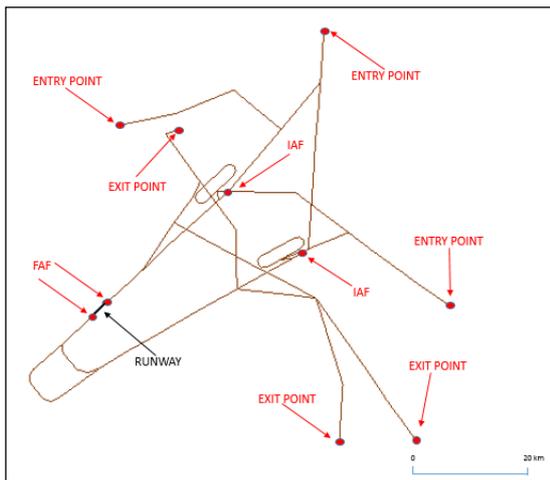


Figure 8: Network representing the routes for Runway 23 and 05

The different scenarios that were tested, were based on different volume of incoming aircraft, and the main results were about number of air traffic movements and number of aircraft delayed into the holding pattern, that are a measure of capacity and level of congestion of the system, respectively. Particularly, looking at the aircraft diverted into the holding pattern it was possible to derive the amount of delay occurred to them by keeping track of the time in which they were flying into the holding pattern.

3. SCENARIO AND RESULTS

The model developed was verified using the public information about the developments that will be implemented. Since it is a future project we did not have historic data to compare to. Therefore the verification was done comparing the obtained values versus the values of a similar airport of the region (Eindhoven). In order to evaluate the performance of the model and make a further analysis, some scenarios were tested. They were based on volume of incoming traffic and number of vehicles available in the turnaround process. Changing the amount of traffic as input we aimed at identify what was the maximum capacity that the system was able to handle without incurring in congestion. In this case congestion is represented by delay occurred in the airspace (delay caused by holding pattern procedure), and delay occurred on the ground (delay due to conflict between aircraft and vehicles in the taxiway). From the model it was possible to obtain results related to number of incoming and outgoing number of aircraft, number of aircraft diverted into the holding pattern (HP), delay due to airspace operations, turnaround time and ground delay due to taxiway congestion.

3.1. Experimental Scenarios

Scenarios were built taking into account two parameters: incoming volume of traffic and number of vehicles involved in the turnaround process. Regarding volume of incoming traffic, three different configurations were tested, in which the traffic was gradually increased. The second parameter used to build the different scenarios was number of vehicles involved in the turnaround process. Two different values were taken into account, a configuration with 8 vehicles available for each type of operation and another scenario with 16 vehicle available. It was simulated one day of operations and were made 10 replications for each scenario, in table 6 are presented the different scenarios that were tested.

Table 6: Scenarios

		Incoming volume of Aircraft		
		Flight schedule 1	Flight schedule 2	Flight schedule 3
Number of Vehicles	8	Scenario 1	Scenario 3	Scenario 5
	16	Scenario 2	Scenario 4	Scenario 6

3.2. Results

The following tables summarize the results from the different scenarios:

Table 7: Results from Scenario 1

Flight Schedule 1 – Number of Vehicles=8			
	Min	Avg	Max
Incoming aircraft Runway 23	21	32,7	48
Incoming aircraft Runway 05	9	17	27
Outgoing aircraft Runway 23	21	33,1	48
Outgoing aircraft Runway 05	8	16,3	26
Aircraft diverted into the HP (05)	1	1	1
Average number of turns made by each aircraft diverted into the HP (05)	1	1	1
Time spent into the HP (05) (min)	4	4	4
Total Time at the stand (avg) (min)	33,57	40,19	52,18
Turnaround Time (avg) (min)	30,54	32,39	35,43
DOBT (avg) (min)	2,06	7,39	19,01

Table 8: Results from Scenario 2

Flight Schedule 1 – Number of Vehicles=16			
	Min	Avg	Max
Incoming aircraft Runway 23	9	27,2	39
Incoming aircraft Runway 05	9	20,8	39
Outgoing aircraft Runway 23	10	28,5	40
Outgoing aircraft Runway 05	8	19,2	38
Aircraft diverted into the HP (05)	1	1	1
Average number of turns	1	1	1

made by each aircraft diverted into the HP (05)			
Time spent into the HP (05) (min)	4	4	4
Total Time at the stand (avg) (min)	33,39	39,15	53,21
Turnaround Time (avg) (min)	31,04	32,25	35,50
DOBT (avg) (min)	2,01	6,49	18,54

Looking at the values of the first two scenarios, it can be seen that the flow of traffic is smooth since there are no aircraft diverted into the holding pattern for runway 23 and there is only 1 aircraft diverted for runway 05. Also turnaround times and DOBT are similar, this means that at this stage, with the given amount of incoming traffic, this output is not affected by the number of vehicles. These values suggest that the system is able to handle this amount of traffic without incurring in congestion situation.

Table 9: Results from Scenario 3

Flight Schedule 2 – Number of Vehicles=8			
	Min	Avg	Max
Incoming aircraft Runway 23	29	47,8	67
Incoming aircraft Runway 05	8	21,33	38
Outgoing aircraft Runway 23	27	46	67
Outgoing aircraft Runway 05	12	21,11	39
Aircraft diverted into the HP (05)	1	1,4	3
Average number of turns made by each aircraft diverted into the HP (05)	1	1,2	2
Time spent into the HP (05) (min)	4	4,48	8
Total Time at the stand (avg) (min)	34,08	46,51	97,38
Turnaround Time (avg) (min)	31,16	32,34	34,19
DOBT (avg) (min)	2,20	13,55	64,09

Table 10: Results from Scenario 4

Flight Schedule 2 – Number of Vehicles=16			
	Min	Avg	Max
Incoming aircraft Runway 23	25	44	59
Incoming aircraft Runway 05	8	23	42
Outgoing aircraft Runway 23	24	43,4	55
Outgoing aircraft Runway 05	11	22,5	40
Aircraft diverted into the HP (23)	1	1	1
Aircraft diverted into the HP (05)	1	2	5
Average number of turns made by each aircraft diverted into the HP (23)	2	2	2
Average number of turns made by each aircraft diverted into the HP (05)	1	1,33	2
Time spent into the HP (23) (min)	8	8	8
Time spent into the HP (05) (min)	4	5,19	8
Total Time at the stand (avg) (min)	32,58	50,21	95,57

Turnaround Time (avg) (min)	30,41	31,49	33,56
DOBT (avg) (min)	2,06	16,40	55,54

In tables 9 and 10 there are the values related to the third and fourth scenario, they show that the number of aircraft diverted into the holding pattern is slightly increased together with the average number of turns made by each aircraft into the holding pattern. The latter, in turn, is translated into delay at landing. Turnaround times are in line with the previous scenarios but values of DOBT are increased, we have in average 13,55 min and 16,40 min for scenario 3 and scenario 4 , respectively.

Table 11: Results from Scenario 5

Flight Schedule 3 – Number of Vehicles=8			
	Min	Avg	Max
Incoming aircraft Runway 23	38	62,20	90
Incoming aircraft Runway 05	16	26,33	51
Outgoing aircraft Runway 23	38	64,4	87
Outgoing aircraft Runway 05	15	24,55	48
Aircraft diverted into the HP (23)	1	1	1
Aircraft diverted into the HP (05)	1	2,25	6
Average number of turns made by each aircraft diverted into the HP (23)	2	2	2
Average number of turns made by each aircraft diverted into the HP (05)	1	1,48	2
Time spent into the HP (23) (min)	8	8	8
Time spent into the Holding pattern (05) (min)	4	5,55	8
Total Time at the stand (avg) (min)	33,37	37,30	45,43
Turnaround Time (avg) (min)	30,33	31,56	33,43
DOBT (avg) (min)	2,34	5,33	13,01

Table 12: Results from Scenario 6

Flight Schedule 3 – Number of Vehicles=16			
	Min	Avg	Max
Incoming aircraft Runway 23	27	46,4	80
Incoming aircraft Runway 05	3	30,4	56
Outgoing aircraft Runway 23	21	44,7	83
Outgoing aircraft Runway 05	1	27,8	52
Aircraft diverted into the HP (23)	1	1	1
Aircraft diverted into the HP (05)	1	2,28	5
Average number of turns made by each aircraft diverted into the HP (23)	1	1,35	2
Average number of turns made by each aircraft diverted into the HP (05)	2	2	2
Time spent into the HP (23) (min)	4	5,24	8
Time spent into the HP (05) (min)	8	8	8
Total Time at the stand (avg)	112,33	117,1	127,40

(min)			
Turnaround Time (avg) (min)	30,11	31,19	32,52
DOBT (avg) (min)	81,55	85,58	96,12

In the last two scenarios presented in tables 11 and 12, under a high traffic inbound, we find again, for both runway 23 and 05, aircraft diverted into the holding pattern. Their amount is similar to the previous scenarios and this means that the system is not affected by the increase of traffic in terms of congestion. A particular focus is put in the value of DOBT for the sixth scenario that is very high unlike the other scenarios. It means that taxiway is over-crowded by vehicles and aircraft at the same time, and so aircraft are stopped at the gate for a long time before reaching the runway while are waiting that the taxiway will be free from other vehicles and aircraft. This fact leads to think about changing the design of the taxiway in order to better manage the traffic in it and avoid blockages.

4. CONCLUSION

In this paper, a simulation model that employs a modular approach is presented. The purpose of this work was to analyze an airport system using an integral approach. The model developed was done using a bottom-up modular approach in which several subsystems were modeled independently and then put together in a model that integrate the different ones airport (Runway, Stand, Airspace). From the results it is possible to verify that this approach suits well for performing data driven decisions since the information extracted from the model allows identifying potential problems in the system once the traffic has reached the levels that saturate the system. In addition, the model-based analysis allows going further into the identification of cause-effect relationships that affect the performance of the system at different levels. Furthermore, the approach allows identifying the subsystem that affect the most the overall performance thus putting more attention in the particular problems that hinder the flow of aircraft.

ACKNOWLEDGMENTS

The authors would like to thank the Aviation Academy of the Amsterdam University of Applied Sciences for the support to perform this study.

REFERENCES

Balakrishnan H., Chandran B. 2010. Algorithms for scheduling runway operations under constrained position shifting. *Operation Research*, 58 (6), 1650-1665.

Banks, J., Carson J.S., Nelson, B., Nicol, D.M., 2010. *Discrete-Event system Simulation*. 5th ed. Upper Saddle River, NJ: Pearson.

Beasley J.E., Krishnamoorthy M., Sharaiha Y.M., Abramson D., 2000. Scheduling aircraft landings—the static case *Transportation Science*, 34 (2), 180-197.

Beasley J.E., Sonander J., Havelock P., 2001. Scheduling aircraft landings at London Heathrow using a population heuristic. *Journal of the Operational Research Society*, 55, 483-493.

Beasley J. E., Krishnamoorthy M., Sharaiha Y. M., Abramson D., 2004. Displacement problem and dynamically scheduling aircraft landings. *Journal of the Operational Research Society*, 55, 54-64.

Bolat A. 2000. Procedures for providing robust gate assignments for arriving aircrafts. *European Journal of Operational Research*, 120 (1), 63-80.

De Neufville R., Odoni, A., 2003. *Airport Systems: Planning, Design and Management*. New York: Mc-Graw-Hill

Dorndorf U., Drexlb A., Nikulin Y., Pesch E., 2007. Flight gate scheduling: State-of-the-art and recent developments. *Omega*, 35 (3), 326-334.

Janic M., 2001. *Analysis and Modeling of Air Transport System: Capacity, Quality of Service and Economics*. London: Taylor and Francis

Khadilkar H., Balakrishnan H., 2014. Network Congestion Control of Airport Surface Operations. *AIAA Journal of Guidance, Control and Dynamics*, 7 (3), 933-940.

Michelin A., Idan M., Speyer J.L., 2009. Merging of air traffic flows. *AIAA Guidance, Navigation, and Control Conference*, August 10-13 August, Chicago (Illinois, USA).

Mujica M., 2015a. Check-In allocation improvements through the use of a Simulation-Optimization Approach, *Transportation Research Part A*, 77, 320-335.

Mujica M., de Bock N., Boosten G., Jimenez E., Pinho J., 2015b. Simulation-Based Turnaround Evaluation for Lelystad Airport. *Air Transport Research Society World Conference*. July 2-5, Singapore (Singapore).

Mujica M. M, Scala P., Boosten G., 2014. Simulation-based capacity analysis for a future airport. *Computer Aided System Engineering (APCASE)*, 2014 Asia-Pacific Conference, pp. 97-101. February 10-12, Bali (Indonesia).

Murca M.C.R., Muller C., 2015. Control-based optimization approach for aircraft scheduling in a terminal area with alternative arrival routes. *Transportation Research Part E*, 73 (C), 96-113.

Narciso M.E., Piera M.A., 2015. Robust gate assignment procedures from an airport management perspective. *Omega*, 50, 82-95.

Pujet N., Delcaire B., Feron E., 1999. Input-output modeling and control of the departure process of congested airports. A collection of technical papers: *AIAA Guidance, Navigation and Control Conference and Exhibit*, Part 3, pp. 1835-1852. August 9-11, Portland (Oregon, USA).

Scala P., Mujica M., Zuniga C.A., 2015. Assessing the future TMA Capacity of Lelystad Airport using Simulation. *Air Transport and Operations Symposium*. July 20-22, Delft (The Netherlands).

- Schiphol Group, 2014. Ondernemingsplan Lelystad Airport. Available from: <http://www.schiphol.nl/InDeSamenleving/ToekomstSchiphol/ActueleProjecten1/OndernemingsplanLelystadAirport1.htm> [Accessed 5 April 2015].
- Simaiakis I., Balakrishnan H., 2014. A Queuing Model of the Airport Departure Process. *Transportation Science*, accepted, July 2014.
- Zuniga C., Delahaye D., Piera M., 2011. Integrating and Sequencing Flows in Terminal Maneuvering Area by Evolutionary Algorithms. *DASC 2011, 30th IEEE/AIAA Digital Avionics Systems Conference*, pp. 1-32, Seattle (Washington, USA).
- Zuniga C.A., Piera M.A., Ruiz S., Del Pozo I., 2013. CD&CR casual model based on path shortening/path stretching techniques. *Transportation Research Part C*, 33, 238-256.

COMPARISON OF OPTIMIZATION METHODS EFFICIENCY IN DIFFERENT DIMENSIONS OF THE SEARCH SPACE OF SIMULATION MODELS

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

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

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

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

ABSTRACT

The paper deals with the comparison of selected optimization methods - Random Search, Hill Climbing, Tabu Search, Local Search, Downhill Simplex, Simulated Annealing, Differential Evolution, SOMA and Evolution Strategy – efficiency used for optimization of four selected testing functions. We tested these optimization methods on different dimensions of the search space to compare their efficiency of finding the global optimum in the search space.

Keywords: simulation optimization, testing function, dimension of the search space

1. INTRODUCTION

Simulation optimization is one possible solution for solving an NP-hard problem. The simulation model is the modelled problem. The problem is to find suitable settings for the simulation model input parameters (input variables):

$$\mathbf{X} = [x_j] \forall j : j = \{0, 1, 2, \dots, n-1\} \quad (1)$$

Where \mathbf{X} denotes the concrete setting of the input parameters; x_j denotes the value of the j -th simulation model input parameter (first simulation model input parameter is indexed by 0 because the first item in the list is commonly labelled by 0 in programming language); n denotes the number of the simulation model input parameters – dimension of the search space. Each possible solution – candidate solution (the concrete setting of the input parameters \mathbf{X}) is a representation of the element of the search space \tilde{X} . The dimension of the search n space equals the number of simulation model input parameters. The search space in the case of Box Constraint can be formulated as follows:

$$\tilde{X} = \prod_{j=1}^{n-1} [a_j, b_j] \forall j : j = \{0, 1, 2, \dots, n-1\} \quad (2)$$

Where a_j denotes the lower bound of the j -th simulation model input parameter; b_j denotes the lower boundary of the j -th simulation model input parameter. We should also say that we obtain one or more simulation model outputs after the simulation run (the result of simulation experiment with a concrete setting) which are inputs of the objective function. The candidate solution in the search space can be evaluated by the objective function value $F(\mathbf{X})$ which represents the quality of candidate solution regarding the specified objective. It is clear that if the number of simulation model input increases, the search space contains a large number of possible solutions.

The basic problem of finding an optimal/suboptimal feasible solution (respecting the defined model constraints) can be formulated as follows:

$$\tilde{\mathbf{X}} = \arg \min_{\mathbf{X} \in \tilde{X}} F(\mathbf{X}) = \{ \tilde{\mathbf{X}} \in \tilde{X} : F(\tilde{\mathbf{X}}) \leq F(\mathbf{X}) \forall \mathbf{X} \in \tilde{X} \} \quad (3)$$

where $\tilde{\mathbf{X}}$ denotes the global minimum of the objective function; $F(\mathbf{X})$ denotes the objective function value of the candidate solution – the range commonly includes real numbers; \tilde{X} denotes the Search space. (Raska & Ulrych, 2013)

The next figure (Figure 1) shows the possible optimization process if the objective function is minimized. We can see that the objective function has a local and global optima if the objective function is maximized – problem of premature convergence. Premature convergence – the global optimization process can converge prematurely to a local optimum because there is no opportunity to examine other areas of a space of possible solutions (currently, only a particular area is examined). Another area of search space exists (which is not currently known) that contains a better solution than the currently known solution. (Weise, 2009). Other basic problems of the optimization methods used in global optimization are (Rockwell Automation, 2014): The whole search space cannot be examined (testing all possible solutions) because of large demands on computer memory, or time

spent on examining the space – nondeterministic polynomial problems; the landscape of objective function – multimodal objective function – premature convergence, objective function smoothness landscape etc.; multi-objective optimization; identification of a suitable method for handling the constraints; specification of appropriate termination criteria; setting the parameters of the optimization method, etc.

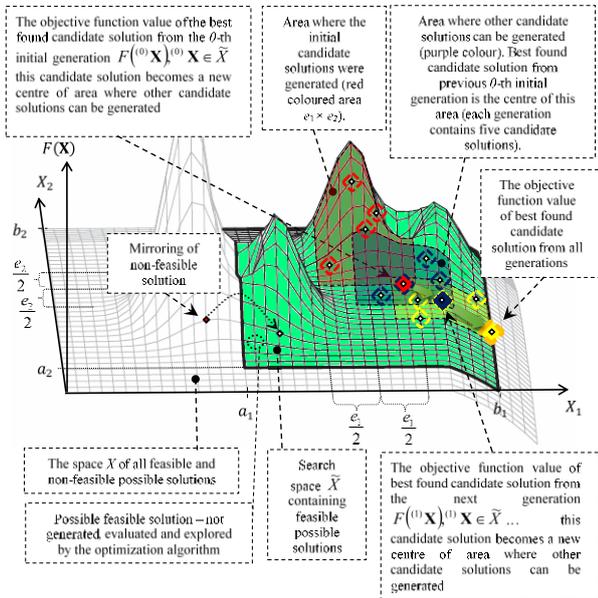


Figure 1: Possible Optimization Process - Objective Function Minimization

The whole space X contains feasible and non-feasible possible solutions when some constraints are specified e.g. Box Constraints. The green area denotes the feasible possible solutions (axes in the search space are indexed from index 1, because it is not common to mark first axes by the index 0. Indexing with index 0 will be used in mathematical notation or in the algorithm). If this candidate solution is generated by the optimization method then it is accepted and evaluated by the objective function value regarding the simulation model output. There are different variants for generating candidate solutions: neighbourhood relation which allows us to define the set of feasible candidate solutions neighbouring a selected candidate solution. Condition of availability is important for a definition of neighbourhood relations, which requires that every feasible solution has to be reached from any other feasible solution by sequential application of neighbourhood relation (Stefka, 2005), generating a random candidate solution using different distribution, genetic operators – e.g. mutation, crossover (Mitchell, 1996; Miranda, 2008; Hynek, 2008), etc.

The process of generating new initial candidate solutions in the specified (red) area is shown in Figure 1.

Optimization methods usually generate more than one candidate solution using the iterations in cycle. These candidate solutions are sequentially placed in the list in

the order which they were generated (in the context of evolutionary algorithms this list is represented by the population - generation). This candidate solution (individual) in the population can be formulated as follows:

$$\mathbf{X}_i = L[i] \forall i : i = \{0, 1, 2, \dots, m-1\} \quad (4)$$

Where \mathbf{X}_i denotes i -th candidate solution – individual; i denotes the index of candidate solution; L denotes the list of candidate solutions - population; m denotes the number of generated candidate solutions - the population size.

The optimization method selects the best candidate solution from this generation (population) regarding the objective function values of candidate solutions. This best candidate solution from the 0 -th initial generation becomes the centre of area where other candidate solutions can be generated. These processes are repeated until some of the termination criteria are satisfied. If the termination criterion is met the optimization method returns one or more best found solutions.

Optimization method can generate the non-feasible solution. Optimization method can use the repair algorithms and special operators (Hynek, 2008). Mirroring of non-feasible solution into a feasible region is shown in the next figure. The non-feasible solution is flipped into the feasible region around the respective edge of the search space at a distance of this non-feasible solution to the edge of the search space (Tvrdik, 2010).

2. SELECTED OPTIMIZATION METHODS

We have transformed some of the selected optimization methods to use the principle of evolutionary algorithms. Different variants of selected optimization methods obtained from a literature review were united into the algorithm.

2.1. Random Search

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

2.2. Downhill Simplex

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

2.3. Stochastic Hill Climbing

Candidate solutions are generated (populated) in the neighbourhood of the best candidate solution from the previous population. Generating new possible solutions

is performed by mutation. This method belongs to the family of local search methods.

2.4. Stochastic Tabu Search

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

2.5. Stochastic Simulated Annealing

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

2.6. Stochastic Local Search

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

2.7. Evolution Strategy

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

2.8. Differential Evolution

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

2.9. SOMA

SOMA is based on the self-organizing behaviour of groups of individuals in a “social environment”. It can also be classified as an evolutionary algorithm, despite the fact that no new generations of individuals are created during the search. Only the positions of the individuals in the search space are changed during a generation, called a “migration loop”. Individuals are generated at random according to what is called the “specimen of the individual” principle. The specimen is in a vector, which comprises an exact definition of all these parameters that together led to the creation of such individuals, including the appropriate constraints of the given parameters. SOMA is not based on the philosophy of evolution (two parents create one new individual – the offspring), but on behaviour of a social group of individuals. (Zelinka, 2004)

3. TESTING FUNCTIONS

Considering the time requirements of testing the behaviour of optimization methods (according to different settings) (Raska & Ulrych, 2015) we substitute the testing on the simulation models (and its objective function) by a different testing function. Implemented optimization methods were tested on four standard testing functions - domain of the function is a defined step for each axis – substitution of the simulation model input parameter (discrete) values of the discrete event simulation model. All testing functions were minimized.

3.1. De Jong’s Function

A convex and unimodal testing function. The function definition: (Pohlheim, 2006)

$$F(\mathbf{X}) = \sum_{j=1}^n x_j^2, \quad (5)$$

$$\forall x_j : \left(x_j - \left\lfloor \frac{x_j}{0.01} \right\rfloor \cdot 0.01 = 0 \right) \wedge (-30 \leq x_j \leq 30),$$

$$j = 1 : n, n \in [2, 10, 20, 30, 44]$$

where $F(\mathbf{X})$ denotes the objective function; j denotes index of control; n denotes the dimension of the search space - the dimension of the search space is 2; 10; 20; 30; 44; x_j denotes the value of control - testing functions (except Michalewicz function) input parameters values are in the range from -30 (lower boundary) to 30 (higher boundary). We substitute the testing on the simulation models by testing on the testing function hence we defined the smallest step that can be performed by the optimization methods is 0.01 for each axis in the search space ($x_j \bmod 0.01 = 0$). The input parameters are not continuous. This resolution represents 36,012,001 possible solutions (combinations of testing function input parameters) in a two dimensional search space and the same resolution represents 1.7452×10^{166} possible solutions in a forty-four dimensional search space. We chose a forty-four dimensional search space regarding the tested discrete event simulation model of automated

guided vehicle transportation with forty-four simulation model input parameters to see the optimization methods' efficiency. To achieve a better idea of the testing functions landscapes the continuous testing functions are shown in the following four figures. De Jong's continuous function is shown in Figure 2.

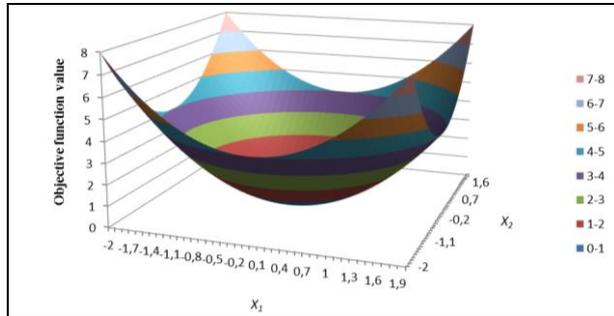


Figure 2: De Jong's Function

3.2. Rosenbrock's Function

Rosenbrock's (Rosenbrock's valley, Rosenbrock's banana) function is unimodal and non-convex testing function. The function definition: (Pohlheim, 2006)

$$F(\mathbf{X}) = \sum_{j=1}^{n-1} 100 \cdot (x_j^2 - x_{j+1})^2 + (1 - x_j)^2, \quad (6)$$

$$\forall x_j : \left(x_j - \left\lfloor \frac{x_j}{0.01} \right\rfloor \cdot 0.01 = 0 \right) \wedge (-30 \leq x_j \leq 30),$$

$$j = 1 : n, n \in [2, 10, 20, 30, 44]$$

The Rosenbrock's continuous function is shown in Figure 3.

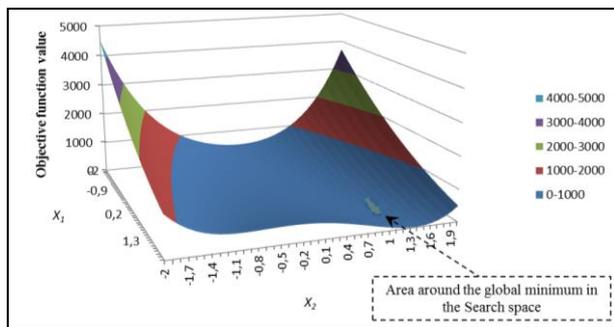


Figure 3: Rosenbrock's Function

3.3. Michalewicz Function

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

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

$$\forall x_j : \left(x_j - \left\lfloor \frac{x_j}{0.01} \right\rfloor \cdot 0.01 = 0 \right) \wedge (0 \leq x_j \leq \pi), \quad (7)$$

$$j = 1 : n, n \in [2, 10, 20, 30, 44]$$

We selected $m=5$ in our simulation model. The Michalewicz continuous function is shown in Figure 4.

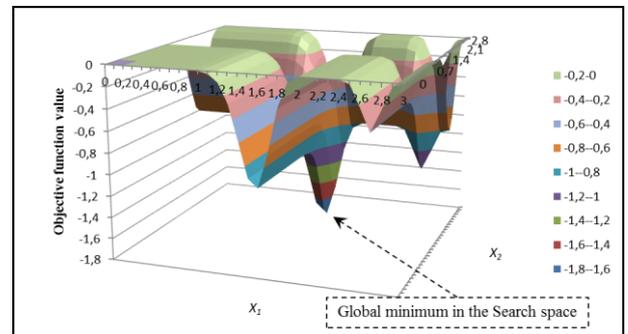


Figure 4: Michalewicz Function

3.4. Ackley's Functions

Ackley's function is a multimodal test function. This function is a widely used testing function for premature convergence. (Tvrđik, 2010)

$$F(\mathbf{X}) = -20 \cdot \exp\left(-0.02 \cdot \sqrt{\frac{1}{n} \cdot \sum_{j=1}^n x_j^2}\right) - \exp\left(\frac{1}{n} \cdot \sum_{j=1}^n \cos(2 \cdot \pi \cdot x_j)\right) + 20 + \exp(1), \quad (8)$$

$$\forall x_j : \left(x_j - \left\lfloor \frac{x_j}{0.01} \right\rfloor \cdot 0.01 = 0 \right) \wedge (-30 \leq x_j \leq 30),$$

$$j = 1 : n, n \in [2, 10, 20, 30, 44]$$

The Ackley's continuous function is shown in Figure 5.

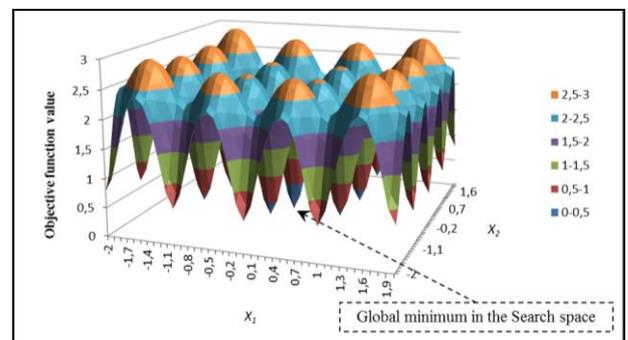


Figure 5: Ackley's Function

4. OPTIMIZATION EXPERIMENTS

We tested optimization methods on the testing functions where the dimension of the search space was 2; 10; 20; 30; 44. We specified the same conditions which had to be satisfied for each optimization method, e.g. the same termination criteria: The optimization method could perform a maximum of $20,000 \times n$ (parameter n denotes the dimension of the search space) simulation experiments to find the global optimum in the search space (Tvrđik, 2010) or the termination criterion was met if the optimum was found - VTR (value to reach); the same search space where the optimization method can search for the global optimum, etc. If the optimization method has the same parameters as

another optimization method, we set up both parameters with the same boundaries (same step, lower and upper boundaries).

We performed optimization experiments with different optimization method settings to find a suitable setting, but the analysis of these results is not the purpose of this paper.

We evaluated these optimization experiments with different settings - series. These series we replicated several times to reduce the random behaviour of the tested optimization methods. The following charts show the average optimization method success of finding optimum (suboptimum if the optimum was not found).

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

$\hat{f}_1 \leftarrow \text{AssignFitnessFindingOptimum}(X^*, X^*, \varepsilon)$	
Function whose output is normalized scalar value in the range $\hat{f}_1 \in [0,1]$. This value represents the failure of finding global optimum by the optimization method in a particular series – value minimization	
Input:	X^* : The list of found optima in each optimization experiment in the series
Input:	X^* : Global optimum X^* in the search space
Input:	ε : Tolerated deviation from the value of the objective function value of global optimum
Data:	$F(X)$: Objective function value
Data:	n_{succ} : Counter of successful finding optimum
Output:	\hat{f}_1 : Standardized scalar value
<pre> 1 begin 2 $n_{succ} \leftarrow 0$; 3 for $i \leftarrow 0$ to $\text{Length}(X^*) - 1$ do 4 if $F(X^*[i]) - F(X^*) \leq \varepsilon$ then 5 $n_{succ} \leftarrow n_{succ} + 1$; (*Optimum or acceptable candidate solution was found *) 6 result $\leftarrow \frac{\text{Length}(X^*) - n_{succ}}{\text{Length}(X^*)}$; (*standardization - % share of unsuccessful series*) 7 end;</pre>	

Figure 6: Pseudopascal Algorithm of First criterion – Finding the Global Optimum or Suboptimum

If the failure is 100[%] the first criterion equals 1 therefore we try to minimize this criterion. Average Method Success of Finding Optimum can be formulated as follows:

$$f_{avg} = \left(1 - \frac{\sum_{i=1}^s f_{i_1}}{s} \right) \cdot 100 [\%] \quad (8)$$

where i denotes the index of one series, f_{i_1} denotes the value of the first criterion (Optimization method success – the best value is zero), s denotes the number of performed series.

The series were also evaluated regarding specified tolerance between the best optimum (suboptimum) found in the series and the specified parameter ε . We initially specified $\varepsilon = 0.001$. The optimization method had to find the candidate solution whose objective function value is nearly the same as the objective function value of global optimum in the search space (the tolerance equals 0.001). This aspect is unfounded in practice. Hence we specified the tolerance to one percent of the difference between the objective function value of the global maximum and the objective function value of the global minimum of the search space:

$$\varepsilon = \frac{|F(\hat{\mathbf{X}}) - F(\check{\mathbf{X}})|}{100} \quad (9)$$

Where ε denotes the difference between the objective function value of the global maximum and the objective function value of the global minimum of the search space; $F(\hat{\mathbf{X}})$ denotes the objective function value of the global maximum of the search space; $F(\check{\mathbf{X}})$ denotes the objective function value of the global minimum of the search space.

The following figures show the average optimization methods success of finding optimum (suboptimum – the best found candidate solution of the series). Chart values of optimization method success are maximized.

We set up the termination criterion in such a manner that the optimization method can perform a maximum of $20,000 \times n$ (parameter n denotes the dimension of the search space) simulation experiments for each series to find the global optimum in the search space. The other part of termination criterion is the value to reach condition – VTR. We tested the optimization method in two; ten; twenty; thirty; forty-four-dimensional search space. We specified two different settings of acceptable tolerance (ε) of best found candidate solution testing (objective) function value from the objective function value of global optimum ($\varepsilon = 0.001$). If the global optimum of testing function is not known (e. g. Michalewicz testing function) the optimum is represented by the best found candidate solution of all performed simulation experiments performed on testing function.

4.1. Two-Dimensional Search Space

We started our simulation optimization experimentation in a two-dimensional search space where tolerated deviation from the value of the objective function value of global optimum is $\varepsilon = 0.001$. Figure 7 shows the average success of finding optimum. All tested optimization methods are successful (except Random Search) when the testing (objective) function landscape is not complicated – De Jong, Gradient optimization methods (Local Search, Hill Climbing and Tabu Search) have a problem to find the optimum if the testing function is multimodal – Ackley’s testing function.

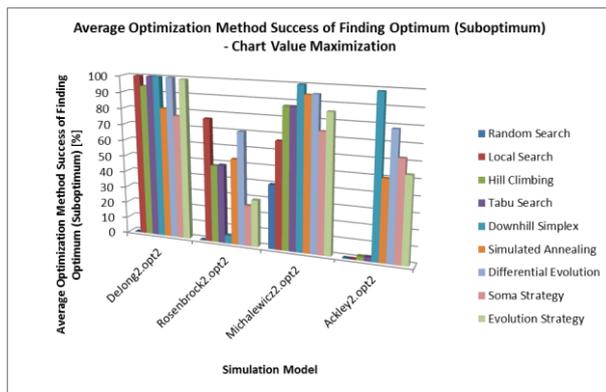


Figure 7: Average Optimization Method Success of Finding Optimum (Suboptimum) - Two-Dimensional Search Space – each series contains 40,000 simulation experiments; $\varepsilon = 0.001$

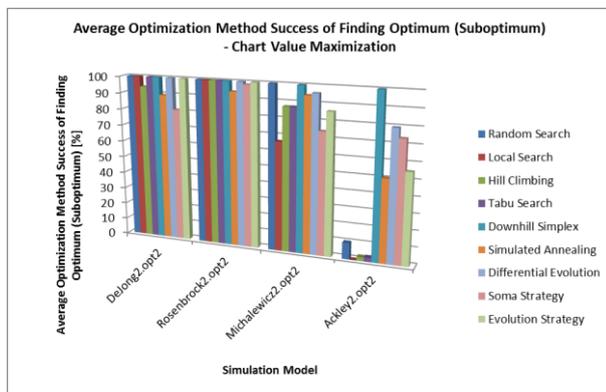


Figure 8: Average Optimization Method Success of Finding Optimum (Suboptimum) - Two-Dimensional Search Space – each series contains 40,000 simulation experiments; $\varepsilon = 1\%$ of objective function value range

If the tolerance is set to one percent of the difference between the objective function value of the global maximum and the objective function value of the global minimum of the search space the optimization methods have no problem to find the global optimum of the testing function in two-dimensional search space - Figure 8. If the function landscape is complicated (especially Ackley’s function and also Michalewicz) the gradient methods are not effective.

4.2. Ten-Dimensional Search Space

Other optimization experiments are performed in a ten-dimensional search space. The finding of global optimum of testing functions is much more difficult - Figure 9. When the testing function landscape is not complicated – De Jong’s convex and unimodal function – some methods - Downhill Simplex, Differential Evolution, Evolution Strategy and eventually SOMA Strategy - are more effective than gradient based methods - Local Search, Hill Climbing and Tabu Search.

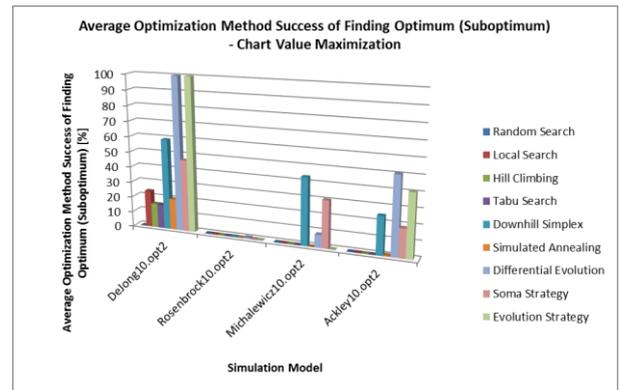


Figure 9: Average Optimization Method Success of Finding Optimum (Suboptimum) - Ten-Dimensional Search Space – each series contains 200,000 simulation experiments; $\varepsilon = 0.001$

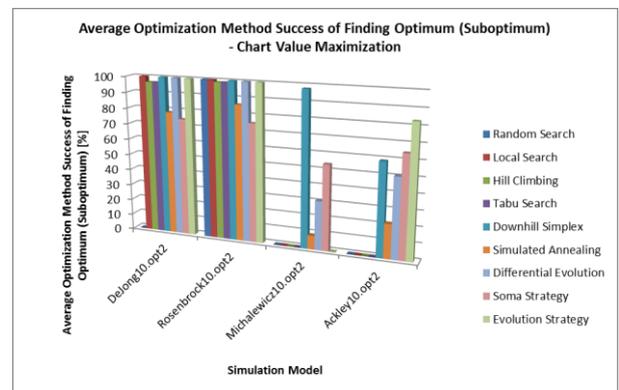


Figure 10: Average Optimization Method Success of Finding Optimum (Suboptimum) - Ten-Dimensional Search Space – each series contains 200,000 simulation experiments; $\varepsilon = 1\%$ of objective function value range

If the user accepts one percent of the difference between the objective function value of the global maximum and the objective function value of the global minimum of the search space, the optimization methods have no problem to find the global optimum of simple testing function – De Jong’s and Rosenbrock’s testing function - Figure 10. If the testing function landscape is complicated – Michalewicz and Ackley’s testing function – gradient methods have a problem with finding the optimum.

4.3. Twenty-Dimensional Search Space

The problem of finding optimum in the search space increases with the higher number of the dimension of the search space. We could see that the optimization method has a problem with a ten-dimensional search space - Figure 9. Gradient based optimization methods are not effective. Local Search reaches 6.6 % success of finding optimum in the search space of De Jong's testing function - Figure 11. Other gradient methods - Hill Climbing and Tabu Search - reach almost the same 2.3 % value of success of finding optimum in the search space of De Jong's function.

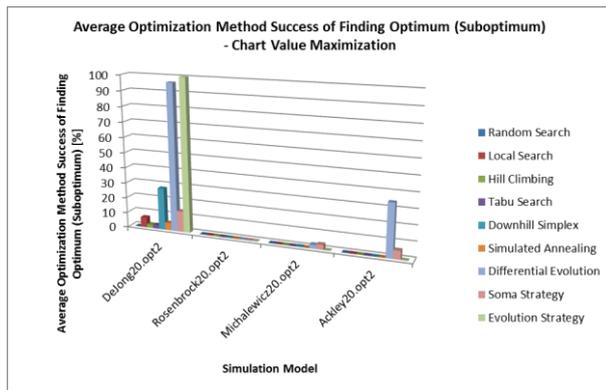


Figure 11: Average Optimization Method Success of Finding Optimum (Suboptimum) - Twenty-Dimensional Search Space – each series contains 400,000 simulation experiments; $\varepsilon = 0.001$

If we compare the efficiency of finding optimum in all searched search spaces of all testing functions when the epsilon is $\varepsilon = 0.001$ we can say that we should favour SOMA and Differential Evolution optimization methods. SOMA is another variation of the Differential Evolution methods.

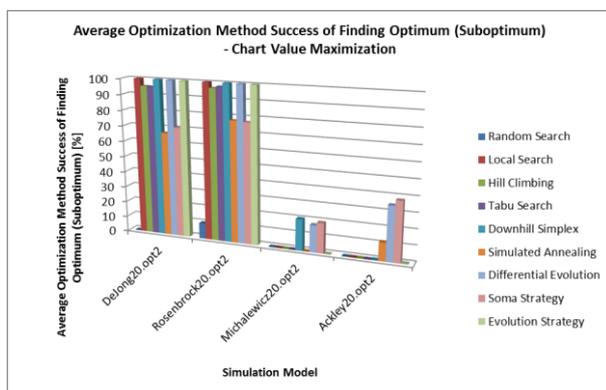


Figure 12: Average Optimization Method Success of Finding Optimum (Suboptimum) - Twenty-Dimensional Search Space – each series contains 400,000 simulation experiments; $\varepsilon = 1\%$ of objective function value range

If we increase the tolerated deviation ε all tested optimization methods (except Random Search) can find the optimum in the search space of simple testing function – De Jong's and Rosenbrock's testing function – Figure 12. Only two methods - Simulated Annealing and SOMA - are less successful. Their average optimization method success of finding optimum varies between 66 % and 77%.

4.4. Thirty-Dimensional Search Space

Differential Evolution is very successful when we optimize De Jong's thirty-dimensional search space - Figure 13.

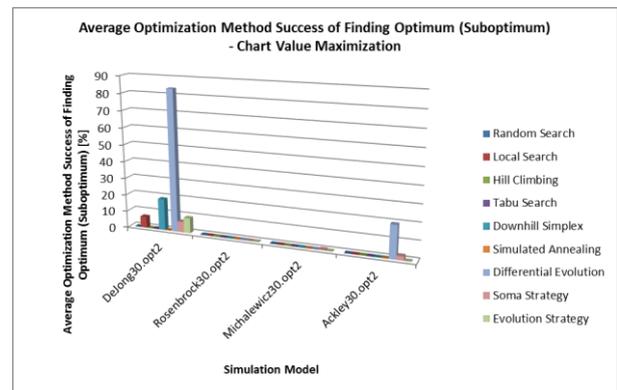


Figure 13: Average Optimization Method Success of Finding Optimum (Suboptimum) - Thirty-Dimensional Search Space – each series contains 600,000 simulation experiments; $\varepsilon = 0.001$

After comparing all charts representing the average optimization methods success we can say that the Differential Evolution is very effective in the case of our tested optimization methods. When ε equals 1 % of objective function value range all optimization methods are able to find the optimum in the simple testing function – De Jong's and Rosenbrock's function - with more than 60 % certainty - Figure 14.

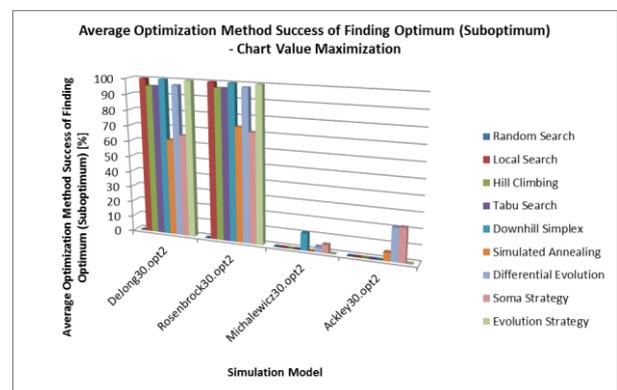


Figure 14: Average Optimization Method Success of Finding Optimum (Suboptimum) - Thirty-Dimensional Search Space – each series contains 600,000 simulation experiments; $\varepsilon = 1\%$ of objective function value range

SOMA, Differential Evolution, eventually Downhill Simplex and Simulated Annealing are able to find the optimum if the testing function landscape is complicated – multimodal test function.

4.5. Forty-Four-Dimensional Search Space

Differential Evolution is also very successful when we want to find the optimum of De Jong’s thirty-dimensional search space. Compared to thirty-dimensional search space Evolution Strategy is not able to find the optimum in forty-four dimensional search space of De Jong’s testing function - Figure 15.

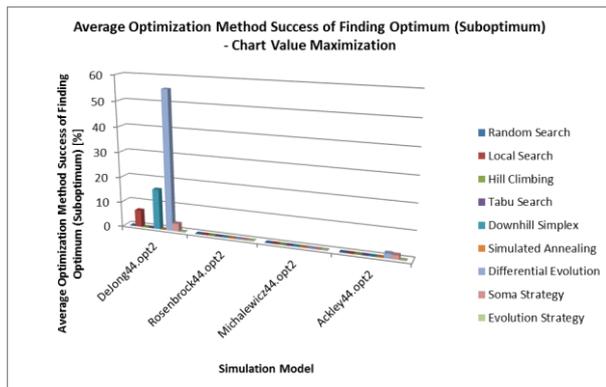


Figure 15: Average Optimization Method Success of Finding Optimum (Suboptimum) - Forty-four-Dimensional Search Space – each series contains 880,000 simulation experiments; $\varepsilon = 0.001$

Gradient based methods are not effective in the case of complicated testing functions with higher dimensions. These methods are useful for finding the optimum of a simple testing function landscape. Differential Evolution is successful optimization for optimization of different dimensional search spaces. This method is the best from our tested optimization methods.

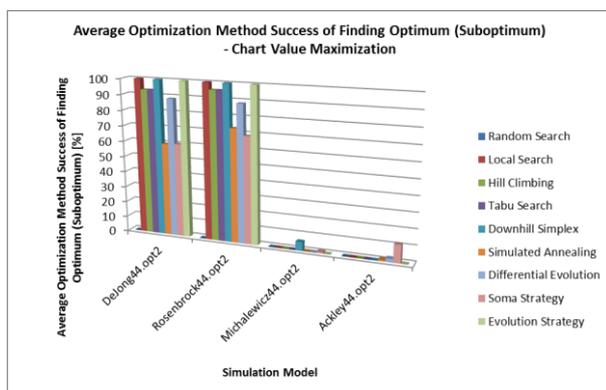


Figure 16: Average Optimization Method Success of Finding Optimum (Suboptimum) - Forty-four-Dimensional Search Space – each series contains 880,000 simulation experiments; $\varepsilon = 1\%$ of objective function value range

SOMA is derived from Differential Evolution. This optimization method is useful when we can accept deviation equals one percent of the difference between the objective function value of the global maximum and the objective function value of the global minimum of the search space. If we do not want to use the methods based on evolution processes we can use Downhill Simplex method or Simulated Annealing.

4.6. Forty-Dimensional Search Space and Higher Number of Performed Simulation experiments

Regarding the large number of possible candidate solutions in the search space we set up the termination criterion in a way that the optimization method can perform a maximum of $100,000 \times n$ (parameter n denotes the dimension of the search space) simulation experiments to find the global optimum in the search space. The next condition of termination criterion is to stop the optimization experiment if the optimum is found - VTR (value to reach). We tested the optimization method on a forty-dimensional search space (i. e. optimization method could perform a maximum of 4,000,000 simulation experiments in each series).

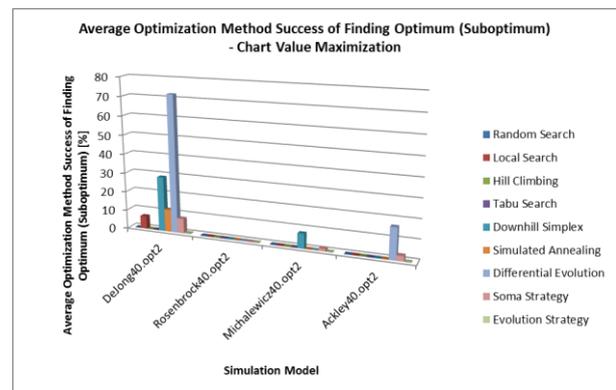


Figure 17: Average Optimization Method Success of Finding Optimum (Suboptimum) - Forty-Dimensional Search Space – each series contains 4,000,000 simulation experiments; $\varepsilon = 0.001$

If we compare the previous chart (forty-four dimension search space – see Figure 15) of success of finding optimum to the following chart (Figure 17) where the same condition of $\varepsilon = 0.001$ is met, it is obvious that the success of finding optimum by SOMA strategy increased the most in the case of Michalewicz, De Jong’s and Ackley’s testing functions. Differential Evolution and Downhill Simplex also improve their efficiency of finding the optimum if the number of performed simulation experiments increases.

If we compare the previous chart (Figure 16) of success of finding optimum to the following chart (Figure 18) where the same condition of $\varepsilon = 1\%$ of objective function value range is met, we can say that the success of finding optimum of Simulated Annealing,

Differential Evolution and SOMA strategy particularly increased in the case of Ackley's function. Downhill Simplex and SOMA method also increase their efficiency of finding optimum of Michalewicz function if the number of performed simulation experiments in one series is five times higher. Gradient based optimization methods like Hill Climbing and Local Search also increase their average success of finding optimum with the higher number of simulation experiments but not as much as previously mentioned methods.

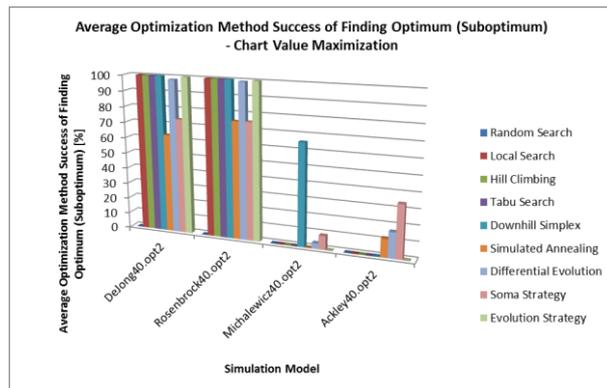


Figure 18: Average Optimization Method Success of Finding Optimum (Suboptimum) - Forty-Dimensional Search Space – each series contains 4,000,000 simulation experiments; $\varepsilon = 1\%$ of objective function value range

5. SUMMARY

The goal of the research is to compare selected optimization methods - Random Search, Hill Climbing, Tabu Search, Local Search, Downhill Simplex, Simulated Annealing, Differential Evolution, SOMA and Evolution Strategy – used for optimization of four selected testing functions. We tested these optimization methods on different dimensions of the search space to compare their efficiency of finding the global optimum in the search space.

We substituted the testing on the discrete event simulation models by a different testing function – De Jong's, Rosenbrock's, Michalewicz and Ackley's function. The dimension of the search space was 2; 10; 20; 30; 44. We specified the same conditions which had to be satisfied for each optimization method. Regarding the large number of possible candidate solutions in the search space we set up the termination criterion in a way that the optimization method can perform a maximum of $20,000 \times$ the dimension of the search space simulation experiments to find the global optimum in the search space. We also tested the optimization method if the number of performed simulation experiments in one series is five times higher.

We initially specified tolerated deviation from the value of the objective function value of global optimum to a specified value ($\varepsilon = 0.001$). The optimization method

had to find the candidate solution whose objective function value is nearly the same as the objective function value of global optimum in the search space. This aspect is unfounded in practice. Hence we specified the tolerance to one percent of the difference between the objective function value of the global maximum and the objective function value of the global minimum of the search space. The success of finding the optimum of the optimization method especially increases when we specified this new value of tolerated deviation (De Jong's and Rosenbrock's testing function).

The success of heuristic optimization methods depends on the objective function landscape. Gradient based methods are not effective in the case of complicated testing functions with higher dimensions. These methods are useful for finding the optimum of simple testing function landscapes. Differential Evolution is successful optimization for optimization of different dimensional search spaces. This method is the best from our tested optimization methods. SOMA is also a useful method for optimization. This method is derived from Differential Evolution.

Regarding the large time demands we would like to test the Client-Server architecture for parallel management of simulation experiments with different optimization methods settings. We also would like to test the use of a knowledge database to increase the speed of finding the optimum. The server sends the information about the setting of the optimization method parameters. The optimization experiment (simulation runs of a discrete event simulation model) is performed on the client PC. The client searches for the result of the simulation experiment with the same simulation model input parameters setting in the knowledge database before the simulation run. If this information is found, the client loads the objective function value from the database. If this information is not found, the client performs the simulation model run and the information about the simulation experiment (the setting of the simulation model input parameters and the objective function value) is sent to the knowledge database where this information is stored.

We would like to test optimization method behaviour and optimization parameters setting on different simulation models (e. g. simulation model of internal company logistics; this discrete event simulation model contains multiple simulation model input parameters).

ACKNOWLEDGEMENTS

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

Koblasa, F., Manlig, F., Vavruska, J., 2013. Evolution Algorithm for Job Shop Scheduling Problem

- Constrained by the Optimization Timespan. *Applied Mechanics and Materials*, 309, 350-357.
- Hynek, J., 2008. Genetic Algorithms and Genetic Programming (in Czech language: Genetické algoritmy a genetické programování). Prague: Grada.
- Miranda, V., 2008. Fundamentals of Evolution Strategies and Evolutionary Programming. In: El-Hawary, M.E., ed. *Modern heuristic optimization techniques*. New Jersey: John Wiley & Sons, 43–60.
- Mitchell, M., 1996. An Introduction to Genetic Algorithms. Cambridge: MA: MIT Press.
- Monticelli, A.J., Romero, R., Asada, E., 2008. Fundamentals of Tabu Search. In: El-Hawary, M.E., ed. *Modern heuristic optimization techniques*. New Jersey: John Wiley & Sons, 101–120.
- Monticelli, A.J., Romero, R., Asada, E., 2008. Fundamentals of Simulated Annealing. In: El-Hawary, M.E., ed. *Modern heuristic optimization techniques*. New Jersey: John Wiley & Sons, 123–144.
- Pohlheim, H., 2006. GEATbx: Example Functions. Available from: http://www.geatbx.com/docu/fcnindex-01.html#P204_10395 [accessed 20 November 2011]
- Raska, P. & Ulrych, Z., 2013. Simulation Optimizer and Optimization Methods Testing On Discrete Event Simulations Models and Testing Functions. ATHENS, GREECE, DIME Università, pp. 50-59.
- Raska, P. & Ulrych, Z., 2015. (in press) Comparison of Optimization Methods Tested On Testing Functions and Discrete Event Simulation Models. *International Journal of Simulation and Process Modelling*.
- Rockwell Automation, 2014. OptQuest for Arena. Available from: <http://www.arenasimulation.com/Products/OptQuest.aspx> [accessed 3 January 2014]
- Stefka, D., 2005. Alternatives to the evolutionary optimization algorithms (in Czech language: Alternativy k evolučním optimalizačním algoritmům) - Diploma Thesis. Prague: CVUT - Czech Technical University In Prague.
- Tvrđík, J., 2010. Stochastic Algorithms for Global Optimization (in Czech language: Stochastické algoritmy pro globální optimalizaci). Available from: http://www1.osu.cz/~tvrdik/wp-content/uploads/STAGO_10.pdf [accessed 5 January 2014]
- Weise, T., 2009. E-Book "Global Optimization Algorithms - Theory and Application" 2nd Edition. Available from: <http://www.it-weise.de/projects/book.pdf> [accessed 2 February 2011]
- Wong, K.P., Dong, Z.Y., 2008. Differential Evolution. In: El-Hawary, M.E., ed. *Modern heuristic*

optimization techniques. New Jersey: John Wiley & Sons, 171–186.

- Zelinka, I., 2004. SOMA — Self-Organizing Migrating Algorithm. *New Optimization Techniques in Engineering Studies in Fuzziness and Soft Computing*. Berlin: Springer Berlin Heidelberg, 167-217.

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.

UTILIZATION OF COMPUTER SIMULATION FOR THE LOCALIZATION OF ROLLING STOCK WITH UTILIZATION TECHNOLOGY ORACLE SPATIAL AND DYNAMIC DATABASE VIEWS

Jan Fikejz^(a), Jan Merta^(b)

^(a) Department of Software Technologies, FEI, University of Pardubice, Pardubice, Czech Republic

^(b) Department of Software Technologies, FEI, University of Pardubice, Pardubice, Czech Republic

^(a) Jan.Fikejz@upce.cz, ^(b) Jan.Merta@student.upce.cz

ABSTRACT

This article deals with utilization of computer simulation for the localization of rolling stock on the railway network using dynamic views. Attention is focused on the description of the location of rolling stock in the designed model of railway network. Further attention is aimed on optimizing search operations ORACLE database and designing of the optimization using dynamic views.

Keywords: Railway infrastructure models, train positioning, railway traffic simulation, database optimization, database dynamic views

1. INTRODUCTION

Rolling stock localization is a constantly discussed topic involving a lot of subjects. The problem of rolling stock localization may be divided in two main areas of interest: localization for the needs of (i) security technology and localization for the needs of (ii) information and telematics systems. In the first case the reliability and safety are emphasized, nevertheless these systems are often connected with higher realization costs due to the fact that 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 on the other hand 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

Rolling stock localization can be divided into three main parts:

- 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

If we use GNSS for various application levels, we need to take into account an indicated position error, which is generally based on the nature of the satellite navigation. If we use systems that operate with the position information only on an informative level, we can tolerate a certain error; however, such inaccuracy is unacceptable in the railway signalling technology. However, various additional systems can be implemented to eliminate the error (completely or at least partially), thus making the position of the tracked object more accurate. The following systems can be listed in this group:

- EGNOS (Senesi 2012)
- Differential GPS (O'Connor 1997)

2.3. GNSS based localization involving additional support systems

As mentioned above, precise localization of rolling stock using GNSS, especially for the needs of signalling technology, is a priori impossible. Nevertheless, the position of a rail vehicle can be put significantly more precisely with the use of additional systems. This concerns especially the solutions using inertial systems

(Standlmann 2006), but also less known systems such as those based on GNSS and contactless eddy current measurement (Becker and Poliak 2008)

3. RAILWAY NETWORK MODEL

Undirected graph, as defined graph theory, is a natural candidate for a railway network model. Based on an analysis of data provided by the company SŽDC-TUDC (consisting of service regulations, passports and codebooks), sets of algorithms were subsequently created, with which it was possible to generate a three-layer model of the rail network (Fikejz and Kavička, 2011). Roughly speaking, the track can be divided into individual so called supertracks, which consist of definition supra-sections (TDNU), where each supra-section contains track definition sections (TUDU) with mileposts (in hectometres). Basic aspects of the description of the rail network are collectively shown in Figure 1.

Mileposts (in hectometres) are shown in Figure with the distance in kilometres and are graphically represented using grey points. TUDU is recorded using a six-digit code (163105, 163106, 16307, 173202) and are graphically represented using solid lines (red, black, orange, brown). Individual supra-sections (CLS 007, CLS008, REG023) are shown in light blue and supertracks (B421021 1 and B421021 1A) are shown in dashed lines. A place significant in terms of transportation (branch line) is symbolized by a green square.

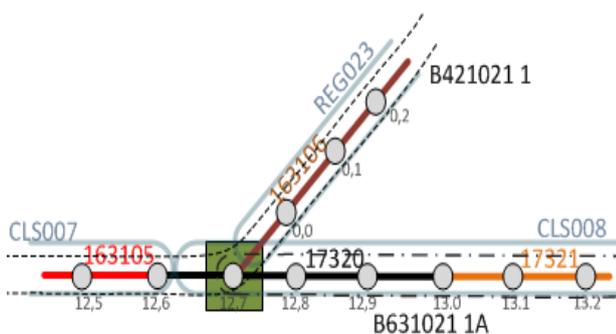


Figure 1: Basic aspects of the description of the rail network

The algorithm of railway network model (Fikejz and Kavička, 2011; Fikejz and Řezanina 2014) was implemented directly on the database level using PL/SQL language. However, the algorithm had to be adjusted and generalized several times since there are various nonstandard conditions in the data, such as jumps in the mileposts (nonlinear growth of the kilometre succession between the mileposts) or change of an increasing kilometre sequence into a decreasing one and vice versa. The final model include three data layers:

- **Data-micro**, consisting of vertices and edges,
- **Data-macro**, containing super-vertices and super-edges,

- **Data-macro**, include mezo-vertices and mezo-edges.

The data structure non-oriented graph was finally implemented directly in the ORACLE database using the ORACLE Spatial Network Data Model (Kothuriat al. 2007) technology. This technology enables the user to build a various network representations, involving also the object scheme and the communication interface API.

The objects scheme includes metadata and network tables. The interface contains on the server side PL/SQL API (an SDO_NET packet) for the creation, control and analysis of the database network, and a middle layer Java API (on client's side) for the network analysis. The actual network is then defined by means of two compulsory tables:

- Node table,
- Link table.

Concept of the technology is described in Figure 2.

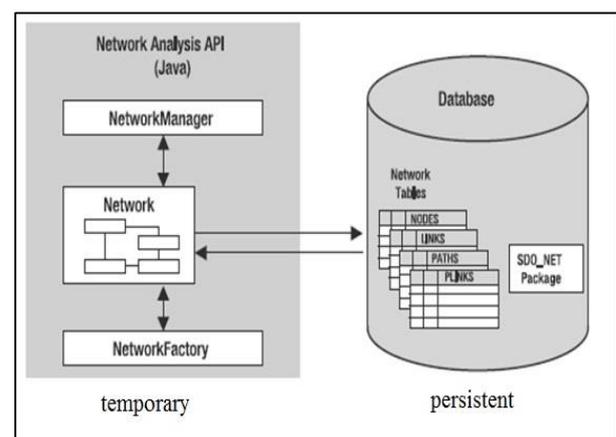


Figure 2: Oracle Network Data Model

For the work with spatial data, ORACLE with Spatial technology defines a special object data type SDO_GEOMETRY, which enables its user to store a number of spatial information and geometric types, such as various points, arcs, linear chains or polygons.

4. VISUALIZATION

For the purposes of visualization of proposed railway infrastructure model visualization tool MapViewer developed in the JAVA language (Murray et al. 2010) was employed. MapViewer represents J2EE service for displaying maps using the spatial data (e.g. using object data type SDO_GEOMETRY) managed by ORACLE Spatial. This technology enables to compose extensive map layers with various levels of details of information displayed (Fikejz and Kavička 2012) of the described information (Fikejz and Kavička, 2011). The basic concept of the MapViewer architecture is demonstrated in Figure 3.

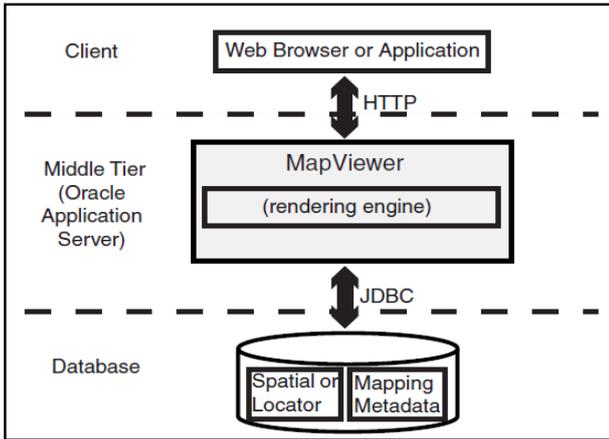


Figure 3: Basic concept of MapViewer service

It is possible to use various operators and functions above spatial objects of ORACLE database with option component Spatial (Kothuriat al. 2007). One of them is SDO_NN (Near Neighbour) that enables to specify the nearest geometry (so called neighbour), in our case the nearest vertex or rather edge of non-oriented graph. If we have GPS information about actual position of a rolling stock, it is possible to use this operator for specifying the nearest vertex/edge and subsequently for example to visualize the location on a map composed by means of MapViewer technology. The figure 4 shows the overall concept of the data and visualization layers of the railway network model.

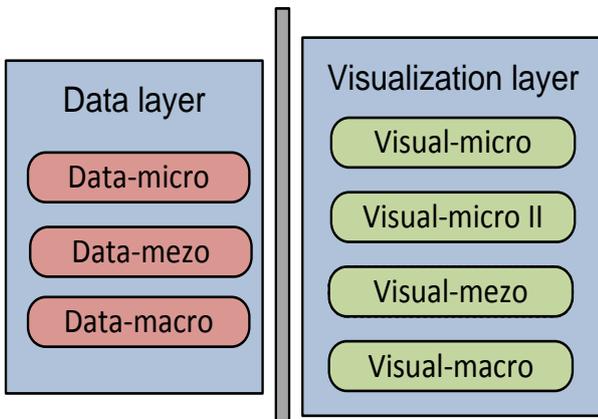


Figure 4: Overall concept of the railway infrastructure network model

5. SIMULATION OF TRAFFIC

The selected train vehicles are equipped with communication terminals, which broadcast data including current GPS coordinates of the rolling stock. When the vehicle is in motion, this communication terminal sends information about its position every 30 seconds. Table 1 contains the data obtained from the communication terminals.

Designed simulation model contains the core of discrete simulation utilizing standard calendar of process messages, which were, during the simulation, executed based on their time stamp. This model was consequently implemented into the software demonstrator InfraRail that is intended for the additional support of dispatching control.

Simulation of traffic of rolling stock can be divided into two parts. Simulation based on:

- Real historical data (emulation of operation),
- Generated data.

The main idea of the localization of rolling stock position is based on the finding the nearest vertex in the data micro layer of model from the current GPS position of moving train. This algorithm include the base tests of GPS validity and is used Oracle SDO_NN operator for the searching the Nearest Neighbour respectively the vertex.

6. SQL OPTIMIZATION QUERIES

If we use the database and SQL language there should be focused on the optimization. In this case we use SDO_NN operator so that we focus on optimization of SQL queries which are called directly from JAVA application. We can use the following:

- Statement Object,
- Prepared Statement Object,
- Calling functions that are stored in the database.

The Statement Object is base of universal query but its main disadvantage is periodical assembling of query. Use of Prepared Statement Object and PL/SQL function brings the use of so called bind variables. In this case the database gets still the same SQL statement (with still same hashcode) but with different values. This means that in the database machine is used always the

Table 1: Recorded time-stamped train movements data

Train number	Latitude	Longitude	Speed	Azimuth	Train vehicle identifier	Time
48701	50.02274	15.33554	78	57	91547123022	14.02.11 04:34:25
48701	50.02654	15.34246	80	43	91547123022	14.02.11 04:34:55
48701	50.03077	15.34873	68	43	91547123022	14.02.11 04:35:25
48701	50.03495	15.35505	61	45	91547123022	14.02.11 04:35:55

same execution plan.

6.1. Design of the optimization using the dynamic views

Despite the advanced optimization techniques of database ORACLE, the question arises as to whether is necessary to perform queries to the whole bases of vertices respectively edges. Whereas we are inquiring on the position of rolling stock (RS) within railway network which from the logical view, are not able to change the position by more than tens of meters, then we can reduce the base of vertices/edges by using the dynamic view. The main ideas are based on the following assumptions:

- For each new RS is created an initial dynamic view,
- Queries are periodically performed into the current view,
- If the RS is not already in the current view then the dynamic view is recalculated regarding to the azimuth of a moving train.

This means SQL query is performed into the relevantly reduced base of vertices/edges. This leads to optimization and the overall time saving of query.

7. TESTING

Testing of proposed optimization of selected search operations should show whether the using of dynamic view are correct. For the purpose of testing there has always been generated thousand points with GPS coordinates in the selected area, wherein the base of data consists of about one hundred thousand entries. For the time measuring of each query there were used information from database system tables ORACLE *v\$sql* and values from column *ELAPSED_TIME*. For secondary comparison there was used time measured directly in application that reflect the time overhead of communication between application and database machine. For each test the time was measured as following:

- Average time,
- Max / min time,
- Median / modus / deviation,
- Total time in database / application.

7.1. Technique of queries

Firstly, there were tested all query techniques without dynamic views, that means:

- Statement Object,
- Prepared Statement Object,
- Calling functions that are stored in the database.

As it was expected, the Statement Object showed approximately 3.5x worse time of processing then

Prepared Statement Object and query in the function. The results are shown in table 2.

Table 2: The comparison of different techniques queries in second

Method	Statement	Prepared Statement	Oracle function
Average	0,1208972	0,0091850	0,0124396
Min	0,059907	0,0035530	0,0047680
Max	3,28629	1,640714	1,607601
Median	0,0994355	0,0066205	0,0097315
Modus	0,066694	0,0040350	0,010057
Diversio	0,1151091	0,0520075	0,0514502
Total time DB	120,897181	9,185022	12,439586
Total time App	195,717069	53,824611	60,93599

7.2. Optimization by the dynamic views

The second area of tests used the dynamic views for Prepared Statement Object and query in the function for different sizes variants of dynamic views. The tests were performed within the three different sized areas:

- 10 x 10 km,
- 40 x 20 km,
- 50 x 50 km.

The comparison of result times for different size views within Prepared Statement shows that using of views brings the better results than without them. The best result comes with using of dynamic views 40 x 20 km. There is the time saving approx. 0.003 second. The all results we can see in the table 3.

Table 3: The results of times for Prepared Statement within different size of area in second

Size of view	10x10	40x20	50x50
Average	0,007647564	0,007537129	0,007978
Min	0,002743	0,002815	0,003319
Max	1,454968	1,157581	1,261327
Median	0,0049415	0,005347	0,006121
Modus	0,004397	0,003494	0,004287
Diversio	0,046213794	0,03691007	0,039899127
Total time DB	7,647564	7,537129	7,977629
Total time App	39,759844446	39,70151928	40,368371538

The next testing was used the function that is stored in the database. In this case was detected as the fastest query of dynamic views for the area 10 x 10 km. In

comparison against PreparedStatement this way brings additional time savings but only on the database level. The total time in the application is on the other hand less favourable. The all results we can see in the table 4.

Table 4: The results of times for function within different size of area in second

Size of view	10x10	40x20	50x50
Average	0,007312738	0,007474627	0,007914767
Min	0,003082	0,003242	0,003108
Max	1,370451	1,354342	1,568563
Median	0,0052415	0,005647	0,005894
Modus	0,003895	0,005491	0,004772
Diversion	0,043302936	0,0427506	0,049460025
Total time DB	7,312738	7,474627	7,914767
Total time App	51,351476774	54,835111842	59,058458159

The next table (Table 5) only summarizes different approaches (Oracle function and Prepared Statement) for using the dynamic database view for the area sized 40 x 20 km.

Table 5: Times of two different approaches for view 40x20 km

Method	Oracle function 40x20	PreparedStatement 40x20
Average	0,007475	0,007537
Min	0,003242	0,002815
Max	1,354342	1,157581
Median	0,005647	0,005347
Modus	0,005491	0,003494
Diversion	0,0427506	0,036910
Total time DB	7,4746270	7,537129
Total time App	54,835112	39,70152

Although the total times in database are quite similar the different situation is in the total time in JAVA application. The ORACLE function against Prepared Statement has significantly worse time. The reason is that the ORACLE function requires additional time to the calling and takeover result.

7.3. Result of tests

The above mentioned tests and their results proved that the set preconditions of time saving with using of dynamic views were correct. On the figure 5 is shown

graph which demonstrated gradual acceleration of SQL queries with utilization of the dynamic views. Also it is shown that the using of Statement queries significantly lags behind the queries with optimization.

The average times for different approaches is shown in Figure 5.

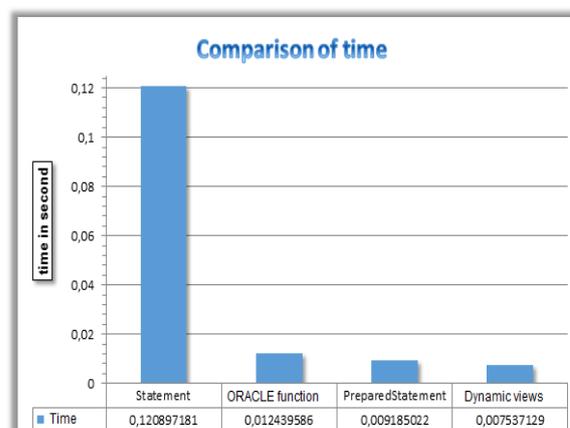


Figure 5: All tested approaches

The optimization of SQL queries based on dynamic views was subsequently integrated into a demonstration application InfraRAIL for localization of rolling stock. The running application using dynamic views within a railway network model is illustrated in Figure 6.

Within the final evaluation of results was found that for hundreds active trains on the railway network when the each train sends information about position each ten seconds is total time savings up to a few seconds.

CONCLUSION

This paper is focused on localization of rolling stock on the railway network and on the optimization of database queries by using dynamic views. In the introductory part, the paper shortly describes the existing systems in the localization area. This entire problem is divided into three parts. Further is described three-layer of railway network model, which is built as undirected graph directly in the ORACLE Spatial Network Data Model technology.

The second part is devoted to describing of different approach of optimization of database queries. Further is there described the design and implementation of the optimization database queries using a dynamic database views. In addition, there are shown the results of various tests which for hundreds active trains on the railway network proves that the using the optimization database queries (by the dynamic views) brings within a running application an additional timesavings.

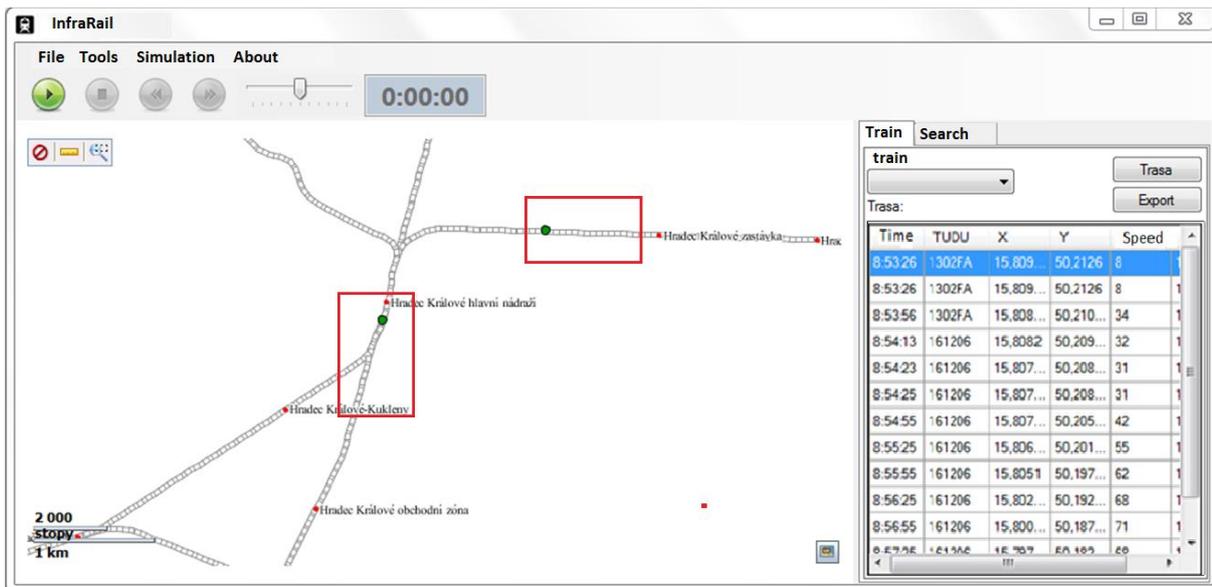


Figure 6: Running application

ACKNOWLEDGMENTS

This work has been supported by the project “SGSFEI_2015008 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 worldwide by 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
- 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].
- 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.

VARIOUS ASPECTS OF MULTI-METHOD MODELLING AND ITS APPLICATIONS IN MODELLING LARGE INFRASTRUCTURE SYSTEMS LIKE AIRPORTS

Barbara Glock^(a), Niki Popper^(b), Felix Breitenecker^(c)

^{(a),(b)}dwh GmbH, dwh Simulation Services, Neustiftgasse 57-59, Vienna 1070, Austria

^(c)Institute for Analysis and Scientific Computing, Vienna University of Technology, Wiedner Hauptstrasse 8-10, Vienna 1040, Austria

^(a)barbara.glock@dwh.at, ^(b)niki.popper@dwh.at
^(c)felix.breitenecker@tuwien.ac.at

ABSTRACT

Large infrastructure systems like airports are complex, consisting of various subsystems that are interconnected dynamically to each other. Modelling only one subsystem, although the research question addresses only that subsystem, can lead to error propagations, as dynamic system effects are underestimated. Extending the system boundaries after identification of the dynamic effects can lead to trade-offs when the chosen modelling method gets to its limits. Combined systems require multi-method modelling where diverse subsystems are modelled with the respectively best fitting modelling method. These subsystems are connected via specifically defined interfaces to get a model that represents the large system. Multi-method modelling allows using all advantages of the different modelling methods by coupling them and give a more natural way of seeing the system. Different coupling methods found in literature are presented and basic concepts for modelling airports and its subsystems are proposed.

Keywords: multi-method modelling, agent-based modelling, system dynamics, airport planning

1. INTRODUCTION

The planning of big infrastructure developments is getting more challenging due to more complex structures and an increased number of construction standards. Large infrastructure systems can be decomposed into a various number of **subsystems** that are somehow interconnected and also correspond on different levels with each other, what makes an analysis more difficult nowadays. Furthermore, there are different views on the system as well, that need to be addressed and satisfied: Stakeholders, planners, consumers, decision makers, etc. The diverse processes going on in these subsystems are connected in some ways that furthermore may not be apparent at first. In general this calls for modelling and simulation (see the real world in Figure 1 a), pictures by Bösch (2013)). These large systems, broken into pieces, consist of

different subsystems with much more detail, each of them with its own dynamic effects.

1.1. Modelling Point of View

On the one hand the modeller can look at the **system as a whole**, decide which method suits best modelling the large system and answering the research questions. So in parts of the model of the large system the modeller has to make some trade-offs, where he can't go too much into detail, where it would have been necessary or where he goes too much into detail, where it was not necessary, see Figure 1b). So with the model, the system may not be represented as realistic as it could have been (*one model with one method for the whole system*). The method may not fit for the system as a whole.

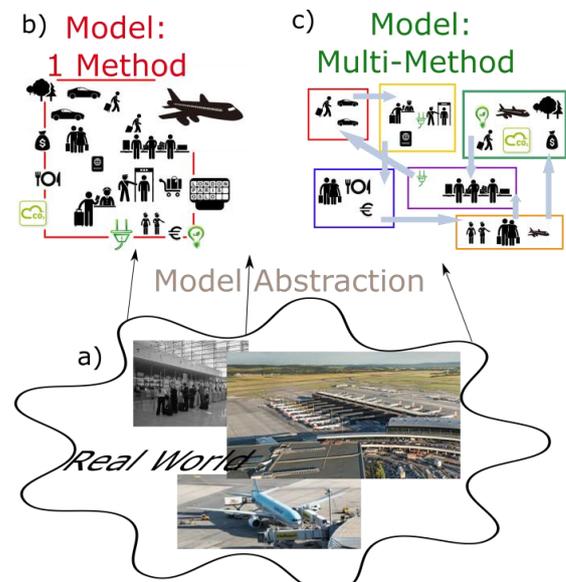


Figure 1: Modelling Large Systems

On the other side, when only modelling a **small part of a large system** with a specific modelling method and not taking into account effects from other subsystems the modeller again makes some trade-offs and propagated errors follow through the model (*one model*

and one method for a subsystem). If the model boundaries only cover the subsystem, dynamic effects from outside that may be relevant are not considered. These effects then accumulate to serious dimensions and have a much stronger negative impact on the system state later on.

When different questions addressing for example the utilization of resources within one subsystem, or waiting times or even the planning process itself, arise, usually different modelling methods are used to model the specific subsystem trying to answer those questions. To get a more realistic model of the large system these different subsystems modelled with different methods can be coupled to get a **multi-method model** and to use all advantages of the methods for each of the subsystems, see Figure 1 c). Complex behaviour arises through the interconnection of different subsystems. Diverse studies found in literature call for use of multi-method modelling. Scholl (2001) even states, regarding Agent-based modelling (ABM) and System Dynamics (SD), that these "...two techniques have a high potential for supporting and complementing each other. Joint ABM and SD research is proposed that may have the capacity for delivering results superior to those based on one technique only."

The coupling mechanisms are diverse and dependent on the used methods. Difficulties and issues can arise. In this paper an overview of classifications on coupling methods found in literature will be given in 3. First, some definitions on terminologies will be provided in 2. In the next section we give some information on why we intend to test the proposed concepts in the area of airport planning.

1.2. Application Area: The Airport City

The concept of multi-method modelling is applied to airports and their decomposed subsystems, because airports or airport cities, as they are called due to the fact that they basically provide everything a real city provides as well, are **large complex socio-technical systems** where interactions between people and technology happens. Furthermore, according to Neufville (2013) aviation passenger and cargo traffic grew remarkably in the last years and will go on growing in the next years. Passenger traffic worldwide increased at an average of about 4 percent per year from 1990 to 2010. In 1990 more than 1000 million passengers were transported worldwide and in 2010 slightly more than 2500 millions, so it more than doubled in the last 20 years. According to the International Civil Aviation Organization ICAO (2015) it will increase in the next years as well, from 2012 4.9% growth rate (for passenger-kilometers performed) to 2015 with 6.3% growth rate. This is due to the fact that flying has become cheaper and safer and therefore planning and utilization of airports with respect to ecologic and economic factors is very important. Furthermore, according to Neufville (2013) there are three dominant trends in the airport and airline industry in the early 21st century:

- **Long term growth:** avg. 4% per year
- **Organizational change:** economic and political deregulation continues to spread worldwide (low-cost and integrated cargo airlines grow, privatization of airlines and airports)
- **Technical change:** in aircraft and air traffic control. The developments increase the efficiency and the capacity of airport facilities and processes. Airports need to adapt these new opportunities as they occur.

Airports, their structures and processes are complex, mostly not easy to understand and multiplex, which can be better analysed by modelling and simulation. In Figure 2 the different subsystems of a traditional airport are shown.

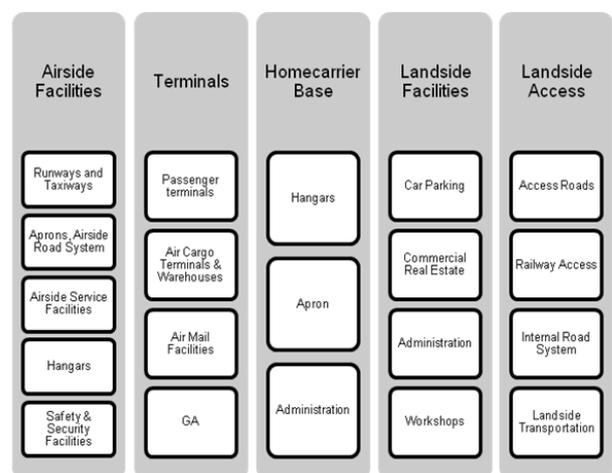


Figure 2: Subsystems of an Airport (Source: Company AI-MS Aviation Infrastructure Management Systems)

The main areas are:

- **Landside:** Passengers arriving by car, train, bus or taxi at the airport; Car parking; Administration;
- **Terminals and Retail:** Passengers going through check-in, passport control and security checks as well as the retail area; Cargo; Air mail;
- **Airside:** Airplanes departing and arriving; Ground handling processes; Airside facilities like runways; Air Traffic Control; Administration; Apron; Hangars; Control Centers;

Not only passengers are part of an airport system, but also personnel resources in different areas and material resources like planes, taxis, catering, cleaning, refuelling, etc. are part of the system, hence a socio-technical system. There are three main transportation goods: passengers, cargo and mail. Furthermore, there are operational aspects and strategic aspects in flight planning and aviation management that need to be addressed that also include meteorological forecasts (for

flight planning and slot management), planning standards (for building or rebuilding airports), safety (which means operating without causing unacceptable harm) and security (which means freedom from unauthorised access, protection against attacks) questions, facilities programming and demand forecasts (ACRP 2010, Sterzenbach 1996, Maurer 2002). According to Neufville (2013) the planning design and management also has to include:

- *Dynamic strategic planning concepts* like the SWOT analysis: include *strengths* of existing site, *weaknesses* of facility, *opportunities* for the region and *threats* to the airport and region.
- *Market Dynamics*: Consumer behaviour, moving preferences, social relevance and also the competition of producers for market share.
- *Environmental impacts*: Aircraft noise, air quality, climate changes, water quality, wildlife and environmental legislation.

Planning, designing and understanding an airport system is complex. By optimizing buildings and especially large buildings in the early stages of planning savings of material and money can be achieved. If space is seen as an endless resource diverse ways of looking at a problem arise. If too many resources were planned the utilization of these resources can be seen as less efficient over time, which leads to unnecessary built space that needs expensive maintenance and inefficiency in the business. Some examples are: increased expenditure of energy or other resources, rising impervious surfaces. Inefficient process on built space can be made visible only by the dynamic utilization of space. With new strategies and intellectual approaches it will be possible to see space, its functionalities and its processes as ecologic relevant resources. The overall aim of modelling and simulation is to reduce the economic expenditures, as well as in increase the positive ecologic aspects. Also accessibility and transit time lengths are in the airport planning and other application areas very important components of the simulation. Usual planning errors like to small turning radius in toilets for wheel chair drives and unacceptable access paths for handicapped persons can be avoided. Different stakeholders need to be satisfied as well. For example projects in the area of airport planning on the development of air infrastructure address not only the planners and airports, but also the effects on industrial and touristic development of the whole region. These different views of planners, architects, airports, ministry, passengers and people who live in that region need to be included in a simulation and the interpretation of the results. This is an example why trade-offs in simulation of large infrastructure developments are not welcome and why a different approach is needed. In 4 a model concept for modelling some airport subsystems and the application of coupling methods as given in 3 will be proposed and discussed.

2. DEFINITIONS

Researching literature showed that there are multiple terminologies for modelling a large system with different modelling methods. Searching databases like *ScienceDirect*, *Scopus*, *Springer Link*, *IEEE Xplore* and *MathSciNet* by using terms and their combinations like “hybrid modelling” and “coupled models” in the first place, showed that other terminologies like “dynamic system modelling”, “hyper modelling”, “interconnected simulation”, “interfaced simulation”, “integrative modelling”, “multi-method modelling” and many more are used for this kind of modelling as well. Swinerd and McNaught (2012), Sargent (1994) and Lätti (2010) refer to it as “hybrid models” or “hybrid modelling”. They also proposed some methods of coupling Agent-based and System Dynamics models (see 3). Scholl (2001) referred to this kind of modelling as “multimethod and integrative approaches” and Schieritz (2003) referred to it as “integration”, which makes it intuitively clearer what is meant than just saying “hybrid” modelling. Fishwick (2012) extends the meaning of “integrative modelling” or “multimodelling” and introduces a new term “hypermodel” to include interaction within models, among models and between human and models. On the other hand some of these terms, like “hybrid” are used for more specific or other interactions, like it is done in *Discrete-Event Modelling and Control of Hybrid Systems* by Nixdorf (2003): “hybrid” has a different definition in the context of modelling and simulation here and means that within one model *discrete and continuous* elements are modelled. Basically said, there are a lot of terms used for what intuitively is best understood as **multi-method modelling**.

Before we go on proposing what a multi-method model in our context is, we need to clear some definitions first: In the real world as seen in Figure 3, a **system** is a *set of interacting or interdependent components forming an integrated whole*.

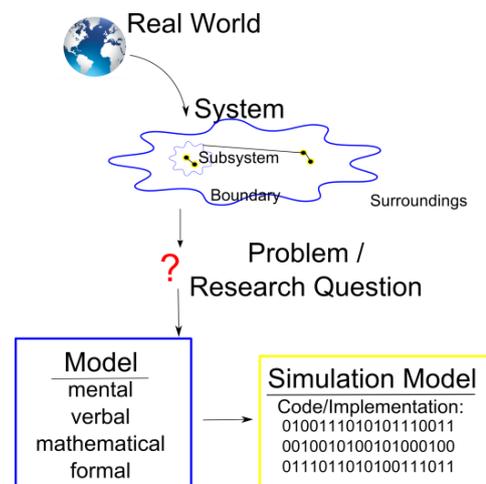


Figure 3: Building a Model out of the Real World

A **subsystem** is a *set of elements, which is a system itself, and a component of a larger system*. According to

the system we look at a **problem** and a **research question** may occur. There are diverse standards how these questions have to be formulated, like in health economics and evidence based medicine the PICO standard (Gerber 2006). After having defined the research question a **model** can be derived. This can be mental, verbal or specific as a formal description. According to Preston White (2009) “*a model is an entity that is used to represent some other entity for some defined purpose. In general, models are simplified abstractions, which embrace only the scope and level of detail needed to satisfy specific study objectives*”. A **simulation model** is the creation of a digital prototype of the model that is executable, also often referred to as a *computer model*.

If such a system is a large system it can be decomposed into **subsystems** where each of them can be modelled with another modelling method, forming a **submodel**, as seen in Figure 4.

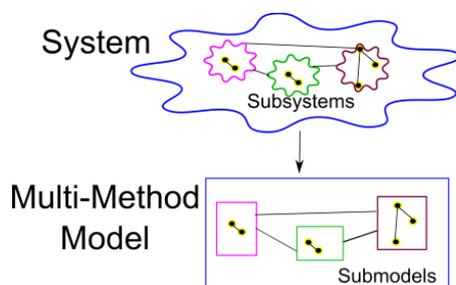


Figure 4: Decomposing the System into Subsystems and then creating Submodels with at least two different Modelling Methods.

The submodels for these parts of the system can be parallel or integrated and they can be on the same level or ordered hierarchically. A **multi-method model** is a model that consists of at least two submodels, where at least two different modelling techniques are used. These submodels exchange information in some way. This process of information exchange is called **coupling**.

2.1. Modelling Methods used here

Multi-method modelling with especially three modelling methods are researched and used as examples in the application of airport planning, because these methods are, according to literature, used most often in this area:

- Agent-based Modelling (ABM)
- System Dynamics (SD)
- Discrete Events (DES)

Agent-based modelling and Discrete Events are micro-based or individual-based modelling methodologies, best suited for modelling systems where the behaviour of (autonomous) individuals determines the system dynamics (Bonabeau 2002, Macal and North 2010). **Agent-based modelling** is a rather new modelling approach for which no consistent definition on what the

modelling instances (agents) have to fulfil exists. **Discrete Events** models are similar, but the entity modelled here is not like an agent autonomous, but is passively led through the system instead. Furthermore, changes in the state of the system happen due to events at discrete points in time (Zeigler 2000). In between two consecutive events the state remains unchanged. This kind of modelling is mostly used in logistics and transportation. Another paradigm is set by **System Dynamics** modelling. Here the point of view is from another level, where only aggregated levels are looked at. It was developed in the 1950s by Jay W. Forrester, who applied it first in management systems (see *Industrial Dynamics* by Forrester (1997) or *Urban Dynamics* by Forrester (1973)). He then transferred this methodology to social systems. Nowadays diverse literature on *System Dynamics* and *Systems Thinking* exists (Sterman 2000). A SD model consists of stocks and flows, which basically is a set of differential equations. The dynamics of the system emerges from causal links of the modelled variables that often form feedback loops. Application areas are economics, health care, policy design.

These two modelling paradigms are different in view and offer different advantages. The one modelling method's advantage is the other ones disadvantage. So why not combine their advantages?

3. CLASSIFICATION OF MULTI-METHOD MODELS

In literature there are some classifications of multi-method models (they usually called it hybrid models) found. In this section a short overview on the (for the further work of this project) most promising definitions of classifications are summarized.

3.1. Original Classification Approach for Analytical and Simulation Models

Sargent (1994) suggested, based on his definition of a “hybrid model”, which “*is a mathematical model which combines identifiably simulation and analytic models*”, four classes of hybrid models:

- Class I – “*A model whose behaviour over time is obtained by alternating between using independent simulation and analytic models.*”
- Class II – “*A model in which a simulation model and an analytic model operate in parallel over time with interactions through their solution procedure.*”
- Class III – “*A model in which a simulation model is used in a subordinate way for an analytic model of the total system.*”
- Class IV – “*A model in which a simulation model is used as an overall model of the total system and it requires values from the solution procedure of an analytic model representing a portion of the system for some or all of its input parameters.*”

This concept was used for actual *analytical models* (for parts of the system where time-dependent behaviour could be completely decomposed) and *simulation models*. This approach although done for analytical models (a set of equations that can characterize a system or a problem entity) and simulation models (in Sargent's (1994) context: a dynamic or an operating model of a system or problem entity that "mimics" the operating behaviour of the system or problem entity and contains its functional relationships) can be applied to simulation models with different paradigms as well, as Swinerd and McNaught (2012) showed for Agent-based and System Dynamics models in their work.

Sargent (1994) already found out that, what we call multi-method modelling, is needed to be researched, because these kinds of models are very useful in different application areas and have a lot of potential in modelling the world in a more realistic and effective way.

3.2. Adapted Classification for ABM and SD models

Swinerd and McNaught (2012) suggest for a system with AB and SD modules (here a module is seen as a submodel as we see it, representing a modelled part of the system) three classifications that are derived from Sargent's (1994) classifications (also see Figure 5):

- **Interfaced** (equivalent to Class I): two submodels with different modelling methods have some point of interaction or communication between elements; the submodel run *alternating* and *independently*;
- **Integrated** (equivalent to Class II): there are three methods suggested:
 - Agents with rich internal structure (also see Figure 6)
 - Stocked agents (also see Figure 7)
 - Parameters with emergent behaviour (also see Figure 8)
- **Sequential** (equivalent to Class III and IV): one submodel needs the output from the other submodel as input.

An **example for an interfaced model** as given in Swinerd and McNaught (2012) is: an AB submodel where a person (agent) walks along a street trying to reach his goal. Public traffic transportation is modelled by another submodel in DES. The agent can decide (within the AB submodel) if he wants to walk (stay in AB submodel) or if he wants to take the bus (DES submodel). He is either in one submodel or in the other, hence independent submodels that run alternating.

An **example for sequential model** according to Swinerd and McNaught (2012) is: "[...] the Land Use Scenario Dynamics (LUSD) model which incorporates SD and CA modules. Within this design concept, the SD module firstly determines national and regional demand for land use based on factors such as land policy, demographics, market demand, the economy, and influence of technology. The CA module then provides a

spatial representation of local land allocation to meet the aggregate demand defined by the SD module and considers local factors such as land suitability, land inheritance and the influence of neighbours. The output from the LUSD model is provided by the CA module and is in the form of a geographic map on which land change is plotted."

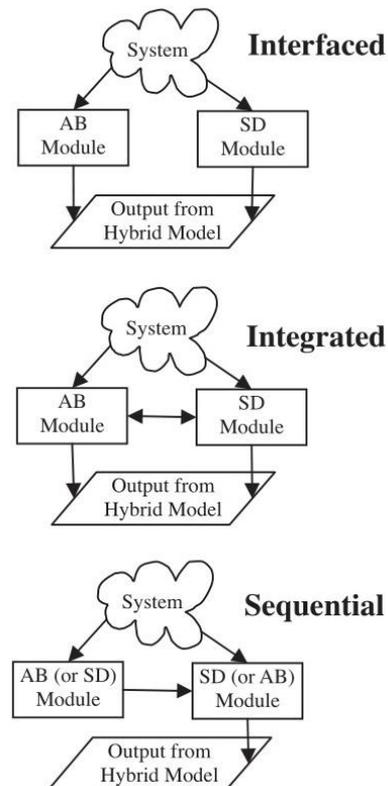


Figure 5: A Classification of Multi-Method Models according to Swinerd and McNaught (2012) for modelling Large Systems

Another example for a sequential multi-method model in airport planning would be an AB landside model that models passengers arriving at the airport by car, bus, train or taxi in order to enter a terminal module that is modelled with DES. First the necessary individual behaviour on the landside where passengers drive cars or arrive by some other vehicle is modelled. The individual behaviour is required for this module for modelling goal seeking strategies of agents that act autonomous and can decide where and when to go and how to react to their environment if something happens (street closed, parking houses closed, traffic jam, etc). In the terminal model (which will be explained more detailed in 4.2) questions addressing resource allocation or waiting times of passengers are answered. Simple server-queue structures call for DES modelling. First the passenger goes through landside and then through terminals. If this is purely sequential depends on what the modeller wants to include in the model. If the modeller also includes passengers that arrive at the airport by plane and proceed to the exit over the landside, this would be an integrated approach, because

interaction during the simulation time will be required as well.

An **example for an integrated model of agents with rich internal structure** is an AB model where each agent contains a SD model, as seen in Figure 6 (by Swinerd and McNaught (2012)). One can think of the SD model of an agent as the agent's "brain" that "tells" him what to do in a dynamic way. Here influences from both sides can be considered: AB submodel passes information to the SD submodel and vice versa.

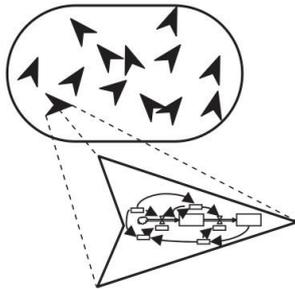


Figure 6: Integrated Multi-Method Model: Agents with Rich Internal Structure

An **example for an integrated model with stocked agents** is "a level within an SD model that is used to bound an aggregate measure of an AB module" (see Figure 7, Swinerd and McNaught (2012)). This could be an SD submodel that calculates production costs on car sales and the influence of fuel price on consumer choice of vehicle technology where consumers are modelled in the AB submodel. Here only influences from SD to AB are modelled, but not the other way round.

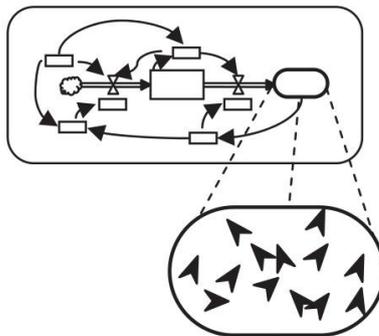


Figure 7: Integrated Multi-Method Model: Stocked Agents

An **example of an integrated model with parameters with emergent behaviour**, as seen in Figure 8, is an AB submodel where demographic development is modelled with agents together with their individual attributes (age, sex, maybe socio-economic factors). Out of these attributes the value of a parameter for a coupled SD submodel is calculated. According to Swinerd and McNaught (2012) this could be the participation of a population to the pension fund of a country that is modelled with SD.

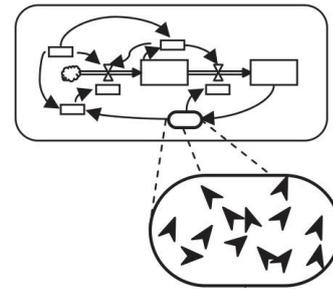


Figure 8: Integrated Multi-Method Model: Parameters with Emergent Behaviour

Basically, there is a fine line between the classes of multi-method models and the modeller has to decide what fits best. It is also dependent on where the system boundaries lie (see comment after example of sequential model above). This classification approach is very useful as a starting point in researching multi-method modelling, since it is also applicable not only for SD and ABM modelling but also for other modelling methods. In the next chapter some examples and concepts of multi-method models in the area of airport planning will be given.

4. APPLICATIONS IN AIRPORT PLANNING

There are many subsystems of an airport, as seen in Figure 2, and in the next three sections a concept for three subsystems and their coupling mechanisms of the multi-method model will be proposed. Some of the subsystems are currently being developed in AnyLogic 7 and others will be implemented during further research of this project. The general advantage of this modular set-up that is included in this multi-method modelling technique. Is that it can be extended for other subsystems in the future to include more influencing parts of the airport city system as well. First some notes on available data and passengers will be given and then existing concepts and existing tools will be explained shortly.

4.1. Some Notes

For most simulations a so-called *Design Peak Day* from the Vienna International Airport, where data for incoming and outgoing flights, their destination respectively origin together with planned and actual arriving respectively departing time and actual time is available. Furthermore, it holds information on how many passengers were transported. According to this Design Peak Day, which represents an optimal day for this airport, number and time of arrival or departure of passengers can be retrieved for the simulation.

There is a main differentiation in passengers: *Tourist Passengers* and *Business Passengers*. Data on who is tourist and who is business can only be estimated, but this is relevant because these different types of passengers have different behaviours in for example arrival time at the airport before departure, amount of luggage carried with them, shopping behaviour in the retail area or travel time.

4.2. Terminal

The terminal area is after the landside area the second area where departing passengers go to. The processes going on include check-in, security controls, passport controls and proceed to gate through retail area (Schulz 2010). In Figure 9 a diagram of the diverse processes according to Schulz (2010), only translated to English can be seen.

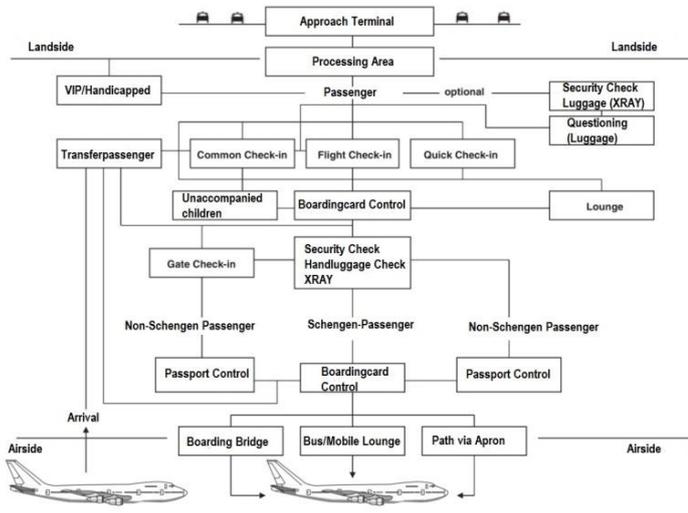


Figure 9: Processes in the Terminal (Schulz 2010)

These are dependent on specific features of the passenger, like if he is *business* or *tourist* (only hand luggage or not), or what his destination is (if within Schengen, then the passenger can proceed without passport control) or if he is handicapped or not. This submodel also includes *transfer* passengers, meaning passengers arriving at the airport by plane, going through passport control if necessary and proceeding to gate after going through retail area. This circumstance shows on the one hand, that this submodel gets input from the landside as well as from the airside and if some delays or other effects happen in the parts of the airport not represented in the terminal submodel it has an effect on the terminal submodel.

The research question in this model is if resources like personnel and number of open counters is enough at each time to maintain the quality standards measured in waiting time of passengers. This being a simple server-queue question is modelled best using **Discrete Events** with counters and personnel being resources and passengers being entities. Like mentioned in 4.1 there is a differentiation in passengers: business and tourist.

A first version of a DES model has already been implemented, as seen in Figure 10. This model includes the basic servers in such a process (check-in, security, passport control and transfer) and distinguishes between tourist and business passengers. In AnyLogic simple blocks for creating (sources), processing (server), and queuing (queue) entities are used. Resources are created via a (scheduled) resource pool. Here, sources and sinks (Exit) build the interface to an adjacent submodel where entities are led through the system.

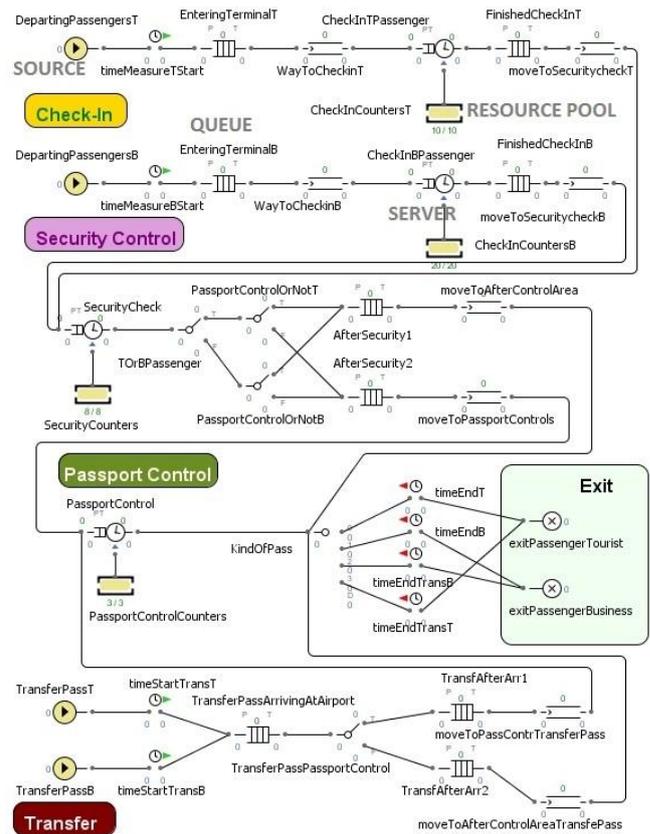


Figure 10: Example of an Implementation in AnyLogic of the Terminal Model.

Now a simple agent based model modelling the landside can be coupled by this rule: every time an agent (passenger) exits the landside model via a port, an entity (it is possible to pass other information as well) in the DES model is generated. On the other side the exit can be seen as an interface as well. Passengers proceed to the gate through the retail area. Every time an entity enters the sink in the DES model an agent will be created in the ABM model of the retail area (which will be explained in the next section).

4.3. Retail Area of an Airport

The retail area is economically seen a very important part of the airport since a large part of the profit is gained by the retail area. The retail area is a shopping area after having passed the controls in the terminal where passengers go through when they proceed to the gate to depart. One main research question in this area is to maximize profit by guaranteeing a specific level of quality standard for passengers like a short way to the gate or attractive sales. What we propose here is a multi-method model itself: **an AB submodel that includes spatial information** (map of the shops) where passengers walk through the retail area as agents. The environment is the retail area with the shops. **Each agent representing passengers consists**, next to some individual parameters, **of an SD submodel** that models the *need to eat* and the *need to buy* other things. In this

case the retail submodel is itself an **integrated form of agents with rich internal structure** as seen in Figure 6. The agent still follows some rules like:

- proceed to gate in time
- if hungry and still time to departure, then look for eating store and eat
- in dependence of attractiveness of store and in dependence of estimated income buy something if there is still time to departure and the need to buy something exceeds a specific threshold

These rules always take into account the by the SD model calculated *need* to buy something. This means there is communication from the SD module to the agent based module (tell him where to go). In return the need is dynamically calculated by the SD module by using individual information from the agent (age, gender, time until departure), but also using information from the environment of the AB module, like the attractiveness of the store that has some “basic attractiveness” and furthermore is calculated by the number of people inside (if no people are inside it may be something wrong with it, if too much people are inside it is overcrowded).

A first version of this submodel is in development (work in progress by M. Obermair and B. Glock), as seen in the 3D version of the animation in Figure 11 implemented in AnyLogic 7. A network is applied to a groundfloor and passenger agents interact on this plan with the shops trying to satisfy their goals.

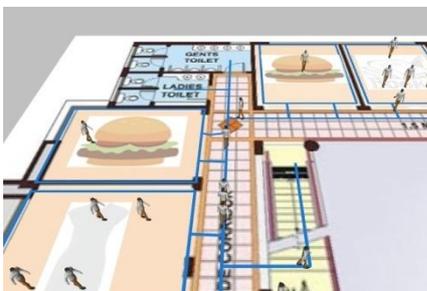


Figure 11: Retail Area modelled with Agents (Screenshot by M. Obermair).

4.4. Airside

The airside of the airport is the part where diverse processes take place that deal with outgoing and incoming planes, like it is shown in Figure 12.

The so called ground handling processes include all processes around the plane. After touchdown (landing of the plane), the taxi arrives to get the passengers. After that the unloading, deboarding and water refillment starts. There are some limitations like cleaning and catering have to start after deboarding or fuelling after unloading luggage that need to be considered as well. In this submodel the spatial context plays an important role since travelling times contribute to quality measurements for passengers and the

calculation of optimizations regarding the ground handling process itself.

The amount of flights is increasing and space on the airside where passengers can board (directly through the gate via boarding bridges or on the apron) is limited.

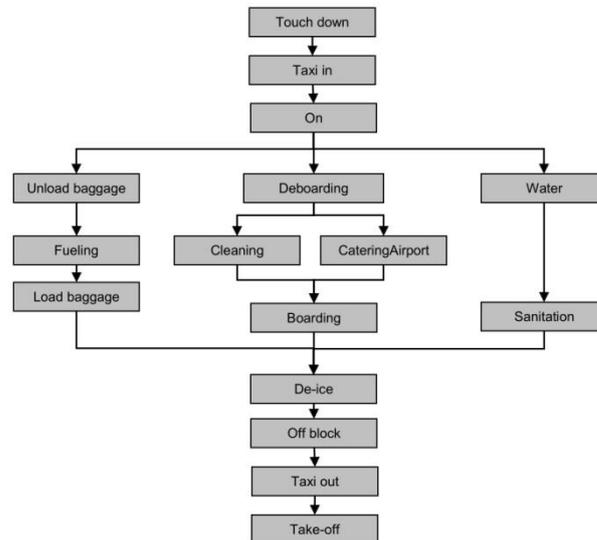


Figure 12: Ground Handling Process (Norin 2012)

Therefore, in this case an **AB submodel** is suggested with a given **network** on which the agents can operate, as seen in Figure 13.

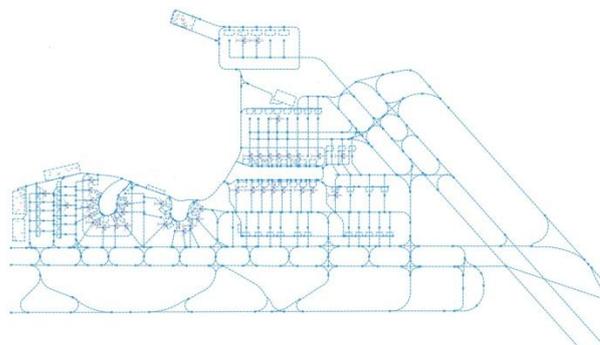


Figure 13: Example of an airside network of the Vienna International Airport.

Different types of agents interact with each other on an environment (network):

- Planes
- Mobile stairs
- Catering vehicle
- Belt loader
- Baggage cart
- Container/pallet dolly
- Container loader
- Tractor
- Lavatory service vehicle
- Refuelling vehicle

- Dispenser vehicle
- Ground power unit
- Air conditioning unit
- Pushback tug

They have one overall goal to get the plane as soon as possible up in the air again.

4.5. Coupling of the modules

Summarizing the proposed submodels as seen in Figure 14 there is a terminal submodel in DES that interacts with the retail submodel in ABM, which is itself a multi-method model due to the “brain” of each agent being a SD submodel. The retail submodel is connected to the airside submodel. If we model only one way of passengers, the departing passengers this would be a sequential approach. If we include a landside submodel and furthermore include passengers arriving at the airport by plane on the airside that then proceed to the landside we would call it an integrated model, because these submodels would have to exchange information (eg. the agent or entity being passed on) as well.

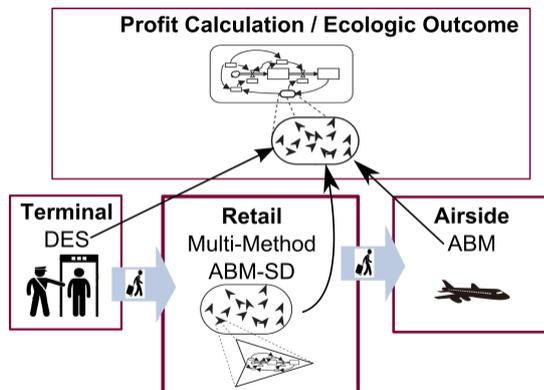


Figure 14: Proposed Multi-Method Model of Parts of the Airport.

Another further possibility as seen in Figure 14 is to build a hierarchically higher SD model that uses input from the “lower” submodels to calculate ecologic outcomes or profit calculations as well.

5. CONCLUSION AND OUTLOOK

Using multi-method models is getting more important every day, since systems and structures get more complex and larger. Errors by using only one method for a large system can accumulate over time and make decision making much more difficult. By using different (best fitting) modelling methods for different subsystems and utilizing all their advantages on the one hand a *more realistic presentation* of the multi-method model can be created (what makes communication to decision makers easier) and on the other hand *accumulating errors can be eliminated* to some extent. Furthermore, *calculation times of the simulation models can be reduced*: If a modeller would have used for example an AB model to model the whole system, but with multi-method modelling now uses only a small AB

module where necessary and in other parts of the model a DES model or a SD model where not so much detail or individual behaviour is needed, the calculation time of simulating the large system can be reduced significantly. In future work of this project the proposed submodels will be finished implementing to get a glance on what may be possible in modelling large infrastructure systems and how methodologies can be refined, researched and made better.

ACKNOWLEDGMENTS

This project is a dissertation within *Talente* funded by BMVIT Ministry of Transport, Innovation and Technology and handled by FFG Austrian research Promotion Agency.

REFERENCES

- ACRP Airport Cooperative Research Program, 2010. ACRP Report 25, Airport Passenger Terminal Planning and Design – Volume 1: Guidebook. Washington, D.C.: Transportational Research Board.
- Bonabeau E., 2002. Agent-based modelling: Methods and techniques for simulating human systems. Arthur M. Sackler Colloquium of the National Academy of Sciences “Adaptive Agents, Intelligence, and Emergent Human Organization: Capturing Complexity through Agent-Based Modeling“. October 4-6, Irvine, CA.
- Bösch R., 2013. VIE Metamorphosis. The Extension and Transformation of Vienna Airport. Wien: Springer.
- Fishwick P.A., 2012. Hypermodelling: an integrated approach to dynamic system modelling. *Journal of Simulation* 6:2-8.
- Forrester J.W., 1973. *World Dynamics*. Cambridge, Mass.: Wright-Allen Press.
- Forrester J.W., 1997. *Industrial Dynamics*. Portland: Productivity Press.
- Gerber A., 2006. *Gesundheitsökonomie und Pädiatrie*. Schattauer.
- ICAO International Civil Aviation Organization. Predicts, 2015. Available from: <http://www.icao.int/Newsroom/Pages/ICAO-predicts-continued-traffic-growth-through-2015.aspx> [Accessed: 14th May 2015]
- Lättilä L., Hilletoft P., Lin B., 2010. Hybrid Simulation Models – When, Why, How?. *Elsevier Expert Systems with Applications* 37:7969-7975.
- Macal C.M., North M.J., 2010. Tutorial on agent-based modelling and simulation. *Journal of Simulation* 4:151-162.
- Maurer P., 2002. *Luftverkehrsmanagement - Basiswissen*. Oldenbourg: Oldenbourg Wissenschaftsverlag.
- Neufville R. De., Odoni A., 2013. *Airport Systems – Planning, Design and Management*. 2nd Edition. United States: The McGraw-Hill Education LLC.
- Nixdorf, B., 2002. *Discrete Event Modelling and Control of Hybrid Systems*. Düsseldorf: VDI-Verl.

- Norin A., Granberg T.A., Yuan D., Värbrand P., 2012. Airport logistics – A Case Study of the turn-around process. *Journal of Air Transport Management* 20:31-34.
- Preston White K., Ingalls, R., 2009. Introduction to Simulation. In: *Proceedings of the 2009 Winter Simulation Conference*, pp.12-23.
- Sargent R.G., 1994. A Historical View of Hybrid Simulation/Analytic Models. In: *Proceedings of the 1994 Winter Simulation Conference*, pp. 383-386. December 11-14, Orlando, Florida.
- Schieritz N., Milling P.M., 2003. Modeling the Forest or Modeling the Trees – A Comparison of System Dynamics and Agent-Based Simulation. In: *Proceedings of the 21st International Conference of the System Dynamics Society*. July 6-9, Waikoloa, USA.
- Scholl H.J., 2001. Agent-based and System Dynamics Modelling: A Call for Cross Study and Joint Research. In: *Proceedings of the 34th Hawaii International Conference on System Sciences*. January 3-6, Maui, USA.
- Schulz A., Baumann S., Wiedenmann S., 2010. *Flughafen Management*. München: Oldenburg Verlag.
- Sterman J.D., 2000. *Business Dynamics: Systems Thinking and Modelling for a Complex World*. Boston, Mass: Irwin McGraw-Hill.
- Sterzenbach R., 1996. *Luftverkehr – Betriebswirtschaftliches Lehr- und Handbuch*. München: Oldenburg Verlag München Wien.
- Swinerd C., McNaught K.R., 2012. Design Classes for Hybrid Simulations Involving Agent-based and System Dynamics Models. *Elsevier Simulation Modelling Practice and Theory* 25:118-133.
- Zeigler B., Praehofer H., Kim T.G., 2000. *Theory of Modelling and Simulation – Integrating Discrete Event and Continuous Complex Dynamic Systems*. 2nd Edition. London:Academic Press.

AUTHORS BIOGRAPHY

Barbara Glock studied Technical Mathematics in Computer Sciences and has done her Master's degree on *Modelling Obesity in Austria with System Dynamics*. She is currently working on her PhD also in the field of modelling and simulation on *Multi-Method Modelling of Large Systems – Use Case: Airports*.

Niki Popper is CEO of dwh GmbH and coordinator of *dexhelpp*, a COMET K-Project funded by BMVIT, BMWFJ and City of Vienna in the Framework COMET – Competence Centers for Excellent Technologies. *dexhelpp* addresses decision support for health policy and planning based on existing health care data by developing methods, models and technologies. Niki Popper has long experience in modelling and simulation of complex systems.

Felix Breitenecker is professor at the TU Wien, head of the group of mathematical modelling and simulation and scientific head of *dexhelpp*. He is active in various modelling and simulation societies: past president of EUROSIM, board member and president of the German Simulation Society ASIM. Felix Breitenecker covers a broad research area, from mathematical modelling to simulator development, from discrete event simulation to symbolic computation.

PERFORMANCE GAIN AND ENERGY SAVINGS USING MULTI-THREADED SOFTWARE ON AN EMBEDDED ASYMMETRIC CHIP MULTIPROCESSOR

Michael Bogner^(a), Christof Steps^(b), Franz Wiesinger^(c)

^{(a),(b),(c)}University of Applied Sciences Upper Austria – Embedded Systems Design
Softwarepark 11, A-4232 Hagenberg, Austria

^(a)michael.bogner@fh-hagenberg.at, ^(b)christof.steps@fh-hagenberg.at, ^(c)franz.wiesinger@fh-hagenberg.at

ABSTRACT

More powerful computers, as well as the rising costs for energy require modern computing systems to be more and more energy efficient. A modern approach in order to solve this issue is the usage of multi-core processors instead of single-core processing units.

This paper focuses on modern ARM multi-core processors, which combine powerful and energy efficient processor cores on a single chip, in order to reduce power consumption and increase performance. To verify how much energy can be saved and how much performance can be gained by using multi-threaded software, we have simulated a various number of different calculation scenarios for a modern ARM System-on-Chip.

The results of these simulations show notably, how advantageous the usage of multi-core processors and power saving processors together on a single chip is. By using intelligent multi-threading concepts, the processor is able to reduce its power consumption by up to 60% compared to a single-threaded execution.

Keywords: multi-core, parallel software, power consumption, embedded asymmetric processor

1. INTRODUCTION

Modern computing systems, especially mobile devices, are getting more powerful every year, which results in an increase in power consumption. Furthermore, modern batteries still don't provide an acceptable power capacity and energy costs are constantly rising. Considering all these factors, processor manufactures tend to develop processors that are more energy efficient.

But being energy efficient alone is not the way to go, if the processors performance is lacking. A modern mobile device for example spends 80% of the time in an idle or standby mode, where only small background tasks are being executed. On the other hand, tasks that require a lot of performance should still be executed as quickly as possible. This trade-off between a low power consumption when doing light work but still providing a great performance when needed is the current aim for modern processor manufacturers.

One way to achieve this is to use multi-core processors. By using more than one processor core, power consumption can be significantly reduced by sharing work across all available cores. Those cores can run at a much lower clock speed and therefore use less power than a single-core processor, whilst still providing the same or even more performance.

Another way to be more energy efficient is to use both high performance and low power processor cores together on a single chip. With this architecture, the processor can save a lot of energy during low performance tasks, but still provide maximum power with its high performance processor cores. An example for this architecture is the NVIDIA Tegra 3 processor. It uses 4 high performance ARM Cortex-A15 processor cores and one low power ARM Cortex-A7 processor core on one chip. This setup is called 4-PLUS-1 (Nvidia Corporation 2011). The advantage of using these cores together is remarkable: During high performance tasks, the Cortex-A15 quad-core unit provides maximum power. While the processor is in idle mode or not doing any heavy work at all, its power consumption can be reduced dramatically by switching to the low power Cortex-A7 core. However, both types of processor cores cannot work together at the same time. This means that the processor never has more than four active cores, because either the Cortex-A15 quad-core, or the single Cortex-A7, so-called "companion core", is active. The operating system itself only sees 4 logical processor cores, without knowing if it's the single companion core or the four high-performance cores (Nvidia Corporation 2013).

The advantages of these setups are huge: By reducing the power consumption, mobile devices will have longer battery lives, less charging cycles and of course smaller batteries, which lead to a thinner and lighter device. Especially modern "internet of things" devices as wells as intelligent embedded systems will profit a lot from a reduced power consumption, because they are independent from constant power supply.

The objective of this study was to simulate different workloads for the NVIDIA Tegra 3 processor, in order to figure out the best way using the processor energy efficiently. This includes single- and multi-threaded

applications as well as simulating heavy workloads and measuring the processors performance, energy consumption and processor allocation. The results are then being compared and discussed.

2. RELATED WORK

Energy efficiency and performance especially on mobile devices have become increasingly important throughout the last years. As a result, many works were published throughout the last years, dealing with multi-threading to reduce power consumption (Youssef, A. et al. 2010, Grant et al. 2006, Arm Ltd. 2013). Those works however differ from this paper, as they are using simulators for the hardware, instead of real hardware and are taking a more theoretical approach. Focusing on using different processor cores on the same chip, there has also been done a lot of work in the past (Nvidia Corporation 2011, Nvidia Corporation 2013, Arm Ltd. 2013). Those works again have a more theoretical approach, as they are lacking tests with real hardware, which is the focus of this paper. Other tests using modern processors for mobile devices as well as real software prototypes have not been conducted yet.

3. TEST ENVIRONMENT

Power consumption and energy efficiency are very important in all modern computing systems, including mobile devices as well as desktop computers. As mobile devices suffer the most from a high power consumption, due to limited battery capacity, these tests focus on an NVIDIA Tegra 3 processor. The Tegra 3 processor is, for example, part of an HTC One X Phone, which was used for these tests.

The used operating system is Android and the simulation software is developed with the Java programming language.

3.1. Clock Speed and Processor Utilization

In order to evaluate the results of the simulation tests, an application to measure the processors clock speed as well as its utilization had to be developed. Unfortunately, the operating system itself does only see four processor cores, even though there are five logical cores. In order to overcome this issue and get the usage and clock speed of all five cores, we took a closer look at the processor cores clock speed: while the high-performance Cortex-A15 operates at a clock speed between 1000 and 1500 MHz, the low-power companion core operates at a clock speed of around 350-750 MHz. By taking this fact into consideration, we can easily determine which processor cores are currently active. For example, if there is only one active processor core, and its clock speed is below 750 MHz, we can assume that we are currently operating on the low-power companion core. On the other hand, if there are more than one cores active, or all of the active cores clock speed is above 1000 MHz, we can assume that the high-performance quad-core is currently enabled.

The processors total utilization can be read from the Android system file “/proc/stat”, as seen in Listing 1. The four values represent time units in which the processor was in a certain state since the device had started. The first value represents the time the processor spent in the user state, then system state, nice state and finally idle state.

Listing 1: Content of the system file /proc/stat
1 cpu 8000 2000 1000 9000

To calculate the relative time the processor was busy and idle, two measurements have to be made and the values have to be subtracted, as shown in Table 2. The Subtracted values now show the relative time the processor spent in a certain state over the last 0.5 seconds.

Table 1: Two measurements of the CPU utilization and the subtracted values over the last 0.5 seconds

	User	System	Nice	Idle
Measure 1	8000	2000	1000	9000
Measure 2	9500	2500	1500	9500
Subtracted	1500	500	500	500

This means that during the last 0.5 seconds the processor spent a total of 3000 time units either working or idle. Using this value and the time the processor spent in each state, the relative processor utilization can finally be calculated:

$$\text{Idle: } 500 / 3000 = 16 \% \quad (1)$$

$$\text{Busy: } 2500 / 3000 = 83 \% \quad (2)$$

To avoid influencing the systems performance by reading the processor utilization too often, the calculations are done every 0.5 seconds. By doing so, the values are still representative and don't influence the systems performance. Figure 1 shows the application on the home screen. The application always stays on top of every other application and shows the clock speed as well as the utilization in real time, updating every 0.5 seconds.



Figure 1: The measuring application on the homescreen.

3.2. Power Consumption

In order to evaluate the power consumption of the processor, an application called “PowerTutor” was used

(Lide Zhan et al. 2010). This application shows the power that certain parts of the device use, such as the processor itself, the Wi-Fi connection or the display. It calculates the processors power consumption approximately using the operating voltage and the current utilization. Figure 2 shows a screenshot of the application. The green chart represents the power consumption of the processor over the last 60 seconds.

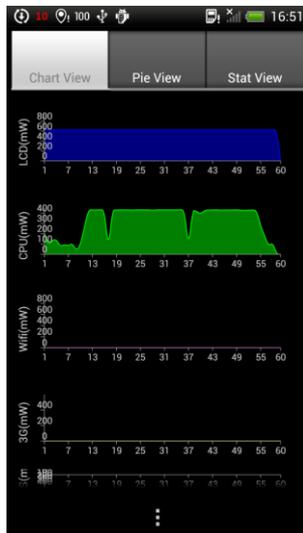


Figure 2: The PowerTutor application.

3.3. Objectives

The objective of this paper is to evaluate the processors multi-threading capabilities as well as its energy efficiency in different scenarios. Thus, we developed several tests to measure the processors multi-threading performance as well as its power consumption and efficiency. Firstly, we took a look at the performance that can be gained when using multi-threading, in order to find out if using more threads will result in an overall greater performance. Secondly, we compared the power consumption of the companion core and the high performance cores. This test aims at finding out if doing more work on the companion core, and therefore sacrificing performance for a lower power consumption, is advantageous in certain scenarios. Next, we analyzed the energy that can be saved when using multi-threading. Furthermore, we tested how the processors dynamic load balancing works. The goal of testing the processors dynamic load balancing, which means its efficiency, is to find the best way for an application to be executed: at high speed with maximum performance but a very high power consumption, with reduced speed but an overall lower power consumption, or a mix of both. The following sections evaluate these tests and discuss the achieved results.

4. PERFORMANCE GAIN THROUGH MULTI-THREADING

This test should show the performance gain when doing heavy work on multiple threads, compared to a single-threaded execution.

4.1. Simulation

In order to test the processors multi-core performance appropriately, an application was developed, that simulates heavy workload. This simulation includes prime number calculations and square root calculations, as those operations require a lot of performance.

The algorithm for the prime number calculation was realized using the well-known division method. In order to check if a certain number is a prime number, it is divided by every number up to the square root of the given number. Is the remainder of one of these divisions zero, the number is not a prime number. The algorithm itself is supposed to have a long execution time, in order to simulate a heavy workload and is therefore not very efficient. The square root algorithm simply involves multiple calls of `Math.sqrt()` on a range of numbers.

In order to get appropriate execution times, those calculations are being called on up to one million values. In the single-threaded test, one thread does the whole calculation. When testing the multi-core performance using more than one thread, the workload is split evenly among all threads.

The multi-core performance is measured using the speedup. The speedup is a mathematical formula that describes the relation between a serial and a parallel execution time of a program.

4.2. Results

The results of the multi-core performance tests are shown in Table 2. The execution times of the algorithms are given in seconds. As expected, the results show that the usage of more threads significantly reduces the execution time of the algorithms. Furthermore, using more than four threads does not speed up the calculation any more, even though the Tegra 3 processor has five logical processor cores. This is due to the fact that, as already mentioned, the Tegra 3 does not use all five processor cores at the same time (Nvidia Corporation 2013). The fifth thread, however, will then be running on one of the main cores instead, resulting in a lower performance as with four threads due to oversubscription.

Table 2: Multi-threaded execution times in seconds

Threads	Prime number	Square root
1	6.550	5.821
2	4.013	2.910
3	2.887	2.002
4	2.125	1.498
5	2.201	1.522

Figure 3 shows the achieved speedup. When using two threads, the program executes approximately 150% faster than on one thread. Three threads are more than 200% faster and four threads are around 300% faster than the single-threaded execution. Using five threads or more does not speed up the calculation any more.

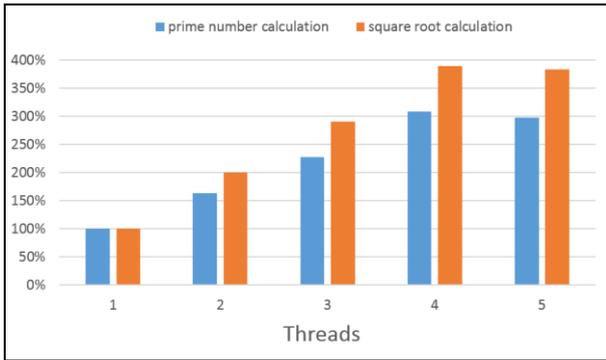


Figure 3: Graphical representation of the achieved speedup in percentage.

5. POWER CONSUMPTION: COMPANION CORE VS HIGH PERFORMANCE CORE

In this test we analyzed the power consumption of the processor in various different scenarios.

5.1. Simulation

The first scenario analyses a single-threaded application with a very low workload. The goal of this is to force the processor to use the companion core. As the operating system does not know anything about the five processor cores (and so the companion core), it can't decide which core to use, because the processor core allocation is done entirely in hardware. But by using a very simple low workload application, the hardware scheduler disables the high performance cores and enables the companion core.

The second scenario analyses a single-threaded application that does a heavy calculation. The goal here is to use one high performance core and measure its power consumption. For this calculation we used the prime number calculation already mentioned in section 4.1.

The last scenario analyses the same calculation as in the second scenario, but the workload is now split among four threads. The power consumption of the processor, as well as the processor core allocation is then being measured for 60 seconds.

5.2. Results

The results of the different scenarios are shown in Table 3. Just as intended, with the low workload, the hardware automatically executed the code on the companion core and therefore saving a lot of power. Over the course of 60 seconds, the processors average power consumption was around 50mW, which can be seen in Figure 4. The power consumption was staggering between 0 and 100mW.

Table 3: The processors power consumption in different scenarios.

Processor	Power Consumption
1x Companion Core @ 475Mhz	50 mW
1x HP Core @ 1500Mhz	380 mW
4x HP Cores @ 1400Mhz	400 mW

The single high performance core, however, needed around 380mW of power on average throughout the 60 seconds, which can be seen in Figure 5. The processors clock speed during the calculation went up to 1500 MHz, which is only possible in single-core mode. When two or more high performance cores are active, their clock speed is limited to 1400 MHz.

As we can see in Table 3, the multi-core power consumption, using all four cores, is around the same as the single-core power consumption. This means that, even though we could increase our performance by around 300% compared to the single-threaded execution, the power consumption still stays the same, no matter how many of the high performance cores are currently active. This is very important point for modern software development: By splitting work among multiple threads, the processor can reduce the clock speed of the processor cores and therefore save a lot of power and still provide the same or even more performance.

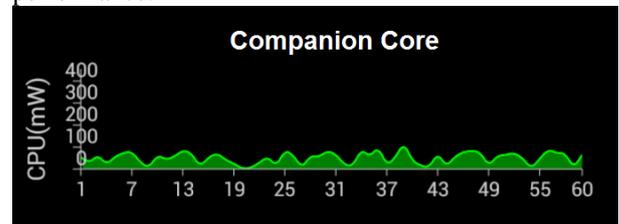


Figure 4: Power consumption of the companion core over 60 seconds.

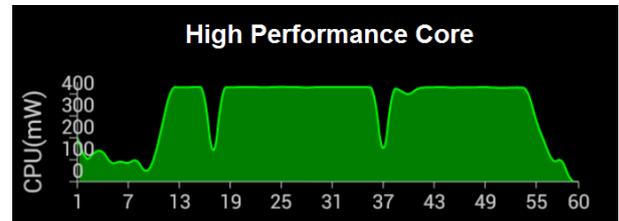


Figure 5: Power consumption of the high performance core over 60 seconds.

6. POWER SAVINGS THROUGH MULTI-THREADING

This test shows how much power can be saved by using a multi-threaded instead of a single-threaded execution.

6.1. Simulation

The simulation involved a prime number calculation up to one million values. The calculation was simulated on one, two, three and four threads, in order to measure the power consumption of the different multi-core scenarios. The workload was split evenly among all available threads. The result we are looking at is the average power consumption over the time that the single-threaded execution took. This means, if the calculation takes 60 seconds on a single-core, we take the average power consumption of these 60 seconds as a reference. Doing the calculation on two cores is faster and therefore takes less time to execute, but we are still taking the full 60 seconds into consideration for the power consumption. This means we are comparing all

scenarios over the full 60 seconds, even though the calculation might finish early in some scenarios.

6.2. Results

As already mentioned, a single-core computation at 1500 MHz has the same power consumption as a multi-core computation at 1400 MHz. Furthermore, we can save power when distributing the workload among multiple cores, as Figure 6 shows. The fact that the multi-core calculation is faster, enables the hardware to disable the high performance cores and enable the companion core earlier. The faster the execution is, the earlier the companion core can be enabled, which saves huge amounts of power, as the results in section 5.2 already showed.

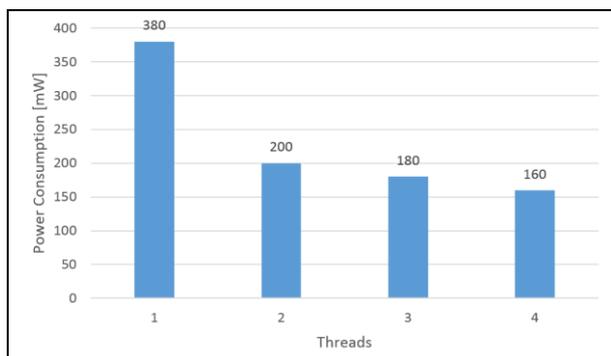


Figure 6: The average power consumption over the time of the execution, with the workload being distributed among all available threads.

Figure 7 shows how long certain processor cores were active during/after the calculation. The single-threaded execution took 60 seconds, so 100% of the time the Tegra 3 processor used the high performance core, which results in a fairly high power consumption.

When using two threads, the calculation only takes half the time and the processor can enable the companion core at around 30 seconds. When using four threads, the calculation finishes after 15 seconds and therefore the processor spends the following 45 seconds on the companion core, saving a lot of power.

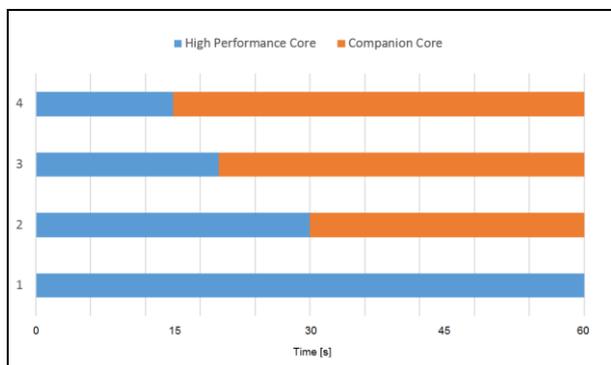


Figure 7: The time the device spent using the high performance cores and the companion core.

When looking at the results in Figure 6, we can clearly see how much power can be saved using more than one

thread. When using two threads, the power consumption can be reduced by 50%, and with four threads it can be reduced by more than 60%.

7. DYNAMIC LOAD BALANCING

This test should show in how far the processor core allocation and the dynamic clock speed can be influenced by software development. We want to find out at which load the hardware scheduler switches from the companion core to the high performance cores, and when the clock speed is increased.

7.1. Simulation

The application for these tests does a square root calculation on a dynamic number of values. Additionally, to dynamically reduce and increase the load, short delays are added between every calculation. The application starts with calculating the square root of 1000 values, including 10 ms break after every calculation. The number of values then has been increased and the delay between the calculations has been decreased after every test, in order to simulate an increasing workload.

7.2. Results

The results of the dynamic load balancing tests can be seen in Table 4. With only 1000 square roots to calculate and 10 ms of breaks in between every calculation, the entire operation is done on the low power companion core. By increasing the number of values and reducing the delay, the hardware scheduler switches to the high performance core with a clock speed of 1 GHz. When further increasing the number of calculations, the processors clock frequency increases to 1.5 GHz with a load of 70%. With 100.000.000 square roots to calculate and no delay, the processors load goes up to 100%, using the maximum performance.

Table 4: Results of the dynamic load balancing test.

Values	Delay	Processor	CPU Load
1.000	10 ms	Companion Core @ 475 Mhz	18%
100.000	100 us	Companion Core @ 475 Mhz	40%
100.000	10 us	Main Core @ 1 GHz	50%
1.000.000	1 u s	Main Core @ 1.5 Ghz	70%
100.000.000	-	Main Core @ 1.5 Ghz	100%

This example shows very well, how the software can influence the processors core allocation as well as the clock speed. By adding short breaks in between calculations or reducing the number of calculations, we can force a complex operation onto the low power companion core, and therefore save a lot of power.

8. CONCLUSION

Based on the results of these tests, a lot of interesting conclusions regarding parallel software development can be made.

Generally, as far as the given algorithms allow it, multi-threading should always be preferred over a single-threaded execution. Using more than one thread on a multi-core processor increases the performance by a lot, and also greatly reduces the power consumption. As the results show, a single-core ARM Cortex-A15 at 1500 MHz clock speed uses the same amount of power as a quad core ARM Cortex-A15 at 1400 MHz. Therefore, it is always better to use multi-threading.

Another factor is dynamic load balancing. When calculation times are shorter, the processor itself can go into an idle state much earlier, and reduce the clock speed drastically or even switch to a low power core to save more energy.

When developing background tasks or tasks that should be executed when the device is in idle or in standby mode, reducing the workload can save a lot of power as well. This can be realized by adding short breaks in between calculations. For example, if you are doing some heavy calculations in a loop, the hardware scheduler will most likely use the high performance core for this operation. But by adding short delays after every loop, we can force the scheduler to stay on the companion core and therefore save power. This can be done if the calculation time is not very important or the result is needed at some time later in the future.

For applications that periodically read data or do calculations, increasing the interval can also reduce the CPU load and therefore force a lower clock speed or different processor cores to be used.

Finally, it can be said that using low power and high performance processor cores together on one chip is definitely the way to go. As the results of these tests show, the advantages that this setup brings to the table are huge, especially for mobile devices. But not only mobile devices profit from a setup like this, the results of these tests can be translated to desktop computers as well. Even though they do not use asymmetric multiprocessing at the moment, dynamic load balancing and multiprocessing in general have been part of desktop processors for years.

All in all, the processor saves a lot of power when not doing any heavy work, but still provides maximum performance when needed. But in order to profit from these technical advantages, the software has to be adapted as well. The software developer has to have knowledge of the underlying architecture in order to profit from all those advantages.

REFERENCES

Lide Zhan et al., 2010, Accurate Online Power Estimation and Automatic Battery Behaviour Based Power Model Generation for Smartphones. Available from:

<http://robertdick.org/publications/zhang10oct.pdf> [accessed 4 May 2015]

Nvidia Corporation, 2013, NVIDIA Tegra 4 Family CPU Architecture. Available from: http://www.nvidia.com/docs/IO/116757/NVIDIA_Quad_a15_whitepaper_FINALv2.pdf [accessed 4 May 2015]

Nvidia Corporation, 2010, The Benefits of Multiple CPU Cores in Mobile Devices. Available from: http://www.nvidia.com/content/PDF/tegra_whitepapers/Benefits-of-Multi-core-CPU-in-Mobile-Devices_Ver1.2.pdf [accessed 4 May 2015]

Nvidia Corporation, 2011, Variable SMP (4-PLUS-1). Available from: http://www.nvidia.com/content/PDF/tegra_whitepapers/Variable-SMP-A-Multi-Core-CPU-Architecture-for-Low-Power-and-High-Performance.pdf [accessed 4 May 2015]

Nvidia Corporation, Nvidia-Tegra-3-Mobilprozessoren. Available from: <http://www.nvidia.de/object/tegra-3-de.html> [accessed 4 May 2015]

Thomas K uneth, Android 4: Apps entwickeln mit dem Android SDK, 2012, Galileo Computing.

ARM Ltd. 2013, Multi-threading technology and the challenges of meeting performance and power consumption demands for mobile applications. Available from: http://www.arm.com/files/pdf/Multi-threading_Technology.pdf [accessed 04.05.2015]

ARM Ltd., ARM Processor Architecture. Available from: <http://www.arm.com/products/processors/instruction-set-architectures/index.php> [accessed 20 March 2014]

ARM Ltd., 2013, big.LITTLE Processing with ARM Cortex-A15 and Cortex-A7. Available from: http://www.arm.com/files/downloads/big_LITTLE_Final_Final.pdf [accessed 20 March 2014]

Youssef, A. et al., 2010 M., On the Power Management of Simultaneous Multithreading Processors, Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.18, no.8, pp.1243,1248, doi: 10.1109/TVLSI.2009.2020727

Grant, R.E.; Afsahi, A. 2006, "Power-performance efficiency of asymmetric multiprocessors for multi-threaded scientific applications," Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International , vol., no., pp.8 pp., 25-29 April 2006

PIPELINE AS HIGH LEVEL PARALLEL COMPOSITION FOR THE IMPLEMENTATION OF A SORTING ALGORITHM

M. Rossainz-López^(a), M. I. Capel-Tuñón^(b), P. Domínguez-Mirón^(a), Ivo H. Pineda Torres^(a)

^(a) Universidad Autónoma de Puebla, Avenida. San Claudio y 14 Sur, San Manuel, Puebla, State of Puebla, 72000, México

^(b) Departamento de Lenguajes y Sistemas Informáticos, ETS Ingeniería Informática y de Telecomunicación, Universidad de Granada, Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain

^(a)rossainz@cs.buap.mx, ^(b)manuelcapel@ugr.es, ^(a)paty.dguez.m@gmail.com, ^(a)ipineda@cs.buap.mx

ABSTRACT

Shown representation by composing Parallel Objects, Pattern communication / interaction PipeLine for implementing a sorting algorithm, through a approach of Structured Parallel Programming using High Level Parallel Compositions or CPANs. Shown the definition, design and implementation of PipeLine as a CPAN under the paradigm of object orientation in order to provide the programmer the ability to reuse the pattern in solving problems, particularly those that deal with management, optimization and information search; CPANPipe are added to a set of restrictions synchronization between processes (maximum parallelism, mutual exclusion, producer-consumer synchronization type). Synchronous communication modes, asynchronous and asynchronous future used. The sort algorithm based on a Pipeline and its design and implementation as shown CPANPipe and finally execution performance is obtained in a parallel 64-processor machine.

Keywords: CPAN Pipeline, Structured Parallel Programming, Parallel Objects, Communication Patterns.

1. INTRODUCTION

Some of the problems of the environments of parallel programming it is that of their acceptance for the users, which depends that they can offer complete expressions of the behavior of the parallel programs that are built with these environments (Danelutto and Orlando 1995). At the moment in the systems oriented to objects, the programming environments based on parallel objects are only known by the scientific community dedicated to the study of the Concurrency. A first approach that tries to attack this problem it is to try to make the user to develop his programs according to a style of sequential programming and, helped of a system or specific environment, this can produce his parallel tally (Akl 1992). However, intrinsic implementation difficulties exist to the definition of the formal semantics of the programming languages that impede the automatic parallelization without the user's participation, for what the problem of generating

parallelism in an automatic way for a general application continues being open (Darlington 1993).

A promising approach alternative that is the one that is adopted in the present investigation to reach the outlined objectives is the denominated structured parallelism (Capel and Troya 1994). In general the parallel applications follow predetermined patterns of execution. These communication patterns are rarely arbitrary and not structured in their logic (Brinch Hansen 1993, Brinch Hansen 1994). The High Level Parallel Compositions or CPANs are patterns parallel defined and logically structured that, once identified in terms of their components and of their communication, they can be taken to the practice and to be available as abstractions of high level in the user's applications within an environment or programming environment, in this case the one of the orientation to objects (Rossainz and Capel 2008, Rossainz 2005). A CPAN has the following properties that can be studied in detail in (Rossainz, Pineda and Dominguez 2014):

1. Capacity of invocation of methods of the objects that contemplates the asynchronous communication ways and asynchronous future (Danelutto and Orlando 1995). The asynchronous way doesn't force to wait the client's result that invokes a method of an object. The asynchronous future communication way makes the client to wait only when needs the result in a future instant of her execution. Both communication ways allow a client to continue being executed concurrently with the execution of the method (parallelism inter-objects).
2. The objects must can to have internal parallelism. A mechanism of threads it must allow to an object to serve several invocations of their methods concurrently (parallelism intra-objects).
3. Availability of synchronization mechanisms when parallel petitions of service take place. It is necessary so that the objects can negotiate several execution flows concurrently and, at

the same time, to guarantee the consistency of their data.

4. Availability of flexible mechanisms of control of types. The capacity must be had of associating types dynamically to the parameters of the methods of the objects. It is needed that the system can negotiate types of generic data, since the CPANs only defines the parallel part of an interaction pattern, therefore, they must can to adapt to the different classes of possible components of the pattern.
5. Transparency of distribution of parallel applications. It must provide the transport of the applications from a system centralized to a distributed system without the user's code is affected. The classes must maintain their properties, independently of the environment of execution of the objects of the applications.
6. Performance. This is always the most important parameter to consider when one makes a new proposal of development environment for parallel applications. An approach based on patterns as classes and parallel objects must solve the denominated problem PPP (Programmability, Portability, Performance) so that it is considered an excellent approach to the search of solutions to the outlined problems.

With the basic set of classes of the model of programming of PO they are possible to be constructed concrete CPANs. To build a CPAN, first it should be had clear the parallel behavior that one needs to implement, so that the CPAN in itself is this pattern. Several parallel patterns of interaction exist as are the farms, the pipes, the trees, the cubes, the meshes, the matrix of processes, etc. So you know what pattern is best suited to a particular application is an important design decision that cannot be fully automated. Once identified the parallel behavior, the second step consists on elaborating a graphic of its representation as mere technique of informal design of what will be later on the parallel processing of the objective system; it is also good to illustrate its general characteristics, etc., and it will allow later to define its representation with CPANs, following the pattern proposed in the next section.

When the model of a CPAN is already had concretized, that defines a specific parallel pattern; say for example, a tree, or some of those previously mentioned ones, the following step would be to carry out its syntactic definition and semantics.

Finally, the syntactic definition previous to a CPAN programmed is translated in the most appropriate programming environment for its parallel implementation. It would be verified that the resulting semantics is the correct one, it would be proven with several different examples to demonstrate its genericity and the performance of the applications would be observed that include it as component software.

The parallel patterns worked in the present investigation have been the pipeline, to be a significant reusable pattern in multiple applications and algorithms. Being used at the moment with different purposes, in different areas and with different applications according to the literature that there is on the topic (Cole 1989, Darlington 1993). In the present investigation the pipeline is constructed as a CPAN to implement a sorting algorithm and solving the problem of generating parallel an increasing sequence of numbers from lowest to highest. Finally performance shown considering measures such as the execution time in seconds, cycles per instruction (CPI), speedup and Amdahl's Law.

2. HIGH LEVEL PARALLEL COMPOSITIONS (HLPC)

A CPAN comes from the composition of a set of objects of three types (see Figure 1): An object manager that it represents the CPAN in itself and makes of him an encapsulated abstraction that it hides their internal structure. The manager controls the references of a set of objects (a denominated object Collector and several denominated objects Stage) that represent the components of the CPAN and whose execution is carried out in parallel and it should be coordinated by the own manager. The objects Stage that are objects of specific purpose, in charge of encapsulating an interface type client-server that settles down between the manager and the objects slaves (objects that are not actively participative in the composition of the CPAN, but rather they are considered external entities that contain the sequential algorithm that constitutes the solution of a given problem). And an object Collector that it is an object in charge of storing in parallel the results that he receives of the objects stage that has connected. That is to say, during the service of a petition, the control flow within the stages of a CPAN depends on the implemented communication pattern.

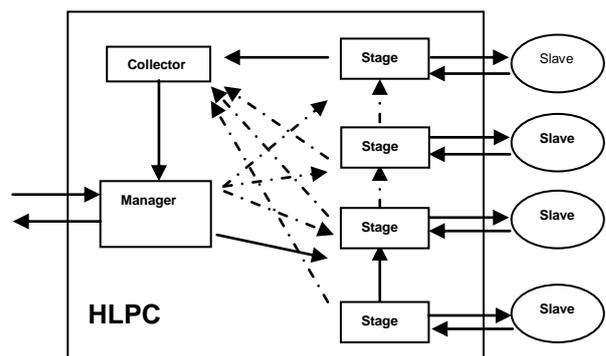


Figure 1: Internal Structure of a HLPC. Composition of its components

The objects manager, collector and stages are included within the definition of Parallel Object (PO) (Corradi 1991). The Parallel Objects are active objects, that is to say, objects that have execution capacity in them (Corradi 1991). The applications within the pattern PO can exploit the parallelism so much among objects

(inter-object) as the internal parallelism of them (intra-object). An object PO has a similar structure to that of an object in Smalltalk, but it also includes a politics of scheduling, determined a priori that specifies the form of synchronizing an or more operations of a class in parallel (Danelutto and Orlando 1995, Corradi 1991). The synchronization policies are expressed in terms of restrictions; for example, the mutual exclusion in processes readers/writers or the maximum parallelism in processes writers. All the parallel objects derive then of the classic definition of “class” more the incorporation of the synchronization restrictions (mutual exclusion and maximum parallelism) (Birrel 1989). The objects of oneself class shares the same specification contained in the class of which you/they are instantiates. The inheritance allows deriving a new specification of one that already exists. The parallel objects support multiple inheritances.

```
CLASS: <class name>
  INHERITS_FROM: <list of class names parents>
  INSTANCE_STATE: < list the names and
                  types of instance>
  INSTANCE_METHODS: <list of public methods>
  PRIVATE_METHODS: <list of private methods>
  SCHEDULING_PART: <list of synchronization
                  restrictions>
END_CLASS <class name>
```

2.1. The abstract class ComponentManager

It defines the generic structure of the component manager of a CPAN, of where will be derived all the manager concretes depending on the parallel behavior that is contemplated for the creation of the CPAN. All concrete instance of a manager accepts as input a list of n-associations. 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 execute through one of their methods.

Once the manager has captured the list of n-associations, this generates the concrete stages, one for each association and then each stage becomes in responsible for an object slave, together with the execution method of this. In turn, the stages are connected to each other, in accordance with the parallel pattern that has been implemented in the CPAN.

Finally the manager carries out the processing of a computation, through the execution of one of his methods (execution). For it is necessary to pass to the method the data input with those that one wants to work. The manager generates a component collector then and it passes to the stages the reference to this collector, as well as the data input. The stages processes them and, according to as be connected some with other, will spend the results that they go obtaining. At the end the collector will pick up the results of the stages to return them to the manager who finally will transmit the results to the exterior of the CPAN (to the code of user's application that uses it). The following

syntactic definitions have been written using the grammar free of context that is in the appendix A of the present document.

```
CLASS ABSTRACT ComponentManager
{
  ComponentStage[ ] stages;

  PUBLIC VOID init (ASOCIACION[] list )
  { ABSTRACT;}

  PUBLIC ANYTYPE execution(ANYTYPE datain)
  VAR
    ComponentCollector res;
  {
    res = ComponentCollector CREATE( );
    commandStages(datain,res);
    RETURN res.get( );
  }

  PRIVATE VOID commandStages(ANYTYPE datain,
                             ComponentCollector res)
  { ABSTRACT; }

  MAXPAR (execution);
};
```

The same manager can be used to carry out more calculations in parallel. The synchronization policies used for it is it the restriction of synchronization of the “maximum parallelism” or MAXPAR applied to the method “execution().”

2.2. The abstract class ComponentStage

It defines the generic structure of the component stage of a CPAN, as well as of their interconnections, of where they will be derived all the concrete stages depending on the parallel behavior that is contemplated in the creation of the CPAN. All concrete instance of a stage accepts as input a list of associations slave_object-method to work with her and it is connected or not with the following stage of the list of associations, depending on the parallel pattern that you/they implement, so that when the manager commands to the stages in parallel, each one of them makes the object slave to carry out the execution of his method, the stage captures the results and it sends them to the following stage or the collector, depending on the implemented structure (everything it within a method called request ()). In turn, each stage can command others in the execution of the computation initiate by the manager.

```
CLASS ABSTRACT ComponentStage
{
  ComponentStage[] otherstages;
  BOOL am_i_last;
  METHOD meth;
  OBJECT obj;
```

```

PUBLIC VOID init (ASOCIACION[] list)
VAR
  ASOCIACION item;
{
  item = HEAD(list);
  obj = item.obj;
  meth= item.meth;
  if (TAIL(list) == NULL)
    am_i_last = true;
}

PUBLIC VOID request (ANYTYPE datain,
                    ComponentCollector res)
VAR
  ANYTYPE dataout;
{
  dataout = EVAL (obj, meth, datain);
  IF (am_i_last)
    TREAD res.put(dataout)
  ELSE commandOtherStages (dataout, res);
}
PRIVATE VOID commandOtherStages (ANYTYPE
                                dataout, ComponentCollector res)
  { ABSTRACT;}
MAXPAR (request);
};

```

Again as in the previous case it can have more than a petition of the operation “request ()” being executed in parallel, for what the synchronization policies that must be used for its correct implementation is the one that provides the synchronization restriction of “maximum parallelism”, applied now to the method “request().”

2.3. The concrete class ComponentCollector

It defines the concrete structure of the component collector of any CPAN. This component implements a buffer multi-item basically, where they will leave storing the results of the stages that make reference to this collector. This way one can obtain the result of the calculation initiate by the manager.

```

CLASS CONCRETE ComponentCollector
{
  VAR
    ANYTYPE[] content;

  PUBLIC VOID put (ANYTYPE item)
    { CONS(content, item); }

  PUBLIC ANYTYPE get( )
  VAR
    ANYTYPE result;
  {
    result = HEAD(content[]);
    content = TAIL(content[]);
    RETURN result;
  }
  SYNC(put,get);
  MUTEX(put);
};

```

```

MUTEX(get);
};

```

In this case the used synchronization restrictions are SYNC and MUTEX to be able to synchronize the communication concurrently among the methods *put()* and *get()* and to provide mutual exclusion.

3. THE PIPELINE AND ITS REPRESENTATION AS CPAN

The Pipeline this compound for a set of interconnected states one after another. The information follows a flow from a state to another.

3.1. The technique of the Pipeline

Using the technique of the Pipeline, the idea is to divide the problem in a series of tasks that have to be completed, one after another. In a pipeline each task can be executed by a process, thread or processor for separate (Robbins and Robbins 1999), (Figure 6):

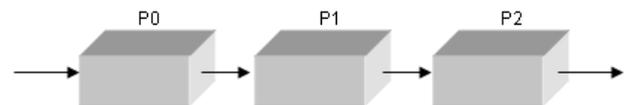


Fig. 2: Pipeline

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.

An algorithm that solves a specific problem can then be formulated as a pipeline if can be divided into a number of functions that could be performed by the stages of pipe (Robbins and Robbins 1999). Suppose, for example, we want to order a disordered set of data from highest to lowest in descending order, but you have already implemented a sorting algorithm ascending; if this sort algorithm were used, it would be necessary to reverse the sequence of already sorted data, which can be carried out by adding an extra stage to the pipeline with an assigned role to carry out the specific process (see Figure 3).

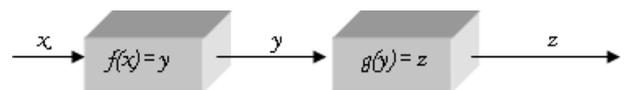


Figure 3: Pipeline seen as functional decomposition

The interpretation of the elements of the previous figure is as follows:

- x represents the initial dataset that we assume is in disarray;

- $f(x)$ represents the function "orders" which receives as input the data set to order and outputs the sort in ascending order (from low to high) dataset that come;
- y represents the output of the function $f(x)$, ie, the data sorted;
- $g(y)$ represents the function "invest" that receives the result of the function "orders" to output the data set previously ordered but reversed in sequence to have a descending order (from highest to lowest);
- z is the output of the function $g(y)$ in the last stage of the pipeline.

Assuming that the data set with which they work in this example are integers, the sequence of results within the pipeline would be as follows, see Figure 4.

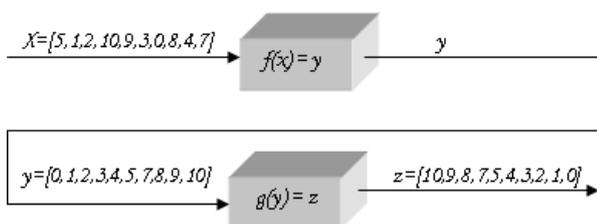


Figure 4: Sequence results in Pipeline

If a problem can be divided into a series of sequential tasks, pipelined approach provides increased speed of execution in the following three types of calculations:

1. When more than one instance of the problem can be run in parallel;
2. Either a series of data can be processed and each of these is used in multiple operations;
3. Or if the information required by the following process to start your calculation happens after the current process has completed all its internal operations.

With this technique many computational problems are performed sequentially can be easily parallelized as a pipeline. Examples of these problems are:

- the sum of numbers,
- the shorting numbers,
- generating prime numbers,
- Solving a system of linear equations.

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

3.2. Representation of the Pipeline as a CPAN

The Figure 5 represents the parallel pattern of communication Pipeline as a CPAN.

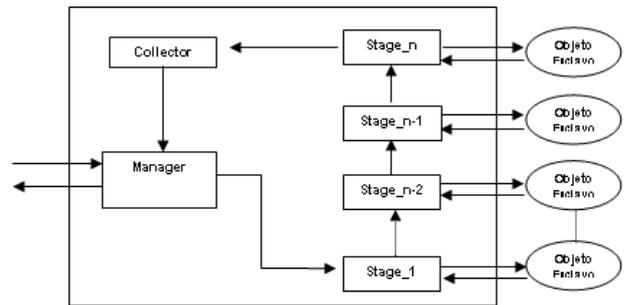


Figure 5: The CPAN of a Pipeline

The objects $stage_i$ and $Manager$ of the graphic pattern of the CpanPipe are instances of concrete classes that inherit the characteristics of the classes ComponentManager and ComponentStage. The Collector is the only object that is an instance of the base class ComponentCollector since this is a concrete class. 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.

3.3. Semantic and Syntactic Definition of the Cpan Pipe

The Cpan Pipe is represented by the class PipeManager that inherits of ComponentManager, and a pattern of communication pipeline implements whose stages is instances of the class PipeStage that inherits of ComponentStage. Any object PipeManager only takes charge of the first stage of the pipeline in its initialization. For details see (Capel and Rossainz 2004). During the execution of a petition of service, the first stage is only commanded.

CLASS CONCRETE PipeManager EXTENDS OF ComponentManager

```
{
  PUBLIC VOID init(ASOCIACION[ ] list)
  {
    stages[0] = PipeStage CREATE( list );
  }

  PRIVATE VOID commandStages (ANYTYPE
  datain, ComponentCollector res)
  {
    THREAD stages[0].request(datain,res);
  }
};
```

The objects of the class PipeStage creates the following stage of the pipeline during its initialization phase. In the execution of their operation request(), an object stage commands directly to the following one and it is the last one that sends the result to the object Collector (instance of the class ComponentCollector) whose reference is transmitted dynamically stage by stage.

```

CLASS CONCRETE PipeStage EXTENDS OF
ComponentStage
{
PUBLIC VOID init (ASOCIACION[] list)
{
stage.init(list);
IF (!am_i_last)
{
otherstages[0] = PipeStage
CREATE(TAIL(list));
}
}

PRIVATE VOID commandOtherStages (ANYTYPE
datain, ComponentCollector res)
{
THREAD otherstages[0].request
(datain, res);
}
};

```

4. 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.

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; Bllloch 1996):

- 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; Bllloch 1996). Suppose that at a certain moment of the execution of the algorithm, the first three stages of the pipeline have a stored integer, while the fourth stage has not yet received any. Suppose the integer 2 is received from the beginning of the pipeline (see figure 6).

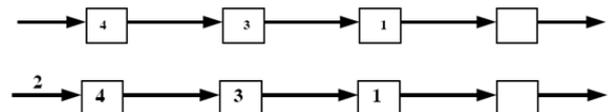
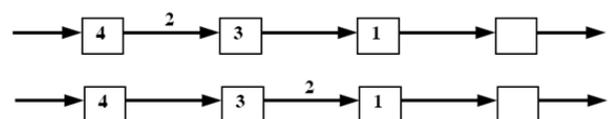


Figure 6: insertion sequence of integers in a pipeline

The first process compares the value received to its stored value, sends the least of them (2) to the next stage of the pipeline and stores the largest (4). Similarly, the second process, after receiving the integer sent by the first (2), sends it to the next stage, keeping stored the integer (3). The third process, however, sends the integer that was stored at the next stage and stores just received (2). The fourth process, not having yet a stored integer, stores the received by the previous stage (see figure 7).



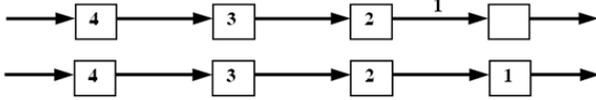


Figure 7: Sort sequence of integers on a Pipeline

The above figure represents the parallel sorting algorithm using a Pipeline. CpanPipe design shown in figure 8.

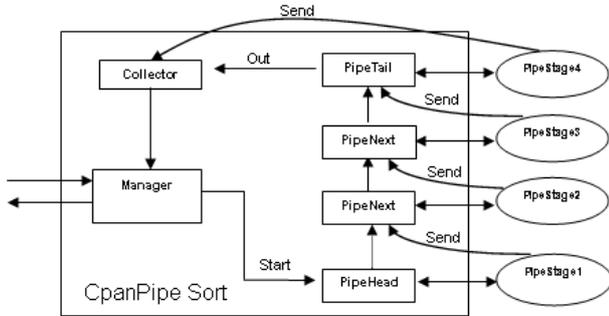


Figure 8: Ordering integers using CpanPipe

5. PERFORMANCE

CpanPipe 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 that solves the problem of sorting is carried out with the following restrictions execution:

- The same sequential algorithm of comparing values in each of the slave objects associated with pipeline stages except the first and last step as described in Chapter IV,
- 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,
- CpanPipe execution for 2, 4, 8, 16 and 32 full-time processors.

Table I and figure 9 show the series of measurements obtained including their corresponding sequential versions, CpanPipe 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 I: Cpan Performance Pipe Parallel to the Management of 50000 Sorting integers

CPAN Pipe	Pipe Secuencial	cpuset2	cpuset4	cpuset8	cpuset16	cpuset32
Runtime in seconds	238.50	127.27	123.83	116.97	115.63	111.03
CPU time in seconds	231.82	224.79	217.80	203.98	198.30	191.89
CPI	1.862	0.909	0.904	0.901	0.886	0.871
Speedup	1.00	1.87	1.93	2.04	2.06	2.15
Amdahl	1.00	1.89	3.39	5.63	8.42	11.19

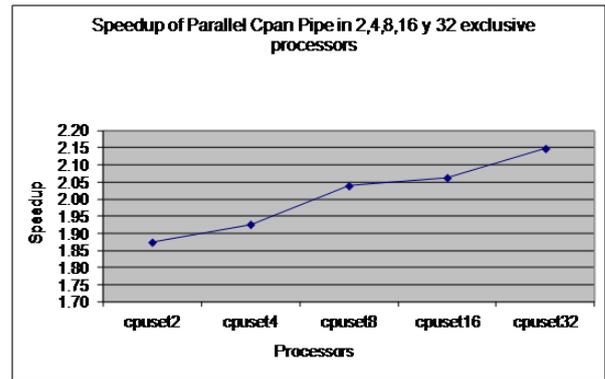


Figure 9: Scaling the magnitude of CpanPipe Speedup for 2, 4, 8, 16 and 32 exclusive processors

6. CONCLUSIONS

It has created a programming methodology based on High Level Parallel Compositions or CPANs (Rossainz and Capel 2008). CPAN Pipeline has been implemented which is part of a class library of parallel objects that provides the user the pattern cited as CPAN and other communication patterns as farms and trees and the details of which are published in (Rossainz and Capel 2012). The CPAN Pipe can be exploited using the approach object-oriented to define new communication patterns based on the already built. For more details see references (Rossainz and Capel 2006, Rossainz and Capel 2012).

Well-known algorithms that solve sequential problems in algorithms parallelizable have transformed and with them the utility of the method has been proven and of the component software developed in the investigation. It has become a sequential sorting algorithm on a parallelizable algorithm using a Pipeline as CPAN. It has implemented the synchronization restrictions suggested by the model Cpan: maximum parallelism (MaxPar), mutual exclusion (mutex) communication and synchronization producer / consumer processes (Sync). Of equal it forms the programming in the asynchronous future communication way for results "futures" within the Cpan it has been carried out in an original way by means of classes.

It has been proven performance Pipe Cpan by metrics Speedup, Amdahl's Law and efficiency to demonstrate that parallel behavior CpanPipe is better than its sequential counterpart.

APPENDIX A. FREE GRAMMAR OF THE CONTEXT FOR THE DEFINITION OF CLASS PO (PARALLEL OBJECTS)

Definición de una clase PO ::= CLASS <tipo de clase> <nombre de clase> [herencia]

```

{
  [<Definición de variables de instancia>]
  [<Definición de métodos públicos>]
  [<Definición de métodos privados>]
  [<Definición de restricciones de sincronización>]
};

```


/MUTEX (<nombre método>, <nombre método>)
 /SYNC (<nombre método>, <nombre método>)
 /MAXPAR(<nombre método>)

REFERENCES

- Akl S.G., 1992. Diseño y Análisis de Algoritmos Paralelos. *Ra-Ma Serie Paradigma*, Madrid.
- 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.
- Brinch Hansen, 1994. SuperPascal- a publication language for parallel scientific computing, *Concurrency: Practice and Experience*, Volume 6, 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.
- Capel, M.; Troya J. M., 1994. An Object-Based Tool and Methodological Approach for Distributed Programming. *Software Concepts and Tools*.
- Capel M., Rossainz, M., 2004. A parallel programming methodology based on high level parallel compositions. *Proceedings of the 14th International Conference on Electronics, Communications and Computers, IEEE CS press*. 0-7695-2074-X.
- Cole, M., 1989. Algorithmic Skeletons: Structured Management of Parallel Computation. *The MIT Press*.
- 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)*.
- 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., 2006. Design and Implementation of the Branch & Bound Algorithmic Design Technique as an High Level Parallel Composition. *International Mediterranean Modelling Multi-conference*. Barcelona, Spain.
- 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., 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.

HEAT TREATMENT PROCESS PARAMETER ESTIMATION USING HEURISTIC OPTIMIZATION ALGORITHMS

Michael Kommenda^(a,b), Bogdan Burlacu^(a,b), Reinhard Holecek^(c),
Andreas Gebeshuber^(c), Michael Affenzeller^(a,b)

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

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

^(c) Rübzig GmbH & Co KG, Schafwiesenstrasse 56, 4600 Wels, Austria

michael.kommenda@fh-hagenberg.at, bogdan.burlacu@fh-hagenberg.at, reinhard.holecek@rubig.com,
andreas.gebeshuber@rubig.com, michael.affenzeller@fh-hagenberg.at

ABSTRACT

We present an approach for estimating control parameters of a plasma nitriding process, so that materials with desired product qualities are created. We achieve this by solving the inverse optimization problem of finding the best combination of parameters using a real-vector optimization algorithm, such that multiple regression models evaluated with a concrete parameter combination predict the desired product qualities simultaneously.

The results obtained on real-world data of the nitriding process demonstrate the effectiveness of the presented methodology. Out of various regression and optimization algorithms, the combination of symbolic regression for creating prediction models and covariant matrix adaptation evolution strategies for estimating the process parameters works particularly well. We discuss the influence of the concrete regression algorithm used to create the prediction models on the parameter estimations and the advantages, as well as the limitations and pitfalls of the methodology.

Keywords: parameter estimation, symbolic regression, genetic programming, heuristic optimization

1. INTRODUCTION

Nitriding is a case-hardening technique where nitrogen is added to the surface of metal alloys to create a thin hardened nitride layer. The increased surface hardness provides high resistance to wear and crack as well as improved sliding and frictional properties and higher fatigue strength values.

The diffusion of nitrogen into the surface of the metal can take place in the presence of ammonia (NH₃) using phosphate activation (salt bath and gas nitriding) or in the presence of plasma (highly ionized gas molecules), where nitrogen ions are accelerated to impinge on the metal surface. Plasma nitriding uses pure nitrogen or a mixture of hydrogen and nitrogen that is ionized in intense electric fields, surrounding the surface to be nitriding with ionized gas molecules.

The advantages of the plasma nitriding compared to salt bath or gas nitriding include:

- Lower process temperature, resulting in very small dimensional deformations and distortions
- Higher surface, case and core hardness
- More precise compared to other methods
- Decreased energy demand
- Non-toxic and environmentally friendly, using H₂ and N₂, with no greenhouse gas emissions

Typically, the treated metals have a predefined set of characteristics that should be reached by the nitriding process:

- Nitriding hardness depth (NHD in mm)
- Surface hardness (SH in HV)
- Thickness of the compound layer (CLT in μm).

The results of the nitriding process depend on the material composition and the process control parameters, such as the temperature, process duration, gas concentrations and the parameters of the plasma ionization. These process parameters have to be adjusted precisely to reach the desired material characteristics. The effect of process parameters on the nitriding results is well studied in the case of gas nitriding and allows a close control of the nitriding results (Mittemeijer 1997). However, such a control mechanism for process parameters to achieve the desired results is not available for plasma nitriding.

In this paper, we present a robust and flexible data-driven heuristic optimization approach for finding the optimal control parameters of a plasma nitriding process. In a first step, we create prediction models for each target characteristic based on the available data containing process parameters and material compositions. Then, we use the generated models to solve the inverse problem of finding combinations of process parameters such that the material characteristics after nitriding reach predefined quality values.

Formally, the inverse optimization problem can be defined as follows: given a (regression) model for each of the target characteristics, and a set of quality values to be achieved, is it possible to find the process parameter values so that for each model the desired output is achieved within a specified confidence bound?

This workflow brings a couple of advantages:

1. It exploits the interpolation abilities of the prediction models so that estimates for unknown data can also be given.
2. It allows us to assess the accuracy of the prediction models (in terms of their estimation of product qualities).
3. The results of the process parameter estimation can be further improved by using more accurate prediction models and/or providing more data for training the prediction models.

The whole workflow for process parameter estimation was implemented in HeuristicLab (Wagner 2014), an open-source optimization environment that offers a wide selection of ready-to-use models and algorithms. Our approach for process parameter estimation supports an arbitrary number of models, as well as any configuration of optimizer, optimization targets and optimization bounds.

The remainder of the publication is organized as follows: In the next Section 2 we give a detailed description of the model generation step and discuss the obtained regression models. Section 3 describes in detail the process parameter estimation methodology, Section 4 discusses the obtained results, while Section 5 is dedicated to the conclusion and final remarks.

2. REGRESSION MODEL GENERATION

The regression models predicting the three target characteristics NHD, SH, and CLT, which are necessary for solving the inverse optimization problem, were obtained using several data-based modeling algorithms on data acquired from a real-world plasma nitriding process. The dataset contains information about the chemical composition of 14 different working materials, 5 different combinations of process parameters and the achieved plasma nitriding results. In total the dataset contains 70 samples and 18 different measurements for each sample. The samples have been divided into a training and test partition such that the distribution of materials and different process parameters is about the same, resulting in 45 training and 25 test samples.

In our experiments, we used genetic programming (GP) with offspring selection (Affenzeller et al. 2009) and constants optimization (Kommenda et al. 2013) to evolve symbolic regression models. The model length was restricted to 25 tree nodes to improve the interpretability of the generated models. Additional experiments were carried out with a fixed model structure, where the GP algorithm was responsible only for finding the correct variables and constant values.

Further on, the models obtained by GP were compared with other regression models obtained via linear regression (LR) (Seber 2012), support vector machines (SVM) (Cortes 1995), Gaussian processes (GPR) (Rasmussen 2004) and random forest regression (RF) (Breiman 2001) in order to determine each algorithm's performance, generalization ability and suitability for process parameter estimation.

For every regression method we used recommended parameter settings and followed best practices such as grid search with crossvalidation where appropriate and multiple repetitions for each method in order to account for stochastic effects. The specific parameter configurations for the optimization methods enumerated above are described in Table 1.

GP	1000 individuals, maximum tree length 25
LR	No parameters
RF	Grid search intervals: <ul style="list-style-type: none"> - Number of trees n between $[2^6, 2^{10}]$ - Feature ratio m in the interval $[0.4, 0.7]$ - Sample size ratio r between $[0.05, 0.66]$
SVM	- NU_SVR regression with an RBF kernel Grid search intervals: <ul style="list-style-type: none"> - Penalty factor C between $[2^{-1}, 2^{12}]$ - Kernel function parameter γ in $[2^{-4}, 2^{-1}]$ - Regularization parameter ν in $[2^{-10}, 1]$
GPR	- Constant mean function - Isotropic squared exponential covariance function

Table 1: Modeling algorithm parameters

The best models for each modeling algorithm and target characteristic were afterwards selected as follows:

- The best genetic programming models were selected according to the lowest mean absolute error and the lowest model complexity.
- Linear regression creates only one model which was selected.
- The best random forest models were selected out of multiple grid search runs according to the lowest out of bag error estimate.
- The best support vector machine models were selected out of multiple grid search runs according to the lowest cross validation error.
- The best Gaussian process models were selected from multiple runs according to the lowest negative log likelihood.

The accuracy of the selected prediction models is stated in Table 2, where for every target characteristic the mean absolute error for the test and training partition is reported. The performance of the different modeling methods varies strongly on the training partition (values inside parenthesis), but with the exception of the linear regression most methods perform similarly on the test partition. This could be explained by the limited amount of data present and we expect that with a larger data basis the modeling accuracies will increase.

	NHD	SH	CLT
GP	0.030 (0.026)	47.65 (51.37)	1.296 (0.772)
LR	0.042 (0.033)	54.42 (42.68)	1.865 (1.140)
RF	0.039 (0.015)	53.36 (17.03)	1.176 (0.627)
SVM	0.031 (0.011)	54.83 (8.370)	1.101 (0.518)
GPR	0.029 (0.014)	52.74 (12.46)	1.208 (0.363)

Table 2: Test and training (inside parenthesis) mean absolute error of the obtained models for each of the target characteristics.

3. PROCESS PARAMETER ESTIMATION

Process parameter estimation is made possible through the interaction of the generated regression models from the previous section with several other components including the dataset, the evaluation function and the optimizer that performs the actual search for real-valued parameter combinations.

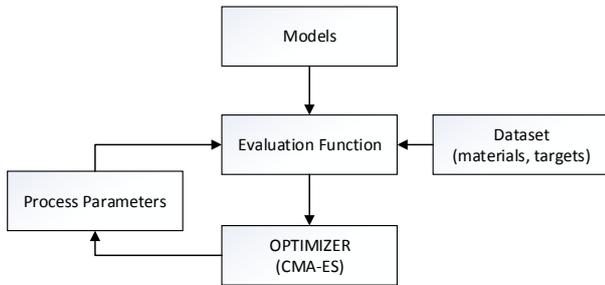


Figure 1: Process Parameter Optimization Flowchart

Figure 1 illustrates the interplay between the individual components when process parameter estimation is performed. The optimization algorithm generates multiple combinations of process parameters within predefined bounds for each parameter. Every parameter combination is evaluated by the aforementioned prediction models including data about the processed material. The deviations between the predicted and desired qualities are passed back to the optimization algorithm, which uses that information to generate new parameter combinations. The optimization algorithm minimizes the deviations by the iterative generation of new process parameter combinations. The procedure is outlined in pseudocode in Algorithm 1.

This general workflow is independent of the concrete data-based modeling algorithm for generating the prediction and the optimization algorithm for tuning the process parameters to reach the desired product qualities. However, we observed that some algorithms are better suited than others for generating the prediction models.

Algorithm 1: Process parameter optimization workflow

Input: material compositions, desired target values T , prediction models M

Output: Optimal parameter combination \vec{p}

- 1 $\vec{p} \leftarrow$ empty parameter vector;
- 2 **while** *optimizer not converged* **do**
- 3 generate new real-valued parameter vector \vec{v} ;
- 4 use models M to get estimated values E with parameters \vec{v} and given material compositions;
- 5 evaluate objective function $f(E, T)$;
- 6 **if** $f(E, T) <$ *best value so far* **then**
- 7 $\vec{p} \leftarrow \vec{v}$
- 8 **return** \vec{p} ;

For example, random forest regression generates highly accurate prediction models, but those models are difficult to use for process parameter optimization as their response can contain instabilities, which makes it harder for the optimization algorithm to find good process parameter settings. In contrast to random forest regression, symbolic regression by genetic programming produces more stable prediction models, especially when the complexity of the models is rather low, and hence it is easier for the optimization algorithm to generate appropriate process parameter settings.

In our experiments three target characteristics (NHD, SH, and CLT) should be obtained by the nitriding process through the estimation of five process parameters: the process duration, temperature, average plasma power, relative amount of nitrogen, and the applied pressure. We used the covariant matrix adaptation evolution strategy – CMA-ES (Hansen 2003) as a real-valued optimization algorithm for generating the estimates for the process parameters.

The objective function to be minimized by the CMA-ES was defined such that each deviation between the desired target value T_i and estimated E_i had an equal influence on the final objective value. This was achieved by scaling the values by the variance of the target characteristic Var_{T_i} over all samples and aggregating the deviations into a single error measure:

$$f(E, T) = \sum_{i=1}^N \frac{(T_i - E_i)^2}{Var_{T_i}}$$

Equation 1: Aggregated error measure

As shown in Figure 1, the evaluation function is arguably the most important component in our optimization approach. In this publication, the aggregated error measure $f(E, T)$ will be used to evaluate the models and investigate their influence on the process parameter estimation.

4. RESULTS

In this section we analyze the influence of the prediction models on the accuracy of the process parameter estimation. We kept the CMA-ES real-valued optimizer (although using other real-valued optimization algorithms is possible as well) for generating the process parameter estimations while exchanging the prediction models for evaluating the process parameters according to the objective function discussed in the previous section and given by Equation 1.

The CMA-ES was set to run for a maximum of 200 generations with a population of 20 individuals, yielding 4000 different process parameter combinations. The algorithm converged in all cases within a third of a second. Although the CMA-ES was able to find parameter combinations that minimized the target for each of the used models, variations due to the scaling of the model targets and the aggregated error measure were present.

The accuracy of the estimated process parameters was evaluated on the test partition of the dataset, which has not been used for model training nor selection. A number of 20 repetitions of the CMA-ES algorithm have been performed to account for the stochasticity of the algorithm and the presence of multiple local optima. Multiple parameter combinations achieve target characteristics equally well, especially when estimating several process parameters simultaneously. Table 3 shows the average absolute deviation of the estimated parameter to the original ones on the training partition and Table 4 shows the same for the test partition.

	GP	LR	RF	SVM	GPR
Duration	5.45	3.75	5.75	2.60	2.66
Temp.	41.72	32.18	31.57	23.14	26.47
Plasma power	16.53	15.52	25.14	16.06	9.69
N ₂	0.15	0.16	0.12	0.11	0.11
Pressure	0.52	0.41	0.53	0.26	0.24

Table 3: Average mean absolute error of the predicted process parameters on the training data.

	GP	LR	RF	SVM	GPR
Duration	4.65	2.91	6.53	4.60	4.18
Temp.	42.05	42.18	35.42	33.55	36.75
Plasma power	22.14	15.64	24.96	13.50	14.43
N ₂	0.19	0.16	0.16	0.17	0.22
Pressure	0.27	0.26	0.62	0.32	0.27

Table 4: Average mean absolute error of the predicted process parameters on the test data.

With the exception of GP and LR, the models perform significantly better on the training partition compared to the test partition. The difference between the training and test qualities of regression models means that some models are slightly overfit on the training set and may exhibit poor interpolation capabilities on new data. Therefore, confidence margins must also be taken into account when evaluating the models, along with their error on the test set.

Confidence margins were calculated by counting how many times the estimated process parameter values were within 10% of the target value. To account for the different value ranges of each process parameter, the deviation was scaled by the range of the parameter and a hit was reported if the scaled deviation is below 10%. Tables 5 and 6 show the 10% hits for each type of models calculated on the training and test partitions.

	GP	LR	RF	SVM	GPR
Duration	28	25	46	34	32
Temp.	26	21	46	30	24
Plasma power	34	36	41	32	40
N ₂	36	23	43	32	30
Pressure	20	25	38	35	34

Table 5: 10% confidence hits on the training data.

	GP	LR	RF	SVM	GPR
Duration	15	15	23	12	13
Temp.	14	9	23	6	10
Plasma power	16	17	19	19	17
N ₂	17	11	18	12	9
Pressure	14	17	17	11	16

Table 6: 10% confidence hits on the test data.

As duration and temperature have the greatest influence on the nitriding outcome, their errors and confidence margins constitute the main criteria for choosing the most suitable models. From this perspective, GP and RF distinguish themselves and will be considered further.

We notice that on the test data RF produces the most confidence hits for both temperature and process duration, with GP on the second place. This result is mainly explained by the large variance of the RF estimation results. The average and standard deviations for the estimations of the process duration and temperature is visible Figures 2 and 3 for GP and in Figure 4 and 5 for RF respectively, where the amplitude of the candlesticks represents the standard deviation of the estimations at the given data row. Therefore, the models produced by GP are more stable and more suitable for the optimization of process parameters.

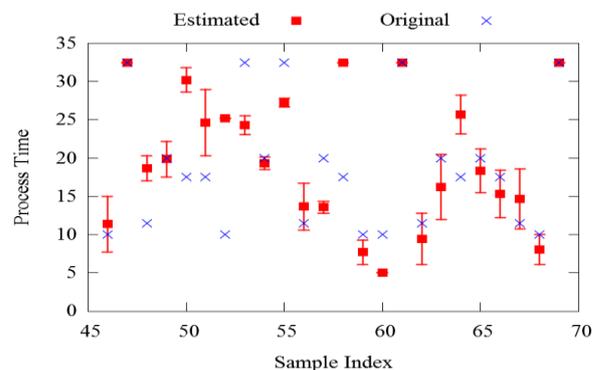


Figure 2: Process duration estimations using GP models on the test data.

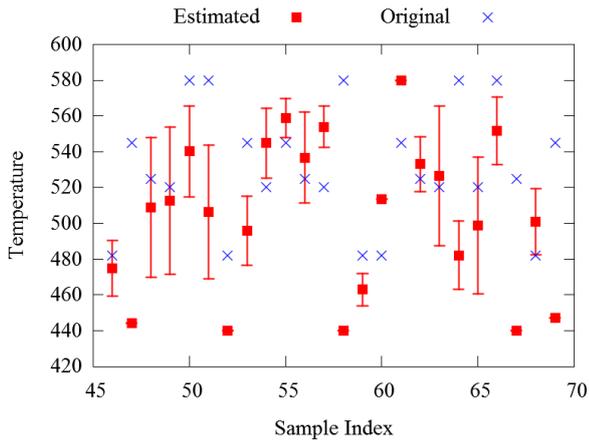


Figure 3: Temperature estimations using GP models on the test data.

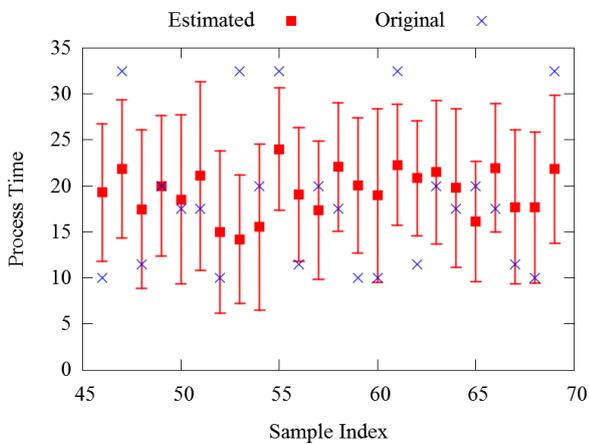


Figure 4: Process duration estimations using RF models on the test data.

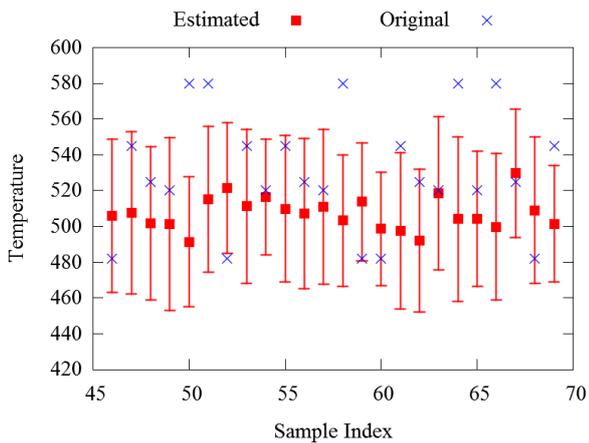


Figure 5: Temperature estimations using RF models on the test data.

Furthermore, the estimation quality of process parameters depends on the number of parameters allowed as inputs to the models. As interdependencies exist between the parameters, more parameters will determine an increased number of local optima and consequently more variance of their values. We illustrate one such interdependency between two antagonistic process parameters:

duration and temperature. From the physical standpoint, a higher temperature allows the nitriding process to complete in less time. Conversely, a lower temperature means that the process will require more time to achieve the desired results.

This relationship should be reflected in the results of our process parameter estimation. To illustrate this, we used the error between the original and the estimated values for duration and temperature. Using the GP models, we notice in Figure 6 that when one of the parameters has a lower value – therefore a negative normal error with respect to the target value, the optimizer compensates by increasing the value of the other. This means that although the estimated values are not entirely accurate on the test data, the optimizer together with the GP models manage to obtain physically valid results, correctly reflecting the inverse physical relationship between duration and temperature. This relationship was not so easily noticeable for the models generated by RF and shown in Figure 7. Therefore, GP models are more suitable for process parameter estimations from both a pure data-mining perspective and from a physical modeling perspective.

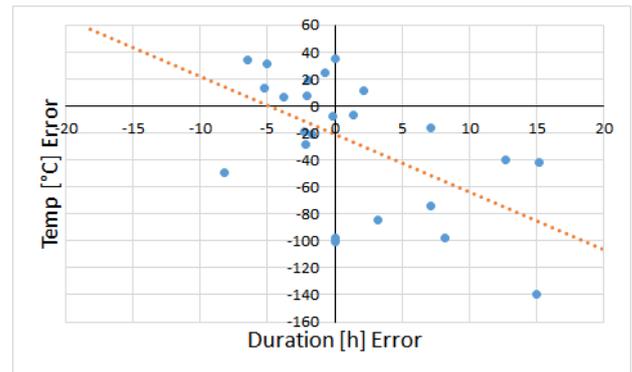


Figure 6: Duration-temperature relationship estimated using GP models

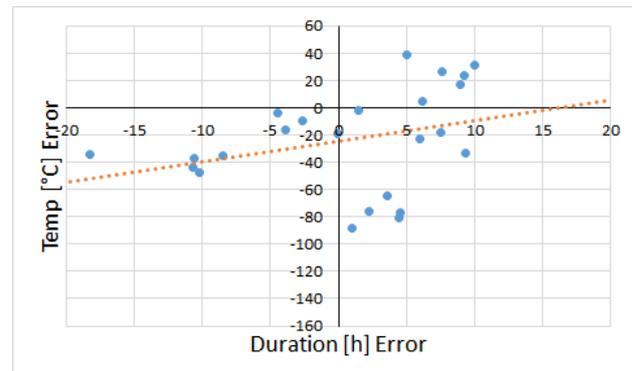


Figure 7: Duration-temperature relationship estimated using RF models

5. CONCLUSION

In plasma nitriding, the goal is to obtain desired material characteristics such as surface hardness, nitriding hardness depth or thickness of the compound layer by tuning the process parameters for each treated material.

In this paper, a two-step optimization approach was presented for finding the optimal control parameters of a plasma nitriding process. In a first step, regression models were generated for each of the target characteristics using a number of different data-based modeling algorithms on data acquired from the process. These models were then used to solve the inverse optimization problem of finding optimal process parameter combinations that would lead to predefined target values for the material characteristics.

The second step of the methodology involved a separate optimization algorithm used to minimize an objective function that aggregated each model's predictions into a unified error measure. To ensure that each model contributes equally to the objective function, for each model the predicted values were scaled according to the variance of the target characteristic.

We performed several experiments to validate the methodology, where we tested GP regression models against models obtained using linear regression, support vector machines, random forests and Gaussian processes. Our models and objective function were incorporated into a CMA-ES optimizer which was used to estimate process parameters on a separate test dataset.

Experimental results have shown that the variance of the estimated values increases with the number of process parameters used as inputs for the regression models. This fact is explained by the increased number of local optima which the optimizer has to overcome when larger combinations of process parameters are estimated. In our experiments we used combinations of five process parameters with the CMA-ES optimizer.

The results concerning two of the most important process parameters, duration and temperature, confirm the fact that GP models with a simpler structure were better suited for solving the inverse optimization problem. Although on average, other models such as RF, SVM and GPR produced better results on the training data, these models proved less effective for the estimation of process parameters on the test data, as shown by the large variance of the estimated values in the case of RF, or the lower number of estimations within the 10% confidence interval in the case of LR, SVM and GPR.

Furthermore, the feasibility of the estimated process parameter combinations was analyzed from a physical perspective. The generated GP models correctly reflected the inverse physical relationship between duration and temperature, which was not as obvious in the case of the other models.

In conclusion, our process parameter optimization approach works reasonably well on real-world data acquired from the chemical process itself. Our optimization setup is flexible as it allows any type of regression model or even regression ensembles to be used for the parameter estimation part, in conjunction with any real-vector

optimizer such as the CMA-ES. Future research directions involve further development of this approach by testing other optimizers and other combinations of regression models.

ACKNOWLEDGMENTS

The work described in this paper was done within DAMOPLAS, a research project financed by the Austrian Research Promotion Agency (FFG Basisprogramm).

REFERENCES

- Affenzeller, M., Winkler, S., Wagner, S., and Beham, A. (2009). *Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications*. Chapman & Hall/CRC. ISBN 978-1584886297.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45 (1): pp. 5-32.
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*. 20: 273-297
- Hansen, Nikolaus, Müller, Sibylle D., and Kourmoutsakos, Petros (2003): Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES), *Evolutionary Computation*, 11(1), pp. 1-18.
- Kommenda, M., Kronberger, G., Winkler, S., Affenzeller, M., and Wagner, S (2013) Effects of constant optimization by nonlinear least squares minimization in symbolic regression. *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*, pp. 1121-1128. ACM, 2013.
- Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. University of Michigan Press.
- Mittemeijer, E. J., & Somers, M. A. (1997). Thermodynamics, kinetics, and process control of nitriding. *Surface Engineering*, 13(6), pp. 483-497.
- Rasmussen, Carl Edward. "Gaussian processes in machine learning." *Advanced lectures on machine learning*. Springer Berlin Heidelberg, 2004. 63-71.
- Seber, George AF, and Alan J. Lee. "Linear regression analysis". Vol. 936. John Wiley & Sons, 2012.
- Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., Dorfer, V., and Affenzeller, M. (2014) Architecture and Design of the HeuristicLab Optimization Environment. In *Advanced Methods and Applications in Computational Intelligence, Topics in Intelligent Engineering and Informatics Series*, pp. 197-261. Springer

FLOW SIMULATIONS IN THE HUMAN CARDIOVASCULAR SYSTEM UNDER VARIABLE EXTERNAL CONDITIONS

R. Adragna^(a), R.C. Cascaval^(b), M.P. D'Arienzo^(c), and R. Manzo^(d)

^{(a),(b)}University of Colorado Colorado Springs, Department of Mathematics

^{(c),(d)}Università degli Studi di Salerno, Dipartimento di Ingegneria dell'Informazione, Ingegneria Elettrica e Matematica Applicata

^(a)radragna@uccs.edu, ^(b)radu@uccs.edu, ^(c)mdarienzo@unisa.it, ^(d)rmanzo@unisa.it

ABSTRACT

In this paper we study the effects of variations in the external conditions (such as pressure changes due to motion in gravitational field) on the cardiovascular system. The goal is to understand how the flow and pressures propagate in the network in such variable conditions and which are the locations where the effects are significant. This is relevant to the auto regulation mechanism of the pressure and flow, designed to maintain the system in a homeostatic condition.

Keywords: vascular network, arterial pressure, blood flow, orthostatic change

1. INTRODUCTION

The primary role of the human cardiovascular system is to transport oxygen and nutrients to all the tissues of the body and to remove carbon dioxide and other harmful waste products of cell metabolism. From a physical point of view, the system consists of two synchronized pumps in parallel that propel a viscous liquid (the blood) through a network of flexible tubes. The heart provides energy to move blood through the circulatory system. It consists of four cavities: two ventricles and two atria, whose size varies during the cardiac cycle due to the activity of the heart muscle. The right heart pumps deoxygenated blood through the *pulmonary circulation* and the left heart pumps oxygen-rich blood through the *systemic circulation*. There are four valves, one at the exit of each heart cavity, which regulate blood flow in the heart and ensure bulk unidirectional motion through both pulmonary and systemic circulation (Formaggia, Lamponi and Veneziani 2006; Ottesen, Olfusen and Larsen 2004). The latter consists of a network of curved and branching vessels whose size decreases in the arteries, arterioles and capillaries and increases in the venules and veins. In particular, arteries distribute blood throughout the body and maintain blood pressure between heartbeats, arterioles transport blood to capillary beds, capillaries diffuse

oxygen and nutrients to cells, and venules collect deoxygenated blood from capillaries and return it to the heart through veins. The topological pattern of the large systemic arteries is primarily a binary tree structure.

In this paper, we consider a simplified arterial network which contains the 55 largest arteries in the human body for the systemic circulation as presented in (Alaustrey, Parker and Sherwin 2012; Alaustrey 2006).

Numerically, we start with the standard hyperbolic system which models area $A = A(x, t)$ and velocity $U = U(x, t)$ in the spatial domain (Cascaval, D'Apice, D'Arienzo and Manzo 2015). At the junctions we use continuity conditions for the Bernoulli pressure and the continuity of the forward and backward characteristics for the hyperbolic system. The outflow is defined using peripheral resistance model. We solve the system with a discontinuous Galerkin scheme, using Gauss-Legendre method to approximate integrals and Newton method to find the solution of the Riemann problem at junctions. The entire system is simulated under variable external conditions such as variations in hydrostatic pressure due to posture changes, variations in respiratory patterns and variations in pressures due to exercise. The heart rate and peripheral resistance are used as controls for maintaining homeostatic conditions, which model the autoregulation mechanism. While an analytical treatment of the optimal control problem is out of reach due to the complexity of the system, numerical simulations of the entire system reveal the nature of the autoregulation mechanism under the considered variable external conditions. The novelty of this paper is to apply the numerical model of the entire 55-edge network to study the effects of the variable environment on the cardiovascular system performance, showing that the obtained results are consistent with the expected ones.

2. NUMERICAL MODEL

We consider the standard hyperbolic system which models area $A = A(x, t)$ and velocity $U = U(x, t)$ in the spatial

domain $\Omega = [0, M]$ (Cascaval, D'Apice, D'Arienzo and Manzo 2015):

$$\frac{\partial A}{\partial t} + \frac{\partial(AU)}{\partial x} = 0,$$

$$\frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x} + \frac{1}{\rho} \frac{\partial P}{\partial x} = f.$$

Here $f = f(A, U)$ is a friction force which models the viscosity of the blood, considered as a non-Newtonian fluid, $P = P(x, t)$ is the hydrodynamic pressure and ρ is the density of the blood. For the pressure we choose the elastic model:

$$P = P_{ext} + \frac{\beta}{A_0} (\sqrt{A} - \sqrt{A_0}),$$

where P_{ext} is the external pressure, A_0 is the cross-sectional area in unstressed conditions, $\beta = \frac{E h_w}{1 - \sigma^2} \sqrt{\pi}$, with σ the Poisson ratio, usually taken to be $\sigma = \frac{1}{2}$, E the Young modulus and h_w the wall thickness.

Inflow conditions (at $x = 0$) are implemented using a valve model, which mimics the real behavior of the physiological system. The opening and closing of valve is determined by the pressure difference between the left ventricle (P_{LV}) and the aortic pressure. More specifically, the valve opens when

$$P(0, t) \leq P_{LV}(t),$$

in which case the pressure at the inflow gets prescribed

$$P(0, t) = P_{LV}(t),$$

and it closes when the velocity becomes negative, in which case the velocity at the inflow is prescribed to be zero:

$$U(0, t) = 0.$$

So:

$$U(0, t) = 0, \quad \text{if } P_{LV}(t) < P(0, t)$$

$$P(0, t) = P_{LV}(t), \quad \text{if } U(0, t) > 0$$

In the simulations, the left ventricular pressure is prescribed equal to

$$P_{LV}(t) = P_{ext} + 3.75 \frac{HR}{75} 10^{-4} \sin \frac{\pi t}{\tau},$$

with HR representing the heart rate and τ the duration of the systole, taken to be a quarter of the heart beat ($\tau = 15/HR$). This model accounts for the fact that peak amplitude of the left ventricular pressure depends on the heart rate.

As terminal condition, we have used a model with *terminal reflection coefficient* R_t , see (Alastruey 2006), which is based on the assumption that W_b is proportional to W_f :

$$W_b = -R_t W_f,$$

where $-1 \leq R_t \leq 1$ and W_f and W_b are the forward and backward characteristics, defined as:

$$W_{f,b} = u \pm (c(A) - c(A_0)), \quad c(A) = \sqrt{\frac{8\beta}{\rho A_0}} A^{1/4}.$$

Note that $R_t = 1$ corresponds to $u = 0$, which means the outflow is completely blocked, while $R_t = -1$ corresponds to $A = A_0$. For general values of $R_t \in [-1, 1]$ the boundary conditions is

$$U = \frac{1 - R_t}{1 + R_t} \sqrt{\frac{8\beta}{\rho A_0}} (A^{1/4} - A_0^{1/4}).$$

The spatial network considered in this paper is the standard 55-edge network representing the large arteries in a human. The labeling of the edges is taken from (Alastruey 2006).

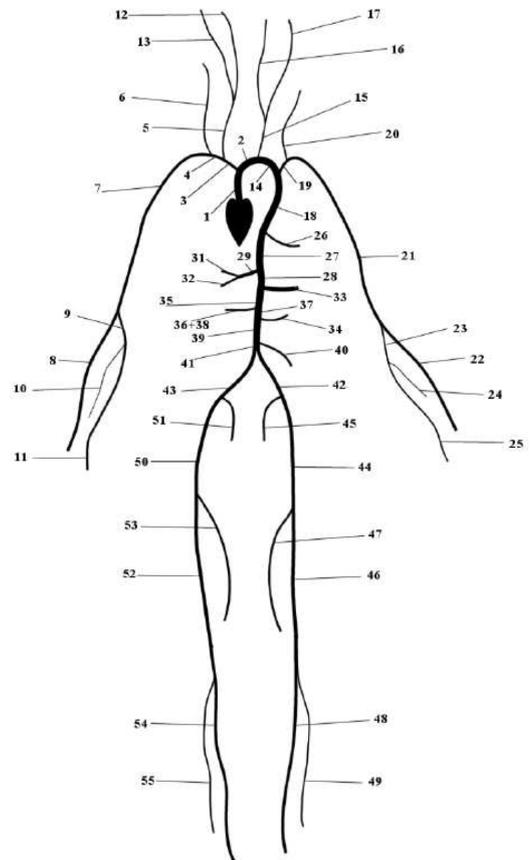


Figure 1: The spatial domain is a 55-edge network.

3. NUMERICAL SOLUTION

In order to find the numerical solution of the problem, we write the system in conservation form:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{F}}{\partial x} = \mathbf{S}, \quad (1)$$

with

$$\mathbf{U} = \begin{bmatrix} A \\ U \end{bmatrix}, \quad \mathbf{F}(\mathbf{U}) = \begin{bmatrix} AU \\ \frac{U^2}{2} + \frac{P}{\rho} \end{bmatrix}$$

$$\text{and } \mathbf{S}(\mathbf{U}) = \left[\begin{array}{c} 0 \\ \frac{1}{\rho} \left(\frac{f}{A} - \frac{\partial P}{\partial \beta} \frac{d\beta}{dx} - \frac{\partial P}{\partial A_0} \frac{dA_0}{dx} \right) \end{array} \right].$$

We solve it using the discontinuous Galerkin scheme as described in (Cascaval, D'Apice, D'Arienzo and Manzo 2015).

We write the weak formulation of the problem, approximate $\mathbf{U}(x, t)$ with its discretized expansion $\mathbf{U}^\delta(x, t)$ and integrate twice by part, so we get:

$$\left(\frac{\partial \mathbf{U}^\delta}{\partial t}, \Phi^\delta \right)_\Omega + \left(\frac{\partial \mathbf{F}(\mathbf{U}^\delta)}{\partial x}, \Phi^\delta \right)_\Omega + [(\mathbf{F}^u - \mathbf{F}(\mathbf{U}^\delta)) \cdot \Phi^\delta]_0^M = (\mathbf{S}(\mathbf{U}^\delta), \Phi^\delta)_\Omega. \quad (2)$$

To simplify the method, we have mapped each elemental region onto the standard element $\Omega_{st} = \{\xi \in \mathbb{R} : -1 \leq \xi \leq 1\}$. This mapping is defined as

$$\chi(\xi) = M \frac{1 + \xi}{2}, \quad \xi \in \Omega_{st},$$

and its inverse is given by

$$\xi = \chi^{-1}(x) = 2 \frac{x}{M} - 1, \quad x \in \Omega.$$

We selected as expansion basis the Legendre polynomials $L_k(\xi)$, with k the polynomial order, because they are orthogonal with respect to the product inner product of L^2 . In this way, the solution is expanded on Ω as

$$\mathbf{U}^\delta(\chi(\xi), t) = \sum_{k=0}^K L_k(\xi) \widehat{\mathbf{U}}^k(t), \quad (3)$$

with $\widehat{\mathbf{U}}^k(t)$ the time-varying coefficients of the expansion. We have chosen Legendre points (which are the zeros of Legendre polynomials) as collocation points. Replacing (3) in (2) and letting $\Phi^\delta = \mathbf{U}^\delta$, we obtain the following system of $2(K+1)$ differential equations to be solved:

$$\frac{d\widehat{U}_i^k}{dt} = \mathcal{F}_i^k(\mathbf{U}^\delta), \quad k = 0, \dots, K, \quad i = 1, 2,$$

where $\widehat{U}_i^k, i = 1, 2$, are each of the two components of $\widehat{\mathbf{U}}^k(t)$ and

$$\mathcal{F}_i^k(\mathbf{U}^\delta) = - \left(\frac{\partial F_i}{\partial x}, L_k \right)_\Omega + \frac{2}{M} [(F_i^u - F_i(\mathbf{U}^\delta))(L_k \circ \xi)]_0^M + (S_i(\mathbf{U}^\delta), L_k)_\Omega.$$

The method is completed with a second-order Adams-Bashforth time-integration scheme:

$$\begin{aligned} (\widehat{U}_i^k)^{n+1} &= (\widehat{U}_i^k)^n + \frac{3\Delta t}{2} \mathcal{F}((\mathbf{U}^\delta)^n) + \\ &\quad - \frac{\Delta t}{2} \mathcal{F}((\mathbf{U}^\delta)^{n-1}), \\ k &= 0, \dots, K, \quad i = 1, 2, \end{aligned}$$

in which Δt is the time step and n the number of every time step. To calculate the integrals we use a Gauss quadrature formula of order $q \geq K+1$.

The upwinded fluxes \mathbf{F}^u are computed solving a Riemann problem that takes into account the characteristic information moving away. At a time t , each interface separates two constant states, (A_L, U_L) and (A_R, U_R) , and we need to determine the two upwinded states, (A_L^u, U_L^u) and (A_R^u, U_R^u) , originated on each side of interface at time $t + \Delta t$. To do this, the following equations are required:

$$\begin{aligned} W_f(A_L, U_L) &= W_f(A_L^u, U_L^u), \\ W_b(A_R, U_R) &= W_b(A_R^u, U_R^u), \\ A_L^u U_L^u &= A_R^u U_R^u, \\ \rho \frac{(U_L^u)^2}{2} + P(A_L^u) &= \rho \frac{(U_R^u)^2}{2} + P(A_R^u). \end{aligned} \quad (4)$$

The first two equations come from the assumption that the flow between two initial states is inviscid, and the forward characteristic information, W_f , and the backward characteristic information, W_b .

4. SIMULATION RESULTS

In this section we describe simulation results obtained by considering the 55-edge network presented above. We have taken input data, such as length, radius, terminal coefficient and β for each edge from (Formaggia, Lamponi, Tuveri and Veneziani 2006).

We perform a 4-second simulation of the entire 55-edge network with variable external pressure (to account for respiration). The timing of the heart beat and respiratory cycle are taken from real data collected on a healthy individual in the physiology lab at University of Colorado (see Figure 2).

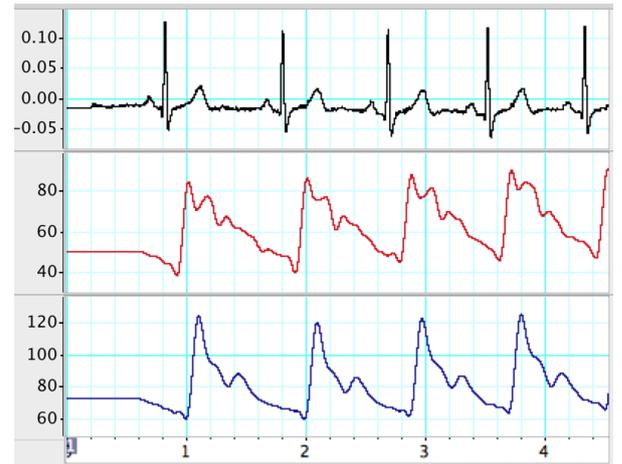


Figure 2: Real data for EKG (black), middle cerebral artery flow velocity in cm/s (red) and arterial blood pressure in mmHg (blue).

Since pressure and flow data are collected only at two sites in the network (radial artery for pressure and middle cerebral artery for flow), the simulation accomplishes to describe the dynamics in all other edges, hence completing the *picture* of the entire network. While the focus was

on developing a working model, no data fitting was done for pressure and flow.

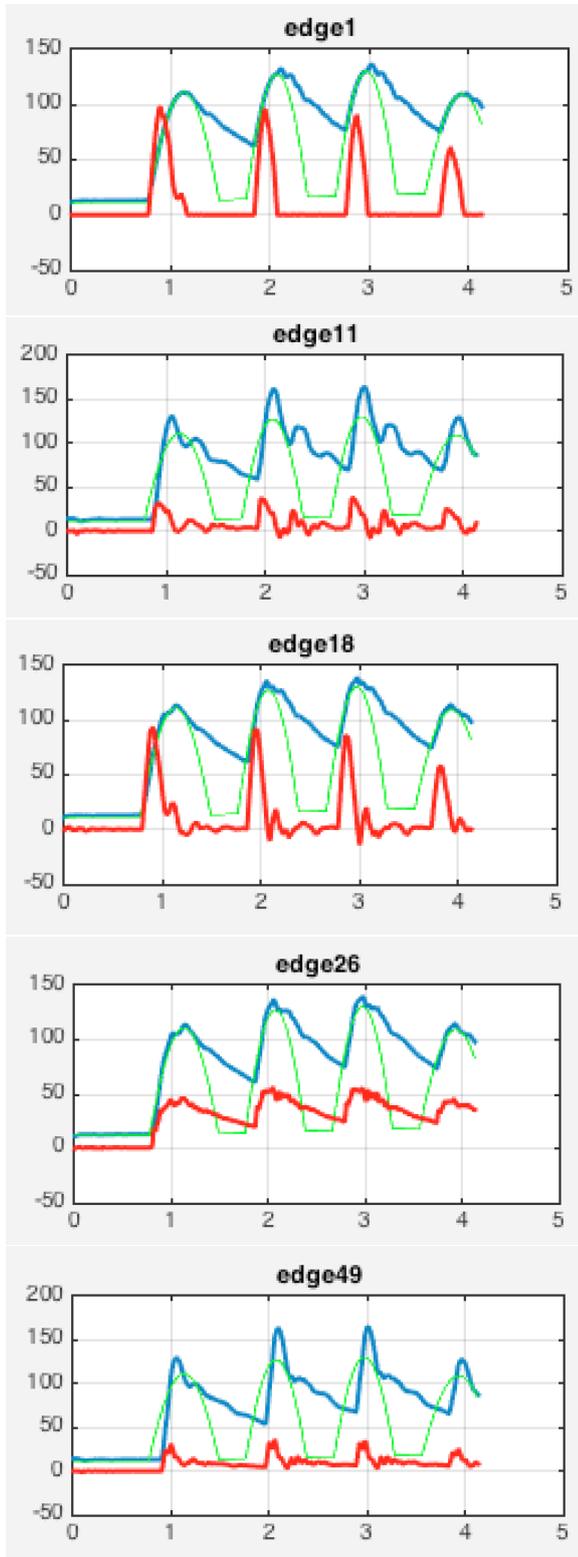


Figure 3: Simulated data in edge 1 (ascending aorta), 11 (right ulnar artery), 18 (thoracic aorta), 26 (intercostal artery) and 49 (anterior tibial artery). Red line is the flow velocity (cm/s), blue is the arterial pressure (mmHg) and green line is the left ventricular pressure.

In the root edge (edge 1) we see the valve in action: the valve is closed hence zero flow goes through it when the left ventricular pressure is lower than the aortic pressure; the valve opens when the left ventricular pressure exceeds the aortic pressure. The timing of the opening of the valve is important, since it determines the total amount of cardiac output. At the other extreme, in edge 49, the pulsatility of the flow is minimal, but still the influence of the respiratory cycle is evident.

There are several noticeable features in this simulation results. The edges depicted in Figure 3 are chosen to illustrate the various aspects of the dynamics. Firstly, the slow variation of the left ventricular pressure (e.g. due to respiration) causes visible variation in the systolic pressure. Secondly, the characteristics of the pressure and flow dynamics is significantly different across the network: in some parts of the network there is small amount of backflow (e.g. edge 11, 18) while in others there is no backflow, (e.g. edge 26). In fact backflow is significant in edges 4, 7, 19 and 21 (only edge 4 is depicted below). Backflow is known to be physiological.

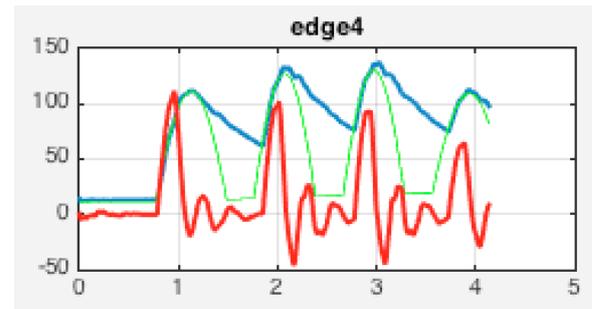


Figure 4: Simulated data in edge 4 (subclavian artery). For the color coding we refer to Figure 3.

For the remaining of this section, we apply the numerical model to a different scenario, that of a tilt test: The body is initially on a horizontal bed, with no orthostatic pressure differences throughout the network. After 2 seconds, a tilt of the table is performed, in such a fashion that the level of the heart remains the same. This means that the majority of the body is sent downward with the exception of the head and shoulders, creating an added orthostatic pressure value in most parts of the network.

Mathematically, this translates to a modification of the external pressure in our working model, to account for the gravitational effect (orthostatic pressure) as it appears during the tilt table test (see e.g. Olufsen, Ottesen, Tran, Ellwein, Lipsitz, and Novak 2005).

$$P_{ext} = P_{ext}(t) = \rho gh(t) = \rho g \Delta h \sin \alpha t$$

where $\alpha = \frac{\pi}{2}$ is the angular velocity, chosen in such a way that after 1 second the tilt table is in upright position, $g = 9.8m/s^2$ is the gravitational constant. Δh is elevation change between the middle of the edge and the heart when the person is upright. Δh can be positive (if

the edge is below the heart) or negative (for edges that lie above the heart level.) Depending on what the center of the tilt is, Δh values would have more negative values.

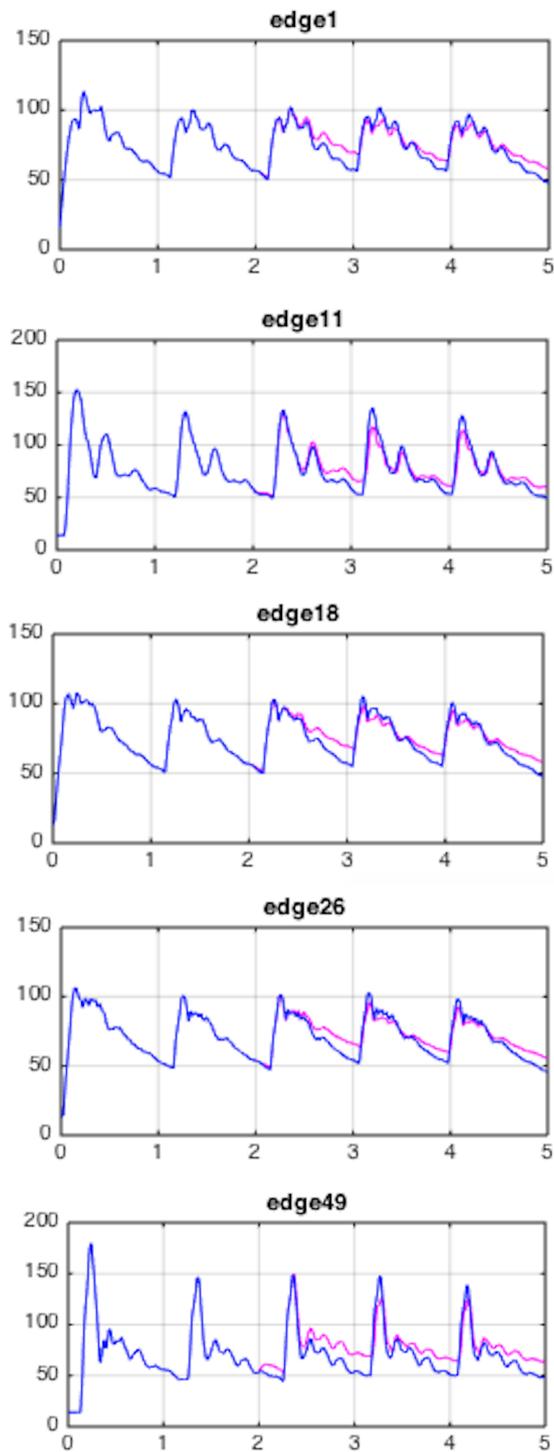


Figure 5: Comparison between the simulated pressure data (in mmHg) in normal conditions (blue) and in presence of the tilt (magenta) in edge 1 (ascending aorta), 11 (right ulnar artery), 18 (thoracic aorta), 26 (intercostal artery) and 49 (anterior tibial artery)

Next we display the simulated data for the flow velocity for both horizontal position and during and after tilt, choosing the same edges as for pressure.

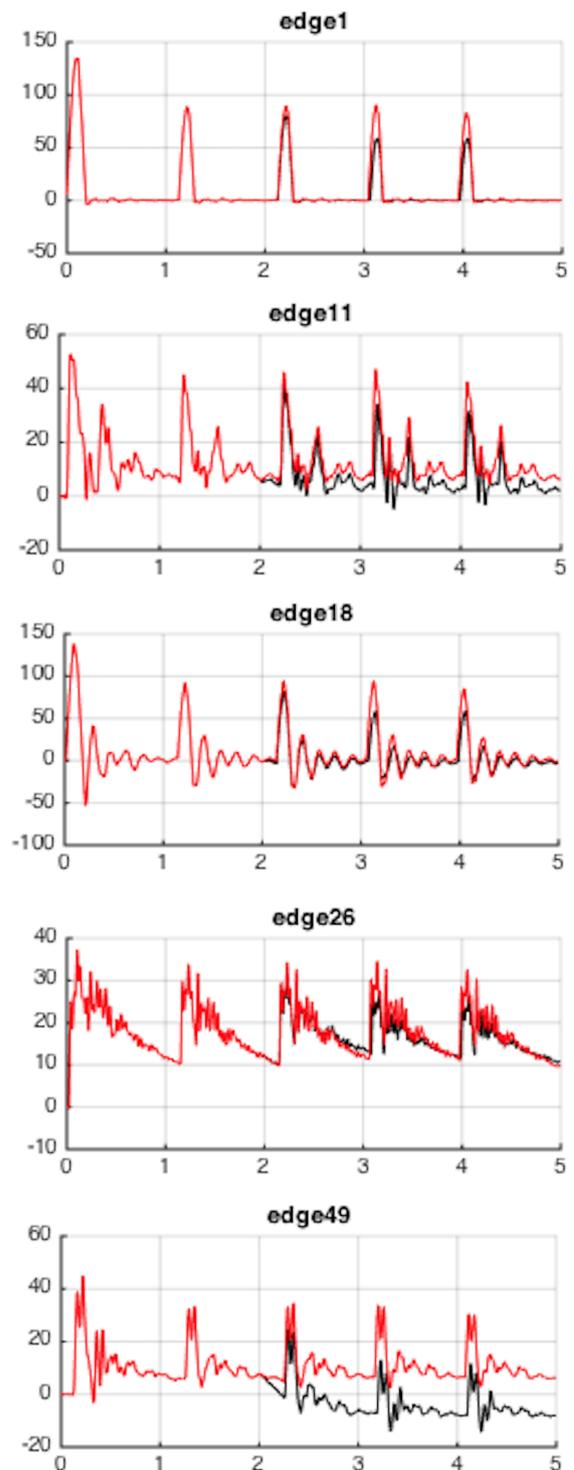


Figure 6: Comparison between the simulated flow velocity data (in cm/s) in normal conditions (blue) and in presence of the tilt (magenta) - same edges chosen for visualization as for pressure on the left

The tilt is performed for 1 second (seconds 2-3 after the beginning of the simulations), time in which the table is raised from horizontal to vertical positions. Then the table is left at that level. We note that several edges come with negative external pressure. Nevertheless, the simulated data shows an increase of diastolic pressure throughout the board. At the same time, the picture with the flow pattern is much more diverse, several edges exhibiting a drop in flow velocity. The simulations are done without changing the Heart Rate calculations, for clarity of the comparison. In reality there is a response of the HR to the change in pressures (baroreceptor control), which would further alter the flow pattern. But even in absence of this complexity, we see that the systolic pressure is decreased immediately after the end of the tilt period, while the diastolic pressure is increased.

5. CONCLUSIONS

A numerical model has been developed here to simulate variable external conditions and their effect on the cardiovascular system, as presented in certain realistic physiological situations. The model includes a valve at the inflow, which accounts for change in the systolic pressure, while the variability of the diastolic pressure is due primarily to the external conditions. The focus has been on comparing the dynamics in various parts of the network in presence and absence of the changes. A real control of the heart rate and peripheral resistance is not present in our paper, but it will be reported elsewhere.

REFERENCES

- Alastruey, J., Parker, K.H., Sherwin, S.J., 2012. Arterial pulse wave haemodynamics. BHR Group's 11th International Conference on Pressure Surges, Lisbon, Portugal, 401–442.
- Alastruey, J., 2006. Numerical Modelling of Pulse Wave Propagation in the Cardiovascular System: Development, Validation and Clinical Applications, PhD Thesis, Imperial College London.
- Cascaval R.C., D'Apice C., D'Arienzo M.P., Manzo R., 2015. Flow Optimization in Vascular Networks, submitted to Math Biosci Eng.
- Formaggia, L., Lamponi, D., Tuveri, M., Veneziani, A., 2006. Numerical modeling of 1D arterial networks coupled with a lumped parameters description of the heart. *Comp. Meth. Biomech. Biomed. Eng.*, 9, 273–288.
- Ottesen, J.T., Olufsen, M.S., Larsen, J.K., 2004. *Applied Mathematical Models in Human Physiology*, SIAM.
- Olufsen, M.S., Ottesen, J.T., Tran, H.T., Ellwein, L.M., Lipsitz, L.A. and Novak, V., 2005. Blood pressure and blood flow variation during postural change from sitting to standing: model development and validation, *J Appl Physiol* 99: 1523–1537.

AUTHORS BIOGRAPHY

Reece Adragna Reece Adragna is a graduate student in the Department of Mathematics at University of Colorado Colorado Springs (UCCS). His e-mail address is radragna@uccs.edu.

Radu C. Cascaval Dr. Cascaval is Associate Professor of Mathematics at University of Colorado Colorado Springs. His main research interests are in mathematical modeling in physiology. He has had several collaborations with exercise physiology faculty at University of Colorado and has recently won a grant from the BioFrontiers Institute at University of Colorado. His e-mail address is radu@uccs.edu.

Maria Pia D'Arienzo Maria Pia D'Arienzo was born in 1989 in Salerno. She graduated in Mathematics in 2012, at University of Salerno, Italy. Now she is a PhD student in Mathematics at University of Salerno. Her main research area is in fluid-dynamic models in physiology. Her e-mail address is mdarlenzo@unisa.it.

Rosanna Manzo Rosanna Manzo is a researcher in Mathematical Analysis at the Department of Information Engineering and Electrical Engineering and Applied Mathematics of the University of Salerno, Italy. She received PhD in Information Engineering from University of Salerno. Her research areas include fluid - dynamic models for traffic flows on road, telecommunication and supply networks, optimal control and queueing theory. Her e-mail address is rmanzo@unisa.it.

DISCRETE-EVENT SIMULATION MODELLING OF PREFABRICATED WALL PRODUCTION LINE

Mohammed Sadiq Altaf^(a), Hexu Liu^(b), Youyi Zhang^(c), Mohamed Al-Hussein^(d), Ahmed Bouferguene^(e)

^{(a),(b),(c),(d)} Department of Civil and Environmental Engineering, University of Alberta, Canada
^(e) Campus St. Jean, University of Alberta, Canada

^(a) msaltaf@ualberta.ca, ^(b) hexu@ualberta.ca, ^(c) youyi@ualberta.ca, ^(d) malhussein@ualberta.ca,
^(e) ahmed.bouferguene@ualberta.ca

ABSTRACT

Simulation is a powerful tool that can be utilized in the production planning and control processes of complex manufacturing and construction facilities. A simulation model can capture complex relationships between different systems and predict possible outcomes in advance. This paper proposes a method to develop a discrete-event simulation (DES) model of a prefabricated wall production line utilizing a special purpose modelling template. Actual production data is collected using a radio frequency identification (RFID) system for the purpose of simulation input modelling. Wall panel information is stored in a database which is connected to the simulation model in order to assign panel attributes to the model entity directly. The proposed method is tested in a wood-framed wall panel production facility and several applications of the simulation model are outlined. The simulation model can be a useful tool to predict total production time, panel sequencing, and performance evaluation.

Keywords: Discrete-event simulation (DES), Prefabricated wall production, Special purpose simulation template.

1. INTRODUCTION

Prefabrication construction is becoming a mainstream construction method due to its superior quality and reduced environmental impact. Simulation models are used as planning tools in various sectors of the prefabrication construction and manufacturing industry. Halpin (1977) created CYCLONE, a simulation environment, which introduced the foundation for the progress of construction simulation. AbouRizk and Hajjar (1998) have proposed a framework for the application of simulation in construction, specifically focusing on construction practitioners. They presented the concept of special purpose simulation (SPS), which is a computer-based environment specially built for experts in the area. AbouRizk and Mohamed (2000) introduced Symphony.NET, an integrated environment to model construction activities. This simulation software supports both DES and continuous simulation. It can provide different model outputs, such as standard statistical averages, resource utilization, standard

deviation, minima and maxima, and charts such as histograms, cumulative density functions (CDFs), and time graphs. Al-bataineh et al. (2013) have presented a case study in which a simulation model for a tunneling project in Edmonton, Canada was developed in Symphony.NET as a decision support system for the project management team. Alvanchi et al. (2012) have developed a DES model of the steel girder bridge fabrication process in order to provide a solution for the complex process of planning off-site girder bridge construction.

In addition, Lu et al. (2008) have developed an automated resource-constrained critical path analysis using DES and particle swarm optimization (PSO). Based on their study, simulation modelling enables engineers to precisely examine different approaches in order to complete a project in the most efficient manner. Performing this type of analysis in advance yields reduced costs, improved quality, and improved productivity (AbouRizk 2010). Liu et al. (2015) introduced a special purpose simulation (SPS) template for the panelized construction process and linked the simulation model with building information modelling (BIM).

This paper develops a detailed DES model of the prefabricated wall production line. The paper identifies different modelling elements for the prefabricated wall production line and outlines the detailed methodology of the simulation model development process using a custom-built special purpose simulation template for the production line process. The simulation input model is designed using actual production data collected through a radio frequency identification (RFID) system. Finally, the application of the simulation model as a production planning and control tool is demonstrated.

2. MEHODOLOGY

Figure 1 shows the methodology of the simulation development process for the prefabricated wall production line. Actual production data is collected from the production floor using the RFID system. RFID antennas are installed at each workstation and an RFID tag is placed on each wall panel at the framing station where the prefabrication process begins. The antenna sensors collect data held within the tag when the panel

moves from one station to another and the timestamp is recorded into the database through the RFID reader. Timestamp data collected from the RFID system is utilized in simulation input modelling for each workstation.

Wall panel information such as length, type, and number of windows/doors are stored in the database from the 3D model and used by the simulation model to determine the process route of each particular wall panel.

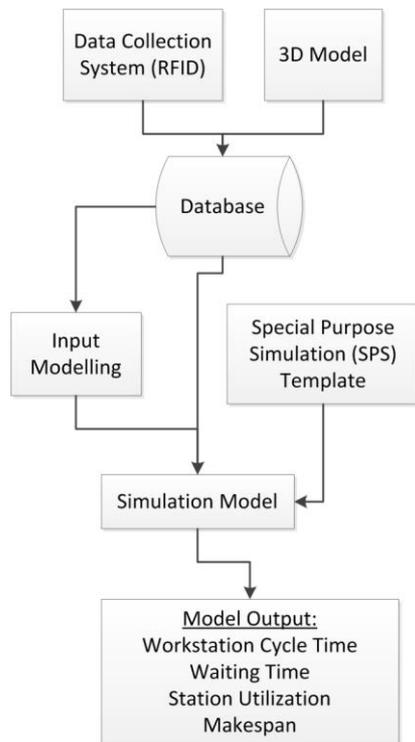


Figure 1: Simulation model development process

A special purpose simulation (SPS) template specifically built for the production line process is used to develop the simulation model. The SPS template contains customized simulation modelling elements such as workstation, storage, and equipment in order to simulate the production line process more efficiently. The simulation model can provide workstation cycle time, waiting time, station utilization, and total makespan for a number of wall panels as model outputs.

3. SIMULATION MODEL DEVELOPMENT

3.1. Production Process

Landmark Group of Builders, a major production homebuilder in Alberta, Canada, has established a production facility in Edmonton where wood-frame wall, floor, and roof sections are prefabricated inside the plant and are then transported to the site for assembly. In this study, the Landmark wall panel production line is simulated using Simphony.NET, a simulation software developed at University of Alberta. Figure 2 shows the prefabrication process of the wall

production line. The wall panel production process begins at the framing station where the studs are nailed to the top and bottom plate along with pre-assembled window and door frames by the computer numerical control (CNC) machine. Multiple single-wall panels are built together in order to maximize the framing station capacity. Next, the multi-wall panel is transported to the sheathing station where OSB sheathing is placed manually and then nailed automatically at the bridge table by another CNC machine. The exterior wall is then moved to the spray-booth in order to spray insulation foam. The multi-wall panels are then cut into single-wall panels and transferred to either the window installation station or window bypass line. After installing the window and door (if required), the wall is placed in the storage line before being loaded onto the transport trailer. Interior wall panels are cut into smaller panel components after the bridge station; the shorter interior wall panels (less than 12 ft.) are transferred directly to the interior wall packaging area from the bridge table while the longer interior walls are transferred to the main storage line through the window bypass line. The following two sections of this paper present the simulation input modelling analysis and special purpose simulation template for the production line facility.

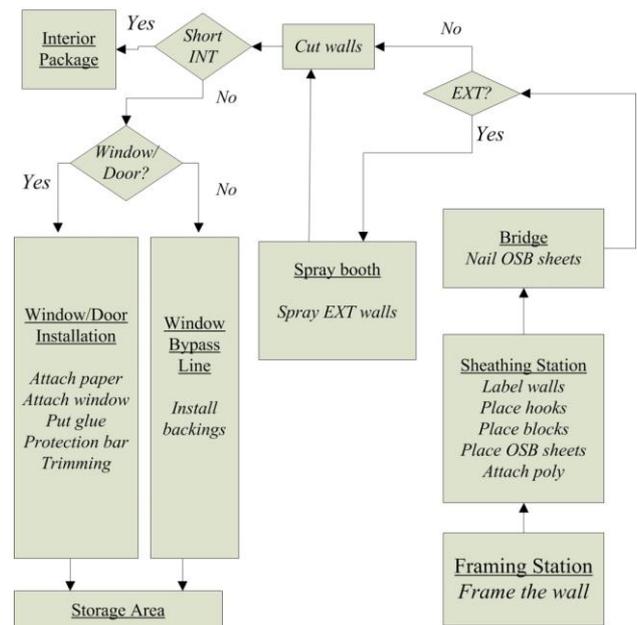


Figure 2: Production process of wood wall panel production line

3.2. Simulation Input Modelling

The RFID system is used in the prefabrication plant to obtain actual production data. Panel processing time at each workstation is collected through the system and stored in the central database. This information is used for distribution fitting using Simphony.NET. Figure 3 shows the probability density function of exterior wall panel processing time at the framing station, which follows a beta distribution.

Similarly, both exterior and interior wall processing time at the sheathing table, nailing station, spray-booth, and window installation line are estimated using actual time data retrieved from the RFID tags. Table 1 summarizes the duration distribution for each workstation.

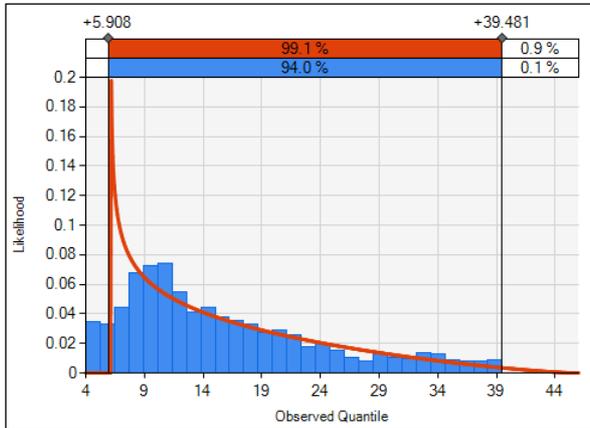


Figure 3: Probability density function of exterior wall processing time at framing station

3.3. Simulation Logic Development using Special Purpose Simulation Template

Simulation models are established using the special purpose simulation template developed by Liu et al. (2015) in Symphony.NET. In the special purpose template, each work station in the production line is represented by a “Station” element; these elements delay the simulation entities that represent building panels for the specific duration of fabrication operations or tasks taking place in each work station. Notably, a work station has a specific capacity limitation and cannot process unlimited building panels simultaneously. As such, the station element provides the user an opportunity to define its capacity in order to control the production flow. Another important element in this new SPS template is the “Entity Router” element. This element routes simulation entities into the correct stations based on the enriched panel information; building panels with various design

features, such as IsExteriorOrNot and HasWindowOrNot, will be routed through various work stations that correspond to their specific features. Further, this SPS template provides another element named “Execute” which enables the user to write custom code in order to enhance the flexibility of simulation modelling. The simulation model is demonstrated in Figure 4. Table 2 tabulates the capacity information for each work station. This information is inputted into the developed simulation model in order to simulate the production flow.

It should be noted that the production line studied in this research produces multi-wall panels in order to optimize the use of the framing station capacity, thereby improving the production productivity. A multi-wall panel herein is defined as a wall panel composed of multiple wall panels and whose length is approximately forty feet, which is similar to the length of the framing table. Table 3 lists examples of single-wall panel information, while some multi-wall panel information is tabulated in Table 4. In work stations such as the framing station, sheathing table, bridge, and spray-booth station, operations/activities are performed on the multiple-wall panels. Interior multiple-wall panels are split into individual single-wall panels after the bridge station, whereas exterior multiple-wall panels are separated into individual single-wall panels at the spray-booth station. This indicates that simulation entities should represent the correct wall panels accordingly in order to mimic the production process. In this context, two execute elements are deployed in the developed simulation model, and user-defined codes (as shown in Figure 5) are embedded into those two elements to solve this problem. Once simulation entities representing multi-wall panels enter the SplitEXT and SplitINT element (as show in Figure 4), the user-defined codes release new simulation entities representing single-wall panels that are contained in the corresponding multi-wall panel and the new simulation entities are sent to the connected simulation element.

Table 1: Duration distribution

Work Station	Wall Type	Sample Size	Distribution Type and Parameter
Framing	Exterior	848	Beta (0.73,2.24,45,6.10)
	Interior	1260	Beta (0.72,6.33,67,3.79)
Sheathing	Exterior	930	Beta (1.26,4.79,54,7.5.96)
	Interior	1273	Beta (0.83,6,67.43,2.83)
Bridge	Exterior	909	Lognormal (2.33, 0.615)
Spray-booth	Exterior	240	Triangular (63.14, 0.76, 46.37)
Window/Door	Exterior	815	Triangular (317, 11, 141)
Window Bypass	Both	1058	Gamma (46.84, 1.79)

Table 2: Work station capacity

Work Station	Capacity	Capacity Measurement	Unit
Framing	1	No. of Panel	Count of Multi-wall Panel
Sheathing	24400	Panel Length	mm
Bridge	1	No. of Panel	Count Multi-wall Panel
Spray-booth	37000	Panel Length	mm
Window/Door	47600	Panel Length	mm
Window Bypass	47600	Panel Length	mm

Table 3: Example of single-wall panel information

Multi-wall Panel	Single-wall Panel	Length	Width	Height	Window	Door	Backing	Floor	Type
EW009061	EW009061-00-110	11564	2772	163	0	0	0	1st	EXT
EW009062	EW009062-00-104	9735	2772	163	0	0	0	1st	EXT
EW009062	EW009062-00-105	2124	2772	163	0	1	0	1st	EXT
EW009063	EW009063-00-112	1362	2622	163	0	0	0	1st	GAR
EW009065	EW009065-00-109	2429	2772	163	0	1	0	1st	EXT

Table 4: Example of multi-wall panel information

Multi-wall Panel	Type	Length	Width	Height	Window	Door	Floor	Num Wall
EW009061	EXT	11564	2772	163	0	0	1st	1
EW009062	EXT	11904	2772	163	0	1	1st	2
EW009063	GAR	9043	2622	163	0	0	1st	3
EW009064	GAR	6890	2622	163	0	0	1st	1
EW009065	EXT	12130	2772	163	2	1	1st	4
EW009066	EXT	6755	2772	163	1	1	1st	3
EW009067	EXT	11748	2467	163	1	0	2nd	2

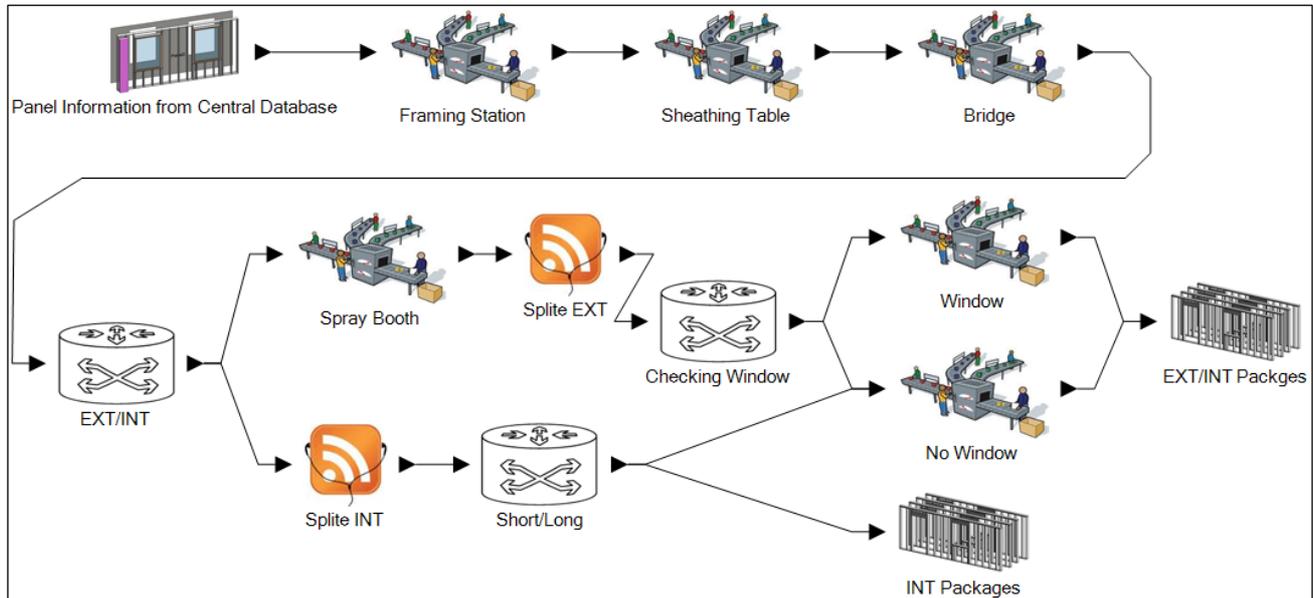


Figure 4: Wall production line simulation model developed in Symphony.NET using SPS template

```

foreach(var newentity in context.CurrentEntity.SubBuildingEntities)
{
    newentity.PreStation = context.CurrentEntity.PreStation;
    newentity.Workpackages.AddRange(context.CurrentEntity.Workpackages);
    windowChecker.InputPoint.TransferIn(newentity);
}

```

Figure 5: Embedded codes to split multi-wall panel

4. RESULTS AND DISCUSSION

The wall prefabrication process for one single-family home is simulated using the SPS template in Symphony.NET. The house requires 56 wall panels, which are produced as 20 multi-wall panels at the framing station. The house has 33 interior wall panels and 23 exterior wall panels. The details for single- and multi-wall panels are read from the 3D model and stored in the database. The simulation model is run fifty times, and Figure 6 shows the cumulative density function of the total makespan for the house. The mean makespan to produce all 56 wall panels ranges from 397 to 644 minutes with a mean value of 490 minutes. Figure 7 shows the cycle time of the spray-booth workstation. The cycle time ranges from 3 to 61 minutes with a mean value of 37.9 minutes. Average cycle time, waiting time, and station utilization for each workstation is summarized in Table 5.

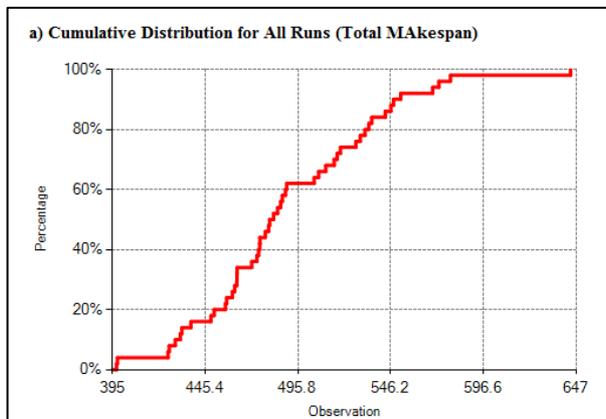


Figure 6: Cumulative distribution function of the makespan

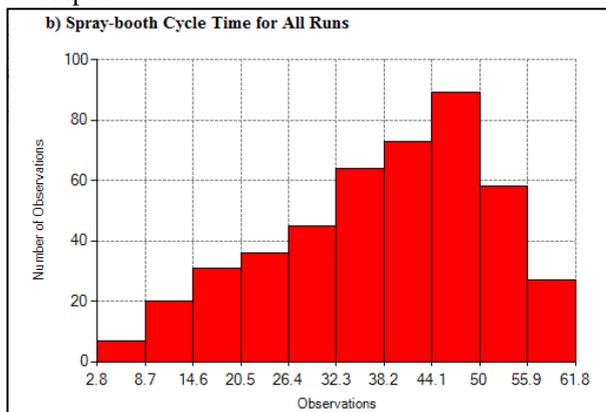


Figure 7: Probability density function of cycle time at Spray-booth

The model results are verified by performing parameter variability-sensitivity analysis and tracing (Sargent 2010). Certain input parameters of the model are adjusted if the model results change based on the expected outcome. For example, if the cycle time of a downstream station increases, the previous station will require a longer waiting time. Also, the number of input wall panels at the framing station should equal the total number of panels in the storage element. The cycle

time, utilization, and waiting time values are also verified by comparing the simulation results with manual calculation. Also, the model entities are traced along the simulation time to determine if the model logic is correct.

Table 5: Simulation results

Work Station	Mean Cycle Time (min)	Mean waiting Time (min)	Mean Utilization
Framing	13.78	0.94	87%
Sheathing	13.25	3.9	63%
Bridge	7.13	0.5	77%
Spray-booth	37.91	0	59%
Window/Door	152	-	47%
Window Bypass	84	-	51%

The simulation model can be utilized in advance for the purpose of production planning and control. A production controller can make changes to the model to analyze different production scenarios. The simulation model can be used to optimize panel sequence in order to estimate production time in advance, evaluate performance, minimize the makespan (Altaf et al. 2014a), and allocate resource more efficiently (Altaf et al. 2014b).

5. CONCLUSION

This paper presents the DES modelling process of the prefabricated wall panel production process using an SPS template. Actual panel production time data is collected through the use of a RFID system and stored in the database for simulation input modelling. Also, necessary wall panel information is read from the 3D file to generate simulation entity attributes. The model is developed in Symphony.NET and actual production data is used for input modelling. The proposed methodology is tested by developing a simulation model for Landmark's prefabricated wall production line facility. The model is run for a single-family home and the results are verified through the parameter variability technique. The use of an SPS template simplified the simulation model development process and various simulation parameters can be adjusted more easily. The production manager can utilize the simulation model to analyze different alternatives, estimate production time, and evaluate actual production performance by comparing them with simulation results.

In future research, the simulation results can be validated using the actual production data collected by the RFID system. Also, a dynamic simulation model can be created by connecting the RFID system with the simulation model to provide actual panel locations prior to starting the simulation model. This will provide more realistic simulation results which can be used as a production evaluation tool on a daily basis.

REFERENCES

- Abourizk, S., 2010. Role of Simulation in Construction Engineering and Management. *Journal of Construction Engineering and Management*, (10), 1140–1153.
- AbouRizk, S.M. & Hajjar, D., 1998. A framework for applying simulation in construction. *Canadian Journal of Civil Engineering*, 25(3), 604–617.
- AbouRizk, S.M. & Mohamed, Y., 2000. Symphony- An integrated environment for construction simulation. In *Proceedings of the 2000 Winter Simulation Conference*.
- Al-bataineh, M. et al., 2013. Using Simulation to Plan Tunnel Construction. *Journal of Construction Engineering and Management*, 139(5), 564–571.
- Altaf, M., Al- Hussein, M., & Yu, H., 2014a. Wood-frame wall panel sequencing based on discrete-event simulation and particle swarm optimization. *The 31st International Symposium on Automation and Robotics in Construction and Mining*, Sydney, Australia.
- Altaf, M., Liu, H., Al-Hussein, M., & AbouRizk, S.M., 2014b. Resource Planning Based On Discrete-Event Simulation For Panelized Construction. *14th International Conference on Construction Applications of Virtual Reality in Construction (ConVR2014)*, Sharjah, UAE, Nov. 16-18.
- Alvanchi, A. et al., 2012. Off-Site Construction Planning Using Discrete Event Simulation. *Journal of Architectural Engineering*, (6), 114–122.
- Halpin, D.W., 1977. CYCLONE: method for modeling of job site processes. *ASCE Journal of the Construction Division*, 103(3), 489–499.
- Liu, H., Altaf, M. S., Lei, Z., Lu, M., and Al-Hussein, M., 2015. Automated production planning in panelized construction enabled by integrating discrete-event simulation and BIM. *Proceedings, International Construction Specialty Conference*, Vancouver, BC, Canada, Jun. 7-10.
- Lu, M., Lam, H.-C. & Dai, F., 2008. Resource-constrained critical path analysis based on discrete event simulation and particle swarm optimization. *Automation in Construction*, 17(6), 670–681.
- Sargent, R. G., 2010. Verification and Validation of Simulation Models. In *Proceedings of the 2010 Winter Simulation Conference* (pp. 166–183).

AUTHORS BIOGRAPHY

MOHAMMED SADIQ ALTAF is a PhD candidate and has been working under the supervision of Dr. Mohamed Al-Hussein since 2009, during which he completed his Master's in Construction Engineering Management. Currently, Mr. Altaf is in the process of completing his PhD, which focuses on construction simulation, modular and panelized construction, production planning, and schedule optimization. His email address is msaltaf@ualberta.ca.

HEXU LIU is a PhD candidate; having completed his Bachelor's and Master's studies in Shaanxi, China, he is now completing his PhD studies at the University of Alberta. His research interests include the integration of discrete-event simulation (DES) and building information modeling (BIM) to advance the practice of construction planning and management and information technology for intelligent construction. His e-mail address is hexu@ualberta.ca.

YOUYI ZHANG is a student who has been under the supervision of Dr. Mohamed Al-Hussein since 2013, and she is currently in the process of completing her Master's degree in Construction Engineering Management. Ms. Zhang's research field is closely related to modular home construction in manufacturing. Improving production line by applying lean construction concepts and BIM through simulation and visualization support are her main focus. Her email address is youyi@ualberta.ca.

MOHAMED AL-HUSSEIN is a professor and Industrial Research Chair (IRC) in the Industrialization of Building Construction at the University of Alberta, and a highly sought researcher and consultant in the areas of lean manufacturing, construction process optimization, CO₂ emission quantification, and building information modelling (BIM), with the development of modular and offsite construction technologies and practices forming the hub of his research.. His email address is malhussein@ualberta.ca.

AHMED BOUFERGUENE is an associate professor at the University of Alberta Saint-Jean campus. Dr. Bouferguene's research focuses on developing computation methodologies in quantum chemistry. Two of his major projects are “first-principles calculations/modeling in heterogeneous catalysis”, and “efficient algorithms for the evaluation of multicenter integrals over Salter Type Orbitals (STOs)”. His email address is ahmed.bouferguene@ualberta.ca.

AMEBA – STRUCTURAL EVOLUTIONARY OPTIMIZATION: METHOD AND TOOLBOX DEVELOPMENT

Marko Corn^(a), Maja Atanasijević-Kunc^(b)

^(a) Ameba d.o.o., Logatec, Slovenia

^(b) University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

^(a) marko.corn@ameba.si, ^(a) maja.atanasijevic@fe.uni-lj.si

ABSTRACT

Evolution algorithms are optimization methods that mimic a process of a natural evolution. Their stochastic properties result in a huge advantage over other optimization methods especially when solving complex optimization problems. In the paper Agent Modelled Evolutionary Based Algorithm (AMEBA) is first presented which was developed and implemented in MATLAB as a Toolbox. AMEBA algorithm has several advantages over other evolutionary algorithms and in this article the advantage of custom designed initial solution is presented. Custom designed initial solution is a solution of the problem that is built on the base of the knowledge of the system and represents a solution which the AMEBA algorithm will try to improve. This capability is presented with the example of an involvement of multivariable controller for the pressure – level system that represents non-linear, multivariable system which is very stiff, with the property of weak inherent coupling. The AMEBA algorithm significantly improved the initial controller solution which shows that classical controller structures can also be automatically altered to increase quality of the solution.

Keywords: evolutionary computation, ameba, pressure-level pilot plant, multivariable control design

1. INTRODUCTION

Design of controllers for complex dynamic systems is usually done with the use of system mathematical model that enables the usage of optimization with which the controllers are optimized. Several types of optimization method can be used to design a proper control system regarding simpler processes. Regarding more complex systems as are for example multivariable, non-linear, time-variant etc. a group of evolutionary computation optimization methods proved to be very efficient (Logar, Dovžan, and Škrjanc 2011; Tomažič et al. 2013).

Evolutional computation algorithms are optimization methods that mimic process of the natural evolution. In general the evolutionary algorithms can be divided into

two major groups: parametrical and structural algorithms.

Parametrical algorithms evolve parameters while structural algorithms evolve structures or mapping functions. For example if we would have to design a controller for the dynamic system, parametric algorithm would demand to define parameters of the chosen controller structure (very frequently a PID controller is used). In contrast to parametrical algorithms structural algorithms do not require predefined form of the controller as they can evolve the whole controller through the evolution process. The most popular parametrical algorithms are genetic algorithms (GA) (Atanasijević-Kunc, Belič, and Karba 2006; David Goldberg 1989), evolutionary strategies (ES) (Beyer 2010), differential evolution (DE) (Storn and Price 1997) and others (Brownlee 2011). Most established structural algorithms are genetic programming (GP) that have multiple implementations from the tree based implementation (Koza 1992) to the grammatically based implementation (Whigham 1992) and the evolutionary programming that is directed into the involvement of finite state machines (Fogel, Owens, and Walsh 1966). New method that supplements the structural group of evolutionary algorithms is Agent Modelled Evolutionary Based Algorithm (AMEBA) (Corn and Atanasijević-Kunc 2011; Corn, Černe, and Atanasijević-Kunc 2012; Corn and Černe 2012). An important property of AMEBA algorithm is that it can include external knowledge in the form of a solution which has some relative advantages while optimal result is still under investigation.

In industrial applications the most widely used controller is PID controller (Vesely and Rosinová 2011) as it is simple to integrate and its control capabilities satisfy technological needs of many processes. This solutions are frequently not optimized which leaves room for the upgrades. AMEBA algorithm can include their current solution and try to improve it by changing the structure of the controller as is demonstrated in this paper.

The paper is organized in the following way. In the next section AMEBA algorithm is presented together with

corresponding toolbox developed in Matlab. In the third section a multivariable pressure-level dynamic system is presented that was used as a testing example. In the fourth section the current solution is presented in a form of multivariable PI controller and an upgraded controller developed by the AMEBA algorithm is described. At the end the conclusions and some ideas for the future work are given.

2. AGENT MODELLED EVOLUTIONARY BASED ALGORITHM - AMEBA

All evolutionary based approaches mimic the process of natural evolution. Natural evolution can be divided in three basic elements: organism, environment and reproduction of organisms. Analogue to the natural evolution, evolutionary computation methods have entity, environment and reproductions of entities.

2.1. Entity - Agent

In AMEBA algorithm entity is represented as an agent. Agent consists of nodes and connections (Figure 1). Each node represents certain mapping function that transforms input data into output data. Nodes are receiving input information from and sending output information to other nodes via connections.

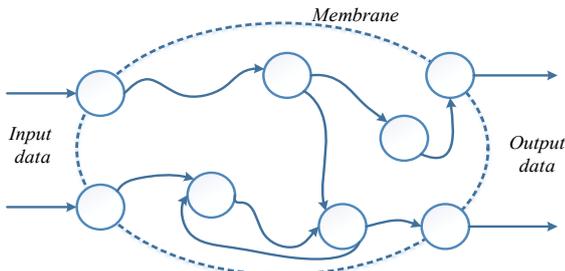


Figure 1. Representation of agent which consists of nodes and connections

In Figure 2 the structure of general node is presented. It consists of nodes inputs (u_1, u_2, \dots, u_n), nodes mapping function f , parameter of node's expression e , and nodes output y .

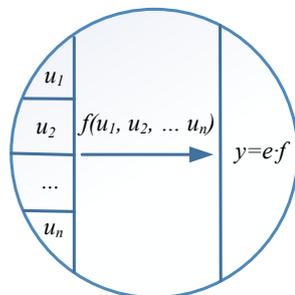


Figure 2: Structure of general node

Main task of the node is transformation of the input signal through mapping function to the output as described in equation (1).

$$y = e \cdot f(u_1, u_2, \dots, u_n) \quad (1)$$

Parameter of expression determines the influence of a node on other connected nodes. Expression parameter can reduce or amplify nodes influence on other nodes depending on the optimization process. Expression parameters initial values are randomly generated and are changing during optimization process. Expression parameter have similar role that is the expression of gens in DNA chain.

Different mapping functions determine different types of nodes (see Table 1) like basic arithmetic, logic or other functions.

Table 1: Nodes of AMEBA algorithm

Function
Input, Output
$y(k) = u_1(k)$
Constant
$y(k) = e \cdot a$
Amplify
$y(k) = e \cdot a \cdot u_1(k)$
Exponent base
$y(k) = e \cdot u_1(k) ^a$
Exponent index
$y(k) = e \cdot a^{u_1(k)}$
Logarithmic
$y(k) = e \cdot \log_a(u_1(k))$
Addition
$y(k) = e \cdot \sum_{i=1}^n a_i \cdot u_i(k)$
Multiplication
$y(k) = e \cdot \prod_{i=1}^n u_i(k)$
Division
$y(k) = e \cdot \frac{u_1(k)}{\prod_{i=2}^n u_i}$
Delay
$y(k) = e \cdot \begin{cases} u_0 & k < n \\ u_1(k - n) & k \geq n \end{cases}$
Integral
$e \cdot \begin{cases} y_0 & k = 0 \\ (k - 1) + k_f \cdot u_1(k) & k > 0 \end{cases}$
Derivative
$e \cdot \begin{cases} 0 & k = 0 \\ k_d \cdot (u_1(k) - u_1(k - 1)) & k > 0 \end{cases}$
High pass filter
$e \cdot \begin{cases} 0 & k = 0 \\ (y(k - 1) + \alpha(u_1(k) - y(k - 1))) & k > 0 \end{cases}$

Low pass filter	
$e \cdot \begin{cases} 0 & k = 0 \\ \alpha \cdot y(k-1) + \alpha(u_1(k) - u_1(k-1)) & k > 0 \end{cases}$	

Special property of AMEBA algorithm is possibility of a definition of nodes with the dynamic mapping function like time delay, integral, derivative and filters. With the usage of dynamic nodes a design of feedback loops is possible and by this the formation of dynamic systems.

2.2. Environment

Natural environment provides a habitat for organisms and the mechanism of selection that determinate which organisms will be reproduced. In artificial evolution natural environment is replaced by simulation environment in which artificial entities are “living” and the selection mechanism is replaced by artificial selection mechanism based on an appropriate fitness function. Fitness function value quantify entities success at solving problem.

In AMEBA algorithm two mechanisms of selection are implemented: the survival of the best and the tournament selection. The survival of the best is a mechanism in which only the best portion of agents in a population are given a chance to reproduce. The tournament selection is a mechanism where two randomly selected agents compete and the winner gets an opportunity to reproduce. These two mechanisms were implemented because they have different effect on population diversity. Selection of the best mechanism strongly reduces diversity of the population as only the best agents survive in contrast to a tournament selection mechanism where also weaker agents can survive.

2.3. Reproduction

With the selection process a group of agents are selected that will be able to reproduce into a new population. In general there are two groups of reproduction mechanisms: mutations and crossovers. Mutation is a reproduction mechanism where an agent replicate itself and during this process slightly change one of its features. Crossover is a reproduction mechanism where two agents combine and reproduce an offspring which possess certain features of both parent-agents. In AMEBA algorithm several reproduction mechanisms are implemented (see Table 2).

Elite division enables cloning of the agent, which means that agent reproduce an exact copy of itself for the new population. By elitism the best solution is always kept trough out the whole optimization process.

Parameter change mutation mechanism reproduces an offspring from parent agent by changing value of nodes parameter if node possesses one. Change can be made from within the node by the equation (2) or with the influence of other nodes regarding the equation (3).

Table 2: Reproduction mechanism of AMEBA algorithm

Group	Reproduction
Mutations	Elite division
	Parameter change
	Add node
	Remove node
	Switch connection at source
	Switch connections at target
Crossovers	Combination of agents parameters
	Combination of agents nodes and connections

$$a' = \begin{cases} a + r \\ r \\ \frac{1}{a} \\ -a \end{cases} \quad (2)$$

$$a' = \begin{cases} b \\ b (+, -, *, /) c \end{cases} \quad (3)$$

Parameter change can be done in four ways. To parameter a can be added random value r , it can be replaced by random value r , it can be transform to its inverse value or it can change its sign.

Parameter change with the influence of other nodes can be done in five ways. It can be switched with parameter from randomly selected node b , or it can be calculated as sum, difference, product or quotient of two parameters b and c from two randomly selected nodes.

Add node reproduction mechanism adds a new node to agent. First the random connection is selected than the random node is generated and inserted in the place of the selected connection. If the new node have free inputs that must be connected to the network they are randomly connected to other nodes in network. In general it is not needed that node is meaningfully connected to the network it can be unconnected which represents a blind node.

Remove node reproduction mechanism removes randomly selected node from parent agent. First the node for removal is selected from the agent than all its input connections are removed. In the next step the output connections are unpinned from the node and randomly connected to some other nodes and then the selected node is removed.

Switch nodes connections reproduction mechanism switches either a source or target ends of two randomly selected connections in the network.

Combination of agents' parameters is a crossing reproduction mechanism and it is very similar to parameter change mechanism from mutations with the difference that nodes for reproduction can be from other agents in the population.

Combination of agent's nodes and connections is a crossing mechanism that combines two parent agents

into new offspring. First randomly selected group of connected nodes is selected and removed from the first parent. Then the same process is done on the second parent. Finally the removed nodes from the first parent are inserted into the second parent and randomly reconnected to the rest of the nodes network. This mechanism is similar to a reproduction mechanism of combining decision trees in genetic programming method with the difference that is made for the graph type of structure.

2.4. Toolbox development

AMEBA algorithm is being developed also as software package with the user friendly graphical interface. The core development is being built in Java programming environment that can be used also with Matlab that have a very strong support in simulation of dynamic systems via Simulink toolbox. Graphical interface is also developed in Matlab for the easier use of AMEBA algorithm.

Toolbox enables settings of the simulation environment with the inclusion of Simulink model as it is shown in Figure 3.

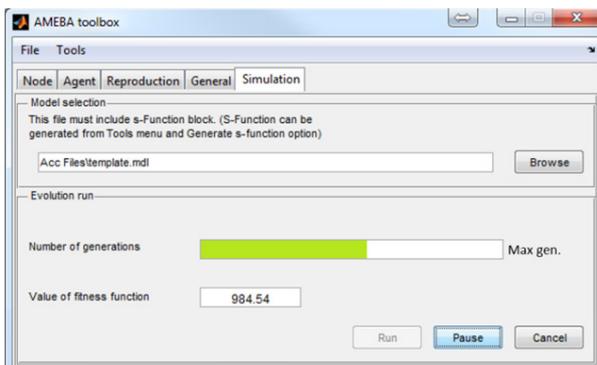


Figure 3: Settings of simulation environment

The agent of AMEBA algorithm is implemented as S-function so it can be included into model as a standard block. Toolbox enables control and monitoring of the optimization process where it displays current generation number and value of the fitness function of the best agent.

Toolbox enables settings of population properties like size of population, size of reproductive population that determines how many best agents will be given opportunity to reproduce, number of elite agents, and other settings that determine the end of optimization process like maximum number of generations and minimum change in fitness function value (Figure 4).

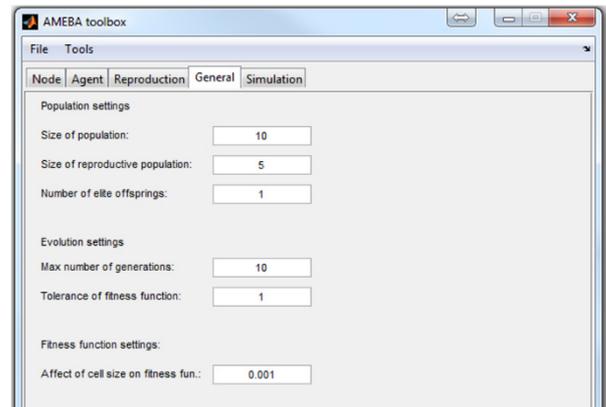


Figure 4: General setting

Number of inputs and outputs of an agent can be set together with the maximum number of nodes that can be generated at the agent's creation (Figure 5).

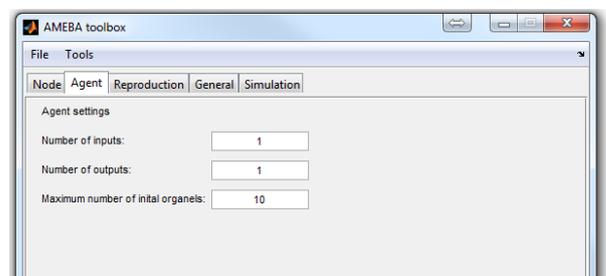


Figure 5: Agent settings

Different types of nodes can be selected from which the algorithm will chose and build agents. Each node has its own settings that determine initial value of the nodes parameter and steepness of change in case node mutates (Figure 6).

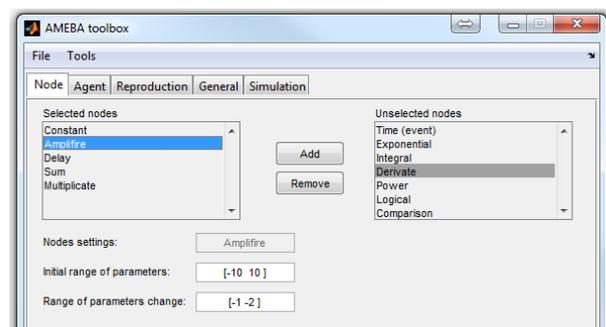


Figure 6: Node settings

Reproduction mechanisms can be set with their parameter of probability. As agents are evaluated and selected for reproduction the reproduction mechanism is randomly selected and the probability parameter determines their possibility of being selected (Figure 7).

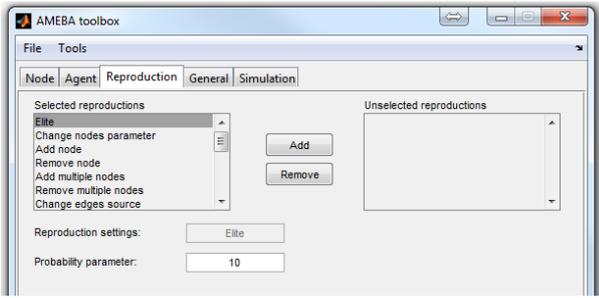


Figure 7: Reproduction settings

Additional functionalities enable better usability of the method such as saving and importing of all setting into file for later use. With this option also the initial population can be imported which enables to inclusion of certain knowledge of the solution into the optimization problem. It is also possible to convert agent into mathematical equation to observe its structure. It can also generate Matlab S-function file for the easier implementation with Simulink tool (Figure 8).

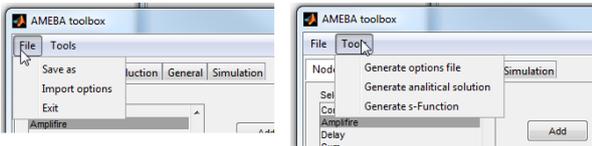


Figure 8: Additions functionalities of Toolbox

3. PRESSURE-LEVEL PILOT PLANT

Optimization is of great importance in many scientific fields where control systems' design is no exception. For testing and comparison purposes we have decided to use mathematical models of pressure-level pilot plant as presented in Figure 9 (Atanasijević-Kunc and Karba 2006; Atanasijević-Kunc, M., Belič, A., and Karba 2007).



Figure 9: Pressure-level pilot plant

Schematically it is illustrated also in Figure 10.

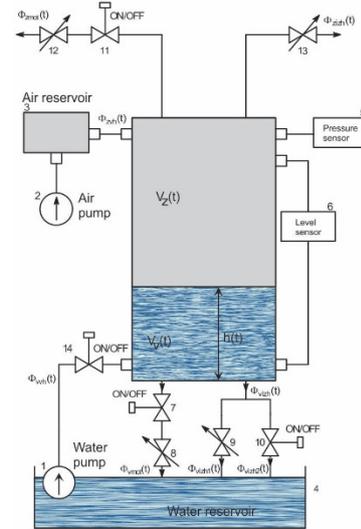


Figure 10: Schematic presentation of pressure-level pilot plant

The central part of the device is a closed tank where air pressure and water level can be controlled through two pumps (1 and 2 in Figure 9). They represent actuators of the system with an input voltage range of 0-10V. The air reservoir (3) increases the time constant of the pressure loop and equalizes pulses in the air flow. The water flow recirculates in the system and therefore there is no need for any external water input. Both outputs of the sensors are also within the 0-10V range. Normal working conditions of the system can be disturbed by changing set-points of corresponding valves, but there is no possibility for direct measurement of these signals.

From the given description it is obvious that the process can be represented by a block diagram as shown in Figure 3 where the following notation is used:

- $u_1(t)$ voltage input to the air pump
- $u_2(t)$ voltage input to the water pump
- $\Phi_{zv}(t)$ air flow to the closed tank
- $\Phi_{vv}(t)$ water flow to the closed tank
- $p_z(t)$ air pressure in the tank
- $h(t)$ water level in the tank
- $y_1(t)$ voltage output from pressure sensor
- $y_2(t)$ voltage output from level sensor

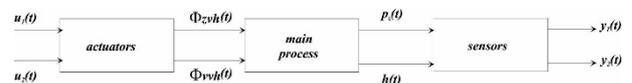


Figure 11: Block diagram of plant components

3.1. Nonlinear model

Model description of this system can be presented regarding the components indicated in Figure 11.

Actuators can be described with the following equations (Atanasijević-Kunc, M., Karba, R., and Zupančič 1997) for the air pump:

$$T_{a11}T_{a12}\ddot{\Phi}_{zv}(t) + [T_{a11} + T_{a12}]\dot{\Phi}_{zv}(t) + \Phi_{zv}(t) = K_{a1}u_1^2(t) \quad (4)$$

and for the water pump:

$$T_{a211} \dot{\Phi}_{vvh}(t) + \Phi_{vvh}(t) = K_{a2} u_2^5(t) \quad (5)$$

Main process equilibrium equations are:

$$\Phi_{zvh}(t) - \Phi_{zizh}(t) = \dot{m}_z(t) \quad (6)$$

where:

$$\Phi_{zizh}(t) = K_{zv} \sqrt{p_z(t) - p_0} \quad (7)$$

Taking into account that air in the reservoir is an ideal gas, specific gas constant is given with:

$$r_z = \frac{p_z V_z}{m_z T_z} = \frac{p_z}{\rho_z T_z} = \frac{10^5 \text{ N/m}^2}{1.3 \text{ kg/m}^3 \cdot 293^\circ} \quad (8)$$

so:

$$m_z(t) = \frac{p_z(t) V_z(t)}{r_z T_z(t)} = \frac{p_z(t)}{r_z T_z(t)} S [H_0 - h(t)] \quad (9)$$

and

$$\dot{m}_z(t) = \frac{S}{r_z T_z} \left[(H_0 - h(t)) \dot{p}_z(t) - p_z(t) \dot{h}(t) \right] \quad (10)$$

The first equilibrium equation of the main process can be presented in the following form:

$$\dot{p}_z(t) = X_1 \frac{K_B}{H_0 - h(t)} \Phi_{zvh}(t) - X_2 \frac{K_B K_{zv}}{H_0 - h(t)} \sqrt{p_z(t) - p_0} + X_3 \frac{1}{H_0 - h(t)} p_z(t) \dot{h}(t) \quad (11)$$

Equilibrium equation of the main process for the lower part can be described with:

$$\Phi_{vvh}(t) - \Phi_{vizh}(t) = \dot{m}_v(t) \quad (12)$$

where:

$$\Phi_{vizh}(t) = \sqrt{K_1 h(t) + K_2 (p_z(t) - p_0)} \quad (13)$$

Taking into account that water is not compressible it can be written:

$$m_z(t) = \rho_v S h(t) \quad (14)$$

and the second equilibrium equation of the main process can be presented in the following form:

$$\dot{h}(t) = \frac{1}{C_2} \Phi_{vvh}(t) - \frac{1}{C_2} \sqrt{K_1 h(t) + K_2 (p_z(t) - p_0)} \quad (15)$$

where the following notation was introduced:

$$C_2 = \rho_v S \quad (16)$$

Both **sensors** are linear and can be described with:

$$y_1(t) = K_{s1} (p_z(t) - p_0) - \bar{y}_{1off} \quad (17)$$

$$y_2(t) = K_{s2} h(t) - \bar{y}_{2off} \quad (18)$$

In all presented equations index z is related with the air and index v with the water. Parameters of this nonlinear

multivariable system were estimated regarding the following conditions: valves V7, V8, V9, V11, V12 closed, valve V10 open, valve V13 open for two revolutions. Working points regarding input voltage signals were chosen as:

$$\bar{u}_1 = 4V$$

$$\bar{u}_2 = 3V$$

and so the input air and water flows at working points were estimated as:

$$\bar{\Phi}_{zvh} = 22 \times 10^{-6} \text{ kg/s}$$

$$\bar{\Phi}_{vvh} = 7 \times 10^{-3} \text{ kg/s}$$

while model parameters are for the actuators:

$$K_{a1} = \frac{\bar{\Phi}_{zvh}}{u_1^2}, K_{a2} = \frac{\bar{\Phi}_{vvh}}{u_2^5}, T_{a11} = 1s, T_{a12} = 0.8s, T_{a211} = 1.5s,$$

for the sensors:

$$K_{s1} = 0.2V / \text{mbar}, K_{s2} = 0.4V / \text{cm},$$

and for the main process:

$$H_0 = 0.3m \text{ [reservoir height]},$$

$$S = 24 \cdot 10^{-4} \text{ m}^2 \text{ [reservoir cross section]},$$

$$p_0 = 76 \cdot 10^3 \text{ N/m}^2 \text{ [normal air pressure]},$$

$$X_1 = X_2 = X_3 = 0.04,$$

$$y_{1off} = -0.0187V \text{ [air sensor off-set]},$$

$$y_{2off} = 1.2524V \text{ [water sensor off-set]},$$

$$K_B = \frac{r_z T_z}{S} = 3.205 \times 10^7 \text{ s}^{-2},$$

$$C_2 = \rho_v S = 2.3958 \text{ kg/m},$$

$$\bar{y}_1 = 0.8157V - y_{1off} \text{ [1}^{st} \text{ output signal at working point]},$$

$$\bar{y}_2 = 3.1133V - y_{2off} \text{ [2}^{nd} \text{ output signal at working point]},$$

$$K_{zv} = \frac{\bar{\Phi}_{zvh}}{\sqrt{p_z - p_0}} \text{ [valve V13 constant]}, \bar{h}_1 = 0.1012m,$$

$$K_1 = \frac{\bar{\Phi}_{vvh}^2}{h_1} \text{ [valve V10 constant]},$$

$$K_2 = \frac{\bar{\Phi}_{vvh}^2 - K_1 \bar{h}}{p_z - p_0} \text{ [valve V10 constant]}.$$

3.2. Linearized model

By the following selection of system states:

$$x_1(t) = \Delta \Phi_{zvh}(t) \dots 1^{st} \text{ state}, \quad (19)$$

$$x_2(t) = \dot{x}_1(t) \dots 2^{nd} \text{ state}, \quad (20)$$

$$x_3(t) = \Delta \Phi_{vvh}(t) \dots 3^{rd} \text{ state}, \quad (21)$$

$$x_4(t) = \Delta p_z(t) \dots 4^{th} \text{ state}, \quad (22)$$

$$x_5(t) = \Delta h(t) \dots 5^{th} \text{ state}, \quad (23)$$

where Δ indicates the variation from chosen working points, the following linearized state-space equations are defined:

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (24)$$

$$y(t) = Cx(t) + Du(t) \quad (25)$$

where state-space matrices are given with:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ -1.25 & -2.25 & 0 & 0 & 0 \\ 0 & 0 & -0.6667 & 0 & 0 \\ 5.7707 \times 10^6 & 0 & 5.7429 \times 10^3 & -0.1782 & -198.6386 \\ 0 & 0 & 0.4174 & -1.8918 \times 10^{-6} & -0.0144 \end{bmatrix} \quad (26)$$

$$B = \begin{bmatrix} 0 & 0 \\ 1.375 \times 10^{-5} & 0 \\ 0 & 0.0078 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (27)$$

$$C = \begin{bmatrix} 0 & 0 & 0 & 0.002 & 0 \\ 0 & 0 & 0 & 0 & 40 \end{bmatrix} \quad (28)$$

$$D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (29)$$

4. CONTROL DESIGN

Control system major tasks were in the presented case focused on two important goals: to overcome system stiffness which in the same time results in better robust behavior and to overcome the property of weak inherent coupling which demands a construction of a dynamic compensator in a combination with the observer in case of the design goal of dynamic system decoupling. This is important for high quality of multivariable system operation as cross-couplings introduce disturbances in reference-tracking operation.

We have tested and compared different control design approaches. During modelling, simulation and analysis phase where also first close-loop solutions were tested certain knowledge of the system was accumulated which we tried to include into the control design phase by prescribing desired responses of the system to the reference signals as illustrated in Figure 12.

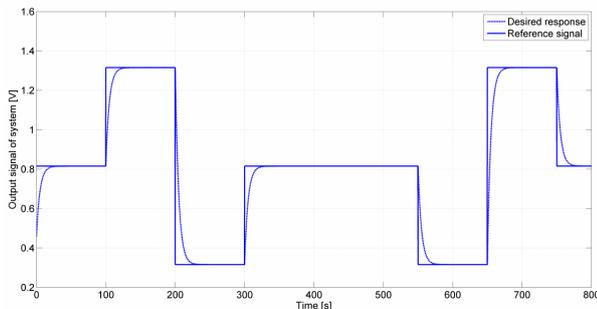


Figure 12: Example of reference signal and desired responses of the system

The velocity of transient responses was estimated regarding control signals which have to be inside the range of actuators. In case of decoupling not only cross-couplings are compensated but also relative orders of each univariable sub-system is correspondingly lower. In our case we have taken into account that approximate performance of the first order can be achieved.

Such system responses were then used in fitness function description through optimization process with which two controllers were designed. First controller is classical multivariable PI or proportional-integral controller and the second one was generated by the AMEBA algorithm.

4.1. Design of the multivariable PI controller

For the system with two input and two output signals multivariable PI controller is described with:

$$\vec{u}(t) = K_p \vec{e}(t) + K_i \int \vec{e}(t) dt$$

We used simplex optimization method from Matlab to calculate the eight parameters of the PI-controller as:

$$\begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix} = \begin{bmatrix} 8.97 & -0.037 \\ 0.71 & 2.55 \end{bmatrix} \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} + \begin{bmatrix} 0.01 & -0.00004 \\ 0.00082 & 0.0029 \end{bmatrix} \int \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} dt \quad (30)$$

Evaluation results are presented in Table 3.

Table 3: Results of PI controller

Error [%]	Activity of the actuators [%]
17%	34%

These results show that controller can control the system with 17% error and with the use of 35% of all capacity of actuators regarding the simulated run. Majority of error is generated by the cross-coupling effect that PI controller could not cope with. System responses are presented in Figures 13 and 14 and corresponding control signals in Figure 15.

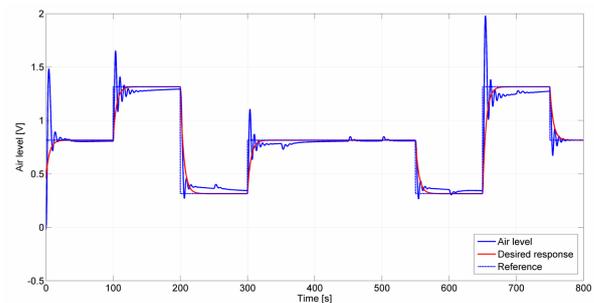


Figure 13: Response of the air-pressure to the change of reference signal

Response of the air pressure is very oscillatory which is not desired and the response of the water level shows that controller doesn't provide good reference tracking as well. Significant is also the consequence of cross-coupling effects. At the same time it is obvious that the response of the air part is much faster as that of water part.

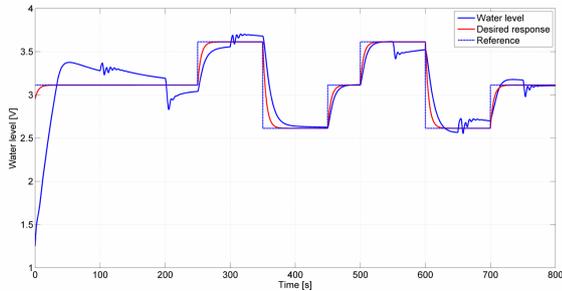


Figure 14: Response of the water level to the change of the reference signal

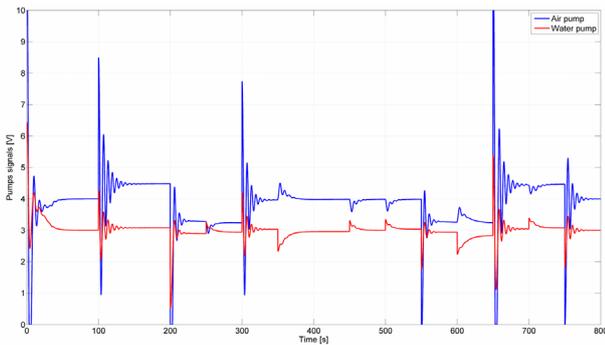


Figure 15: Activity of the air and water pumps

It is important to mention that such oscillatory responses are not desired also because of the actuators as in such cases the lifespan is expected to be reduced.

4.2. Design of the controller with the AMEBA algorithm

AMEBA algorithm has a very useful capability that it can include external knowledge of the problem into its solution. In this case we have used multivariable PI-controller that was designed through the first experiment as a starting point for the AMEBA algorithm. The agent representation of the multivariable PI-controller is presented in Figure 16.

The presented PI-controller was included as a member of the initial population in optimization process. During the optimization or evolution process there were no limitations regarding agent's structure, nor nodes' properties.

Control design results of AMEBA algorithm are presented in Table 4.

The results show an improvement of 11% in error reduction with practically no increase in actuators activities which represents significant improvement over previously presented multivariable PI-controller. The difference can be observed also in Figures 17, 18, and 19. Agent that was generated is presented in Figure 20.

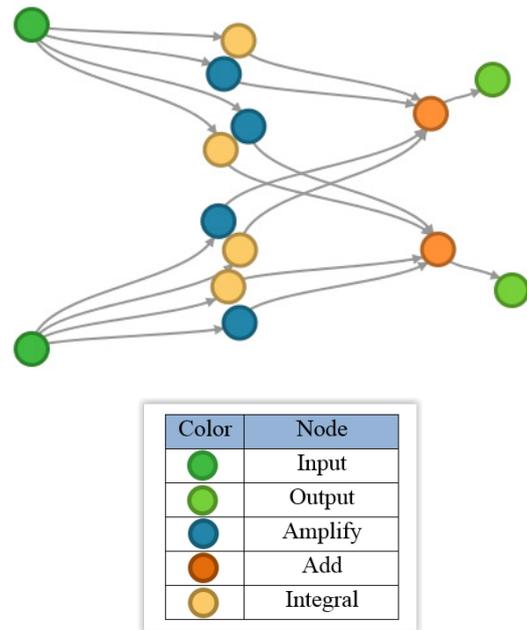


Figure 16: Representation of multivariable PI-controller as an agent of AMEBA algorithm

Table 4: AMEBA algorithm control design results

Error [%]	Activity of the acutators [%]
5,7%	36%

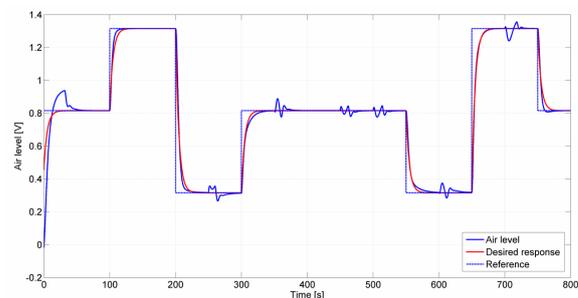


Figure 17: Response of the air pressure to the change of reference signal

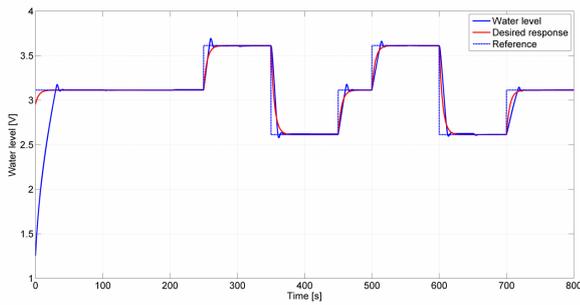


Figure 18: Response of the water level to the change of reference signal

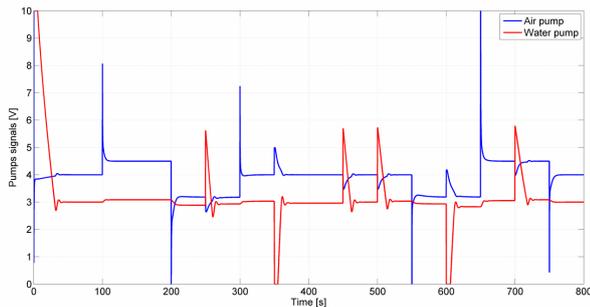
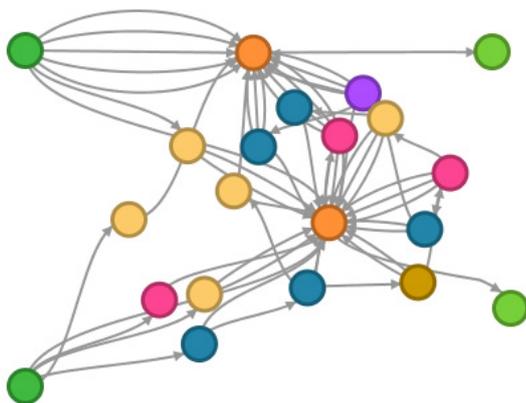


Figure 19: Activity of the air and water pumps



Color	Node
Green	Input
Light Green	Output
Blue	Amplify
Yellow	Integral
Pink	Derivative
Purple	Delay
Brown	Flow pass filter

Figure 20: Controller generated by AMEBA algorithm

5. CONCLUSIONS

We have presented new optimization algorithm AMEBA. Its efficacy was illustrated through the improvement of an existing multivariable PI-controller

for the pressure-level pilot plant. The upgraded effects were especially evident in reduction of the cross-couplings and in reduction of the oscillations which can extend actuators' lifespan and increase control quality. Future work on AMEBA algorithm development will be focused on optimization process as we will explore the impact of various effects on the quality of the solution and on the convergence rate of optimization process like effect of size of population, effect of suppression of the agents with large number of nodes, effect of using multiple environments at once and similar, of course in comparison with other optimization approaches. Special attention will be devoted to the so called smart optimization where additional knowledge from chosen area can be taken into account to improve searching efficacy.

REFERENCES

- Atanasijević-Kunc, M., Belič, A., and Karba, R., 2006. Optimal Multivariable Control Design Using Genetic Algorithms. *5th Vienna Symposium on Mathematical Modeling*, Vienna University of Technology.
- Atanasijević-Kunc, M., and Karba, R. 2006. Multivariable Control Design with Expert-Aided Support, *WSEAS Trans. Syst* 5(10):2299–2306.
- Atanasijević-Kunc, M., Belič, A., and Karba, R. 2007. Optimal Multivariable Control Design Using Genetic Algorithms. *EUROSIM simulation news Europe* 17:41–45.
- Atanasijević-Kunc, M., Karba, R., and Zupančič, B. 1997. Multipurpose Modelling in the Evaluation of Laboratory Pilot Plant. *Simulation Practice and Theory* 5:751–76.
- Beyer, H. G. 2010. *The Theory of Evolution Strategies (Natural Computing Series)*. Springer, Retrieved March 14, 2012 (<http://www.amazon.com/Theory-Evolution-Strategies-Natural-Computing/dp/3642086705>).
- Brownlee, J. 2011. *Clever Algorithms: Nature-Inspired Programming Recipes*. Swinburne University in Melbourne, Australia.
- Corn, M. and Atanasijević-Kunc, M., 2011. Cell Based Genetic Programming Toolbox. 20th ERK Conference, Portorož, pp.295–98.
- Corn, M. and Černe, G., 2012. A Graph-Based Evolutionary Algorithm: Cell Based Genetic Programming. *Proceedings of the Fifth International Conference on Bioinspired Optimization Methods and their Applications, BIOMA*, Kranjska gora, pp. 163–72.

- Corn, M., Černe, G., and Atanasijević-Kunc, M.. 2012. Balance Group Model with Smart Grid Elements. *7th Vienna Conference on Mathematical Modeling: MathMod 2012*. Vienna, February 15.-18.2012.: Vienna University of Technology.
- Goldberg, D., 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st ed. Massachusetts: Addison Wesley.
- Fogel, L. J., Owens, A. J., and Walsh, M. J., 1966. *Artificial Intelligence through Simulated Evolution*. John Wiley.
- Koza, J. R. 1992. *Genetic Programming On the Programming of Computers by Means of Natural Selection*. 6th ed. Cambridge, Massachusetts, London, England: MIT Press.
- Logar, V., Dovžan, D., and Škrjanc, I., 2011. Mathematical Modeling and Experimental Validation of an Electric Arc Furnace. *ISIJ International* 51(3):382–91.
- Storn, R., and Kenneth P., 1997. Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* 11:341–59.
- Tomažič, S. et al. 2013. “Indoor-Environment Simulator for Control Design Purposes. *Building and Environment* 70(0):60–72. Retrieved (<http://www.sciencedirect.com/science/article/pii/S0360132313002527>).
- Veselý, V., and Rosinová, D., 2011. Robust PSD Controller Design. *18th International Conference on Process Control*. Tatranská Lomnica, Slovakia, pp. 479–84. Retrieved (<https://www.kirp.chof.stuba.sk/pc11/data/papers/077.pdf>).
- Whigham, P. A. 1992. Grammatically-Based Genetic Programming. *Workshop on Genetic Programming: From Theory to Real-World Applications*. Tahoe City, California: Department of Computer Science, University College, University of New South Wales, pp. 33–41.

A FUZZY COGNITIVE MAPS MODEL TO DEVELOP A RISK ANALYSIS MODEL THROUGH THE IDENTIFICATION OF CRITICAL HUMAN FACTORS

Fabio De Felice^(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, ^(b) antonella.petrillo@uniparthenope.it, ^(c) agostino@itim.unige.it, ^(d) francesco.longo@unical.it

ABSTRACT

Over the past years, technological developments have led to a decrease of accidents due to technical failures through the use of redundancy and protection. However, the "human factor" contributes significantly in accident dynamics, both statistically and in terms of severity of consequences. In fact, estimates agree that the errors committed by man are causes over 60% of accidents and for the remaining part the causes are due to technical deficiencies. This paper deals with various aspects of human behavior that can influence operator reliability, considering the environment in which operator is working. The aim of the research is to propose a fuzzy cognitive map capable to represent human knowledge where the relationships are difficult to describe in mathematical terms.

Keywords: human error, risk analysis, fuzzy cognitive maps, cognitive model

1. INTRODUCTION

Nowadays, the analysis of human factors constitute a highly interdisciplinary field of study not yet well defined, therefore, a complete and universally accepted taxonomy of different types of human errors and causes determining them, does not exist. The objective difficulties of governing the human factor and human error have made many experts believe that the conduct of prevention and safety were related to a person's intrinsic characteristics, such as personality traits. (D'Elia *et al.*, 2013).

In the literature, there are different risk-based approaches reported, ranging from the purely qualitative to the quantitative (Bruzzone *et al.*, 2014 a). Many authors used probabilistic risk assessment. A vast majority of tools and techniques available for the Human Reliability Analysis (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 the literature, some papers proposed the Fuzzy cognitive maps (FCMs) approach in the field of risk analysis.

In fact, human element in risk analysis cannot be structured in a hierarchical way, as it involves

interaction and dependence among factors under various heads, like organizational, personal, design related, task related, and so on. However, this can be analyzed with a network structure, like the use of cognitive maps.

FCMs have gained considerable research interest and widely used to analyse complex systems and making decisions. Recently, they found large applicability in diverse domains for decision support and classification tasks. FCM is an efficient knowledge representation and reasoning method, which is based on human knowledge and experience (Kandasamy and Smarandache, 2003).

This paper deals with various aspects of human behavior that can influence operator reliability, considering the environment in which operator is working (Bruzzone *et al.*, 2014 b). The focus is on understanding the nature of human performance variability and eventually, how to describe and analyze it. The proposed fuzzy cognitive map is capable of representing human knowledge where the relationships are difficult to describe in mathematical terms.

The applicability of the methodology is demonstrated through a case study. The paper is organized in the following sections: section 2 presents general features of fuzzy cognitive maps approach; section 3 analyzes human factors in the context of risk analysis; section 4 elaborates a fuzzy cognitive model to evaluate factors influencing human reliability; finally, section 5 presents results and conclusion.

2. FUZZY COGNITIVE MAP APPROACH

Fuzzy Cognitive Maps (FCM) have found favor in a variety of theoretical and applied contexts that span the hard and soft sciences. Given the utility and flexibility of the method, coupled with the broad appeal of FCM to a variety of scientific disciplines, FCM have been appropriated in many different ways and, depending on the academic discipline in which it has been applied, used to draw a range of conclusions about the belief systems of individuals and groups (Groumpos, 2010).

FCM encompasses a wide range of applications including: risk assessment (Hurtado, 2010) work efficiency and performance optimization strategic deterrence and crisis management, scenario/policy

assessment (Amer *et al.*, 2011) spatial suitability and prediction mapping and environmental modeling and management (Adriaenssens *et al.*, 2004). FCMs are most often employed in participatory planning and management and/or environmental decision-making contexts, and are primarily used to gain an understanding of how stakeholders internally construct their understanding of their world or a particular issue of interest (Kontogianni *et al.*, 2012).

For the above reasons Fuzzy Cognitive Maps (FCMs) constitute an attractive modeling technique for complex systems (Jose, 2010).

The term fuzzy cognitive map (FCM) was coined in (Kosko, 1986) to describe a cognitive map model with two significant characteristics:

- *Causal relationships between nodes are fuzzified.* Instead of only using signs to indicate positive or negative causality, a number is associated with the relationship to express the degree of relationship between two concepts.
- *The system is dynamic involving feedback,* where the effect of change in a concept node affects other nodes, which in turn can affect the node initiating the change. The presence of feedback adds a temporal aspect to the operation of the FCM.

FCM has its roots in concept and cognitive mapping. Concept maps are graphical representations of organized knowledge that visually illustrate the relationships between elements within a knowledge domain. By connecting concepts (nodes) with semantic or otherwise meaningful directed linkages, the relationships between concepts in a hierarchical structure are logically defined (Novak and Cañas, 2008; Papageorgiou and Salmeron, 2011).

In other words, FCM is a soft computing technique that follows an approach similar to the human reasoning and decision-making process. An FCM consists of nodes which illustrate the different aspects of the system behavior.

These nodes (concepts) interact with each other, illustrating the dynamics of the model. FCM is a complex form of data collection where study participants are asked to develop qualitative static models which are translated into quantitative dynamic models.

A cognitive map can be thought of as a concept map that reflects mental processing, which is comprised of collected information and a series of cognitive abstractions by which individuals filter, code, store, refine and recall information about physical phenomena and experiences.

Fuzzy cognitive map uses two types of elements: concepts, represented by nodes; and causal beliefs, represented by weighted edges connecting the nodes. The graphical representation clearly depicts interaction of concepts and their degree of dependence/interdependence (Kumar *et al.*, 2013). Figure 1 illustrates an example of an FCM with seven concepts. Concepts characterize the behavior of the system under

consideration, and it can be an action, variable, or state, and so on whose values change with time.

The FCM structure can be viewed as a recurrent artificial neural network, where concepts are represented by neurons and causal relationships by weighted links or edges connecting the neurons.

By using Kosko's conventions, the interconnection strength between two nodes C_i and C_j is e_{ij} , with e_{ij} , taking on any value in the range -1 to 1. Values -1 and 1 represent, respectively, full negative and full positive causality, zero denotes no causal effects and all other values correspond to different fuzzy levels of causal effects. In general, an FCM is described by a connection matrix E whose elements are the connection strengths (or weights) e_{ij} . The element in the i^{th} row and j^{th} column of matrix E represents the connection strength of the link directed out of node C_i and into C_j . If the value of this link takes on discrete values in the set $\{-1, 0, 1\}$, it is called a simple FCM. The concept values of nodes C_1, C_2, \dots, C_n (where n is the number of concepts in the problem domain) together represent the state vector C .

An FCM state vector at any point in time gives a snapshot of events (concepts) in the scenario being modelled.

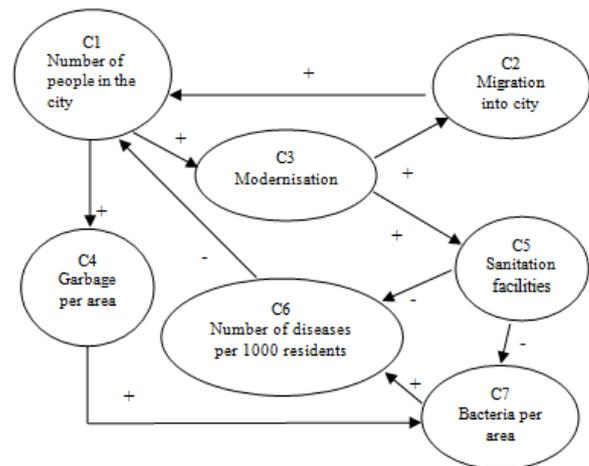


Figure 1: Example of fuzzy cognitive maps

The nodes represent concepts or variables relevant to a given domain. The causal links between these concepts are represented by the edges. The edges are directed to show the direction of influence. Apart from the direction, the other attribute of an edge is its sign, which can be positive (a promoting effect) or negative (an inhibitory effect). Cognitive maps can be pictured as a form of signed directed graph.

The construction of a cognitive map requires the involvement of a knowledge engineer and one or more experts in a given problem domain. In a cognitive map, the effect of a node A on another node B, linked directly or indirectly to it, is given by the number of negative edges forming the path between the two nodes. The effect is positive if the path has an even number of negative edges, and negative otherwise.

3. ROLE OF HUMAN FACTORS IN RISK ANALYSIS

With the development of nuclear power, and the increasing complexity of automated manufacturing processes, the magnitude and significance of accidents have greatly increased. The Chernobyl nuclear power plant disaster in the Ukraine and the Challenger spacecraft explosion are vivid examples.

In order to ensure effective prevention of dangerous events, the role of humans in accident dynamics should be considered during risk assessment processes (D'Elia *et al.*, 2014).

The researchers' great efforts to propose models of human behavior (French *et al.*, 2008) favoring numerical values of error probability in order to predict and prevent unsafe conduct are clearly evident.

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). Human Reliability Analysis (HRA) grew up in the 1960s, with the intention of modeling the likelihood and consequences of human errors (Sharit, 2012).

Human reliability is a crucial element in ensuring plant performance, and it is critical to situations or activities where the operator interacts directly with the system the stress factors include physiological and psychological stressors like fear, monotonous workload, overload, and so on. There are not many studies dedicated to the assessment of performance shaping factors (PSFs) on human reliability.

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).

Risk assessment is a systematic process for assessing the impact, occurrence and the outcome of human activities involving products or systems with hazardous characteristics. Three questions need to be answered: "What can go wrong?"; "How likely is it that this will happen?"; "If it does happen, what are the consequences?" (Kaplan and Garrick, 1981).

Risk identification usually involves specifying one or more scenarios of risks. A risk scenario describes an interaction between a person and a system or product that possesses hazardous characteristics. It describes the activity of the person(s) involved, the hazard(s), the external factors of the situation and the potential injury.

Methods for risk assessment need to be as univocal and precise as possible to differentiate the risk level of various activities. In this light, we need to reduce the uncertainty in measurement of qualitative attributes such as "severity" and "likelihood" (Carter, 2003).

The focus is on understanding the nature of human performance. The likelihood of human errors in a specific condition depends upon the combined effects.

4. IMPLEMENTATION OF A FUZZY COGNITIVE MODEL TO EVALUATE HUMAN INFLUENCE IN RISK ANALYSIS

Given the basic outline of risk assessment, this paper explores the factors and parameters that need to be considered in order to design a good quality risk assessment.

It is important to realise that perceived risks and benefits are important, and that these may differ from person to person.

The proposed model deals with various aspects of human behavior that can influence his/her reliability, considering the environment in which he/she is working and the nature of task.

The main objective of building a cognitive map around a problem is to be able to predict the outcome by letting the relevant issues interact with one another. These predictions can be used for finding out whether a decision made by someone is consistent with the whole collection of stated causal assertions.

In the proposed research Mental Modeler software is used. Mental Modeler is a fuzzy-logic cognitive mapping software (<http://www.mentalmodeler.org/>), allows you to build Fuzzy-logic Cognitive Maps easily and intuitively (Gray *et al.*, 2013). Once models are built, increasing or decreasing the components included in the model allows you to examine different scenarios of change. Because of their flexibility, FCM have been used in a range of scientific disciplines, from political science to economics to ecology.

Based in Fuzzy-logic Cognitive Mapping (FCM), users can easily develop semi-quantitative models of environmental issues, social concerns or social-ecological systems in Mental Modeler by:

- Defining the important components of a system
- Defining the relationships between these components
- Running "what if" scenarios to determine how the system might react under a range of possible changes

Here below case study framework is presented.

Step #1: Problem description.

In the present case study a chemical company is evaluated. The data collection phase followed the normative literature for conducting fieldwork in risk management. The authors drew upon their own extensive industrial experience to inform and guide the design of an interview protocol that defined the scope of the necessary data to explore.

Semi-structured interviews were conducted with the Managing Director (MD), Production Director (PD), and Production Manager (PM) in order to:

- determine or describe the context;
- describe the possible error modes;
- describe the probable causes;
- perform a more detailed analysis of main task steps.

Step#2: Factors influencing human performance.

Based on the literature review and personal experience, the human error in risk analysis is highly dependent on state and condition of the personnel engaged in the work. Major factors, which characterize the behavior of the personnel are the following:

- Emotional stability C1
- Knowledge and skill C2
- Attention and alertness C3
- Perception and memory C4
- Motivation C5
- Workplace environment C6

- Clarity of instructions C7
- Communication C8
- Time pressure C9

The above factors formed our fuzzy cognitive map.

Step#3: Fuzzy cognitive map model development

The factors that influence others, along with their linguistic degree of causation are determined as shown in Figure 2. While in Figure 3 fuzzy cognitive map's metrics are shown.

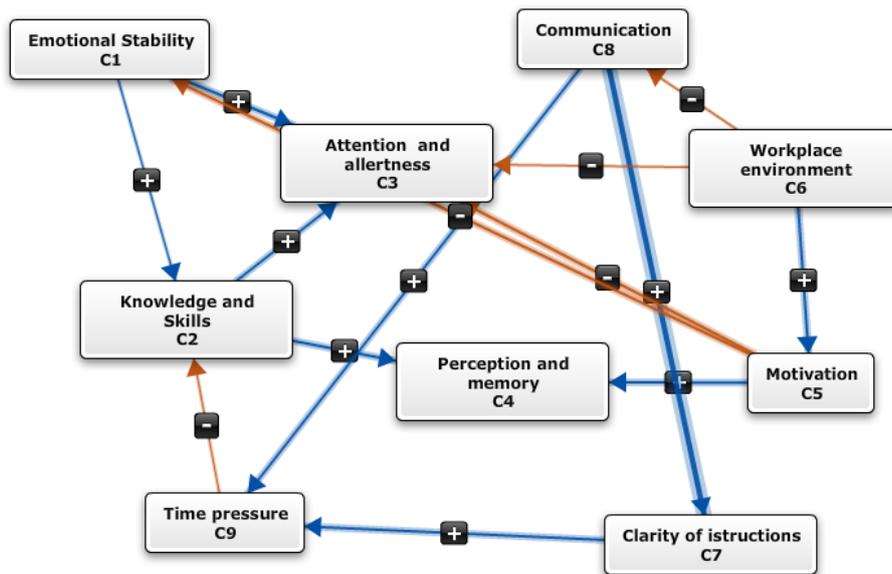


Figure 2: Fuzzy cognitive map

Number of Concepts	9	Highest Centrality Variables	
Number of Connections	14	Motivation\nC5	2
Density	0,1728395062	Attention and alertness\nC3	1,75
Number of Connections / Components	0,6428571429	Communication\nC8	1,75
Driver Components	Attention and alertness\nC3 1,75 Perception and memory\nC4 1	Knowledge and Skills\nC2	1,5
Receiver Components	Workplace environment\nC6 1	Clarity of instructions\nC7	1,5
Number of Ordinary	6	Emotional Stability\nC1	1,25
Complexity Score	0,5	Time pressure\nC9	1,25
Highest Centrality Variables	Motivation\nC5 2 Attention and alertness\nC3 1,75 Communication\nC8 1,75 Knowledge and Skills\nC2 1,5 Clarity of instructions\nC7 1,5 Emotional Stability\nC1 1,25 Time pressure\nC9 1,25 Perception and memory\nC4 1 Workplace environment\nC6 1	OutDegree Values	Attention and alertness\nC3 1,75 Perception and memory\nC4 1 Clarity of instructions\nC7 1 Time pressure\nC9 1 Emotional Stability\nC1 0,5 Knowledge and Skills\nC2 0,5 Motivation\nC5 0,5 Communication\nC8 0,25 Workplace environment\nC6 0
		InDegree Values	Motivation\nC5 1,5 Communication\nC8 1,5 Knowledge and Skills\nC2 1 Workplace environment\nC6 1 Emotional Stability\nC1 0,75 Clarity of instructions\nC7 0,5 Time pressure\nC9 0,25 Attention and alertness\nC3 0 Perception and memory\nC4 0

Figure 3: Fuzzy cognitive map's metrics

The degree of relationships among the considered factors is established following consensus group method. Here, each member contributes to the discussion, and the group as a whole then arrives at an estimate upon which all members of the group agree. After concepts in the model have been determined, relationships between concepts can be added by using directional arrows which indicate the amount of influence one component can have on another, called

edge relationships. Concepts included in the model can have positive (high, medium, or low), negative (high, medium, or low) or no (no relationship defined) edge relationships.

The qualitative weights of edge relationships (i.e. “fuzzy” approximation of influence) between components are then translated into the quantitative values between -1 (high negative) to 1 (high positive) used in the matrix interface (Table 1).

Table 1: The qualitative weights of edge relationships

	C1	C2	C3	C4	C5	C6	C7	C8	C9
C1	0	0,5	0,25	0	0	0	0	0	0
C2	0	0	0	0	0	0	0	0	0
C3	0	0,5	0	0	0,5	0	0	0	0
C4	-0,5	-0,5	0	0	0,5	0	0	0	0
C5	0	0	0	0	0	0	0	0	0
C6	0	-0,25	0	0,5	0	0	-0,25	0	0
C7	0	0	0	0	0	0	0	1	0,5
C8	0	0	0	0	0	0	0	0	0,5
C9	0	0	-0,25	0	0	0	0	0	0

Mental Modeler also includes a Matrix interface that converts the concept map built in the Concept Mapping interface into a structural matrix. The Matrix

interface can easily be revised based on the original concept map once the users familiarize themselves with the structure of the tool (Figure 4).

Effect	Value	Emotional St...	Attention a...	Knowledge...	Motivation\...	Perception a...	Workplace e...	Communicat...	Clarity of istr...	Time pressu...
H+	1		M+	L+						
M+	0,5									
L+	0,25			M+		M+				
	0		M-	M-		M+				
L-	-0,25									
M-	-0,5			L-		M+		L-		
H-	-1								H+	M+
										M-
					L-					

Figure 4: Matrix

Step#4: Scenario analysis

Changing the weight of the components selected from each of the scenarios, it is possible to estimate how the system reacts to changes induced described. In other words, the scenario interface indicates the amount of relative change in the components included in the model based on the edge relationships defined in the Concept Mapping interface for the chosen scenario. It is possible to decide what scenario to run based on probable, improbable, gradual and extreme changes to the system.

It is important to note, that the results presented here must be read in a comparative way, not absolute, that is used to check which components, given the

structure of the map, are affected more than the other simulated change and with what results. This type of analysis is useful to deepen the knowledge of the system and to develop strategies and assumptions useful to react to the occurrence of events simulated. Scenario analysis was discretized into three modes in ascending order of control and performance reliability, and thus in descending order of human failure probability: scrambled, tactical and strategic, as defined below:

- *Scrambled scenario*: choice of forthcoming action is unpredictable or haphazard;
- *Tactical scenario*: performance typically follows planned procedures while some ad-hoc deviations are still possible;

- *Strategic scenario*: plentiful time is available to consider actions to be taken in light of wider objectives to be fulfilled and within the given context.

Figures 5, 6 and 7 show the three different scenario.

In the **first scenario**, suppose the C1, C2, C3 and C6 “high”, according to the adopted linguistic scale. The output indicates that a positive change influences positively the

factors C5 and C4. In the **second scenario** suppose the C6, C7 and C8 “high”, according to the adopted linguistic scale. The output indicates that a positive change influences positively the factors C5, C4 and C9. In the **third scenario** suppose the C1, C6 and C9 “high”, according to the adopted linguistic scale. The output indicates that a positive change influences positively the factors C2, C3, C4 and C5.

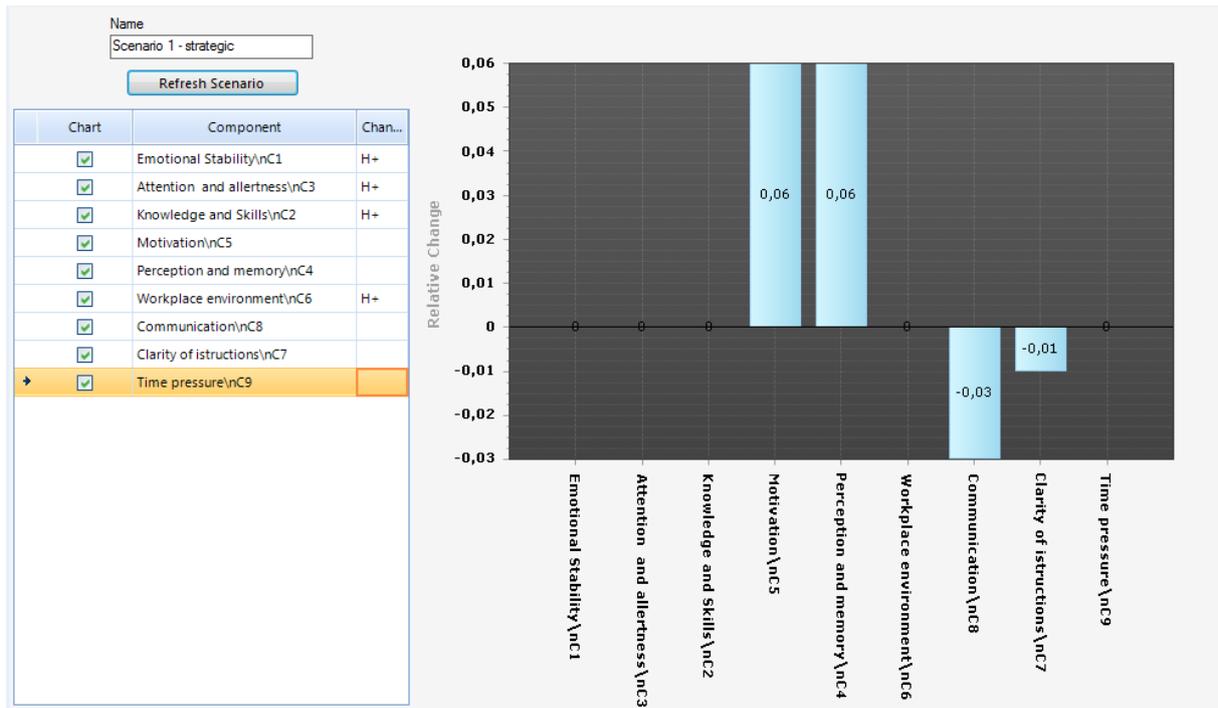


Figure 5: Strategic scenario

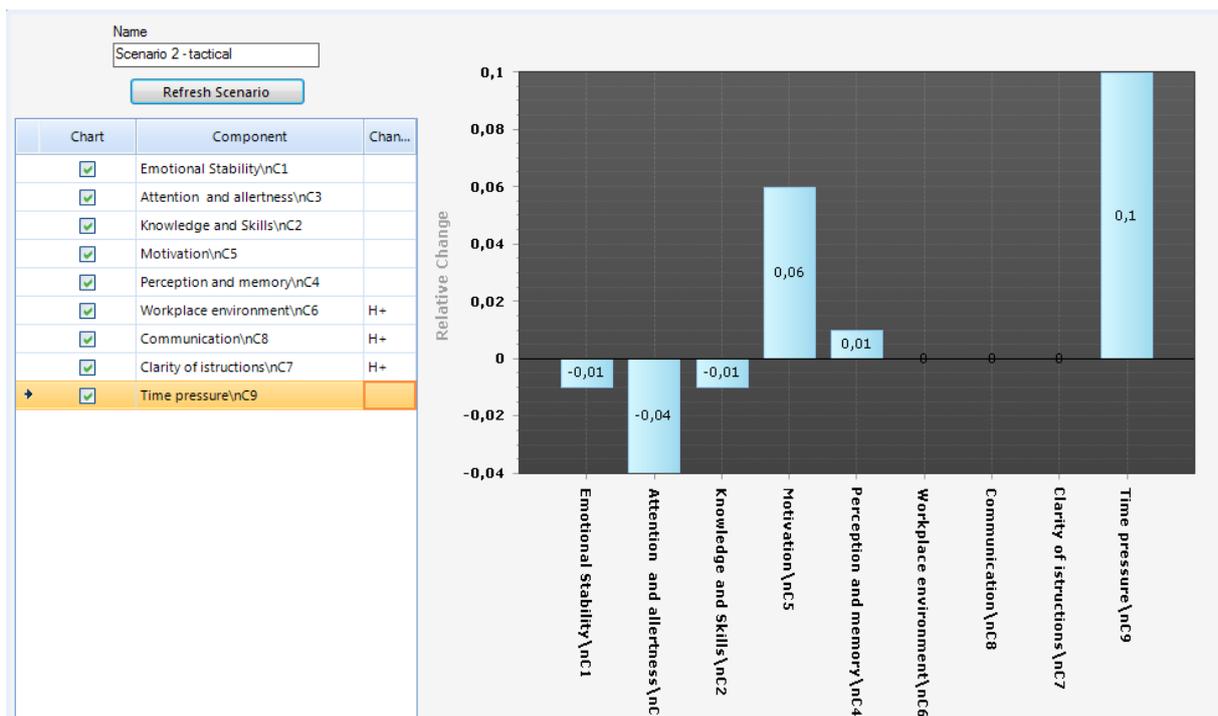


Figure 6: Tactical scenario

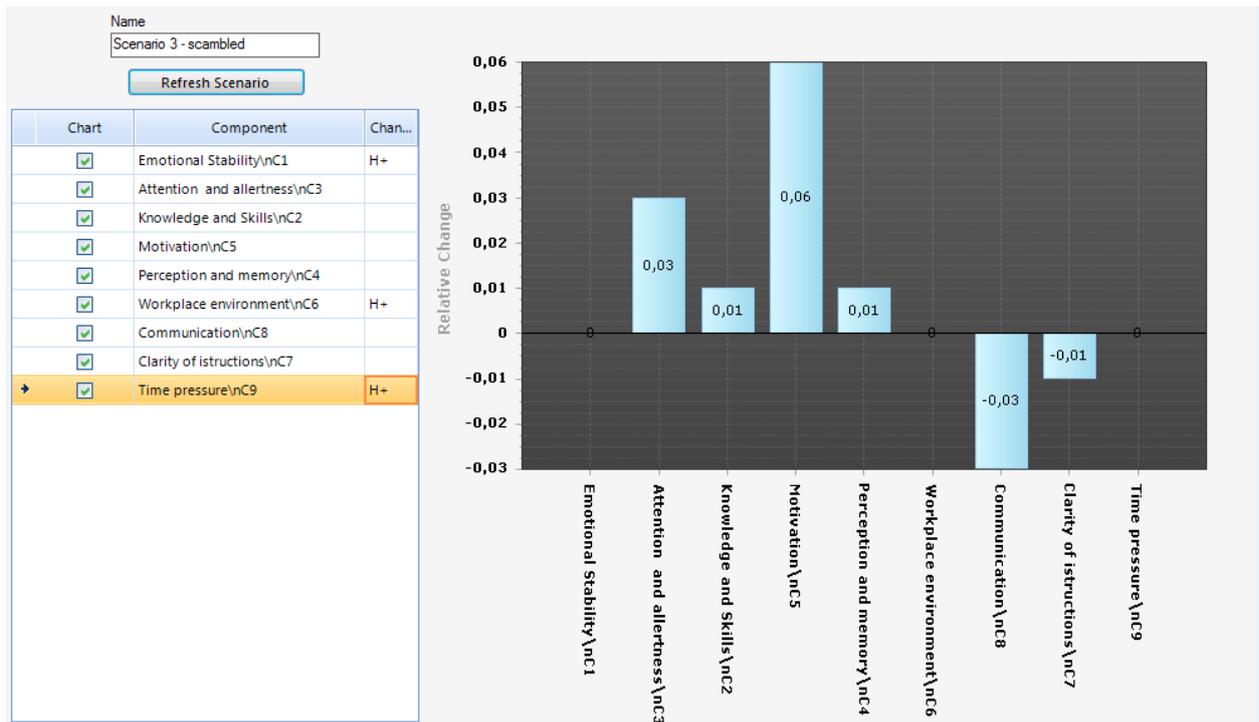


Figure 7: Scrambled scenario

6. CONCLUSION

The main question of risk evaluation is ‘‘how safe is safe enough?’’

An fuzzy cognitive model was applied to manage the human factors in order to improve risk analysis.

Fuzzy systems in general have already proven themselves capable of dealing with the inherent imprecision encountered in real-world problem domains. Fuzzy cognitive maps combine this advantage with the well-established utility of cognitive maps as a decision support tool, and the dynamic and adaptive nature of artificial neural networks.

The proposed FCM methodology evaluates the influence of error inducing factors on human reliability and has potential to examine a scenario, where it is not easy to describe in terms of relations/mathematical formulas.

Future research will focus in the fuzzy quantification of the human error probability within the framework of the Cognitive Reliability and Error Analysis Method (CREAM) method and other methods within HRA approach in order to compare different results and approach.

REFERENCES

- Adriaenssens V., De Baets B., Goethals P.L.M., De Pauw, N., 2004. Fuzzy rule-based models for decision support in ecosystem management. *Sci. Total Environ*, 319(1–3), 1–12.
- Amer M., Jetter A., Daim T., 2011. Development of fuzzy cognitive map (FCM)-based scenarios for wind energy. *Int. J. Energy Sect. Manage*, 5(4), 564–584.
- Bruzzone A.G., Frascio M., Longo F., Chiorco A., Zanoni S., Zavanella, L., Fadda P., Fancello G., Falcone D., De Felice F., Petrillo A., Carotenuto, P., 2014 (a). Disaster and emergency management simulation in industrial plants. *Proceedings of the International Conference on Modeling and Applied Simulation (EMSS 2014)*. September 10-12 2014, Bordeaux, France, pp 649-656. ISBN 978-88-97999-34-8 (Paperback) ISBN 978-88-97999-40-9 (PDF).
- Bruzzone A.G., Longo F., Massei M, Nicoletti L., 2014 (b), Human modeling for multi coalition joint operations Summer Computer Simulation Conference, SCSC 2014, Part of the 2014 Summer Simulation Multiconference, SummerSim 2014; Monterey; United States; 6 July 2014 through 10 July 2014.
- Cacciabue P.C., 1996. Understanding and modelling man-machine interaction. *Nuclear Engineering and Design*, 165:351–358.
- Carter D.A., Hirst I.L., Maddison T.E., Porter S.R., 2003. Appropriate risk assessment methods for major accident establishments. *Trans IChemE* 81B, 12–18.
- D’Elia C., De Felice F., Mariano P., Petrillo A., Ruscino, S., 2014. Probabilistic approach as a support system for safety-critical environments by knowledge driven bayesian networks. *Proceedings of the International Conference on Modeling and Applied Simulation (EMSS 2014)*. September 10-12 2014, Bordeaux, France, pp 280-288. ISBN 978-88-97999-34-8 (Paperback) ISBN 978-88-97999-40-9 (PDF).
- 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. Code 100159
- French et al., 2009. Human Reliability Analysis: A Review and Critique. Manchester Business School Working Paper.
- Steven A. Gray, Stefan Gray, Linda J. Cox, Sarah Henly-Shepard, 2013. Mental Modeler: A Fuzzy-Logic Cognitive Mapping Modeling Tool for Adaptive Environmental Management, 46th Hawaii International Conference on System Sciences.
- Groumpos P., 2010. Fuzzy cognitive maps: basic theories and their application to complex systems. In: Glykas, M. (ed.) Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications, pp. 1–22. Springer, Berlin.
- 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; 1–6.
- Hollnagel E., 2005. Human reliability assessment in context. Nuclear Engineering and Technology, 37(2):159–166.
- Hurtado S.M., 2010. Modeling of operative risk using fuzzy expert systems. In: Glykas, M.(ed.) Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications. pp. 135–159. Springer, Berlin.
- Jose A., 2010. dynamic fuzzy cognitive maps for the supervision of multiagent systems. In: Glykas, M. (ed). Fuzzy Cognitive Maps: Advances in Theory, Methodologies, Tools and Applications. Springer, Berlin. pp. 307–324.
- Kandasamy WBV, Smarandache, (2003). F. Fuzzy Cognitive Maps and Neutrosophic Cognitive Maps. Xiquan: Phoenix, USA.
- Kaplan, S., Garrick, B.J., 1981. On the quantitative definition of risk. Risk Analysis 1, 11–27
- Kontogianni A., Papageorgiou E., Salomatina L., Skourtos M., Zanou B., 2012. Risks for the black sea marine environment as perceived by ukrainian stakeholders: a fuzzy cognitive mapping application. Ocean Coast. Manag. 62, 34–42.
- Kosko B., 1986. Fuzzy cognitive maps. International Journal of Man-Machine Studies, 24(1):65–75.
- Aju kumar V.N., Gandhi M.S., Gandhi O.P., 2015. Identification and Assessment of Factors Influencing Human Reliability in Maintenance Using Fuzzy Cognitive Maps. Quality Reliability Engineering International, 31 169–181.
- Novak J.D., Cañas A.J., 2008. The theory underlying concept maps and how to construct and use them. Technical Report IHMC CmapTools 2006–01 Rev 01–2008. Institute for Human and Machine Cognition. Florida. pp. 36.
- Papageorgiou E. I., Salmeron J.L., 2011. A review of fuzzy cognitive map research during the last decade. IEEE Trans. Fuzzy Syst. (IEEE TFS) 99, pp.1–14.
- Sharit J., 2012. Human error and human reliability analysis. In Handbook of Human Factors (4th edn). Salvendy G (ed). John Wiley and Sons, Inc.: Hoboken, New Jersey, 737.

NUMEQUARES — ONLINE SIMULATION TOOL FOR EDUCATION AND ACADEMIC RESEARCH

Stepan Orlov^(a), Nikolay Shabrov^(b)

^{(a),(b)} Computer Technologies in Engineering dept.,
St. Petersburg State Polytechnical University,
St. Petersburg, Russia

^(a)majorsteve@mail.ru, ^(b)shabrov@rwwws.ru

ABSTRACT

A new online simulation tool, NumEquaRes, is presented. While other simulation tools focus on helping to formulate model equations and provide a limited number of analysis types, NumEquaRes focuses on the formulation of the analysis algorithm to be applied to the model. The model and the algorithm together make a data processing system that can be run in order to obtain numerical results. We have found that some simple algorithms, e.g., generating Poincaré map, require too much programming when implemented using popular tools such as MATLAB, Mathematica, Maple, and others. This motivates the development of yet another simulation tool providing much more flexibility for simulation algorithms without having to code them. The paper presents a set of example simulations that illustrate the ability to perform numerical investigations of dynamic systems. NumEquaRes is an open source system available for use and further extension to everyone. It is hosted at <http://equares.ctmech.ru>.

Keywords: simulation, web application, numerical research.

1. INTRODUCTION

In this work we present a new Web application, NumEquaRes (the name means “Numerical Equation Research”). It is a general tool for numerical simulations available online. Currently, we are targeting small systems of ordinary differential equations (ODE) or finite difference equations arising in the education process, but that might change in the near future — see Sections 8, 9.

The reasons for developing yet another simulation software have emerged as follows. Students were given tasks to deduce the equations of motions of mechanical systems — for example, a disk rolling on the horizontal plane without slip (Routh 1905), or a classical double pendulum (Meirovitch 1986), — and to try further investigating these equations. While in some cases such an investigation can more or less easily be done with MATLAB, SciLab, or other existing software, in other

cases the situation is like there is no (freely available) software that would allow one to formulate the task for numerical investigation in a straightforward and natural way.

For example, the double pendulum system exhibits quasi-periodic or chaotic behavior (Meirovitch 1986), depending on the initial state. To determine which kind of motion corresponds to certain initial state, one needs the Poincaré map (Teschl 2012) — the intersection of phase trajectory with a hyperplane. Of course, there are ODE solvers in MATLAB that produce phase trajectories. We can obtain these trajectories as piecewise-linear functions and then compute intersections with the hyperplane. But what if we want 10^4 – 10^5 points in the Poincaré map? How many points do we need in the phase trajectory? Maybe 10^7 or more? Obviously, the simplest approach described above would be waste of resources. A better approach would look at trajectory points one by one, test for intersections with hyperplane, and forget points that are no longer needed. But there is no straightforward way to have simulation process like this in MATLAB.

Of course, there is software (even free software) that can compute Poincaré maps. For example, the XPP (X-Window PhasePlane) tool (Ermentrout 2002) can do that. But what we have learned from our examples is that we need certain set of features that we could not find in any existing software. These features are as follows:

- ability to explicitly specify simulation algorithm;
- reasonable computational performance;
- ease of use by everyone, at least for certain use cases;
- extensibility by everyone who needs a new feature.

The first of these features is very important, but it is missing in all existing tools we tried (see Section 7). It seems that developers of these tools and authors of this paper have different understanding of what a computer simulation can be. Common understanding is that the goal of any simulation is to reproduce the behavior of system being investigated. Numerical simulations

therefore most often perform time integration of equations given by a mathematical model of the system. In this paper, we give the term *simulation* a broader meaning: it is data processing. Given that meaning, we do not think the term is misused, because time integration of model equations often remains the central part of the entire process. Importantly, researcher might need to organize the execution of that part differently, e.g., run initial value problem many times for different initial states or parameters, do intermediate processing on consecutive system states produced by time integrator, and so on.

Given the above general concept of numerical simulation, our goal is to provide a framework that supports the creation of data processing algorithms in a simple and straightforward manner, avoiding any coding except to specify model equations.

Next sections describe design decisions and technologies chosen for the NumEquaRes system (Section 2); simulation specification (Section 3) and workflow semantics (Section 4); performance, extensibility, and ease of use (Section 5); examples of simulations (Section 6); comparison with existing tools (Section 7); next steps to achieve interoperability with modeling and simulation tools (Section 8). Section 9 summarizes current results and presents a roadmap for future work.

2. DESIGN DECISIONS AND CHOICE OF TECHNOLOGIES

Keeping in mind the primary goals formulated above, we started our work.

Traditionally, simulation software have been designed as desktop applications or high performance computing (HPC) applications with desktop front-ends. Nowadays, there are strong reasons to consider Web applications instead of desktop ones, because on the one hand, main limitations for doing so in the past are now vanishing, and, on the other hand, there are many well-known advantages of Web apps.

For example, our “ease of use” goal benefits if we have a Web app, because this means “no need for user to install any additional software”.

Thus we have decided that our software has to be a Web application, available directly in user’s Web browser.

Now, the “extensibility by everyone” goal means that our project must be free software, so the GNU Affero GPL v3 license has been chosen. That should enforce the usefulness of software for anyone who could potentially extend it.

The “Reasonable performance” goal has determined the choice of programming language for software core components.

Preliminary measurements have shown that for a typical simulation, native code compiled from C++ runs approx. 100 times faster than similar code in MATLAB, SciLab, or JavaScript (as of JavaScript, we tested QtScript from Qt4; with other implementations, results might be different). Therefore, we decided that the simulation core has to be written in C++. The core is a

console application that runs on the server and interacts with the outer world through its command line parameters and standard input and output streams. It can also generate files (e.g., text or images).

JavaScript has been chosen as the language for simulation description and controlling the core application. However, this does not mean that any part of running simulation is executing JavaScript code.

The decision to use the Qt library has been made, because it provides a rich set of platform-independent abstractions for working with operating system resources, and also because it supports JavaScript (QtScript) out of the box.

Other parts of the applications are the Web server, the database engine, and components running on the client side. For the server, we preferred Node.js over other technologies because we believe its design is really suitable for Web applications — first of all, due to the asynchronous request processing. For example, it is easy to use HTML5 Server Sent Events (W3C 2014) with Node.js, which is not the case with LAMP/WAMP (Wikipedia 2015).

The MongoDB database engine has been picked among others, because, on the one hand, its concept of storing JSON-like documents in collections is suitable for us, and, on the other hand, we do not really need SQL, and, finally, it is a popular choice for Node.js applications.

As of the client code running in the browser, the components used so far are jQuery and jQueryUI (which is no surprise), the d3 library (Bostock 2015) for interactive visualization of simulation schemes, the marked (Jeffrey 2015) and MathJax (Krautzberger 2014) libraries to format markdown pages with $\text{T}_\text{E}_\text{X}$ formulas.

In the future, we are planning to add 3D visualization using WebGL.

3. SIMULATION SPECIFICATION

The very primary requirement for NumEquaRes is to provide user with the ability to explicitly specify how data flows are organized in a simulation. This determines how simulations are described. This is done similarly to, e.g., the description of a scheme in the Visualization Toolkit (VTK) (Kitware 2010), employing the “pipes and filters” design pattern. The basic idea is that simulation is a *data processing system* defined by a scheme consisting of *boxes* (filters) with *input ports* and *output ports* that can be connected by *links* (pipes). Output ports may have many connections; input ports are allowed to have at most one connection. Simulation data travels from output ports to input ports along the links, and from input ports to output ports inside boxes. Inside each box, the data undergoes certain transformation determined by the box type.

Typically boxes have input and output ports, so they are *data transformers*. Boxes without input ports are *data sources*, and boxes without output ports are *data storage*.

Simulation data is considered to be a sequence of *frames*. Each frame can consist of a scalar real value or

one-dimensional or multi-dimensional array of scalar real values. The list of sizes of that array in all its dimensions is called *frame format*. For example, format {1} describes frames of scalar values, and format {500,400} describes frames of two-dimensional arrays, each having size 500×400. The format of each port is assumed to be fixed during simulation.

Links between box ports are logical data channels, they cannot modify data frames in any way. This means that data format has to be the same at ports connected by a link. Some ports define data format, while some do not; instead, such a port takes format of port connected with it by a link. Thus, data format *propagates along links*. Furthermore, data format can also *propagate through boxes*. This allows to provide quite flexible design to fit the demands of various simulations.

4. SIMULATION WORKFLOW

This section explains how simulation runs, i.e., how the core application processes data frames generated by boxes.

Further, the main routine that controls the data processing is called *runner*.

4.1. Activation notifications

When a box generates a data frame and sends it to an output port, it actually does two things:

- makes the new data frame available in its output port;
- *activates* all links connected to the output port. This step can also be called *output port activation*.

Each link connects an output port to an input port, and its activation means sending notification to input port owner box. The notification just says that a new data frame is available at that input port.

When a box receives such a notification, it is free to do whatever it wants to. In some cases, these notifications are ignored; in other cases, they cause box to start processing data and generate output data frames, which leads to link activation again, and the data processing goes one level deeper. For example, the `Pendulum` box has two input ports, `parameters` and `state`. When a data frame comes to `parameters`, the activation notification is ignored (but next time the box will be able to read parameters from that port). When a data frame comes to `state`, the activation is not ignored. Instead, the box computes ODE right hand side and sends it to the output port `orderhs`.

4.2. Cancellation of data processing

Link activation notification is actually a function call, and the box being notified returns a value indicating success or failure. If link activation fails, the data processing is *cancelled*. This can happen when some box cannot obtain all data it needs from input ports. For example, the `Pendulum` box can process the activation of link connected to port `state` only if there are some parameters available in port `parameters`. If it is so,

the activation succeeds. Otherwise, the activation fails, and the processing is canceled.

If a box sends a data frame to its output port, and the activation of that output port fails, the box always cancels the data processing. Notice that this is always done by returning a value indicating activation failure, because the box can only do something within an activation notification.

4.3. Data source box activation

Each simulation must have at least one *data source* box — a box having output ports but no input ports. There can be more than one data source in a simulation.

Data sources can be *passive sources* or *generators*. A generator is a box that can be notified just as a link can be. A passive data source cannot be notified.

A passive data source produces one data frame (per output port) during the entire simulation. The data frame is available on its output port from the very beginning of the simulation.

4.4. Initialization of the queue of notifications

When the runner starts data processing, it first considers all data sources and builds the initial state of the *queue of notifications*. For each generator, its notification is enqueued. For each passive data source, the notification of each of its links is enqueued.

4.5. Processing of the queue of notifications

Then the queue is processed by sending the activation notifications (i.e., calling notification functions) one by one, from the beginning to the end. If a notification call succeeds, the notification is removed from the queue. Otherwise, if the notification call fails (i.e., the data processing gets canceled), the notification is moved to the end of the queue, and the process continues.

The runner processes its queue of notifications until it becomes empty, or maximum number of activation notification failures (currently 100) is exceeded. In the latter case, the entire simulation fails.

4.6. Post-processing

When the queue of notifications becomes empty, the runner can enqueue *post-processors* before it stops the data processing. The only example of a post-processor is the `Pause` box. Post-processors, like generators, are boxes that can receive activation notifications.

4.7. User input events

The above process normally takes place during the simulation. In addition, there could be events that break the processing of the queue of notifications. These events are caused by *interactive user input*. Once a user input event occurs, an exception is thrown, which leads to the unwinding of any nested link activation calls and the change of the queue of notifications. Besides, each box gets notified about simulation restart.

The queue of notifications is changed as follows when user input occurs. First, the queue is cleared. Then one of two things happens.

- If the box that threw the exception specifies which box should be activated after restart, the notifications for that box are enqueued (if the box is a generator, its activation notification is enqueued; otherwise, the activation notifications of all links connected to its output ports are enqueued). An input box can only specify itself as the next box to activate, or specify nothing.
- If the box that threw the exception specifies no box to be activated after restart, the standard initialization of the notification queue is done.

After that, the processing of notification queue continues.

There is an important issue that must be taken care of. Simulation can potentially be defined in such a way that its execution leads to an infinite loop of recursive invocation of activation notifications. This normally causes program to crash due to stack overflow. In our system, however, some boxes (not all, but only those activating outputs in response to more than one input notification) are required to implement counters for recursive call depth. When such a counter reaches 2, simulation is considered to be invalid and is terminated. This allows to do some kind of runtime validation against recursion at the cost of managing call depth counters.

It should be noticed that theoretically, simulations that we are dealing with here are a subclass of discrete event systems (Zeigler, Kim, and Praehofer 2000) with discrete time; one time step corresponds to the returning from an activation call.

5. PERFORMANCE, EXTENSIBILITY, AND EASE OF USE

As stated in Section 1, computational performance and functional extensibility are considered important design features of the NumEquaRes system. This section provides technical details on what has been done to achieve performance and support extensibility. Last subsection highlights design features that make system easier to use.

5.1. Performance

To achieve reasonable performance, it is not enough to just use C++. Some additional design decisions should be made. Most important of them are already described above. The ability to organize simulation workflow arbitrarily allows to achieve efficient memory usage, which is illustrated by an example in Section 1. A number of specific decisions made in the design of NumEquaRes core are targeted to high throughput. They are driven by the following rules.

- Perform simulation in a single thread. While this is a serious performance limitation for a single simulation, we have made this decision because the simulation runs on the Web server, and parallelization inside a single simulation is likely to impact the performance of server, as it

might run multiple simulations simultaneously. And, on the other hand, single thread means no synchronization overhead.

- No frequent operations involving interaction with operating system. Each box is responsible for that. For example, data storage boxes should not write output data to files or check for user input frequently. The performance might drop even if the time is measured using `QTime::elapsed()` too frequently.
- No memory management for data frames within activation calls. In fact, almost 100% of simulation time is spent in just one activation call made by runner (during that call, in turn, other activation calls are made). Therefore, memory management outside activation calls (e.g., the allocation of an element of the queue of notifications) is not a problem. Still some memory allocation happens when a box writes its output data, but this is not a problem as well, since such operations are not frequent.
- No movement of data frames in memory. If a box produces an output frame and makes it available in its output port, all connected boxes read the data directly from memory it was originally written to. This item and the previous one both imply that there are nothing like queues of data frames, and each frame is processed immediately after it is produced.
- No virtual function calls within activation calls. Instead, calls by function pointer are preferred.

A simple architecture of classes has been developed to comply with the rules listed above and, in the same time, to encapsulate the concepts of box, port, link, and others. These classes are split into ones for use at the initialization stage, when simulation is loaded, and others for use at simulation run time. First set of classes may rely on Qt object management system to support their lifetime and the exposure of parameters as JavaScript object properties. Classes of the second set are more lightweight; their implementations are inlined whenever possible and appropriate, in order to reduce function call overhead.

Although NumEquaRes core performance has been optimized in many aspects, it seems impossible to combine speed and flexibility. Our experience with some examples indicates that hand-coded algorithms run several times faster than those prepared in our system.

5.2. Extensibility

The functionality of NumEquaRes mostly resides in boxes. To add a new feature, one thus can write code for a new box. Boxes are completely independent. Therefore, adding a new one to the core simply boils down to adding one header file and one source file and recompiling. The core will be aware of the presence of the new box through its box factory mechanism. Next

steps are to support the new box on server by adding some meta-information related to it (including user documentation page) and some client code reproducing the semantics of port format propagation through the box. The checklist can be found in the online documentation.

Some extensions, however, cannot be done by adding boxes. For example, to add 3D visualization, one needs to change the client-side JavaScript code. We are planning to simplify extensions of this kind; however, this requires refactoring of current client code.

5.3. Ease of use

First of all, NumEquaRes is an online system, so user does not have to download and install any software, provided user already has a Web browser. All user interaction with the system is done through the browser. To formulate a simulation as a data processing algorithm, user composes a scheme consisting of boxes and links, and there is no need to code.

Online help system contains a detailed documentation page for each box; it also explains simulation workflow, user interface, and other things; there is one step-by-step tutorial.

To prepare a simulation, user can find a similar one in the database, then clone it and modify. User can decide to make his/her simulation public or private; public simulations can be viewed, run, and cloned by everyone. To share a simulation with a colleague, one shares a hyperlink to it; besides, simulations can be downloaded and uploaded.

Currently, user might have to specify part of simulation, such as ODE right hand side evaluation, in the form of C++ code. We understand this might be difficult for people not familiar with C++. To mitigate this problem, there are two features. Firstly, each box that needs C++ code input provides a simple working example that can be copied and modified. Secondly, NumEquaRes supports the concept of *code snippets*. Each piece of C++ input can be given a documentation page and added to the list of code snippets. These snippets can easily be reused by everyone. See also Section 8.

Another feature that plays a role similar to debugger's is the visualization of simulation data flows. The feature is currently under development and not available through the Web interface.

6. EXAMPLES OF SIMULATIONS

This section lists several examples of simulations.

Figure 1 shows one of the simplest simulations — the one that plots a single phase trajectory for a simple pendulum. The ODE system is provided by the `ode` box. NumEquaRes has a number of options for user to supply equations. In particular, it is possible to provide C++ code that computes ODE right hand side. Such code compiles and runs on the server, if it passes a security check. The ODE right hand side depends on the state variables and the vector of parameters. They are supplied through input ports. Parameters are specified in the `odeParam` box. State variables come from the

`solver` box. The solver performs numerical integration of the initial value problem, starting from the user-specified initial state (the `initState` box). The solver can be configured to perform a fixed number of time steps or to run until interrupted by a data frame at its `stop` port. Each time the solver obtains a new system state vector, it sends the vector to its `nextState` port. Once the solver finishes, it activates the `finish` port to let others know about it. In this simulation, consecutive system states are projected to the phase plane (the `proj` box) and then rasterized by the `canvas` box. Finally, the data comes to the `bitmap` box that generates the output image file. Notice that this simulation has three data sources, `odeParam`, `solverParam`, and `initState`.

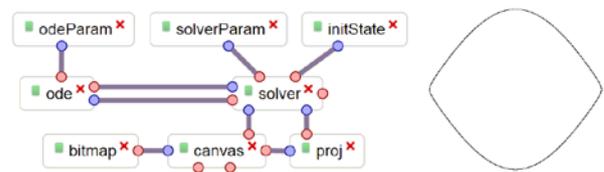


Figure 1: Single phase trajectory

From this simplest example one can see how to construct simulation scheme from boxes and links that computes what user needs. Other examples are more complex, but they basically contain boxes of the same types, plus probably some more. So far, there are 40 different box types in NumEquaRes, and it is beyond the scope of this article to describe them all. Further, we will just focus on some of them to show how simulations work.

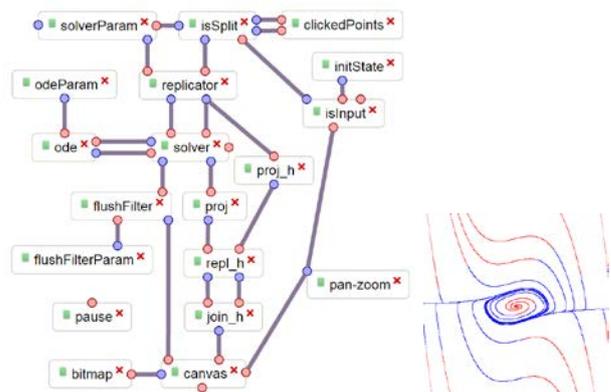


Figure 2: Interactive phase portrait

An important aspect of a simulation is its ability to *interact with the user*. There are a few boxes that transform various kinds of interactive user input (clicking, moving sliders, rotating mouse wheel, etc.) into numerical values. These boxes usually act as simple filters of data frames; they replace some components of data frames with values obtained from user. Figure 2 shows an example of interactive simulation: it generates phase trajectories going through points on plane — the ones user has clicked with the mouse. The box `isInput` is responsible for that kind

of input. Each generated phase curve has two parts: blue in the time-positive direction (with resp. to the clicked point) and red in the time-negative direction.

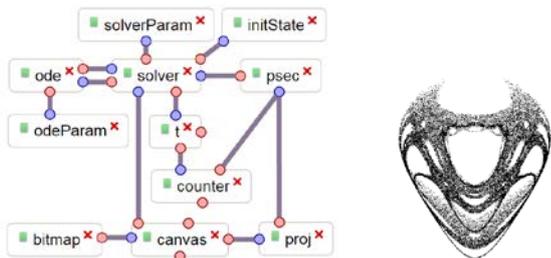


Figure 3: Double pendulum, Poincaré map (50000 points, 28.5 s)

Figure 3 shows the Poincaré map for the classical double pendulum system. Importantly, there is no need to store phase trajectory or individual points of intersection of the trajectory with the plane during simulation. The entire processing cycle (test for intersection; projection; rasterization) is done as soon as a new point of the trajectory is obtained. After that, we need to store just one last point from the trajectory. Simulations like this are what we could not do easily in MATLAB or SciLab, and they have inspired us to develop NumEquaRes.

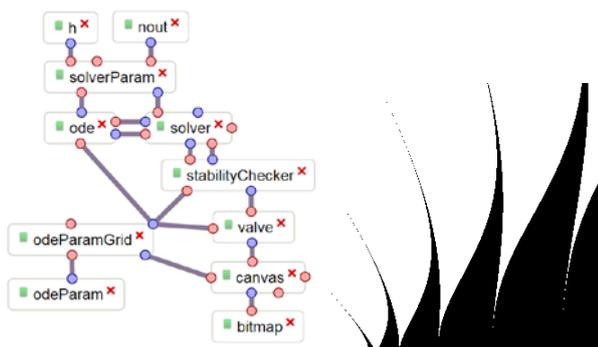


Figure 4: Ince-Strutt stability diagram (500x500 points, 6.3 s)

Figure 4 shows a simple simulation that allows one to obtain a stability diagram of a linear ODE system with periodic coefficients on the plane of parameters. Here the picture on the right is the Ince–Strutt diagram for the Mathieu equation (Abramowitz and Stegun 1972). People who have experience with it know how difficult it is to build such kind of diagrams analytically, even to find the boundaries of stability region near the horizontal axis. What we suggest here is the brute force approach — it is fast enough, general enough, and it is done easily. The idea is to split the rectangle of parameters into pixels and analyze the stability in the bottom-left corner of each pixel (by computing eigenvalues of the monodromy matrix (Teschl 2012)), then assign pixel color to black or white depending on the result. In this simulation, important new boxes are `odeParamGrid` and `stabilityChecker`. The former one provides a way to generate points on a multi-dimensional grid, and the latter one analyzes the

stability of a linear ODE system with periodic coefficients.

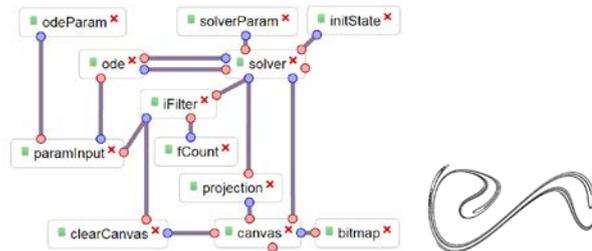


Figure 5: Strange attractor for forced Duffing equation (interactive simulation)

Figure 5 shows another application of the Poincaré map, now in the visualization of the strange attractor arising in the forced Duffing equation (Bender and Orszag 1999). User can change parameters interactively and see how the picture changes. This simulation is simpler than the one shown in Figure 3, because to obtain a new point on canvas, one just needs to apply time integration over known time period of system excitation.

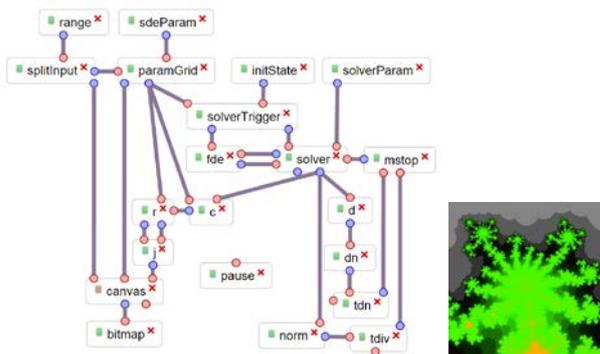


Figure 6: Colored Mandelbrot set (interactive simulation)

Figure 6 shows an interactive simulation of the Mandelbrot set (Mandelbrot 2004). User can pan and zoom the picture using the mouse. Importantly, we did not have to develop any new box types in order to describe the logic of convergence analysis for sequences of complex numbers generated by the system. This is done with general purpose boxes `d` (computes differences of subsequent data frames), `dn` (computes vector norm), and `tdn` (detects if a scalar value exceeds some threshold). Pixel colors depend on how many iterations passed (box `c` is a counter, its output value is joined with pixel coordinates at box `j` and sent to `canvas`).

7. COMPARISON WITH OTHER TOOLS

Direct comparison between NumEquaRes and other existing tools is problematic because all of them (at least, those that we have found) do not provide an easy way for user to describe the data processing algorithm. In some systems, the algorithm can be available as a predefined analysis type; in others, user would have to

code the algorithm; also, there are systems that need to be complemented with external analysis algorithms.

Let us consider example simulations shown in Figures 3, 4, 5, and try to solve those using different free tools; for commercial software, try to find out how to do it from the documentation. Further in this section, figure number refers to the example problem.

Table 1: Comparison of NumEquaRes with other tools

Name	Free	Web	Can solve	Fast
Mathematica	no	yes	3, 4, 5; needs coding	n/a
Maple	no	no	3, 4, 5; needs coding	n/a
MATLAB	no	no	3, 4, 5;	no
SciLab	yes	no	needs even more coding	no
OpenModelica	yes	no	none	could be
XPP	yes	no	3, 5	yes
InsightMaker	yes	yes	none	n/a

In Table 1, commercial proprietary software is limited to most popular tools — Mathematica, Maple, and MATLAB. In many cases, purchasing a tool might be not what a user (e.g., a student) is likely to do.

All of the three example simulations are solvable with commercial tools Mathematica, Maple, and MATLAB.

In Mathematica, it is possible to solve problems like 3,5 using standard time-stepping algorithms since version 9 (released 24 years later than version 1) due to the `WhenEvent` functionality. Problem 4 can also be solved. All algorithms have to be coded. Notice that Wolfram Alpha (Wolfram, 2015) (freely available Web interface to Mathematica) cannot be used for these problems.

Maple has the `DEtools[Poincare]` subpackage that makes it possible to solve problem 3 and others with Hamiltonian equations; problems 4, 5 can be solved by coding their algorithms.

With MATLAB or SciLab, one can code algorithms for problems 4, 5 using standard time-stepping algorithms. For problem 3, one needs either to implement time-stepping algorithm separately or to obtain Poincaré map points by finding intersections of long parts of phase trajectory with the hyperplane. Both approaches are more difficult than those in Mathematica and Maple. And, even if implemented, simulations are much slower than with NumEquaRes.

OpenModelica (Fritzson 2004) is a tool that helps user formulate the equations for a system to be simulated; however, it is currently limited to only one type of analysis — the solution of initial value problem. Therefore, to solve problems like 3, 4, 5, one has to code their algorithms (e.g., in C or C++, because the code for evaluating equations can be exported as C code).

XPP (Ermentrout 2002) provides all functionality necessary to solve problems 3, 5. It contains many

algorithms for solving equations (while NumEquaRes does not) and is a powerful research tool. Yet it does not allow user to define a simulation algorithm, and we have no idea how to use it for solving problem 4.

Among other simulation tools we would like to mention InsightMaker (Fortmann-Roe 2014). It is a free Web application for simulations. It has many common points with NumEquaRes, although its set of algorithms is fixed and limited. Therefore, problems 3, 4, 5 cannot be solved with InsightMaker.

Concluding this section, we have to state that other tools either provide a fixed set of data processing algorithms or require them to be coded by users.

8. STEPS TO ACHIEVE INTEROPERABILITY WITH MODELING AND SIMULATION TOOLS

8.1. Importing model equations

NumEquaRes focuses on the development of simulation algorithm and the execution of it. Simulations are dedicated to investigation of certain systems that are described by some equations. Currently, these equations have to be coded in C++ by hand as parameters of some boxes, such as `CxxOde`. We realize that it's easy only for very simple systems. We also realize that there are tools, such as OpenModelica (Fritzson 2004), capable of generating C-code plus some meta-data that describe models specified by user in a more simple and natural way.

To utilize the power of modeling tools in NumEquaRes, it is not necessary to create a monolith system integrating all pieces together. Instead, one could export model equations from a modeling tool and use them in a NumEquaRes simulation. Before year 2010, it was questionable which data format to use for the export of model equations, and how to enable the export feature in a modeling tool. In 2010, the Functional Mock-up Interface (FMI) (Modelica 2015) has been proposed, and since then has been adopted as the model exchange format by many modeling and simulation tools. The format is suitable for describing a model of dynamical system or its part; the system can be described by differential, algebraic, and discrete-time equations with possible state transitions due to events. Therefore, FMI is perfectly suitable for use to import model equations in NumEquaRes.

Using the FMI, a modeling tool exports the equations of a model in the form of Functional Mock-up Unit (FMU). It is quite easy to develop a box in NumEquaRes that wraps an arbitrary FMU and exhibits all necessary parameters and variables through its input and output ports.

Next step is to have a time integrator box in NumEquaRes that could be used for solving the initial value problem for an FMU (currently, we only have explicit Runge — Kutta scheme of fourth order with fixed step-size, which is insufficient). To accomplish this step, we are planning to use the CVODE (Woodward, Reynolds, Hindmarsh, and Banks 2015)

solver, providing its functionality through a box. Other existing freely available time integrators can be used as well in a similar manner.

8.2. Using NumEquaRes core library

The other way to interoperate with NumEquaRes is to use its core library in a simulation tool. The developer of the tool can supply model equations to NumEquaRes through the CxxOde box or develop a new one that better integrates with the tool. It might also be necessary to wrap some additional functionality in new boxes. Any tool can use NumEquaRes, provided that the usage conforms to the GNU Affero General Public License, version 3.

9. CONCLUSION AND FUTURE WORK

A new tool for numerical simulations, NumEquaRes, has been developed and implemented as a Web application. The core of the system is implemented in C++ in order to deliver good computational performance. It is free software and thus everyone can contribute into its development. The tool already provides functionality suitable for solving many numerical problems, including the visualization of Poincaré maps, stability diagrams, fractals, and more. NumEquaRes lacks any modeling capabilities, since model equations have to be coded by hand. However, there are ways to interoperate with modeling tools using FMI. Besides, NumEquaRes core library can be used in a simulation tool.

The algorithm of simulation runner implies that the order of activation calls it makes is not important, i.e., does not affect simulation results. While this is true for typical simulations, counter-examples can be invented. Further work is to make it possible to distinguish such simulations from regular ones and render them invalid. NumEquaRes is a new project, and the current state of its source code corresponds more to the proof-of-concept stage than the production-ready stage, because human resources assigned to the project are very limited. To improve the source code, it is necessary to add developer documentation, add unit tests, and deeply refactor both client and server parts of the Web interface.

Further plans of NumEquaRes development include new features that would significantly extend its field of application. One of them is a box wrapping arbitrary FMUs containing model equations; the other one is a box wrapping the functionality of the CVODE solver.

Another set of planned features aims to enhance the level of presentation of simulation results (currently, it is quite modest). Among them is 3D visualization and animation.

REFERENCES

Abramowitz M., Stegun I., 1972. Handbook of mathematical functions, pp. 721–750. Washington DC: Dover Publications Inc.

Bender C.M., Orszag S.A., 1999. Advanced mathematical methods for scientists and engineers

I: Asymptotic methods and perturbation theory, pp. 545–551. New York: Springer-Verlag.

Bostock M., 2015. Data-driven documents. Available from: <http://d3js.org> [Accessed June 2015].

Ermentrout B., 2002. Simulating, analyzing, and animating dynamical systems: A guide to XPPAUT for researchers and students. SIAM.

Fortmann-Roe S., 2014. Insight Maker: A general-purpose tool for web-based modeling & simulation. Simulation modelling practice and theory, 47:28–45.

Fritzson P., 2004. Principles of object-oriented modeling and simulation with Modelica 2.1. Wiley-IEEE Press, New Jersey.

Jeffrey C., 2015. A markdown parser and compiler. Built for speed. Available from: <https://github.com/chjj/marked> [Accessed June 2015].

Kitware, Inc., 2010. VTK user's guide. Kitware.

Krautzberger P., 2014. MathJax — beautiful mathematics on the web. Available from: <http://pkra.github.io/slides-mathjax> [Accessed June 2015].

Mandelbrot B.B., 2004. Fractals and chaos: the Mandelbrot set and beyond. New York, Berlin, Paris: Springer-Verlag.

Meirovitch, L., 1986. Elements of vibration analysis. New York: McGraw-Hill.

Modelica Association, 2015. Functional mock-up interface. Available at <https://www.fmi-standard.org> [Accessed June 2015].

Routh E.J., 1905. The advanced part of a treatise on the dynamics of a system of rigid bodies. London: Macmillan.

Teschl G., 2012. Ordinary differential equations and dynamical systems. Available from: http://www4.ncsu.edu/~schechter/ma_732_sp13/teschl_ode.pdf [Accessed June 2015].

W3C, 2014. Server-sent events (second edition). Available from: <https://w3c.github.io/eventsource> [Accessed June 2015].

Wikipedia, 2015. LAMP (software bundle). Available from: [https://en.wikipedia.org/wiki/LAMP_\(software_bundle\)](https://en.wikipedia.org/wiki/LAMP_(software_bundle)) [Accessed June 2015].

Wolfram Research, 2015. WolframAlpha — computational knowledge engine. Available at <http://www.wolframalpha.com> [Accessed June 2015].

Woodward C.S., Reynolds D.R., Hindmarsh A.C., Banks L.E., 2015. SUNDIALS — suite of nonlinear and differential/algebraic equation solvers. Available at <https://computation.llnl.gov/casc/sundials/main.html> [Accessed June 2015].

Zeigler B.P., Kim, T.G., Praehofer H., 2000. Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems. Amsterdam, San Diego (Calif.), London: Academic Press.

EXTENDING THE RETSIM SIMULATOR FOR ESTIMATING THE COST OF FRAUD IN THE RETAIL STORE DOMAIN

Edgar Alonso Lopez-Rojas

Blekinge Institute of Technology
Department of Computer Science and Engineering

edgar.lopez@bth.se

ABSTRACT

RetSim is a multi-agent based simulator (MABS) calibrated with real transaction data from one of the largest shoe retailers in Scandinavia. RetSim allows us to generate synthetic transactional data that can be publicly shared and studied without leaking business sensitive information, and still preserve the important characteristics of the data.

In this paper we extended the fraud model of RetSim to cover more cases of internal fraud perpetrated by the staff and allow inventory control to flag even more suspicious activity. We also generated sufficient number of runs using a range of fraud parameters to cover a vast number of fraud scenarios that can be studied.

We then use RetSim to simulate some of the more common retail fraud scenarios to ascertain exactly the cost of fraud using different fraud parameters for each case.

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

1. INTRODUCTION

Fraud is an important problem in a number of different situations, and more specifically in retail stores is a very common problem in all countries. The economic impact for the loses can be substantial, this is why many major store retailers invest in security. However, how much to invest could be a political decision since many times it is hard to calculate the cost of possible fraudulent behaviour due to the multiple possible fraud causes. Once the fraud problem and cost of it is identified, it is easier for managers to take the prevention measures to lower the loses. For example, in one recent case the major US home improvement chain *Home Depot* was the target of a fraudulent return scam where two perpetrators netted several thousand dollars before being caught (FBI, 2013). Return fraud, i.e. the defrauding of a retail merchant by abusing the return process, alone is estimated to cost US retailers about 9 billion dollars yearly. To further illustrate the seriousness of the problem and try and combat it both EU and US recently started to mandate the use of fraud detection as one part of the minimum security requirements for financial services (ECB, 2013; Council, 2011).

Our approach to this problem makes use of the RetSim simulator as an strategy to measure and estimate the cost of loses and as a experimental laboratory for managers to apply measures to detect and discourage internal fraud. RetSim is a multi agent-based simulator (MABS) of a shoe store based on the transactional data of one of the largest retail shoe sellers in Sweden (Lopez-Rojas et al., 2013). RetSim uses this real data to develop and calibrate the model through statistical and a social Network Analysis (SNA) of the relations between staff and customers and generates a synthetic data set similar to the original one but without any unwanted disclosure of private information about either the staff or the customers.

Therefore, the aim of RetSim is the generation of synthetic data that can be used for fraud detection research and prognosis of fraud scenarios. With the RetSim simulator researchers and managers can test and measure in different fraud scenarios the cost and performance of fraud detection methods as simple as thresholds or even more elaborated such as triage models bases on thresholds, machine learning algorithms and others. The initial purpose of RetSim was to model a realistic data set that resembles the distribution of financial transactions found in an original given data set. Now that we have that we added fraudulent behaviour to study the cost of fraud in retail store.

In this paper we extended the fraud model of RetSim to cover more cases of internal fraud perpetrated by the staff and allow inventory control to flag even more suspicious activity. We also generated sufficient number of runs using a range of fraud parameters to cover a vast number of fraud scenarios that can be studied.

Our ultimate goal is for RetSim to be usable to model relevant scenarios to generate realistic data sets that can be used by academia, managers, security inspectors, students and others, to develop and reason about fraud detection methods without leaking any sensitive information about the underlying data. Synthetic data has the added benefit of being easier to acquire, faster and at less cost, for experimentation even for those that *have* access to their own data. We argue that RetSim generates data that usefully approximates the relevant aspects of the real data.

Outline: The rest of this paper is organised as follows: Section 2. introduce the topic of fraud detection for retail stores and present previous and related work on simulators. Section 3. describes the problem, which is the generation of synthetic data of a retail store system for estimating the cost of fraud. Sections 4. and 5. present our implementation and results of a MABS for our domain and shows the description of the extended fraud scenarios implemented on RetSim. We finish with a discussion and conclusions, including future work in section 6..

2. BACKGROUND AND RELATED WORK

Simulations in the domain of retail stores have traditionally been focused on finding answers to logistics problems such as inventory management, supply management, staff scheduling and customer queue reductions (Chaczko and Chiu, 2008; Schwaiger and Stahmer, 2003; Bovinet, 1993). We find no research focusing on simulations generating fraud data to be used for fraud detection in retail stores. Therefore, we recently introduced RetSim with the purpose of fraud detection research. In this article we built upon previous version of RetSim to study the cost of specific fraud scenarios, including agents using known fraud behaviour patterns (Lopez-Rojas et al., 2013).

Anonymization techniques have been used to preserve the privacy of sensitive information present in data sets. But de-anonymizing data sets is not an insurmountable task, far from it (Narayanan and Shmatikov, 2009). For this reason we have decided to use simulation techniques to keep specific properties of the original data set, such as statistical and social network properties, and at the same time providing an extra layer of insulation that pure anonymization does not provide.

There are tools such as IDSG (IDAS Data and Scenario Generator (Lin et al., 2006)) that were developed for the purpose of generating synthetic data based on the relationship between attributes and their statistical distributions. IDSG was created to support data mining systems during the testing phase, and it has been used to test fraud detection systems. Our approach differs in that we are implementing an agent-based model which is based on agent micro behaviour rather, than a fixed statistical distribution of macro parameters.

With the current popularity of social networks, such as *Facebook*, the topic of Social Network Analysis (SNA) has seen interest in the research community (Alam and Geller, 2012). Social Network Analysis is currently being combined with *Social Simulation*. Both topics support each other in the representation of interactions and behaviour of agents in the specific context of social networks. However, there is no work addressing the question of customer/salesman-interaction, that we are aware of.

Other methods to generate the necessary fraud data have been proposed by (Yannikos et al., 2010; Lundin, 2002; Kargupta et al., 2003; Evfimievski et al., 2002; Agrawal and Aggarwal, 2001). The work by (Yannikos

et al., 2010) lets the user specify the assumptions about the environment at hand; i.e., there is no need for access to real data. However, this will certainly affect the quality of the synthetic data. The work by (Lundin, 2002) makes use of a small sample of real data to generate synthetic data. This approach is similar to ours. However, the direct use of real data to prime the generation of synthetic data is limited in that it makes it harder to generate realistic data with other characteristics that those of the original real data (Yannikos et al., 2010). The work by (Kargupta et al., 2003) focused on privacy-preserving methods for data mining. However, that method also does not have the possibility of generating realistic data with other characteristics than those of the original data. In our work, we use social simulation, which makes it possible to change the parameters of the agents in the model to create realistic synthetic data, potentially producing emergent behaviour in the logs which is hard to produce in other ways.

3. PROBLEM

The RetSim simulator was previously used to solve the problem of finding a synthetic data set that realistically represent a given real data set without disclosing any specific information about customers or staff members. Now with this paper we aim to solve the question of how much are the loses due to fraud giving a specific scenario of fraud in a retail store. This problem is of particular interest for managers of stores, mainly because the investment on security is limited, therefore the impact of each of these fraud models and expected fraud scenarios will give an idea to a manager of the cost of fraud. Historically on a real store the cost is only estimated but can not be exactly calculated due to the disguise nature of the fraud.

By using simulators, we can be certain of the profit of each fraudulent agent over the time. We generate synthetic data that contains flagged fraud behaviour, therefore it simplifies the process of estimating the cost of losses due to specific fraud. These estimations can be used to take informed decisions because they are a good approximation of a real scenario.

4. MODEL AND METHOD

RetSim uses the MABS toolkit MASON version 17 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 RetSim (Railsback et al., 2006).

Our model contains the following entities and behaviours. The *Store* is the main entity of the simulation, it contains all the variables and states required to run the simulation such as: *Salesmen*, *Customers*, *Products*, *Frequencies* and other parameters used to calibrate the model.

During the initialization of the simulation a store load all products, set up an inventory and creates the salesmen

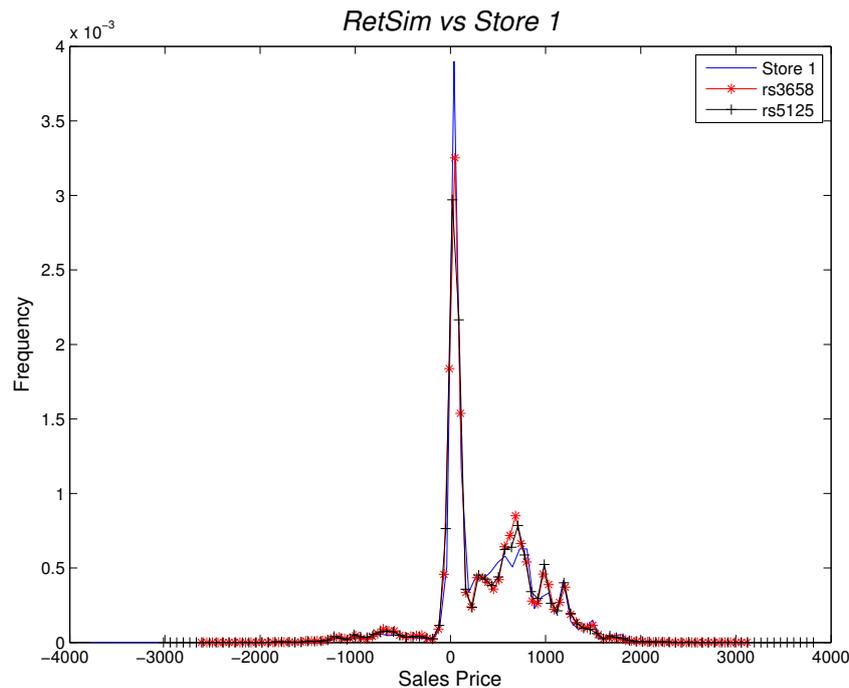


Figure 1: Overlap of Two Runs of RetSim vs Real Data

according to the parameters of the store. We calibrated RetSim to consider each day as a step in the simulation. During a single step of the simulation some of the salesmen became active and others inactive or not working that day. The salesmen are distributed in a virtual geographical space on the store. Once the day starts an average of customers are instantiated and distributed around the geographical space.

The *basic principle* of this model is the concept of a commercial transaction. The process of creating financial transactions starts when a salesman sense nearby customers in the *need-help* state and offers help. There are two different outcomes: either a transaction takes place or not, similar to the real world.

There are no known instances of fraud in the real data (as certified by the data owner). So we modelled malicious agents with fraud behaviour, by programming agents that behave according to some known or hypothesised retail fraud case.

The following retail fraud scenarios are based on selected cases from the Grant Thornton report (Member and Council, 2009). As can be seen below, we extended the RetSim model by adding different fraud scenarios. The initial purpose of RetSim was to model a realistic data set that resembles the distribution of financial transactions found in an original given data set. Now that we have that we added fraudulent behaviour to study the cost of fraud in retail store.

The *Refunds fraud scenario* is perhaps the most common cause of fraud and includes cases where a salesman creates fraudulent refund slips, keeping the cash refund for him- or herself.

In terms of the object model used in RetSim, the refund scenario was simulated by estimating the average number of refunds per sale and the corresponding standard deviation. We used these statistics to simulate refunds in the RetSim model. Fraudulent salesmen will perform normal refunds, as well as fraudulent ones. The volume of fraudulent refunds was modelled using a salesman specific parameter. The "red flag" for detection would in this case be a high number of refunds for a salesman.

The *Coupon reductions/discounts fraud scenario* is perhaps the second responsible for loses and includes cases where the salesman registers a discount on the sale without telling the customer; i.e., the customer pays the full sales price, and the salesman keeps the difference.

In terms of the object model used in RetSim, the coupon reduction/discounts scenario was implemented by estimating the average number of cancellations per sale and the corresponding standard deviation. Using these statistics we simulated discounts in the RetSim model. Fraudulent salesmen performed normal discounts, as well as fraudulent ones. The volume of fraudulent discounts was modelled using a salesman-specific parameter. The "red flag" for detection would in this case be a high number of discounts for a salesman with a relatively low number of average sales.

We extended the original RetSim by adding inventory control over products. Our initial assumptions were that the replenish of the inventory was performed automatically and therefore we did not focus on this aspect. But in reality inventory control is one of the more effective ways to detect fraud. Unfortunately performing inventory control is an expensive procedure in most of the stores due to

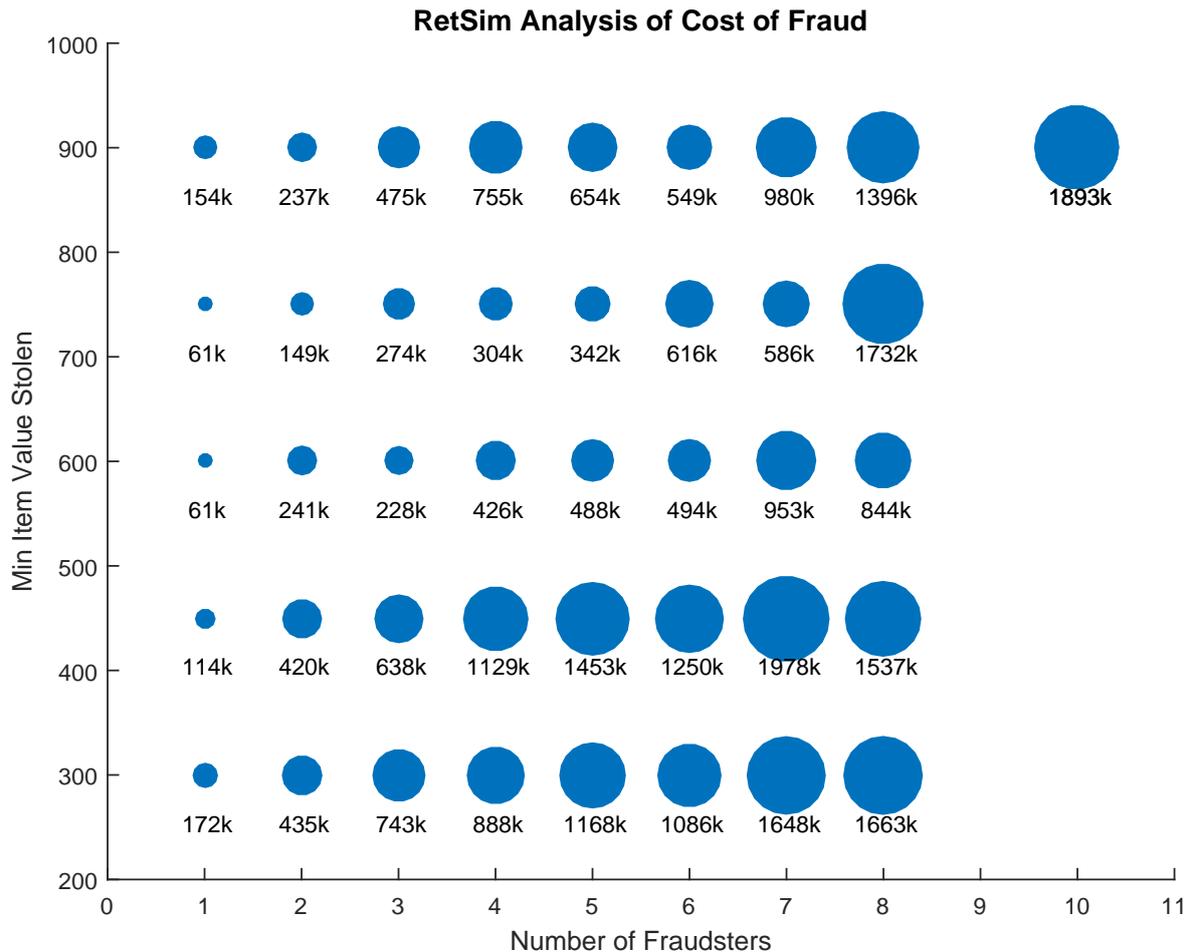


Figure 2: Analysis of Cost of Fraud

manual counting.

There are other possible scenarios, but as mentioned in the introduction return fraud (both by customers and sales staff alike) is a major problem, so we have chosen to focus on return fraud and the structurally similar discount fraud, as these are common and serious.

5. RESULTS

We built upon previous successful simulations that were calibrated from a Store taken from the original data set of our retailer. Figure 1 shows the results from a comparison of sales frequency for items of 2 runs of RetSim versus the original store (named Store 1). We can clearly see that our simulator successfully models the sales of Store 1. We can analyse that items that cost less are the most wanted products by the customers. For convenience and protection of disclosure of the business we decided to use a fictitious monetary unit that we will name just *units*.

From an analysis of figure 1, a manager can design experiments to study the phenomenon of fraud performed by the staff in the previous scenarios explained (Refunds fraud and discounts fraud).

Our experiment aims to calculate the total cost of losses if one or many employees are responsible for fraud. One common behaviour is of fraudsters is to mentally set some

limits about the min or max value of the item stolen. Many fraud controls nowadays are performed using simple thresholds around the prices. One way to avoid these controls is by carefully selecting items with value below. We will assume that the fraudster mentally set up a minimum value for stealing to make it worth the risk of being caught by the manager.

We setup our experiment to run 100 times under a specific setting. We iterated around minimum values starting from 300 *units* up to 900 *units*. We use a probability of 10% of the time that a salesmen should do the checks for performing fraud. In total we collected 500 runs of RetSim and summarize our cost analysis in figure 2.

Figure 2 shows one way to visualize the total cost of fraud performed by a different number of staff members. This figure allows a manager to study better the phenomenon of fraud in a store. One can notice for instance that the most of the items sold are around the 450 *units* price. So the profit of a fraudster that steals around this minimum value will be in many cases higher than other that decide to steal only high value articles. This information is particularly important when we require to implement controls over the refunds of items that cost more or less than 450 *units*. We can also see that below this value the amount of losses is considerable when then number of

staff members involved in fraud increases.

Besides the controls and thresholds that can be used for fraud, a major contribution of the RetSim to the managers is to measure the total cost of fraud. With this information a manager can decide how much to invest in any required fraud detection control. In many cases it is "acceptable" for managers to deal with low losses if the investment in fraud is higher than the fraud that can be prevented.

6. CONCLUSIONS

RetSim is a simulator of a retail store that generates transaction data set of diverse fraud scenarios, in this paper we extend the fraud model and measure the cost of total losses of each scenario. The cost of fraud in different scenarios can be estimated by summing the profit from each malicious agent (usually known as fraudsters). These estimations can be used to take informed decisions about how much should be invested in fraud controls, because they are a good approximation of real fraud scenarios.

Synthetic data sets generated with RetSim can aid academia, managers of companies and governmental agencies in testing their methods, in exploring the cost of different fraud scenarios and the performance of different fraud detection methods while maintaining similar conditions by using the same test data set, or in generally reasoning about the limits of effectiveness of fraud detection.

Future work on RetSim may include the addition of more relevant fraud models as well as tools for setting up experiments with higher degree of parametrization that includes more variables such as variations (increase or decrease) of sales with respect to the original store.

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. The author thanks Dr. Stefan Axelsson, main supervisor of doctoral studies, for encouraging this solo research as a preparation for upcoming doctor degree next year.

REFERENCES

- Dakshi Agrawal and CC Aggarwal. On the design and quantification of privacy preserving data mining algorithms. *PODS '01 Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2001.
- SJ Alam and Armando Geller. Networks in agent-based social simulation. *Agent-based models of geographical systems*, pages 77--79, 2012.
- JW Bovinet. *RETSIM: A Retail Simulation with a Small Business Perspective*. West Pub. Co., Minneapolis/St. Paul, 1993. ISBN 0314016708.
- Z. Chaczko and C.C. Chiu. A smart-shop system - Multi-agent simulation system for monitoring retail activities. *20th European Modelling and Simulation Symposium*, pages 20--26, 2008.
- FFIE Council. Supplement to Authentication in an Internet Banking Environment. URL: <http://www.ffiec.gov/pdf/Auth-ITS-Final>, pages 206--222, 2011.
- (European Central Bank) ECB. Recommendations for the Security of Internet Payments. Technical Report January, 2013.
- Alexandre Evfimievski, Ramakrishnan Srikant, Rakesh Agrawal, and Johannes Gehrke. Privacy preserving mining of association rules. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '02*, page 217, 2002. doi: 10.1145/775079.775080.
- FBI. Ticket Switch Fraud Scheme at Home Depot, 2013. URL <http://www.fbi.gov/buffalo/press-releases/2013/mexican-man-charged-in-ticket-switch-fraud-scheme-at-home-depot>.
- H. Kargupta, S. Datta, and Q. Wang. On the privacy preserving properties of random data perturbation techniques. *Third IEEE International Conference on Data Mining*, pages 99--106, 2003. doi: 10.1109/ICDM.2003.1250908.
- P.J. Lin, B. Samadi, and Alan Cipelone. Development of a synthetic data set generator for building and testing information discovery systems. In *ITNG 2006.*, pages 707--712. IEEE, 2006. ISBN 0769524974.
- Edgar Alonso Lopez-Rojas, Stefan Axelsson, and Dan Gorton. RetSim: A Shoe Store Agent-Based Simulation for Fraud Detection. *The 25th European Modeling and Simulation Symposium*, 2013.
- S. Luke. MASON: A Multiagent Simulation Environment. *Simulation*, 81(7):517--527, July 2005. ISSN 0037-5497. doi: 10.1177/0037549705058073.
- Emilie Lundin. A synthetic fraud data generation methodology. In *Information and Communications Security*, pages 265--277. Springer, Berlin Heidelberg, 2002.
- Associate Member and Advisory Council. Reviving retail Strategies for growth in 2009 Executive summary, 2009. URL http://www.grantthornton.com/staticfiles/GTCom/files/Industries/Consumer&industrialproducts/Whitepapers/Revivingretail_Strategiesforgrowthin2009.pdf.
- Arvind Narayanan and Vitaly Shmatikov. De-anonymizing Social Networks. *2009 30th IEEE Symposium on Security and Privacy*, pages 173--187, May 2009. doi: 10.1109/SP.2009.22.
- S. F. Railsback, S. L. Lytinen, and S. K. Jackson. Agent-based Simulation Platforms: Review and Development Recommendations. *Simulation*, 82(9):609--623, September 2006. ISSN 0037-5497. doi: 10.1177/0037549706073695.
- Arndt Schwaiger and B. Stahmer. SimMarket: Multiagent-based customer simulation and decision support for category management. *Multiagent System Technologies*, pages 74--84, 2003.

York Yannikos, Frederik Franke, Christian Winter, and Markus Schneider. 3LSPG : Forensic Tool Evaluation by Three Layer Stochastic Process-Based Generation of Data. In *4th International Workshop in Computational Forensics*, pages 200--211, Tokyo, Japan, 2010.

AUTHORS BIOGRAPHY

MSc. Edgar A. Lopez-Rojas

Edgar Lopez is a PhD student in Computer Science and his research area is 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.

NAVAL OPERATIONS' ASSESSMENT THROUGH A MULTI-PURPOSE HLA SIMULATION FRAMEWORK

Davide Tozzi, Federica Valdenazzi, Aldo Zini

CETENA S.p.A.

davide.tozzi@cetena.it, federica.valdenazzi@cetena.it, aldo.zini@cetena.it

ABSTRACT

Ship design is a complex process that needs a wide range of technologies and methodologies to be accomplished, from computational calculations to physic based analysis. The main lack of this process is that ship's performances and capabilities are difficult to be evaluated during early design phase. This is the reason why a simulation tool for the test and validation of naval operations procedures in different environmental conditions could fill this gap. Usually these operations involve more entities than the single ship, so it becomes important to arrange a synthetic environment in which different simulators and simulations are able to cooperate for demonstrating complex operations. Besides this, the "system of systems" approach finds its applicability in the naval field where a ship can be represented as the cooperation of several subsystems. HLA distributed simulations represent a good solution for creating and simulating complex scenarios, allowing several simulators and partners to work together just exchanging inputs and outputs, but keeping safe intellectual properties rights. CETENA's approach to deal with the evaluation of complex operations through the use of a multi-purpose HLA based simulation framework will be presented, with particular emphasis on the main simulation activities led by CETENA in the last years like Replenishment At Sea, Landing Craft maneuverability, Vertical Take Off and Landing, Small Craft Launch and Recovery.

1. INTRODUCTION

The ship is a very complex system of systems and interacts with an even more complex environment which influence and is influenced by the ship behavior. Usually the ship is the only prototype of herself, this is the reasons why simulation, physic based analysis and computational calculations have been largely used since long time in ship design process.

These tools and methodologies are usually used separately, in different ways and by different experts, while the final user is rarely involved in the process.

These are some of key points that bring the design approach toward the use of Virtual Reality environment and High Level Architecture (HLA) based simulation in order to develop a distributed and interactive synthetic environment as much realistic as possible.

One of the most significant purposes is to provide building companies and Navies a tool to test the overall functionality of the ship by the interoperation of multiple simulators. Virtual Reality techniques are becoming commonplace in the field of industrial design since they can help reducing time and cost of building processes, as well as providing a higher quality of the final product [1].

CETENA's approach in the realization of a multipurpose HLA [2] (High Level Architecture – IEEE 1516) based simulation framework for ship's capabilities assessment is presented in this paper.

2. VIRTUAL PROTOTYPING IN SHIP DESIGN

A prototype is any preliminary working implementation of a product, component or system and often it is less detailed than the final version. Two main classes of prototypes are used in design processes: physical prototypes and virtual prototypes.

A physical prototype is a physical model of a product, component or system. Physical prototypes requires weeks-to-months fabrication time and their modification could require days-or-weeks.

A virtual prototype is a computer simulation model of a final product, component or system. A virtual prototype can be used in a design process specifically to explore design alternatives, demonstrate design concepts, test for requirements satisfaction and/or correctness.

In practice a virtual prototype is mainly composed by three elements and/or features:

- Visualization: virtual reality can be used for navigating inside the model. The user can walk

through the complex project viewing exactly how the product will look like. Since the environment in which he is moving is all virtual and controlled by the computer, the user can easily query physical, topological and geometrical characteristics to any system or component.

- Simulation: a static 3D view or a flythrough can be insufficient to point out problems deriving from the movement of objects on deck or interaction between correlated components. Virtual reality can help in keeping together simulation results and real time photo-realistic visualization in synthetic environments.
- Interactivity: the capability to consider the human interaction with the object we are designing may represent a considerable step forward in the ergonomic field or in stressing operational efficiency during design time. The “man in the loop” represents an important design aspect rarely or difficulty taken into account in standard design processes.

Virtual prototyping is a good way to synthesis and visualize all design efforts both for the designer and for the client who can be more directly involved in strategic decisions from the very beginning since they can view and choose between different solutions.

A virtual prototype can be configured more quickly and cost-effectively than a physical one; it can be more abstract and can be invoked earlier in the design process. Moreover a single simulator can improve its capabilities and usefulness if included in a common environment where more simulators interoperate each other to reproduce the behavior of a more complex system.

Great effort is necessary to combine design expectation (simplicity, integration, etc.) with physical and mathematical modeling and simulation. Some interaction effects which are automatically accounted for in physical models need to be consciously accounted for in numerical modeling.

Validation, verification and accreditation of simulation will be the next important step for virtual prototyping credibility and interoperability.

1.1 A laboratory for Virtual Prototyping

During the years, CETENA has developed a set of tools and software in Virtual Prototyping for Ship design. CETENA's Virtual Integrated Ship Laboratory (VISLab) is a place where experts from different ship design fields, along with IT experts, created an architecture devoted to incorporate and reuse all available software tools (COTS or in-house developed) normally used for different purposes with the objective of building a common environment for studying, using and operating virtual prototypes of existing and under design ships [3].

The underlying approach is HLA (High Level Architecture – IEE1516) which let the development of complex distributed simulations.

Being the ship an intrinsic system of systems, the HLA approach results well suited for modeling and building federations which are able to have different levels of details in the implementations of complex behaviors.

In particular this approach let us create simulation systems which can easily updated by the insertion of both new and more detailed physic effects and new interacting entities in the synthetic environment.

2.2 SAND: CETENA's Ship Simulator

The core of the VISLab consists in the CETENA's interactive ship simulator system called SAND.

SAND is a software for maneuvering simulation and training. It is compliant with the HLA 1516 architecture and it includes a ship console and high quality visualization system (Figure 1). A dedicated computing station is used to set the simulation conditions, in terms of scenario, environmental conditions, ship characteristics; the controls on the console can be easily set up through reconfigurable LCD touch panels in order to reproduce those of the simulated ship. During each simulation, all relevant data are stored for subsequent debriefing sessions.



Figure 1: SAND and 3D visualization system

In the tack of reusing all existing tools and competences and thanks to more than twenty years of R&D in ship design and consultancy for harbor designers and Port Authorities, SAND has rapidly reached a very high level of confidence in the representation of the ship behavior.

The CETENA's ship mathematical model included in SAND is based on a Maneuverability and Seakeeping model integration, that allows to calculate the six degrees of freedom (Surge, Sway, Heave, Roll, Pitch, Yaw) and has been validated during the years either from the use and sensitiveness of the Port pilots who used it during the port assessment activities and,

moreover, from the huge real data coming from the experiment activities CETENA performs onboard ships. The simulation model is designed to be “open” and configurable for all type of ship (cruise, bulk carrier, petrol, LNG, container ship) and for all environmental scenarios (map, sea state, wind, sea current, time of day).

Different type of vessels can be easily configured using wide range of characteristics like:

- hull data
- engine type (diesel, diesel-electrical, turbine, or every combination of them)
- propulsion (azimuth, fixed/controllable pitch propeller, POD, water-jet)
- thruster
- rudder (simple type, compound butt, under hung deep horn, shallow horn, spade)

The physical model is completely modular allowing the implementation of add-ons to take into accounts new or more detailed environmental and interaction effects.

The user can choose the environmental condition through a set of sea state, wind intensity and sea current and the software compute the interactive ship behavior for very accurate and realistic vessel simulation taking into account several aspects like:

- shallow water effects
- navigation in narrow channel
- collisions
- anchors and chains

SAND can be used either as a stand-alone system or as a federate, therefore as part of a more complex HLA federation. This can be useful when other simulation entities has to be considered as actors in the same simulation environment. For this reason a basic standard HLA federation (Figure 2) was designed and developed within the VISLab and it is structured as follows:

- *Execution Manager federate* is used to set up the whole federation and to specify all the initial parameters.
- *Data Logger* records all the data send through the RTI, in order to be able to replay the federation in offline mode.
- *3D Visualization federate* visualize the simulation scenario in a 3D environment.
- *Environment federate* sets all the parameters concerning the marine and atmosphere conditions, as sea state, current, wind.

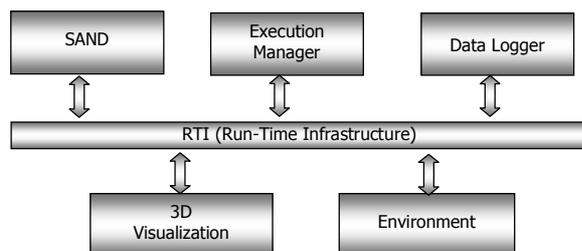


Figure 2: CETENA's basic standard HLA Federation.

In order to be compliant with HLA distributed simulations standards, the above mentioned federation was developed using the VS-FOM (Virtual Ship – Federation Object Model) [4], deriving from the RPR-FOM (Real-Time Platform Reference – FOM).

Multiple entities of SAND can be run in a federation allowing different ships interact in the same synthetic environment.

A portable version of the SAND simulator (see Figure 3) was recently developed in order to easily transport the simulation system into possible partners' facilities and to participate to any other HLA based federation.



Figure 3: Portable SAND

3. APPLICATIONS IN EVALUATING NAVAL OPERATIONS

As previously stated, the driving idea in the VISLab activity is that interactive real-time simulation can be used to investigate complex ship operations for verifying performances and receiving feedback for the design phase.

The implemented architecture and the open structure of the SAND “federate” are the basic ingredients for implementing complex distributed federation simulating operations involving multiple ships or multiple other entities.

In all the following presented experiences the least common denominator is the use of the previously described framework for multidisciplinary topics, with the aim of verifying contractual requisites and also extreme conditions scenarios which may be dangerous to verify in reality.

3.1 Replenishment at Sea (RAS) Operation

Replenishment At Sea is one of the more complex operations performed by modern navies. The operation is executed in three main phases:

- The vessels approach each other and sail alongside.
- The transfer system is established between the vessels using a cable system and the fuel is transferred.

- The transfer system is removed and the ships depart (breakaway).

In all these phases there is a risk of ship collision, especially in high sea states, and when small receiving vessels are involved. A simulation of the RAS operation will allow safe operating conditions to be determined and may influence the design of the supply ship, receiving ship and transfer rigs.

Primary goals of the simulation scenario are:

- to simulate Replenishment operations in open seas with different meteorological conditions;
- to verify operation constraints (ship speed, distance, etc.);
- to analyse different RAS devices in term of position, operative behaviour and efficiency;
- to study refuelling operation (no solid objects transfer).

Basically the operation is focused on the capability of the two ships to navigate parallel for a enough long period of time avoiding collisions and early breakaway. Studies of port manoeuvrability or approaching operations also in open seas are similar to RAS operations and the implemented federation reused all these experience improving the mathematical model of interactions between the two ships.

The two ships are interactively manoeuvrable and the systems reacts in real-time to any change in speed and direction of each ship. In particular the federation can take into account problems deriving from a malfunction in the manoeuvring devices (rudder, propeller...) of one of the two ships in order to accordingly evaluate the capability of the other ship to avoid collision and to continue in the parallel course. This may be very interesting in the design assessment of either the supply and the receiving ships.

From the implementation point of view, two SAND federates were involved in the federation in order to simulate both the supply and the receiving ships. In addition to the standard CETENA's federation two federates were developed and inserted in the federation: a *RAS equipment federate* and an *Interaction federate*.

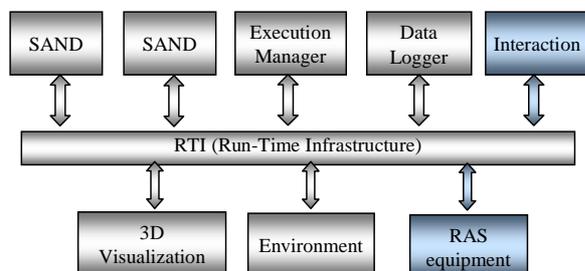


Figure 4: RAS federation

RAS equipment federate was design for the calculation of the forces applied to the cables used for the fuel

transferring operation. In case the RAS cables were subjected to a tension exceeding the maximum allowed, federate is able to give a warning representing the unsuccessfully result of the operation.

Interaction federate was designed in order to simulate the forces due to the side by side shipping; it was developed using mathematical models deriving from towing tank tests and it can be configured to simulate this specific behaviour for different kind of ships.



Figure 5: Screenshot of the RAS simulation

The implemented simulation system is completely configurable and parameterized so that it can be used to simulate ship-to-ship transfer of liquids for:

- Any pair of ships
- Any sea state
- Any type of transfer rig
- Any type of liquid
- Any flow rate of liquids transfer.

3.2 Load Transfer operation

The load transfer between two ships in navigation is part of the NOSE (Naval Operations Simulation Environment), a project led by the Italian Navy and implemented by CETENA.

This operation is similar to the RAS operation but in this case a new simulator has been design in order to analyse the behaviour of the heavy load during its transfer from a ship to the other one. The above mentioned scenarios were developed with the same architecture actually in force at the CETENA' VISLAB facilities, the HLA (High Level Architecture), so that different software modules can be shared among different projects and also among different partners.

In this case the simulation framework previously described has been revisited in order to be compliant with the latest STANAG on Virtual Ship.

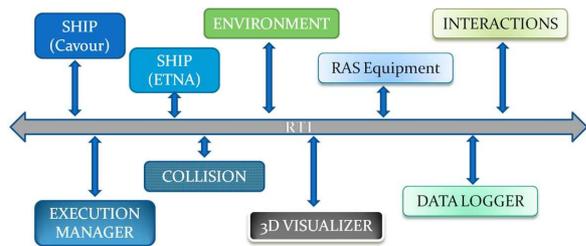


Figure 6: RAS federation

RAS equipment federate has been adapted in order to calculate the forces applied to the cables used for the load transfer operation. The Early Prototype version of the RAS federate can simulate the motion of the load while it is transferred along the cable from a ship to the other. To do that a dynamics simulator based on open-source physical libraries was developed in order to reproduce the behaviour of an object moving on a cable through the use of a system composed by two spherical joints. In the Early Prototype federation the cable among the two ships has been supposed to be linear.

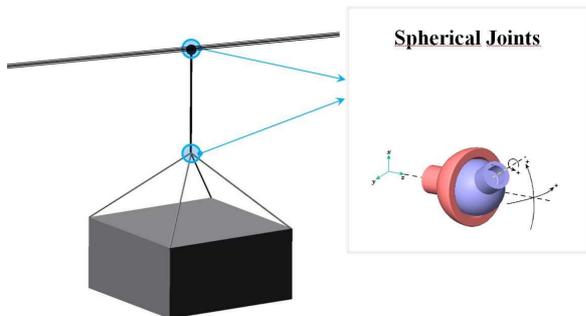


Figure 7 – Heavy Load for RAS simulation

Interaction federate has been reused from the previously described experience, since the aim is still to simulate the forces due to the side by side shipping.

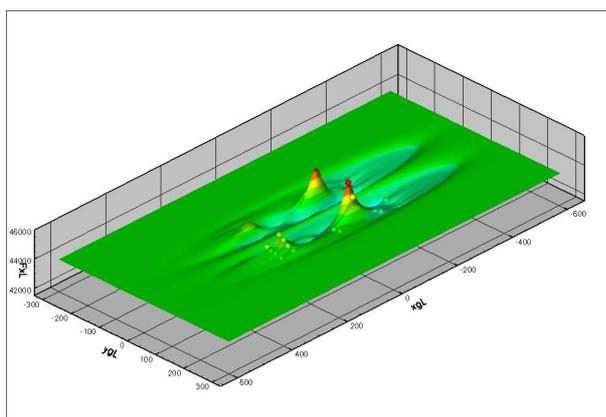


Figure 8: Graphical representation of the Interaction module's results

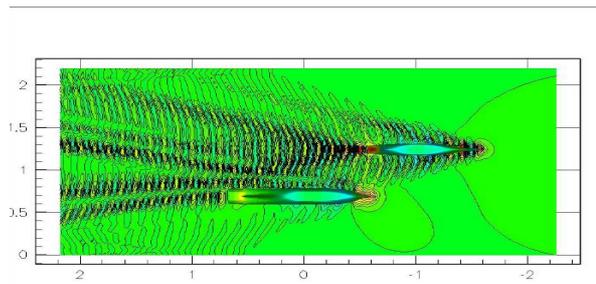


Figure 9 - Graphical representation of the Interaction module's results



Figure 10 – Tank test for interaction forces

The adopted architecture allows the reuse of each federate in other federations and the improvement of the mathematical model for any of the effects to be taken into account. This improvable approach will allow CETENA to easily maintain and upgrade the developed software for future purposes. Some of the federates have been recently used in an international project and simulation results have been validated with sea trials data and towing tanks tests results.



Figure 11: Screenshot of the RAS simulation

3.3 Landing Craft Unit approaching manoeuvre to an LPD ship

In this case an interactive simulation system was developed with the purpose of allowing the Italian Navy to define and verify the operational limits related to the manoeuvrability of a Landing Craft Unit (LCU) while it is approaching the stern of a LPD ship. This study is part of the NOSE project too.

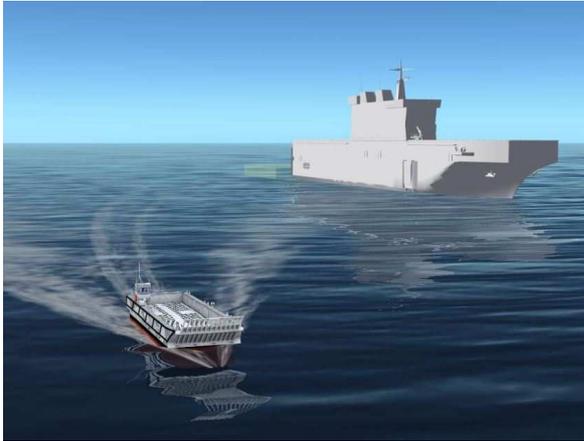


Figure 12 – Screenshot of LCU simulation

All the entities in the scenario are completely interactively manoeuvrable.

Again, the LPD ship is implemented using the SAND federate while the LCU manoeuvrability and seakeeping model is under development in order to take into account the effects caused by the particular type of hull and propulsion and by the stern wake effect generated by the mother ship.

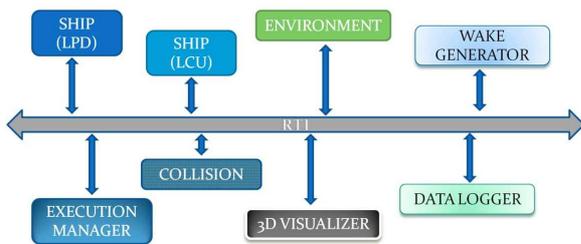
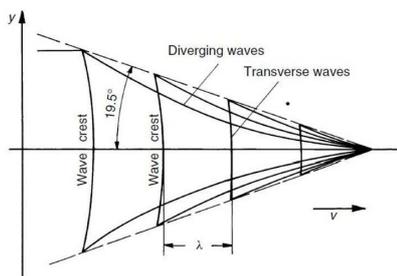


Figure 13: federation for LPD-LCU simulation

The focus of the complete federation will be to accurately simulate the behaviour of the LCU while it is located in the space behind the stern of the mother ship, where the influence of the stern wake is more significant.



For this reason a hydrodynamic study was conducted with the purpose of creating the bases and the knowhow to improve the LCU hydrodynamics model that will be implemented in the final stage of the project.

The manoeuvre can be divided in three phases:

1. The LCU is far from the LPD ship and it is not affected by the waves in the stern of the LPD ship.
2. The LCU is in a region where it is affected by the LPD stern wake but it is still far from the LPD stern door.
3. The LCU is close to stern door and is about to enter the wet dock of the LPD.

The NOSE project focuses on the second phase of the above mentioned approaching manoeuvre. In this phase the LCU is moving in waves that are the result of the superposition of the sea waves and the LPD stern wave.

The hydrodynamic study has been performed for the case in which the LPD ship has a forward speed of 5 knots and travels in sea state 4 waves... In these conditions, the wave generated by the LPD can be modelled as a Kelvin wave and its effects can be superimposed to those of the sea waves.

The motions of both LPD and LCU have been computed using a BEM based seakeeping code which, in addition to providing the ship motions, allows to predict the interaction between the sea waves and the LPD and the shielding effect exerted by the LPD on the LCU for head seas, when the LCU is close to the entrance of the wet dock.

The comments that can be made as a result of this study are the following:

- The LCU motions are not affected by the waves generated by the mother ship (in the speed and sea state conditions considered in this study).
- The shield effect of the LPD ship on the LCU when the waves are coming from the bow is relevant for the last phase of the approaching manoeuvre.
- The LPD motions in sea state 4 are within the limits permitted for the performing of the manoeuvre.

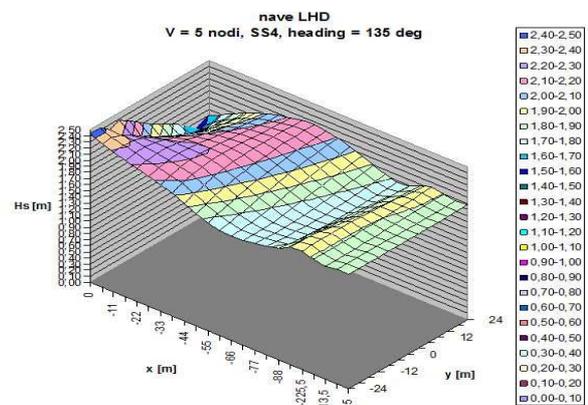


Figure 15 – Numerical results of LPD stern waves

The results of the hydrodynamics study will be used to improve the LCU seakeeping model that will be integrated in the corresponding federate.

All these developments will be included as supporting effects to the main SAND engine and will be reused in future federations.

3.4 Landing Craft Unit operations inside amphibious ship

Another example of complex operation involving multiple interactive but independent entities is the federation we designed and built for reproducing the manoeuvrability of a landing craft (LCU) inside the wet dock of an amphibious ship (LPD) and the transfer of a tank from the Landing Craft to the dock of ship.

Again, the ship is implemented using the SAND federate while the Landing Craft Unit is a completely new mathematical model due to the particular type of hull and propulsion and moreover due to the shallow water effect inside the dock.

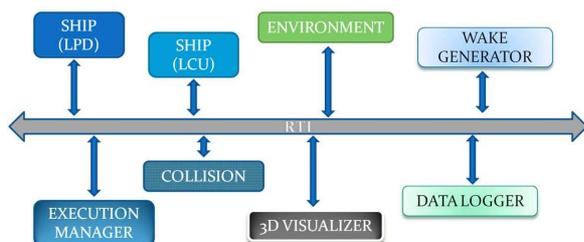


Figure 16: federation for LPD-LCU simulation

Four additional federates were developed in order to simulate the following entities:

- The Landing Craft maneuverability: this model was developed using Physx SDK, an open source physics simulator [1]. With Physx it is possible to simulate the dynamics of one or more objects with different geometries and characteristics. Moreover, Physx allows to interact with the simulated objects by applying to them some forces and moments.
- The Shallow Water effect: the hydrodynamic forces due to the shallow water inside the dock were applied to the Landing Craft. The Tank: when the Landing Craft approaches to the dock and it is stable, the tank can be moved from the landing craft itself to the LPD dock.
- The Collision Detection: when the Landing Craft is inside the LPD dock it is important to monitor the clearance among the craft and the dock. At the end of the simulation the federate can produce a report with the list of collisions detected during the running.

Particular attention has been put in the sloshing effect which is very complicate due to the fact that it is generated by the LPD Ship and is characterized by the volume of the water in the wet dock. Experimental test have been performed to tune the mathematical model.

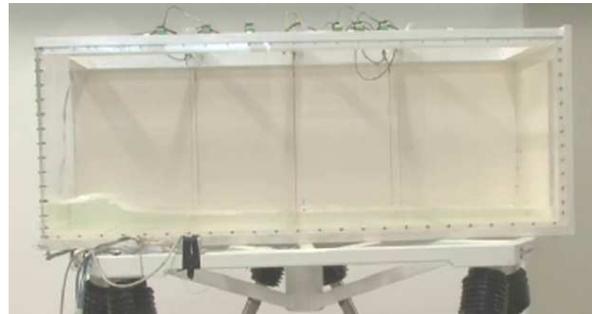


Figure 17: Sloshing experimental tests

The implemented federation has been used during the design of a possible LPD ship and one of the asked question was to determine which was the optimal trim for the internal wet deck in order to have enough ware to avoid collisions with the bottom of the LCU during movements inside and especially loading and unloading operation of the tank.

During the development also the external approach phase of the LCU has been simulated with a very simplified but effective mathematical model. Even if the main purpose is for inside operations, also external behavior of the ship and the LCU are simulated



Figure 18: Screenshot of the Landing Craft simulation

Now the work is going on and the federation will be updated with a more sophisticated physic model of the behavior of the LCU behind the Ship. The LCU travels in a wave field that is the superposition of the LPD's stern wave and the sea waves modified by the presence of the LPD.

Hydrodynamic motion and forces due to the wave actions are now accurately predicted in frequency domain and time domain module accounting for complete wave actions is under development.

All these development will be included as supporting effect to the main SAND engine and will be reused in future federations requiring them.

3.5 Vertical Take Off and Landing (VTOL)

Take-off and landing operations involves different issues related to the interaction between the ship and the air vehicle.

One of the problems is related to the forces during touch down: the ship moves and air vehicle approaches her choosing the best moment for touching the deck. The forces involved are the ones coming from the ship motions and the ones coming from the drop of the vehicle. Especially in heavy sea state the sum of the forces can be very high and can damage the air vehicle landing devices [3].

This effect has been studied in one of the first federation developed inside an international group studying the best landing period for an helicopter to land on a frigate [4].

Besides this impact effect, another important aspect to take care of is the turbulence generated by superstructure in the surrounding of the landing spots. This aspect involves an heavy interaction between fluid dynamic calculations (CFD) for simulating the flows around the structures (Air Wake) and the relative position of the two moving entities: the helicopter and the ship. As in other experiences, these two separated effects were already studied in CETENA and we have the competences and codes for simulation the two effects separately.

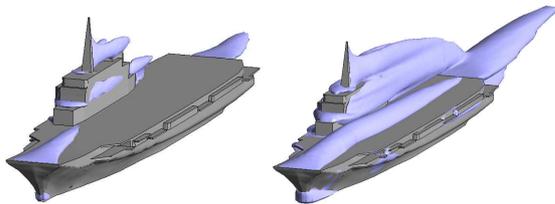


Figure 19: CFD results for VTOL simulation

Again, reusing these knowledge and making the relative tools interoperable, has brought to the design and implementation of a federation simulate the interactive scenario of an helicopter landing on a ship dynamically taking into account in real-time of the air-wake effect and the relative movements and forces between the two moving entities.

The developed system was composed by the standard basic federation for the ship with the addition of some new federates: the helicopter and the airwake.

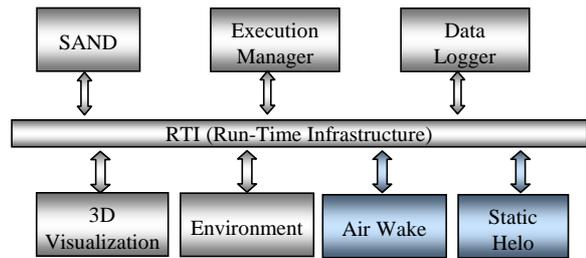


Figure 20: VTOL federation

The helicopter federate is in reality a ultra simplified federate which simulate a point in the space and collect information from the flows deriving from the airwake and the distance from the ship.

The airwake federate is basically a sophisticated interpolator of pre calculated values coming from offline CFD processing. Depending on the relative position from the ship, the speed, the wind direction and other physical parameters, the federate determine the flow pressure and speed in a certain point.



Figure 21: Screenshot of the VTOL simulation

As a matter of fact, since the federation was developed in compliance with both the VS-FOM and the RPR-FOM, a whatever Helicopter federate compliant with the above mentioned FOMs can be easily integrated

3.6 Small Craft Launch and Recovery

CETENA, as a member of CRS (Cooperative Research Ships) group, participated to a Simulation Group focused on the design and implementation of a HLA federation to be used to simulate Small Craft Launch and Recovery (SCLaR) operations.

Also in this context many of the expertises and algorithms developed in previous studies has been reused and integrated in the complete federation.

In particular, in the international project, CETENA provided its VISLab laboratory to integrate the whole federation structured as described in Figure where the “blue” federates were developed by CETENA and the “yellow” ones were developed by other partners.

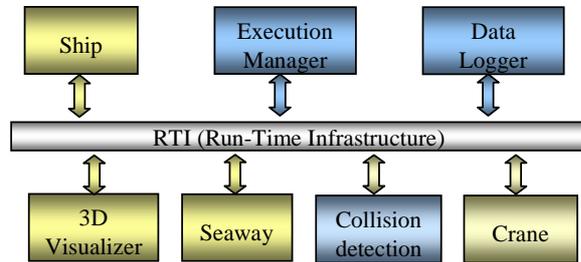


Figure 22: SCLaR Federation

The federation was run in our VISLab and a screenshot is visible in figure Figure .

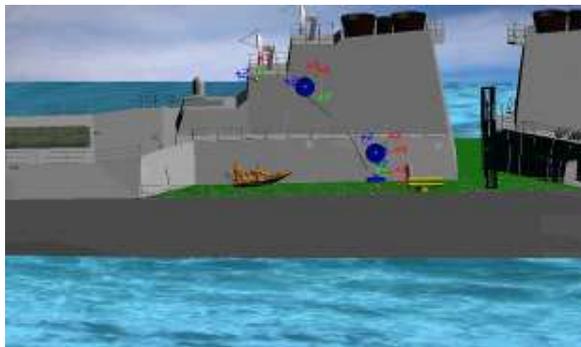


Figure 23: SCLaR screenshot of the small craft launching phase

Primary scopes of this application were the following:

- To share with CRS-SIM group members the knowledge in the development of distributed simulations.
- To set up a common framework able to be improved with detailed simulation models, for the following purposes:
 - To test the winch system in different weather conditions
 - To monitor the waves contact on the small craft during the launching phase
 - To measure the time spent to perform the whole Launch/Recovery operation.

Many of the software and solutions identified in the development of the project have been reused in other applications during the years.

4. INTERNATIONAL COOPERATION

The use of a distributed simulation system utilising the High Level Architecture (HLA) allows the reproduction of a complex scenario in a synthetic environment, permitting several simulators and partners to work

together through the simple exchange of inputs and outputs, whilst protecting confidential information. International Replenishment At Sea (IRAS) Project conducted through a 5-nation Project Arrangement under the Virtual Ships Memorandum of Understanding is an example of an international cooperation in which Italian Navy and CETENA, as technical supporter, is currently involved. The IRAS Project Arrangement encompasses the design, development and validation of a distributed simulation of Replenishment at Sea (RAS) using HLA. The simulation architecture has been developed with the intention also of exercising the draft Virtual Ships (VS) Standardisation Agreement (STANAG) and corresponding Allied Naval Engineering Publication (ANEP) [8]. Relevant pre-existing HLA compliant simulations developed by the Project partners have been collected and revised as necessary to enable them to take part in the federated simulation system for comparison and validation purposes. The verification and validation process has been carried out using both sea trials and tank test data. The final aim of this project is both to create and share a reliable, reusable simulation framework which can be adopted by the partners’ Navies and design contractors during the first design phase of future vessels, and to permit the practicalities of international RAS operations to be assessed. The significant results obtained in this project demonstrate that the same approach could be adopted for a wide range of different naval operations.

5. CONCLUSIONS AND FUTURE DEVELOPMENTS

New technologies, higher computational resources and a new cultural way in the relationship with the computer (especially derived by video games) have foster the idea that something new could be done to face the above mentioned distortions.

Tools, systems and standards are growing. Now it is our turn to put all these staff together and build the Virtual Ship. Two major obstacles on the horizon: costs and culture.

The first one is mainly evident when simulation is seen only as a tool for design purposes and no “reuse” is foreseen. In this case, yes, simulation is a cost and the return of the investment is very small. Making simulations and simulator reusable for other purposes and for other project is a solution to overcome the problem. In this perspective if most of the models and algorithms developed in the design phase can be reused for instance in a training environment, the ROI is shifted in this second phase which can start before the ship is built and can be a benefit for the customer who can train the crew in advance but also for the shipbuilder who has the chance to expand its business also in the post sale

area of the ship and can build an even more strong relationship with the ship-owner.

Italian Navy, with the technical support of CETENA, is now working on the definition of another Project Arrangement to be established with other nations entitled "International Launch And Recovery" project, concerning cooperation in simulation-based design and virtual prototyping for naval ships. This Project Arrangement concerns the design, development and validation of a distributed simulation framework of launch and recovery of small boat using the High Level Architecture (HLA). Two kind of launch and recovery approaches shall be considered, both the stern and the side one.

REFERENCES

- [1] Kanerva M., "Virtual Reality – 4D Product Modelling Tool for Efficient Shipbuilding Process".
- [2] Department of Defense,. *High Level Architecture*, 1988.
- [3] Nilsson, Per-Olof: *The digital product model – a valuable tool for shipbuilders and ship-owners*, Scandinavian Yearbook of Maritime Technology, pp. 37-38, 2000.
- [4] NATO STANAG 4684 Virtual Ship
- [5] Physx, see NVIDIA web site.
- [3] B. Ferrier, O. Le Bihan: *The use of simulation tools in the calculation of aircraft-ship interface optional limits*, 20th Congress on the International Council of the Aeronautical Sciences, Italy 1996
- [4] E. Budde: *The Multi-National HLA Federation Development Supporting Simulation Based Acquisition*, International Conference on HLA Simulation Methodology and Applications, 2005
- [8] NATO STANAG 4684 Virtual Ship

CONNECTION OF MICROCONTROLLER AND MICROCOMPUTER TO DISTRIBUTED SIMULATION

Martin Jakeš^(a), Josef Brožek^(b)

^{(a),(b)} Department of Software technologies
University of Pardubice
Pardubice, Czech Republic

^(a)jakesmar@gmail.com, ^(b)mail@jobro.cz

ABSTRACT

The paper addresses a complementary part of the problem about making of a sturdy trainer which emulates the dispatching center. A specific problem that is solved, is the need for create new controls for interactive simulation. Controls that physically look like a device from real dispatch center must be connected with a computer simulation. The article is focused on a collection of data from various sensors through a microcontroller, or alone microcomputer. After like further processing, either directly in a microcontroller or through a traditional computer as IBM PC. Both of these methods are compared with each other and with focus on regard to the advantage and disadvantages of each.

The simulation is implemented by using online simulation techniques and interactive simulations. The entire simulator is built on the HLA architecture, because it is a modern and widely used standard (IEEE1516: 2010), with elements of Distributed interactive simulation (IEEE1278).

Keywords: simulation, HLA, microcontroller, distributed simulation, IEEE1516, IEEE1278, HLA convertor, signal processing, interactive simulation.

1. INTRODUCTION

When you are creating simulation systems of some classes it is necessary to take into account user preferences; this is especially true for applications such as trainer. For this type of application it is assumed, that the user effectively handles control of the input device. That simultaneously serves as the input/output means of simulation. That is the reason, why the success of the simulator depends on its appearance. Virtualization environments on a computer is often insufficient, due its low user-friendly operation. In a real situation, there is necessary to create a realistic looking environment, in which it is possible to emulate, through the simulator, various scenarios.

2. STATE OF THE ART

The extensive exploration of solutions to similar issues, we could not find similarly oriented study describing the formation of the sensor as a separate federate care

only to provide information about the real world or receiving states of the simulated system and screening of these changes into the real world. One of the exceptions was very outdated publications (3) that it does not capture the current state of the HLA software or hardware that are available today. Another work was found linking minicomputers Raspberry Pi with a mobile device using Android via the RTI DDS technology (2). In the vast majority of publications any real-world sensors are connected to a computer which already represents a functional unit in which the simulation corresponds to a motor vehicle, aircraft or railway station. Or, see some publications performs simulating sensors (4).

The above-mentioned DDS middleware is a competitor HLA. Abbreviation for RTI with DDS means (Real-Time Innovations), which is the name of the company. In HLA RTI means (Run-time infrastructure is a middleware that is required when implementing the High Level Architecture) therefore, these shortcuts do not have nothing in common (for more information and comparisons these technologies resource (1))

2.1. Connecting hardware to computer

For connecting sensors to a computer are commonly used dedicated chips that connect via USB. These chips provide easy connection between computer and hardware without having to know the implementation of the USB communication. Probably the best-known manufacturer is company FTDI. Communication between these chips and hardware (possibly sensors) are mainly serial but chips with a parallel communications also exists (William J. 2002). Using them for interconnection with microcomputer Raspberry Pi is possible but unlike conventional computers, this microcomputer has also dedicated pins witch provide communication directly without the required chip.

3. HLA

High Level Architecture (hereinafter referred as HLA) is a very complex methodology for the creation and operation of distributed simulations. It does not limit the application domain or modelling method (i.e. it is

possible to model through the logical processes, using the principle of decomposition of logical processes and it is possible to create monolithic models, agent-based models, etc.). In addition to the architectural level it specifies many methods at lower levels. There are specified: method of communication between nodes of distributed simulation and the requirements placed only on the actual transmission format (in terms of communication via XML). The major benefit makes possibilities to create heterogeneous simulation models (e.g. using of JAVA and C++ in one simulator). One of the most valuable benefits of HLA is the possibility to adapt an existing simulator for cooperation in HLAs distributed simulation.

A very large standard with a lot of partial documents (IEEE1516: 2010 IEEE1516.1: 2010 IEEE1516.2: 2010 IEEE1516.3: 2010 IEEE1516.4 : 2010) makes it difficult to enter into the HLA problematics.

HLA was originally created for the army of the United States, specifically it has been developed for the Defense Modeling and Simulation Office, on order from the United States Department of Defense. There was a motivation for the interconnection of different simulators, respectively distributed simulation models to more complex simulator clusters. Today HLA it used also in the academic and commercial sector. Most of the non-traditional input and output devices are still the domain of the army only. This is reason, why we focused on developing them.

4. SENSOR / DRIVER CONNECTION PRINCIPLE

Existing IT solutions (e.g. USB signal converter) are not appropriate. As one purpose it could serve to connect the sensors to the computer, but the connection to the computer is not a sufficient condition for the proper functioning of the input device in a distributed interactive simulation.

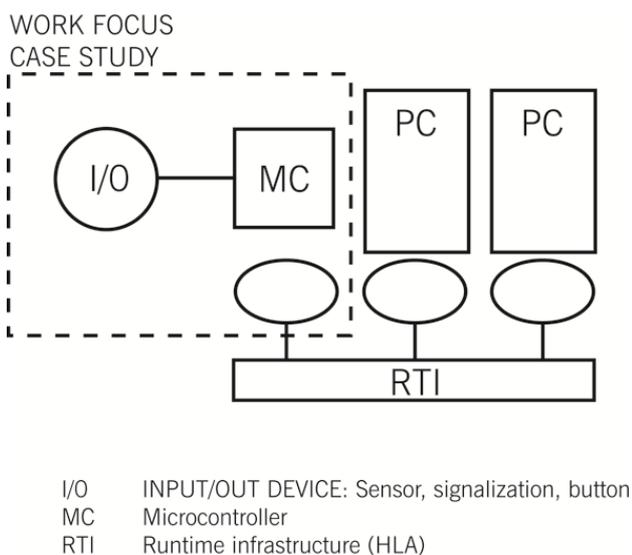


Figure 1: First alternative of situation

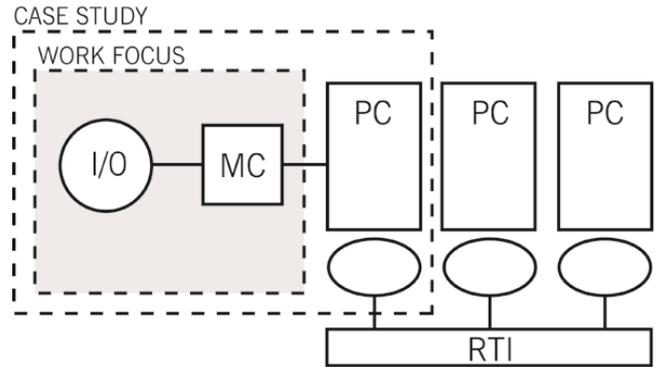


Figure 2: Second alternative of situation

A microcontroller itself is capable of processing various signals from input devices and sensors. The main part of developing is creation of bridge between distributed simulation and microcontroller. The various types of input and output devices are taken into scope in final paper, and in particular to account for their speed and complexity of data processing from those. The text is, rather focused on techniques how to get signals from the microcontroller and awards after-processing of data to possibilities to use them in a distributed interactive simulation.

Comparison of optimality of direct connection (a microcomputer connected to the LAN as an element of distributed simulation), with solution, which is realized by connecting microcomputer to another computer, will be part of the full text of the article. It will also include focus in regards to the advantage of each of those solutions and their measurement.

5. HARDWARE LAYER

For the needs of this publication, hardware layer is understood as a computing hardware serving as a pre-processor or an input-output data processor. Solution to be generally described as the use of a computer is too simplistic, because except of those complying with IBM-PC standard, there are numerous other devices considered to be a computer.

To connect external hardware (input-output device) with the simulator, there exist two basic ways. The first method is taking the input-output device and its encapsulating hardware as one object, while this solution has a certain advantage especially in the event of its engagement in distributed simulation. Alternatively, an ideal solution (namely in the programming point of view) is to separate both devices from each other followed by the formation of independent logic with the option to connect the given device and also its analogues.

5.1. Personal computer

In order to a comparison was good enough and had sufficient informative value personal computer was also included to it. Given that the development of sensors connected to the simulation is usually performed using a PC, it's another reason why incorporate computers.

To simplify the comparison of the computer will perform calculations solely on the CPU (computing on graphics card are not considered) Core2Duo. Fitted with memory in dual-channel RAM and has a size of four gigabytes. For connecting the input output devices it was used only USB port. However, many devices at the hardware level using only the USB emulation through a serial communications port. The reader should therefore always be careful whether (if necessary) device uses USB full capacity.

5.2. Mobile device

Besides standard computer, we can choose to use a mobile device (tablet, mobile phone) as a hardware interlayer. Despite it does not seem quite useful at first sight, it is essential to mention the advantages of mobile devices: mobility and price, low overhead cost, simple programming and especially large number of interfaces. Mobile device have USB ports and often also Wi-Fi, Bluetooth and other technologies. The development may be as easy as that of PC, however, thanks to the portability and price, or due to the implicit network support, solution in certain events appears to be much more effective.

Another advantage, not yet mentioned, which can be used mainly in online situations, is the permanent battery operation. If a device, for example a pre-processor for sensor network data processing, is connected we can collect data even during a power failure because mobile devices have their back up source.

A drawback of this solution is a lower computation output. Despite the price is lower than that of the previous solution, it is far from being ideal.

5.3. Raspberry PI

Raspberry PI (Figure 3) is an interesting solution that, for the price around €39, gives options similar to those of a standard computer or mobile device. Raspberry has standard USB ports (for connecting commercial devices) and ports enabling specific programming options. Feeding is carried out via USB, optionally from a mobile phone Power Bank. Certain drawback of Raspberries is a limited choice of operating system selection: the inherent system is implicitly used. In any case, such hardware platform presents a great potential for the formation of simulator input-output devices. Its small size, low consumption, a relatively high computation output and a wide selection of different ports makes Raspberry a perfect interlink between any (commercial or inherent) hardware device and computer simulation.

Additional Information: After making a comparison test it was released a new market model Raspberry PI,

which has all the advantages mentioned in this article, but there has been a significant increase in computing power. Comparison can be found for example on the (Benchoff Brian 2015). But modern equipment for the purposes of this paper was not included for reasons of state of the article.



Figure 3 Raspberry PI

5.4. Microcontroller

It is only a few years ago that non-standard devices were connected to a PC via microcontrollers.

Original requirement for programming in specific programming language which is different for various chips (e.g. Assembler), the programming language C became a standard method that opened micro-controllers to the community not specializing in the lowest solution levels.

In any event, this device becomes obsolete and even though there exist classes of tasks, where it still can be used, they are not suitable for applications in modern simulators. You can see (Richardson 2012, Shawn, 2012) for more information.

5.5. Comparison of solutions

Comparison is based on own research in (Brozek, Jakes and Gago, 2014) and a few prestige sources as (Huddleston, 2007; Schön, 2013; Schönborn 2013).

5.5.1. Price

Taking in view sufficiently subsidized simulations such as Hi-tech segment that do not have to consider the price differences within percent units, it will be quite interesting to look at Table 1 which illustrates current prices, and what is more, the price is compared with the computation output.

Computational efficiency of microcontroller is under focused scope.

Table 1 Price comparison

Device	Price [EUR]	Power [GFLOPS]	Power / price [GFLOP / EUR]
PC (2x AMD Opteron 6344, 64 GB RAM, GTX 660)	2 000	10 000	5
Raspberry PI B+ (ARM 900 MHz, 1 GB RAM)	40	1	0,025
Dell Tablet (Intel Atom 1,6 GHz, 2 core, 512 MB RAM)	56	2	0,036

5.5.2. Computation output

However, computation output (Figure 4) may not play its role. Dependence of the whole solution on the computation output relates the way the unit is used. Therefore, if the task is a mere collection of interactive data or providing interactive outputs, all the compared solutions have a sufficient computation output. If we take in consideration the fact that besides a I/O message, a unit may be the main node of a particular simulation, or one of the computation nodes for a distributed simulation, then the absolute value becomes important. Major effect of computation output is monitored in extensive online simulator characterized by a large number of very fast processes.

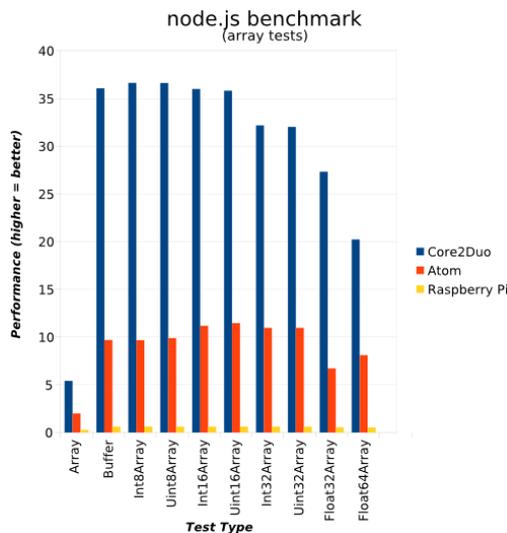


Figure 4 Computation output [Henniger, 2014]

5.5.3. Energy requirements

Comparison of energy requirements is illustrated in Table 2. Microcontroller is a clear winner. On the other hand, when returning to computation output, it is obvious that computation output conversion into an electric energy output of Watt is not as favourable as it seemed (the value is poor), and micro-controller excels only optically. The worst values appear at computers, which is not a surprise.

Table 2 Energy consumption

Device	Run on battery 5AH [hours]	Power [GFLOPS]
PC (2x AMD Opteron 6344, 64 GB RAM, GTX 660)	<1	10 000
Raspberry PI B+ (ARM 900 MHz, 1 GB RAM)	17	1
Dell Tablet (Intel Atom 1,6 GHz, 2 core, 512 MB RAM)	6	2
Micocontroller (AT Mega 16 series)	48+	<0.001

5.5.4. Results of comparison

Based on several independent measurements and also on the team’s experience, we can demonstrate that Raspberry can be considered an optimal device for the formation of interlinks for input-output device. On the other hand, in terms of a simple solution, especially if mobility is not required, users prefer a more conservative way – connection to a PC.

6. OUR OWN SOLUTIONS

To test different variants, has been created demonstrator simulator with an intuitive rules. Distributed simulation has three key parts: The first part is the software driver for the entire solution, the second part visualizer and archival core. The last, third part of the Federation, is the one, who in various ways connected hardware solution.

6.1. Issue conception

The simulator is intended to reflect operation of the simple logistics system: A test rail circuit. The principle is that, there is a railroad track with two different circuits. Track is used to test the basic handling characteristics of the train. Demonstrator software can test only characteristic as speed, stability. It also can show the possibility of inclination in dependence on the speed or electrical properties, as speed to input electrical power (more specifically, on set of input voltage and supply current).

In the simulator is a sequence of simulation calculation evolution. First is practical management (interactive Federate). The orders are send to second part: i/o federate, which operates physically at model railways and at the same time collect data (eg. A journey time of the train, information on how the error occurred, or report time of crashes the train). A demonstrator is so interesting example of simulation that is not only software, but it is a combination of software and physical model.

6.2. Electric signal / federate simulation

Physical Model Railroad (trains size TT) is controlled by voltage. The basic source of tension is a certified laboratory source, which is connected via other electronics to railway. The electronic varying according to the method of controlling: a computer / mobile devices / raspberry / Microcontroller with the service circuits.

The most complex development required the involvement of microcontroller. It was difficult especially due needs of very complex wiring, realized by circuits on a non-solder field. Simultaneously, programming was very non trivial due necessity to work with a low level instructions. Sad finding was the fact, that implementation at a low level does not improve performance. Hardware of Microcontroller has been generally insufficient neither for basic data processing in distributed simulation HLA (ie the translation of key data in XML format).

From the development perspective, microcomputers were a very pleasant surprise, thanks to its easy setup. Raspberry PI has hardware pins, which are very easy to address and operate with them. Programming interface for these input / outputs is very easy.

6.3. Other federates

The other two federate types are standard implemented as standard full software solution. They are created by programmatically in Java. Operation (communication and synchronization) is provided with standard library classes of HLA. RTI is used from the Pitch Company (Pitch RTI).

Management federate is implemented a standard interactive simulator. User entry is written to the calendar of events by introduced NextEvent. Events calendar make synchronization with the rest of the federation.

7. DISCUSSION

The results of our study reflects the scientific trends in the world. Based on testing it is necessary to declare that the involvement of microcontroller in distributed simulation provides no added value. At the same time, the development of microcontroller is the most challenging of all tested technology. Ten years ago, it might be interesting to use microcontrollers for economic reasons. The current market situation, however, offers a much better solution (raspberry) with similar price. In further discussion and microcontrollers are taken into consideration only marginally.

Chapter expands the comparison of the various technologies about results of our own research and testing.

7.1. Sophistication

An important parameter when considering whether to include a technology into our own solutions is difficulty of implementation (Respectively, difficulty of implementation in relation to the added value for a solution. But since in this paper are examined only solution with the same added value that we can remove this influence and focus only at difficulty of implementation).

Implementation time was examined according to the same solution implementation tasks for all examined systems. The measured values were man hour.

All four solutions that were tested are characterized by distinct complexity implementation. Among the simplest solution are a standard computer programming and programming for Rawsberry PI. Difference between these two technologies is very small. This is due to very good support for high-level programming languages, and a large variety of supported drivers.

The third solution in difficulty is in our comparison of programming for mobile devices (mobile telephony or tablets). Moderate disadvantage is the requirement for a limited number of environments and difficult access to the drivers. Overall, the programming for mobile devices is about 2 times more difficult than the case of computers or microcomputers.

Far more complex (10x implementation demanding than solutions for mobile devices and therefore 20x implementations demanding than for computer and microcomputer) became implementations for microcontrollers. In addition, it was found that there are some problem sets that cannot be solved through Microcontroller because of hardware limitations.

7.2. Mobility

One of the key characteristics in simulators may be mobility solutions. For example: In the simulator real environment, it is necessary to attach some interface directly to the soldier in the simulator is completely inappropriate to connect a standard PC with a weight of several kilograms, and the requirement for permanent power connection.

In the tests were the key requirements for ease of installation, size and method of power.

Although at first sight could be natural that those tests prevails mobile devices, it is not. The real winner was the microcomputer, which (in conjunction with power pack) managed to work much longer than mobile devices. Weight and dimensions should take better character than with mobile device.

In second place could be placed Microcontroller. But, because it is necessary to take into account also ancillary circuits and specific power requirements and

extreme demands on the communication scatters, was very high discomfort when trying to build an "island simulator".

Mobile devices have the disadvantage of required imaging unit. Calculations in sleep mode, or turn off the display is very difficult to implement and that are characteristics that strongly supported Discomfort in solving.

Through actual miniaturization of computers with a standard computer is totally unsuitable as a competitive candidate in this discipline.

7.3. Usability and potential

As mentioned in the introduction, the potential for the Microcontroller is currently very small. Old solutions can still deploy, but the development of innovations already lacks effectiveness.

Other technologies are advised mobile technology. Although all tests achieved average results, general this is a solution that lacks the potential (except for their use in distributed interactive simulations, as explained in Brozek, Jakes, Gago (2014).

The computer has a big advantage if it is used in stationary solutions. A great advantage in this case is the number of inputs that reaches many dozens. While potential in processing speed and ease of maintenance is very high.

Microcomputers excel in this type of task, where the requirements are very similar to those on a standard PC, but also require mobility solution of. Likewise, it is win solutions where we get by with limited computational power but is paramount for us economical solution. Then they are an ideal example of emulation, live simulation and input output units simulators.

8. CONCLUSION

A key objective in this work was to test different ways of microcontroller participate in distributed simulation. The second primary objective was to compare different methods of involvement with alternative solutions (computer, mobile device, I microcomputer).

The findings showed that the trends that were still in the last decade obvious are overcome. Using of microcontrollers in complex systems can no longer be recommended. Claims for developers and the extreme need for efficient optimization function in a complex system cannot compensate for the benefits of microcontrollers.

The second set goal (compared with alternative technologies) revealed the extreme potential of microcomputers in distributed simulation. And not just as potential candidates for compute nodes, but also as signal converters - input and output.

The paper describes the development in IT industry. Previously, cheap and efficient, Microcontroller is thanks similar affordability, but far greater user-

friendliness with programming is replaced by the microcomputer.

Paper resume is clear: microcontrollers are overcome. Let's concentrate on the microcomputer.

REFERENCES

- BENCHOFF, Brian. BENCHMARKING THE RASPBERRY PI 2. Hackaday [online]. 2015 [accessed 2015-07-22]. Available from <http://hackaday.com/2015/02/05/benchmarking-the-raspberry-pi-2/>
- Brozek J., Jakes M., Gago L. Using tablets in distributed simulation. Pardubice, 2014. ISBN 978-889799932-4. Conference Paper. EMSS 2014 Proceedings, University of Bordeaux.
- BY TERO KARVINEN, Kimmo Karvinen. Make sensors: a hands-on primer for monitoring the real world with Arduino and Raspberry Pi. 1. Aufl. MakerMedia/KM/Feb: O'Reilly, 2014. ISBN 978-144-9368-104.
- Chemical sensors: simulation and modeling. Editor Ghenadii Korotcenkov. New York: Momentum Press, c2012, xxiv, 440 s. Sensor technology series. ISBN 978-1-60650-309-6.
- FRADEN, Jacob. Handbook of modern sensors: physics, designs, and applications. 4th ed. New York: Springer, 2010, xv, 663 s. ISBN 978-1-4419-6465-6.
- Free software foundation, Inc, 2002, CERTI - Summary [Savannah]. Available from: <http://savannah.nongnu.org/projects/certi> [accessed 22 April 2014].
- FTDI [online]. [accessed 2015-07-22]. Available from: <http://www.ftdichip.com/>
- Fujimoto, Richard M., c2000, Parallel and distributed simulation systems. New York; John Wiley & Sons. ISBN 04-711-8383-0.
- Geeknet, Inc., 2005, SourceForge.net: Open HLA - Project Web Hosting - Open Source Software. Available from: <http://ohla.sourceforge.net/> [accessed 22 April 2014].
- Huddleston C. Intelligent sensor design using the microchip dsPIC. Amsterdam: Elsevier/Newnes, c2007. xix, 283 p. Embedded technology series.
- IYENGAR, S, L PRASAD a H MIN. Advances in distributed sensor integration: application and theory. Upper Saddle River, N.J.: Prentice Hall PTR, c1995, xvii, 273 p. ISBN 01-336-0033-5.
- KARLSSON, Mikael; NÄSSTRÖM, Fredrik; HAAPALAHTI, Gustav. Building a modular multi-sensor simulation framework using COTS software and HLA. In: Proceedings of the IASTED International Conference. 2009. p. 025.
- Knight, Pamela, Corder, Aaron, Liedel Ron, 2001, Evaluation of Run Time Infrastructure (RTI) Implementations. U.S. Army SMDS. Available from: <https://www.scs.org/confernc/hsc/hsc02/hsc/papers/hsc017.pdf> [accessed 15 July 2014].

- Kovač, Pavel, 2012. Reflection: Where are we going tablets?. Available from: <http://www.svethardware.cz/uvaha-kam-smerujitablety/34506-3> [accessed 15 July 2014].
- Ku, Andrew, 2012. Lenovo's ThinkPad X230T Tablet PC, Tested And Reviewed. Available from: <http://www.tomshardware.com/reviews/thinkpadx230t-review-benchmark,3229-2.html> [accessed 15 July 2014].
- Kuhl, Frederick, Dahmann, Judith, Weatherly Richard, The institute of electrical and electronics engineers, Inc, c2000, Upper Saddle River, NJ; Prentice Hall PTR. ISBN 01-302-2511-8.
- Manlig, František, 1999. Computer simulation of discrete events. Available from: <http://www2.humusoft.cz/www/archived/pub/witness/9910/manlig.htm> [accessed 15 July 2014].
- Mocný, Ondřej, 2009. Real-time physics simulation for mobile device. Bachelor thesis. CHARLES UNIVERSITY.
- NGOC THANH NGUYEN ... [ET AL.], Ngoc Thanh Nguyen ... [et al.(eds. Agent and multi-agent systems technologies and applications: second KES international symposium, KES-AMSTA 2008 : Inch'ön, Korea, March 26-28, 2008 : proceedings. Berlin: Springer, 2008. ISBN 978-354-0785-828.
- Pi4J [online]. 2012, 18.4.2015 [cit. 2015-07-22]. Available from: <http://pi4j.com/>
- Pitch technologies ab, 1993, Products – Overview, Available from: <http://www.pitch.se/products/products-overview> [accessed 22 April 2014].
- Rabelo, Luis, Sala-Diakanda, Serge, Pastrana, John, Marin, Mario, Bhide, Sayli, Joledo, Oloruntomi, Bardina, Jorge, 2013. Simulation Modeling of Space Missions Using the High Level Architecture. Available from: <http://www.hindawi.com/journals/mse/2013/967483/>[accessed 15 July 2014].
- RAJIVE, Joshi a Castellote GERARDO-PARDO. A Comparison and Mapping of Data Distribution Service and High-Level Architecture [online]. 2006 [accessed 2015-07-21]. Available from: <https://www.rti.com/docs/Comparison-Mapping-DDS-HLA.pdf>
- Richardson M., Wallace S.P. Getting started with Raspberry Pi. 1st ed. Sebastopol, CA: O'Reilly Media, 2012, xiii, 161 p. Make: projects. ISBN 14-493-4421-6.
- RTI Connex DDS Android-Raspberry Pi Demos [online]. MEENA, Ashish. 2014 [accessed 2015-07-21]. Available from: <https://community.rti.com/howto/rti-connex-dds-android-raspberry-pi-demos>
- Schön O., 2013. How powerful is your PC? New 3DMark test measures the ability of PC and tablets and phones. Available from: <http://tech.ihned.cz/hry/c1-59252020-3dmark-propc-i-iphone-a-android> [accessed 15 July 2014].
- Schön, Otakar, 2013. How powerful is your PC? New 3DMark test measures the ability of PC and tablets and phones. Available from: <http://tech.ihned.cz/hry/c1-59252020-3dmark-propc-i-iphone-a-android> [accessed 15 July 2014].
- Schonborn T. SoC Shootout: x86 vs. ARM. In: [online]. 2013 [accessed 2015-03-15]. Available from: <http://www.notebookcheck.net/SoC-Shootout-x86-vs-ARM.99496.0.html>
- Sensors. 2011, vol. 11, issue 12. ISSN 1424-8220. Available from: <http://www.mdpi.com/1424-8220/11/4/4385/>
- The institute of electrical and electronics engineers, Inc, 2010, IEEE1516:2010: IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-Framework and Rules. New York; IEEE. ISBN 978-0-7381-6251-5.
- The institute of electrical and electronics engineers, Inc, 2010, IEEE1516:2010: IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-Object Model Template (OMT) Specifications. New York; IEEE. 2010.ISBN 978-0-7381-6249-2.
- The institute of electrical and electronics engineers, Inc, 2010, IEEE1516:2010: IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-Federate Interface Specification. New York;IEEE. ISBN 978-0-7381-6247-8.
- The simulation interoperability standards organization,. Independent Throughput and Latency Benchmarking for the Evaluation of RTI Implementations, 2001, The Simulation Interoperability Standards Organization. Fall. DOI: SISO-01F-SIW-033.
- Using HLA to Transmit Real-Time Sensor Imagery [online]. In: SILBERT, Mark, William SCHIBLER a Gordon CURRAN. 2000 [accessed 2015-07-21]. Available from: https://www.researchgate.net/publication/235185975_Using_HLA_to_Transmit_Real-Time_Sensor_Imagery
- Vtmäk, 2014, HLA RTI – Run Time Infrastructure – MÄK RTI. Available from: <http://www.mak.com/products/link/mak-rti.html> [accessed 22 April 2014].
- William J. Serial and Parallel Transmission. THE GALE GROUP INC. Encyclopedia.com [online]. 2002 [accessed 2015-07-22]. Available from: <http://www.encyclopedia.com/doc/1G2-3401200276.html>

SIMULATION-BASED SET-UP TIME OPTIMIZATION USING SIM# AND HEURISTICLAB

Johannes Karder^(a), Andreas Scheibenpflug^(b), Andreas Beham^(c), Stefan Wagner^(d), Michael Affenzeller^(e)

^{(a),(b),(c),(d),(e)}Heuristic and Evolutionary Algorithms Laboratory
University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg, Austria
^{(b),(c),(e)}Institute for Formal Models and Verification
Johannes Kepler University, Altenbergerstraße 69, 4040 Linz, Austria

^(a)jkarder@heuristiclab.com, ^(b)ascheibe@heuristiclab.com, ^(c)abeham@heuristiclab.com, ^(d)swagner@heuristiclab.com,
^(e)maffenze@heuristiclab.com

ABSTRACT

Model building is a fundamental task in simulation-based optimization. In this paper we demonstrate the application of Sim# in combination with HeuristicLab to optimize set-up times of arbitrary machinery. On top of Sim#, custom simulation extensions have been implemented and are used to create a simulation model of real world machinery. These extensions enable the design of simulation components that can be reused within different simulation models. This allows to easily create multiple model implementations that reflect different designs of a machine by using a combination of already existing and adapted components. The resulting model is used as evaluation function for the optimization inside HeuristicLab.

Keywords: simulation-based optimization, set-up time optimization, sim#, heuristiclab

1. INTRODUCTION

Sim# (Beham et al. 2014) is a .NET port of SimPy (Matloff 2008), a process-based discrete event simulation framework. Implemented in the C# programming language, it can easily be integrated in other .NET applications such as HeuristicLab (Wagner et al. 2014). It provides an intuitive way to simulate processes as events. Sim# contains an event queue which allows fast evaluation and computation of processes that make up the simulation model. Simulation-based heuristic optimization methods often require many fitness evaluations that employ a simulator. Therefore, fast simulations have to be implemented so that evaluation phases can be executed in a reasonable amount of time.

HeuristicLab (HL) provides different ways of integrating external software such as simulation models in optimization algorithms, e.g. by communicating with other applications using Google's Protocol Buffers (<https://developers.google.com/protocol-buffers/>) or by using other plug-ins specifically designed for this purpose (e.g. MATLAB and Scilab problems). In this publication the authors created a new customized plug-in for simulation-based set-up time minimization. It features a

simulation model which is created with a new simulation architecture based on Sim#.

The focus is to extend Sim# in such a way that it is possible to model reusable software components for simulation models and use those components to create similar but individual model implementations which represent e.g. different machine designs.

In Section 2, previous related work is presented. Section 3 explains the architecture of a general simulator. In Section 4, insights into the actual implementation of the used simulation model are given. The optimization is explained in Section 5 and the achieved results are presented in Section 6. Section 7 concludes the paper with gained domain knowledge and an outlook about future work.

2. RELATED WORK

Previous work including the use of HeuristicLab has been done by Affenzeller et al. (2007) to tune simulation parameters by applying simulation-based and heuristic optimization methods. Bruzzone et al. (2011) conducted a real case study about short period production planning in manufacturing systems by combining a flexible simulation model, genetic algorithms and dispatching rules. Beham et al. (2012) also worked on alternative ways for information exchange between simulation and optimization processes. In (Beham et al. 2014), Sim# was used to implement a simple supply chain simulation. HeuristicLab was then used to optimize the simulation's parameters. A similar set-up time optimization in the same domain has already been done by Karder et al. (2015) without using an underlying simulation framework such as Sim#. Here the simulator was written and continuously extended with new functionality without reusability in mind. This made it very difficult to implement additional specifications and new features which were added occasionally. A specific optimization problem was also created for HeuristicLab, in which three different optimization aspects, namely the job execution order, the storage organization and different (un-)loading strategies were optimized. Set-up times could be reduced by approximately 20.21 %.

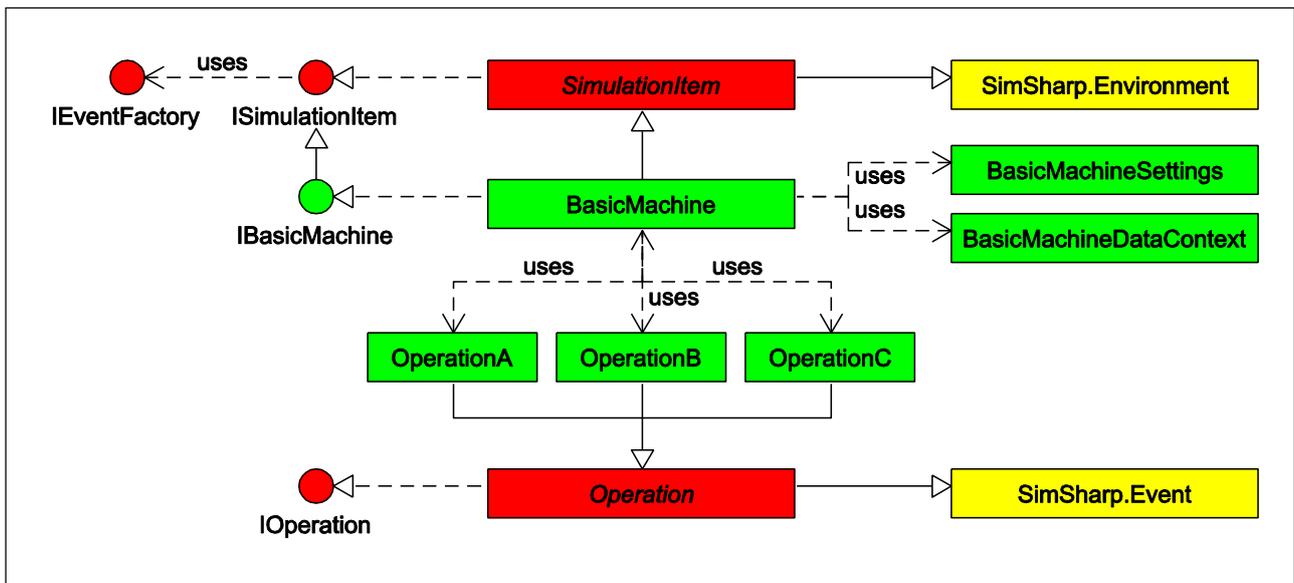


Figure 2: The simulator's architecture showing parts of Sim# (yellow), classes from the implemented extensions (red) and from the implemented simulation (green).

3. SIMULATOR ARCHITECTURE

The simulator is built with modularity in mind. With Sim# as underlying simulation framework, the focus of the implementation is not the creation of a simulation framework itself, but rather the application and extension of already existing simulation infrastructure and the design of modular components which can be reused in different versions of a simulation that belong to a certain kind of simulation domain. Instead of creating three separate implementations, a basic version of the machine can be modeled and extended with different functionality by deriving or adding new operations. This leads to a three-layered architecture, where a concrete simulation builds upon specific Sim# extensions, which again build upon the simulation framework itself. This architecture is shown in Figure 1. Using this concept, the changes in the machine's designs should only be minor, since atomic operations are often implemented for a specific kind of machinery.

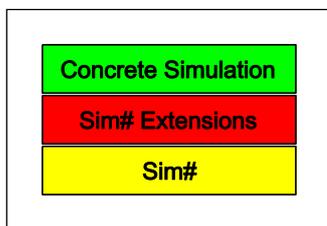


Figure 1: The three simulation layers.

Each layer's most important classes/interfaces and their relations are shown in Figure 2. The base for any simulated machine is a **SimulationItem**, which represents the execution environment and is shown in Listing 1.

```
public abstract class SimulationItem
: SimSharp.Environment, ISimulationItem {
private readonly IEventFactory eventFactory;
public IEventFactory EventFactory
{ get { return eventFactory; } }
```

```
protected SimulationItem() {
eventFactory = new EventFactory();
}

public SimulationResult Run() {
var op = CreateSimulationOperation();
Process(op);
base.Run();

var result = new SimulationResult {
Duration = Now - StartDate
};
return result;
}

public abstract IEnumerable<SimSharp.Event>
CreateSimulationOperation();
}
```

Listing 1: The abstract base class **SimulationItem** from which all simulated environments should derive.

It implements the **ISimulationItem** interface, which provides access to an **EventFactory** and a **Run** method. The **EventFactory** can be used to register to different events and allows to execute according event handlers after an operation, its dependencies or its sub-operations have been executed. This concept allows to easily extend a simulator with debugging, logging or visualization features. The **Run** method is called to execute the simulation.

Any machine deriving from **SimulationItem** owns a context and has specific settings. Contexts represent the current state of a machine and its components and are composites of multiple properties and sub-contexts, which are accessible and can be modified by the machine and its operations. Operations are always created with the specific context they modify. Each operation also has access to the global context of the environment. Settings define different properties of the machine, e.g. sizes and speeds. Unlike the context, settings are read-only and therefore cannot be modified during a simulation run.

Simulated operations are always designed for a specific environment. By implementing (and deriving from) an `ISimulationItem`, all operations from the simulated environment have to be defined. In the presented architecture it is possible to reuse operations defined for environment A in another environment B. The limiting factor here is that an operation can only be reused in B, if B is a subtype of A. This constraint has to be satisfied, because the operation was originally designed for A and possibly requires contexts and settings that have also been designed for A.

Each machine operation is represented by a separate class that derives from `Operation` which implements `IOperation`. All operations have to declare their dependencies (i.e. operations which have to be executed before the actual operation) and sub-operations (i.e. all operations that are executed after the actual operation). By calling `Operation.Execute`, only the actual operation is executed and its dependencies and sub-operations are ignored. `Operation.ExecuteAll` calls all dependencies, the actual operation and all sub-operations (in that order). Different information such as the start time, the duration and the description are set by all operations and passed to the `EventFactory` in form of `OperationInformation` instances. Each `SimulationItem` defines an abstract `CreateSimulationOperation` method that represents the entry point of the simulation. This method is called within the `Run` method and builds the operation tree, i.e. yields all necessary operations as well as its dependencies and sub-operations. Listing 2 shows the `Operation` class.

```
public abstract class Operation
: SimSharp.Event, IOperation {
    private readonly IDataContext ctx;

    protected SimulationItem Env
    { get { return (SimulationItem)Environment; } }

    public IDataContext DataContext
    { get { return ctx; } }

    public virtual IEnumerable<SimSharp.Event>
    Dependencies { get { yield break; } }
    public virtual IEnumerable<SimSharp.Event>
    SubOperations { get { yield break; } }

    public OperationInformation
    DependencyOperationInformation { get; set; }
    public OperationInformation
    OperationInformation { get; set; }
    public OperationInformation
    SubOperationInformation { get; set; }

    protected Operation(SimSharp.Environment env,
        IDataContext ctx = null)
    : base(env) {
        this.ctx = ctx;
        DependencyOperationInformation = new ...
        OperationInformation = new ...
        SubOperationInformation = new ...
    }

    public virtual IEnumerable<SimSharp.Event>
    Execute() { yield break; }
}
```

```
public virtual IEnumerable<SimSharp.Event>
ExecuteAll() {
    var depOpStart = Env.Now;
    yield return Env.Process(Dependencies);

    DependencyOperationInformation.Duration =
    Env.Now - depOpStart;
    Env.EventFactory
    .FireOperationsDependenciesExecuted(this);

    OperationInformation.StartTime = Env.Now;
    yield return Env.Process(Execute());

    OperationInformation.Duration =
    Env.Now - OperationInformation.StartTime;
    Env.EventFactory.FireOperationExecuted(this);

    var subOpStart = Env.Now;
    yield return Env.Process(SubOperations);

    SubOperationInformation.Duration =
    Env.Now - subOpStart;
    Env.EventFactory
    .FireOperationsSubOperationsExecuted(this);
}
}
```

```
public abstract class Operation<TEnvironment>
: Operation
where TEnvironment : SimulationItem {
    protected new TEnvironment Env
    { get { return (TEnvironment)base.Env; } }

    protected Operation(TEnvironment env,
        IDataContext ctx = null)
    : base(env, ctx) { }
}
```

Listing 2: Implementation of the base class for each specific operation.

The resulting operation tree is traversed and executed by the environment. Each operation is defined in its own class. This allows operations to be used in more than one environment. Each environment (i.e. each machine) contains methods to create the operations it can execute. This architecture allows developers to build simulations with single components in a modular way. The idea is to separate simulation definition (i.e. what operations are available) and operation definition (i.e. what is done when executing an operation). In Figure 2 a sample machine is described by the `IBasicMachine` interface, which defines all operations, the context and the settings that can be used. Each operation (A, B, C) represents actions and their necessary dependencies/sub-operations. The `BasicMachine` class implements `IBasicMachine` and is responsible for creating operation instances. By extending from `BasicMachine`, each operation creation can be overridden and alternative operations that represent different features can be included.

4. MODEL IMPLEMENTATION

The real world machinery is transformed into a static, deterministic, discrete simulation model. All machine operations as well as their dependencies and sub-operations are implemented to provide a detailed model of the real world system. The operations are built with reusability in mind. It should be possible to easily create alternative

model implementations with different features by extending a basic version of the simulation and adding specialized operations.

A new BasicMachine derives from SimulationItem and implements the IBasicMachine interface. Listing 3 shows the BasicMachine class.

```
public class BasicMachine
: SimulationItem, IBasicMachine {
// ...
public override IEnumerable<SimSharp.Event>
CreateSimulationOperation() {
return new SimulationOperation<BasicMachine>
(this).ExecuteAll();
}
// ...
public virtual IEnumerable<SimSharp.Event>
MoveShuttleHome(
BasicComponentDataContext dataContext) {
return new MoveShuttleHome<BasicMachine>
(this, dataContext).ExecuteAll();
}
// ...
}
```

Listing 3: An excerpt of the BasicMachine class showing the creation of a MoveShuttleHome operation.

The interface specifies methods to create all Operations that are available in the basic implementation of the machine, e.g. the MoveShuttleHome operation that is shown in Listing 4.

```
public class MoveShuttleHome<TEnvironment>
: Operation<TEnvironment>
where TEnvironment : BasicMachine {
protected new
BasicComponentDataContext DataContext {
get { return
(BasicComponentDataContext)base.DataContext;
} }

public MoveShuttleHome(TEnvironment env,
BasicComponentDataContext ctx)
: base(env, ctx) { }

public override IEnumerable<Event> Dependencies {
get { yield return Env.Process(
Env.SwitchMagnetState(DataContext,
MagnetState.MagnetOff));
} }

public override IEnumerable<SimSharp.Event>
Execute() {
var shuttle = DataContext.Shuttle;

if (shuttle.Position.IsAlmost(0.0)) {
shuttle.Position = 0.0;
yield break;
}

yield return Env.Timeout(
Env.Settings.ShuttleMove.CalculateMovementTime(
shuttle.Position));
OperationInformation.Distance = shuttle.Position;
shuttle.Position = 0.0;
}
}
```

Listing 4: Implementation of the MoveShuttleHome operation that shows how dependencies and operation logic can be added.

All methods are implemented in BasicMachine and return an according instance of the required operation class. This creates two possibilities for extending the machine design. On the one hand it is possible to derive from BasicMachine and override particular methods to return new operations. Those extended operations should derive from the original operation and can introduce new dependencies, sub-operations and logic by overriding as required. On the other hand new interfaces can be added to a new class derived from BasicMachine which introduce new methods for creating new operations.

The operation sequence of a simulated environment is represented by the operation tree. This tree is constructed by the operations and their dependencies/sub-operations themselves. The entry point of the simulation is the Run method defined in the ISimulation interface. It creates a SimulationOperation (root operation) and generates the operation tree by calling ExecuteAll on that operation. The operation tree is traversed when the environment processes all resulting operations.

An improved machine could for example require additional steps before the shuttle is moved to its home position and therefore, an AdvancedMoveShuttleHome operation could be derived. The new operation would then be returned by the respective environment method and yield additional dependency operations.

5. OPTIMIZATION

The optimization focuses on the set-up time minimization of a machine that sequentially processes multiple jobs. Each job requires a certain set of tools that are stored in a storage area and must be set up in the machine's working area before processing can be started. The machine automatically loads and unloads required components.

A new HeuristicLab problem which is targeted to work with already existing evolutionary algorithms such as genetic algorithms is used to optimize the job sequence and the storage organization. The problem derives from HeuristicLab's BasicProblem, which makes it easy to implement the necessary evaluation and analysis logic without the need to create custom operators. The required problem data is stored in a TJOData instance. A TJOData object represents the scenario of the machine, i.e. available tools and jobs as well as the machine settings. This TJOData instance is then passed on to each solution evaluation. Solutions are represented by a multi-encoding that represents the identified optimization aspects. It consists of three permutations that are used in combination with the TJOData to construct the actual solution. The first permutation represents the order of jobs the machine has to process. Depending on the processing order of the jobs, the simulated execution time can increase or decrease. The other two permutations are used to create the storage layout of the machine. The way that components are sorted plays an important role when executing job after job. The indices of the tools are fixed within the TJOData of the problem instantiation. Each storage rack is assigned exactly one tool (or none). The actual storage

configuration can be created with the respective permutation by rearranging the tools by their index in the permutation. Figure 3 shows how the actual storage configuration is constructed by combining the respective permutation and the data that is available within the TJOData object. In this example tool 0 (T_0) was originally assigned to rack 0 (index 0 contains 0). The upper rack assignments permutation now dictates that in rack 0 tool 3 can be found (index 0 contains 3). If there are less tools (n) than racks (m), all numbers greater than n represent void tools that are used to create empty racks.

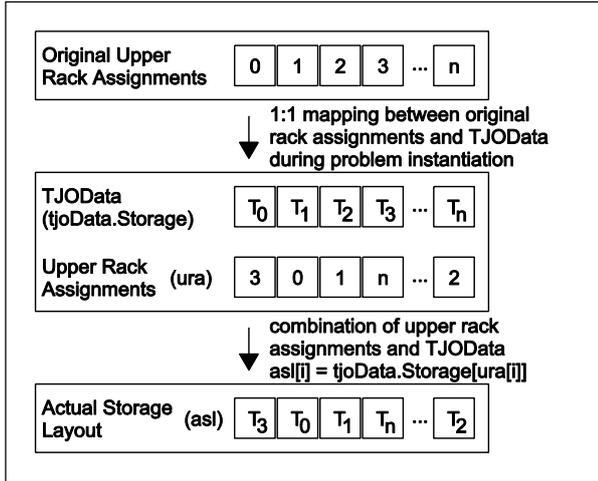


Figure 1: The procedure for creating the actual storage configuration.

For all experiments, a genetic algorithm with offspring selection (Affenzeller and Wagner 2005) is used. Selection scheme and replacement strategy are set to linear rank selection (LR) and 1-elitism, respectively. For all permutations, crossover and mutation operators have been selected carefully. For recombination, the OrderCrossover2 (OX2) (Davies 1985; Affenzeller et al. 2009) was chosen since it is known to preserve order rather than position within a permutation. For mutation, the InsertionManipulator (IM) was used.

Each time a solution has to be evaluated, a new simulation instance is created specifically for this evaluation to allow parallel execution of optimization algorithms. Therefore, efficient algorithm usage is assured. Quality is measured in total seconds of simulation time, i.e. the simulated time the machine took to load and unload tools for all jobs.

6. RESULTS

Three different problem configurations were tested. In the first configuration, five different jobs were defined. This scenario can be found in rather small manufacturers or manufacturers with low fluctuation in production cycles. The second configuration represents middle class manufacturers and contains ten different jobs together with a medium filled storage. This scenario should simulate a daily production cycles of medium sized manufacturers that are not restricted to just five different jobs. The last configuration features 20 different jobs and a

storage that contains a large variety of different tools. This scenario represents large manufacturers or manufacturers with a high fluctuation in production. In each problem configuration, only the least amount of tools and the machine's default settings were used. Each problem instance was tested within an experiment containing a batch run with 10 repetitions. The results of the best parameter combinations are shown in Table 1.

Table 1: Selected algorithm parameters and achieved results for each problem instance.

Parameter	Prob. 1	Prob. 2	Prob. 3
Population Size	1000	1000	1000
Max. Generations	100	100	100
Elites	1	1	1
Selection Operator	LR	LR	LR
Crossover Operator	OX2	OX2	OX2
Mutation Operator	IM	IM	IM
Mutation Prob.	5 %	5 %	5 %
Comp. Factor	1	1	1
Success Ratio	1	1	1
Max. Sel. Press.	100	200	200
Result			
Worst Quality	546.041	944.004	2629.342
Best Quality	476.568	832.955	2354.767
Avg. Best Quality	476.621	833.449	2356.019
Avg. Eval. Sol.	248700	611490	830640

The results shown in Table 1 were selected based on the number of runs that achieved the best found solution in each configuration and the number of evaluated solutions. The only difference between the configurations that is worth mentioning is the change in the number of evaluated solutions. The achieved quality values were quite similar in each algorithm setup. The simulated processing times for problem instance 1, 2 and 3 could be reduced by approximately 12.72 %, 11.76 % and 10.44 %, respectively.

7. CONCLUSION AND OUTLOOK

Despite the machine's current constraints, set-up time improvements of more than 10 % have been achieved. By extending the machine in its design, more optimization potential can be created and set-up times could be reduced even more. Currently, only one tool can be stored in one storage rack. This limits the number of different storage layouts and the number of tools that can be used. A redesign could e.g. allow multiple tools to be stored within one storage rack. Adding new unloading strategies would allow to move the tools in different ways, e.g. back to their original location or to other positions. This approach can be extended by using priority rules to select a specific rack for each tool that needs to be unloaded.

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. *Adaptive and Natural Computing Algorithms*, 218–221. Springer.
- Affenzeller, M., Kronberger, G., Winkler, S., Ionescu, M., Wagner, S., 2007. Heuristic Optimization Methods for the Tuning of Input Parameters of Simulation Models. *Proceedings of I3M 2007*, 278–283. October 4–6, Bergeggi, Italy.
- Affenzeller, M., Winkler, S., Wagner, S., Beham, A., 2009. *Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications*. Chapman & Hall/CRC. ISBN 978-1584886297.
- Beham, A., Pitzer, E., Wagner, S., Affenzeller, M., Altendorfer, K., Felberbauer, T., Bäck, M., 2012. Integration of Flexible Interfaces in Optimization Software Frameworks for Simulation-Based Optimization. *Companion Publication of the 2012 Genetic and Evolutionary Computation Conference, GECCO'12 Companion*, 125–132. July, Philadelphia, PA, USA.
- Beham, A., Kronberger, G., Karder, J., Kommenda, M., Scheibenpflug, A., Wagner, S., Affenzeller, M., 2014. Integrated Simulation and Optimization in HeuristicLab. *Proceedings of I3M 2014*, 418–423. September 10–12, Bordeaux, France.
- Bruzzzone, A., Longo, F., Nicoletti, L., Diaz, R., 2011. On the short period production planning in industrial plants: A real case study. *Proceedings of the 23rd European Modeling and Simulation Symposium EMSS 2011*, 782–791. September 12–14, Rome, Italy.
- Davis, L., 1985. Applying Adaptive Algorithms to Epistatic Domains. *Proceedings of the International Joint Conference on Artificial Intelligence*, 162–164. August, Los Angeles, CA, USA.
- Karder, J., Scheibenpflug, A., Wagner, S., Affenzeller, M., 2015. Optimizing Set-up Times using the HeuristicLab Optimization Environment. Accepted to be published in *Proceedings of the Computer Aided Systems Theory – Eurocast 2015*, February 8–13, Las Palmas de Gran Canaria, Spain.
- Matloff, N., 2008. Introduction to Discrete-Event Simulation and the SimPy Language. *Davis, CA. Dept of Computer Science. University of California at Davis*. Available from: <http://heather.cs.ucdavis.edu/~matloff/156/PLN/DESIntro.pdf> [accessed 13 July 2015]
- 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. *Advanced Methods and Applications in Computational Intelligence, Topics*

in Intelligent Engineering and Informatics Series, 197–261. Springer.

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 development 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.

ANDREAS BEHAM received his Master in computer science in 2007 from from the Johannes Kepler University Linz, Austria, and is a research associate at the Research Center Hagenberg. His research interests include metaheuristic methods applied to combinatorial and simulation-based problems. 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, Campus Hagenberg. 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, Campus Hagenberg.

HARDWARE LIBRARIES FOR ONLINE CONTROL OF INTERACTIVE SIMULATIONS

Josef Brozek^(a), Martin Jakes^(b)

^{(a),(b)}Department of software technologies, University of Pardubice, CZ

^(a)mail@jobro.cz, ^(b)jakesmar@gmail.com

ABSTRACT

The article presents possible ways of connecting hardware devices with computer simulation.

Today we often see interactive simulators that truly copy the modelled system. Such simulators require extensive know-how in terms of methods enabling data transfer between the running simulation core and the controlling device that is a mere electromechanical emulation of the real system.

Our libraries are largely applicable in monolithic simulations, distributed simulations, in drive-simulators design and in simulation-based decision making. Their simple application enables each reader of this article effectively employ different input-output devices for their own simulators.

An integral part of the article is an introduction of a case study aiming at consistent verification of methods used for the libraries development and for verification of the libraries as such.

Keywords: HLA, simulator, simulation, Java FX, Java, HLA-VA, Rawsberry, hardware, microcontroller, army simulator, car simulator.

1. STRUCTURE OF THE ARTICLE

The first part of the article defines the problem by introducing signals to be processed. Readers may feel that this chapter is rather abstract due to the lack of specification of input-output devices to be used. However, the article is structured in a way that introduces particular devices in chapters four and five.

Chapter four deal with the software layer. Selected programming languages are introduced, advantages and disadvantages presented in terms of particular hardware component controls and an introduction to a more complex distributed simulation concept is presented. HLA was selected for distributed simulation: this chapter gives reasons why HLA and how its properties were used in individual libraries design. Readers may find out more about the ways of use of our library in the environment of implicit and other programming languages.

In the course of development, individual solutions were designed, where the task was to connect the input-output devices with the simulator; therefore it was inevitable to form simple application (or rather application-hardware) prototypes. Chapter five describes individual prototypes with the aim to extend the basic reader's notion of possible ways our library can be used.

Besides individual prototypes, we designed an extensive case study which is presented in Chapter six. The sixth chapter summarizes the whole solution and presents the premises for possible future development.

2. SITUATION SEARCH

In current situation exists models, which using different hardware input devices for control of simulators. But publications about it, focuses primarily on the major topic of simulation (submarine, train) and do not help community to understand, how connect hardware drivers into their simulations.

We used space between current simulation teams focus, and made very simple libraries, which should enable of simple connection hardware devices with HLA (or different) simulator.

Then, with our scope there is done relatively original work, which can be used by many developer teams for their works. Simply put, our work is professionally profiled on the border between software engineering and simulation architecture. It publishes real libraries, which after installation transfers data directly from I / O devices into the simulator (if you want to use in the HLA, you still need to adjust OMT).

3. PROBLEM DEFINITION

Recently designed interactive simulators show very good results in real visualization. Besides the standard computer simulators, drive-simulators or emulators have been often used. The difference between a high quality simulation and a high quality emulator rests in the fact that the whole simulation runs on a computer, including optional user interactions (i.e. parameterization of simulation calculation under the running simulation). The emulator carries out the

calculations through a computer; however, their input-output devices are a true imitation of real input-output devices. An automobile emulator thus has a real steering wheel, pedals and gear shift and is explicitly controlled by these devices. An automobile simulator may have a very sophisticated physics and dynamics of the drive, etc.; however, as it is only controlled by a keyboard, it is a mere simulator.

Besides the above mentioned, there is another specific simulation type called “live simulation” which is mostly dealt with in the military sector. This highly sophisticated simulator type enables VDU and vibration units (and other components) wearing users real engagement in virtual reality. At all events, our libraries have been optimized for a live simulation use.

Simulation, live simulation and emulators can be simply named as simulators. Simulators are in principle used for two main purposes: decision-making and training. The existence of input-output devices corresponding to the true image of professionally trained personnel is of the key importance for the success of simulators. You can access [Sinclair, 2013] for more information.

3.1. Types of input-output values

We can briefly characterize the types of values that a simulator is expected to work with. They are divided according to the direction, type and the processing demand. Before focusing on individual types, it is essential to determine the actual system to be observed. We can explain the preceding proposition on an example: a simple steering wheel can be approached from several different angles.

The basic view is a very high abstraction defining the steering wheel as a device in the state “steer angle”. However, we would fail in a real system since we must take in consideration also the steering dynamics, i.e. the monitored system except the steer angle must be extended by e.g. monitoring of dynamic change, because the steering speed can be of the key importance in certain parts of the simulation.

If we extend the view, the same steering wheel can be simultaneously used as an output device – the simulator can automatically manipulate such motions as the vehicle yank in a road wash-out, etc.

3.2. Characterization by direction

The essential data characterizing feature is direction, i.e. the determination whether input or output data are in question. Many devices at real simulators are input and output elements at the same time.

The following are simple examples:

- Input device: a button for the simulated industrial operation start up
- Output device: an indicator displaying danger for train operation on railway
- Combined device: a steering wheel taking the input data from the user; however, at the same time enables an

output in the form of variable resistance in turning in different terrain or according to the selected simulated vehicle.

3.3. Data types

The basic types of data are the following: continuous, digital and logic. Continuous data in driving trainers are always transferred into digital data by means of a/d converters. It is important to mention that minimum deformation occurs because even the most basic converter that transfers continuous data into a 1-byte value achieves acceptable accuracy. In our example, the continuous value is transferred to one of 256 values – a relative deviation of any step is 0.4%, which is much a higher accuracy than a person can achieve with his/her physical engagement.

Digital value processing in information technology environment is very easy and simple to use in abstraction. It is convenient (already in the problem analysis) to select a specific data type, i.e. logical (Boolean) variables that might reach only the “true – false” values. Logical value processing not only is fastest but also easy to integrate in input-output devices of the lowest level.

3.4. Processing demands

The data must be also characterized in terms of their processing demands due to close relation to quality validation and verification. In the event that an input-output device is formed on the lowest level and all data administered by a software architect, then the processing demand, as well as testing, are very high.

In addition, there exists a real risk of a blind point in the testing (i.e. situation where everything appears to be functioning until a very specific situation occurs in which the solution is non-functional / inaccurate).

A suitable way of reducing the processing demands is the use of already existing solutions and software libraries, mainly of a commercial nature. The best case of such commercial solutions is affordable price focusing on the segment highly similar to that of training devices (especially computer games segment).“

4. OWN SOFTWARE

It is inevitable to concentrate on the next solution level, i.e. the software level. Despite the article focuses on a very low level of simulation (in the following order: simulation theory – practical solution – architecture of simulators – libraries for simulators) and the hardware aspect is dealt with as a mere overview, the principles used in software are described with more details. The reason for this is the fact that another potential user should know when the solution can be really used.

4.1. Software description

Major part of this work is software-based, then, there is need define parts of used software.

4.1.1. HLA-VA

Java FX was selected as the basic programming language because it provides a quality support for the scene animation and the code can be performed from Java programming language. This aspect is of a high importance because the HLA-VA framework in the basic scheme is implemented directly for Java and will be used in our study. Since the objective is to employ also distributed simulation, we can use the fact that pRTI provides a library for Java language.

To perform the code, a virtual machine is required (JVM); however, this is a standard approach in modern languages (most other languages use .NET virtual machine). Languages of C/C++ family are an exception: they are in fact inapplicable for high quality visualization without extensions and frameworks they use through virtual libraries. More information at Brozek, Onggo and Kavicka (2004).

For further versions and quality 3D graphic supported views can be used options of OpenGL, including its approaches (non-object programming) and specific programming languages.

4.1.2. Use of drivers

Certain hardware types have drivers installed by the manufacturer, which can be successfully used. The programming principle is described in Figure 2 The application logic itself can lean on the driver interface, therefore in the event of hardware replacement for a device whose driver has a corresponding interface (e.g. joystick), the solution will keep functioning correctly. For more information see e.g Henninger, Cutts, Loper (2014)

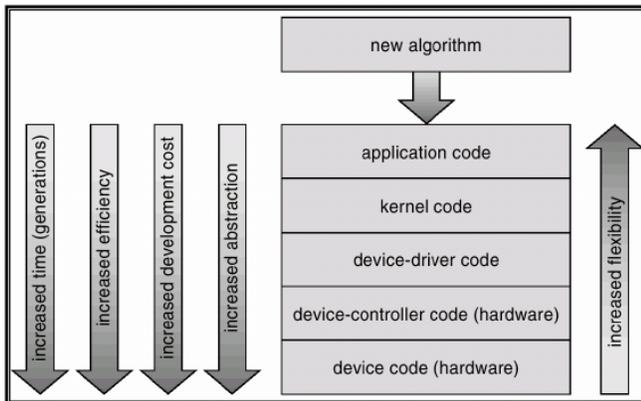


Figure 1: Programming levels [5]

4.1.3. Driver-free solution

Driver-free programming can be approached in several different ways, where the most effective is formation of a generic scheme that will later be capable of parametrization. All of principles are low-level and are under scope of this article. But principle which demonstrates that each individual output of the device can be set up as input - output and, at the same time, data

collection can be characterized (upon request / after a change, etc.) from Huddleston (2007).

4.1.4. Interface of our solution

Interface of individual libraries is designed in an open way, but intuitively.

For the connection to actual simulator, the designer is only required to connect the hardware device, link the library and implement interface methods for the appropriate device.

The input -output interface within the library is provided independently. Should the application require a combined device (i.e. both input and output), it is essential to implement two interfaces.

Complications may occur in the event that connection of several input or output devices of the same type is needed (with the same drivers) in one application (e.g. connection of two steering wheels). This solution is currently not applicable within our libraries, because the construction does not allow for the use of parallel devices (the same solution used by most commercial software types).

4.2. Use of our solution in SW

In the event that JAVA language is used, the link of drivers coded in the same language is very simple for desktop, website and mobile applications – i.e. mobile device solutions.

Our libraries can be used in other programming languages; however, the use is slightly more complicated (although simpler than implementation from zero).

5. PROTOTYPES

The existence of input devices, their drivers and control libraries are of abstract nature and very difficult to grasp. An integral part of individual libraries formation is, among others, formation of simple software prototypes whose objective was a thorough testing of the appropriate libraries. Validation and verification of a particular library was carried out simultaneously.

5.1. Prototype 1: Automobile

At first, a simple prototype was designed. A set of steering wheels and pedals was used that was originally designed for computer games. This solution facilitated successful validation and verification of a library for PC steering wheels connected by a USB, which is very important especially in regard to the fact that it is a wide spectre of applicable devices that makes the library very useful. The library is limited to the use of basic functions (steering wheel turn, pedal operations, gears changing and six additional buttons at the most). An advantage of the library is implementation of the feedback interface, which transforms a simple input device a combined one.

The prototype was used for the needs of our case study (see Chapter 6).

5.2. Prototype 2: Heavy gun sighting system

Library for driving a joystick was tested on an example of a machine gun turret control. As in the previous example, the connection was made via USB. The same prototype was run on a tablet, which tested the libraries on another platform. The prototype itself was formed with the stress on its integration in a more extensive case study, which was performed later (see Chapter 6).

5.3. Prototype 3: Connecting online simulation, sensor network

Besides the distributed interactive simulation, our solution also aims at online simulations. We decided to carry out two case studies within the prototypes, where the first one is more extensive. Generally speaking, sensor network online reading is concerned, where actual input devices were used (buttons and levers) connected to individual pins of a Raspberry PI. The prototype (for complex connection see Sinclair (2001) and Richardson (2012)) serves for demonstration of the approach trivialities in the event that our libraries are used.

An input device can be formed only by means of a PI Raspberry, our library and almost any connective devices (even by two wires connected to the Raspberry). This prototype presents a key factor for the formation of any type of input-output devices. USB feeding (optionally from a mobile phone power bank) facilitates the creation of a super-portable computer that can be used in a live simulation since its weight (including power bank) does not exceed 400g.

The most complex solution made within the prototype was sensor network. Its objective was to emulate the security device for operation on railway. The total of 10 sensors with logical value can be evaluated and the data immediately transferred to the simulation. This solution can be used especially in decision-making simulations.

5.4. Prototype 4: Connection of online simulation, status indicators

The last prototype (chronologically developed as the first one) is a solution that employs low-level approaches. Microcontroller AtMega was used for data processing and input devices and mainly output devices were connected directly. Although it is the most economical solution, its implementation complexity outweighed all the others. In addition, the libraries written for the needs of this solution were eventually eliminated because their function required a specific sensor type. The competitive price cannot compensate the lack of comfort in further development. Although the prototype with a microcontroller can be a functioning solution, it is currently considered to bring more drawbacks than benefits. The article lists this solution rather for the purpose of complexity; the recommended option is clearly the use of Raspberry PI.

6. CASE STUDY

Simple prototypes cannot fully demonstrate the suitability of solutions, therefore a more extensive case study was carried out.

For more information about construction of HLA simulation models see Kuhl, Dahmann, Weatherly (2000), Manling (1999) or Rabelo and col. (2014).

6.1. Demonstration

Model for the case study was a military vehicle displayed in Figure 3. The vehicle has two important parts: the first one is driver station, the second is shooter station – the latter usually sits on the front- passenger seat and the heavy gun located on the armoured vehicle roof is controlled electronically.



Figure 2: Single car from case study

Figure 3 shows the vehicle drawing and its division into individual simulators. The first – driver simulator – contains a standard steering wheel and pedals. The second simulator is designed for the shooter and its position is emulated through joystick. Distributed interactive simulation thus demonstrates two persons in one vehicle, while each person can have his/her own point of view. One person controls the vehicle and the other manipulates the gun. The persons are independent of each other (but both depend on the vehicle). The third federate (we are unable to say that it is the simulator as such) takes care of the environment (maps) and relating calculations. The motion of the vehicle itself is not solved until the third federate. Only two computers are required for the run of the whole simulator (one for each user). Execution of the third federate in view of the simulation study is not important.

6.2. Selection of technologies

For the above indicated reasons, HLA was used for distribution of the solution. Since the remote control libraries and also HLA-VA framework supply a sufficient software solution for the formation of individual federates, the most important part of the development was visualization. Open GL was selected as visualization technology. It is essential to say that in the selection of technology, the application is not expected to have sophisticated graphics as that of PC games. Graphic parameters must comply with the requirements of distributed interactive simulation (it is not a distributed

interactive simulation in terms of the DIS standard, i.e. IEEE1278). For more information about accessible technologies see for DIS IEEE1278, for HLA IEEE1516:2010, Fujimoto (2000) or SISO (2001).

6.3. Case study architecture

The case study is relatively simple: a simple system of permitted and enabled operations is placed on the HLA-VA framework for both of the federates. The federation diagram is relatively simple due to the use of a large number of libraries; it is demonstrated in Figure 4.

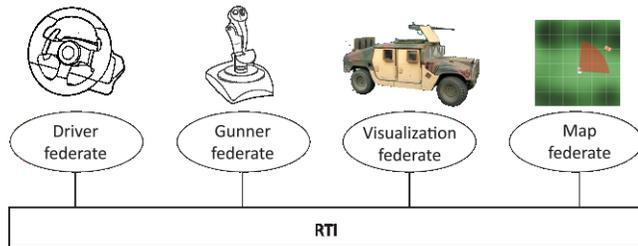


Figure 3: Real federation architecture

Few parts as Visualization federate can be run on mobile devices, see Brozek, Jakes, Gago (2014) for more information about this principle.

There are differences in federation architecture for direct method (when hardware control devices is connected to computer), and spited method (when hardware control device is connected to simpler device without own monitor). This differences are at figure 5.

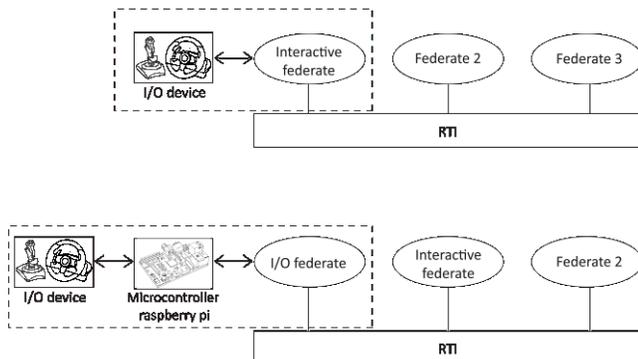


Figure 4: Two principles connection HW to HLA simulation

6.4. Testing method

Two users have been selected for testing, while each of them had the task to take up one position in the vehicle. One target was determined within the environment to be eliminated by users. All testing was carried out in relation to this target (a shooter's target). For example, in one of the situations the driver's task was to rotate around the target while the shooter's task was to keep the target in the view finder of the gun (this way tested the coordination of both the federates against each other).

Since the testing used the pRTI solution by Pitch technologies, an integral part of RTI is a high quality solution for monitoring the whole running federation data. This method enables us to monitor actual synchronization as well as the values of most of the attributes. Therefore validation and verification could be performed from an independent point through the third party's software. The running application in testing and the view of the monitoring software are displayed in Figures 6, 7 and 8.



Figure 5: Drivers federate

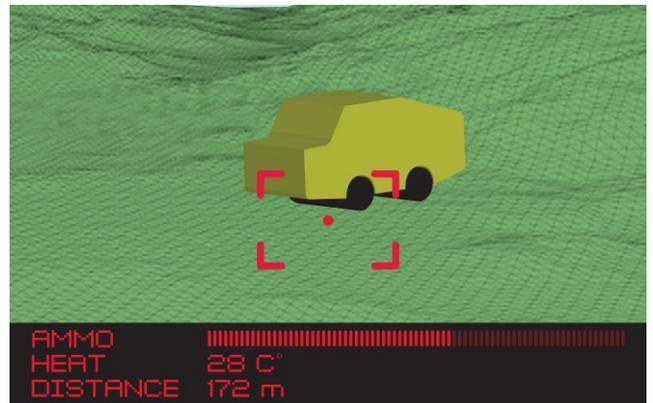


Figure 6: Gunners federate

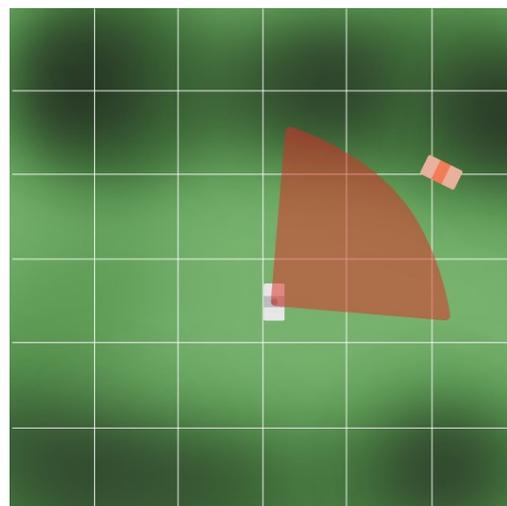


Figure 7: Visualization federate (two cars mode)

7. CONCLUSION

The article reviewed the problem of connecting input-output devices to a computer simulation and presented new solutions based on the reusability principle. Encapsulating the solution in libraries and release of this method to the simulation community results in the fact that every person within their commercial, academic or educational activity can develop simulators that - besides keyboard and mouse - can in reality use any input-output device.

Libraries as such open the way to building live simulators or distributed simulation models connected into an online simulation.

7.1. Potential for future development

The research was finalized and the designed libraries were tested. In view of this fact, the development has been completed. On the other hand, results published in this article present only one of partial tasks to be carried out to develop a high quality simulator for railway staff training. Despite the development of libraries was finished, further steps within our workplace consist in their application.

ACKNOWLEDGMENTS

Expenses related with publication of article were funded from SGS announced by University of Pardubice at 2015, in section Application section software technology.

REFERENCES

- Brozek J., Jakes M., Gago L. Using tablets in distributed simulation. Pardubice, 2014. ISBN 978-889799932-4. Conference Paper. EMSS 2014 Proceedings, University of Bordeaux.
- Brozek J., Onggo B.S., Kavicka A. High level architecture virtual assistant framework. Pardubice, 2014. ISBN 978-889799932-4. Conference Paper. EMSS 2014 Proceedings, University of Bordeaux.
- Fujimoto R.M., c2000, Parallel and distributed simulation systems. New York; John Wiley & Sons. ISBN 04-711-8383-0.
- Henninger A.E., Cutts D., Loper M. Live Virtual Constructive Architecture Roadmap (LVCAR) Final Report, Institute for Defense Analysis, Sept, 2014, Available from:
<http://www.msco.mil/files/MSCO%20Online%20Library/LVCAR%20-%201%20of%205%20-%20Final%20Report%20-%2020090814.pdf>
- Kuhl F., Dahmann J., Weatherly R., Creating Computer Simulation Systems: An Introduction to the High Level Architecture, c2000, Upper Saddle River, NJ; Prentice Hall PTR. ISBN 01-302-2511-8.
- Manlig, F., 1999. Computer simulation of discrete events. Available from:
<http://www2.humusoft.cz/www/archived/pub/witness/9910/manlig.htm> [accessed 15 July 2014].
- Rabelo, L., Sala-Diakanda S., Pastrana J., Marin M., Bhide S., Joledo O., Bardina J., 2013. Simulation Modeling of Space Missions Using the High Level Architecture. Available from:
<http://www.hindawi.com/journals/mse/2013/967483/> [accessed 15 July 2014].
- Richardson M., Wallace S.P. Getting started with Raspberry Pi. 1st ed. Sebastopol, CA: O'Reilly Media, 2012, xiii, 161 p. Make: projects. ISBN 14-493-4421-6.
- Sinclair I. Sensors and transducers. 3rd ed. Oxford: Newnes, 2001, xiv, 306 s. ISBN 07-506-4932-1.
- The institute of electrical and electronics engineers, Inc, 2010, IEEE1516:2010: IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-Framework and Rules. New York; IEEE. ISBN 978-0-7381-6251-5.
- The institute of electrical and electronics engineers, Inc, 2010, IEEE1516:2010: IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-Object Model Template (OMT) Specifications. New York; IEEE. 2010.ISBN 978-0-7381-6249-2.
- The institute of electrical and electronics engineers, Inc, 2010, IEEE1516:2010: IEEE Standard for Modeling and Simulation (M&S) High Level Architecture (HLA)-Federate Interface Specification. New York;IEEE. ISBN 978-0-7381-6247-8.
- The institute of electrical and electronics engineers, Inc, 2010, IEEE standard for distributed interactive simulation application protocols. New York: Institute of Electrical and Electronics Engineers, 2002. ISBN 07-381-0992-4.
- The simulation interoperability standards organization, Independent Throughput and Latency Benchmarking for the Evaluation of RTI Implementations, 2001, The Simulation Interoperability Standards Organization. Fall. DOI: SISO-01F-SIW-033.

A DYNAMIC MULTICOMMODITY NETWORK FLOW PROBLEM FOR LOGISTICS NETWORKS

Sebastian Raggel^(a), Judith Fechter^(b,d), Andreas Beham^(c,d)

^{(a)(b)(c)} Heuristic and Evolutionary Algorithms Laboratory
School of Informatics, Communication, and Media
University of Applied Sciences Upper Austria,
Softwarepark 11, 4232 Hagenberg, Austria

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

^(a)sebastian.raggel@fh-hagenberg.at, ^(b)judith.fechter@fh-hagenberg.at, ^(c)andreas.beham@fh-hagenberg.at

ABSTRACT

We present a dynamic multicommodity minimum cost network flow problem with storage at the nodes and optimal supply for modeling operations within a logistics network.

The model can be used to evaluate critical business decisions such as the amount of required resources for storage transportation as well as optimal supply policies.

A generator for test instances was written in order to evaluate the performance of the different solution strategies. Using a number of differently sized randomly generated problem instances we compare the execution time and the memory demand of two methods for solving the problem. The first is solving the whole problem formulation directly using general purpose linear programming solvers implemented in IBM Ilog CPLEX. In the second approach we attempt to split the model into two parts and link them together in an optimization network. We analyze the quality of the link and propose possibilities to improve the two step approach through input parameter variation

Keywords: Dynamic networks, Minimum cost flow problem, Multicommodity flows, Dynamic Lot Sizing Problem

1. INTRODUCTION

Network flow problems have been used to model a wide variety of systems and therefore network flow literature distinguishes a large number of problems and variants. A good overview over problems as well as algorithms can be found in (Ahuja et al. 1993, Kotnyec 2003).

When modeling a logistics network there will likely be more than one commodity that has to be transported across the network and the transport capacities must be shared between all the commodities. This kind of problem is known in the literature as a multicommodity

network flow problem as opposed to the classical single-commodity network flow problem.

Another major distinction is between static and dynamic networks. Dynamic network flows were first introduced by (Ford and Fulkerson 1958). They described the problem of maximum flows with transport time on the arcs. Their solution was to transform the dynamic network into a so called time-expanded network then solve this problem using classical static methods and finally transform the solution back into the original dynamic form. It has been shown that the same approach is also valid for dynamic minimum cost multicommodity flows (Fonoberova 2010). In this case the problem that has to be solved on the time-extended network is a minimum-cost multicommodity network flow problem (Assad 1978, McBride 1998, Castro and Nabona 1996, Karakostas 2008).

Time is an essential factor in logistics networks because of a number of reasons. While the classical static problem only considers the transport capacities and costs on the arcs the dynamic problem can model storage capacity as well as costs at the nodes which are equally important. In order to make meaningful statements about the network the model must take into account that supplies and demands that change over time. Supply policies that are implemented in many companies, e.g. using stock reorder and order-up-to levels govern the time and amount of items that have to be purchased. These policies, if parameterized wrongly, put a large burden on the efficiency of the network. In the presented case, the model considers both the movement as well as the acquisition of commodities.

The network flow problem presented in this paper is a dynamic multicommodity minimum cost network flow problem with storage at the nodes and optimal supply. The model also allows for multiple modes of transports

as well as multiple sinks and sources for each commodity.

2. OPTIMIZATION MODEL

The goal of the model is to describe operations in supply chains and optimize the associated costs arising out of the acquisition, storage, handling, and transportation of commodities. In general, a logistic network can be described as a set of transport modes M and a graph $G = (V, E)$ with nodes (V) and edges (E). An edge (u, v, m) with $u, v \in V$ and $m \in M$ describes a link between two nodes making use of a certain transport mode. The notation $E^+(u)$ and $E^-(u)$ is used to describe the set of all the edges going to and coming from a certain node u . The logistic network is used to execute the flow of a set of commodities C . The operations are described as a series of discrete time steps T . In this model all commodities that are to be purchased emerge in a virtual node “ q ” from which it must be distributed. This option was chosen to let the model decide, taking into account logistic cost and capacities, the delivery node for a certain purchase.

Figure 1 shows a small exemplary network with two large storage nodes and three smaller sink nodes. The supply appears in the virtual source node and because there is no storage capacity in this node it has to be delivered to one of the two warehouses immediately. When there is a demand at a store the network transports the commodities from the warehouses to this store.

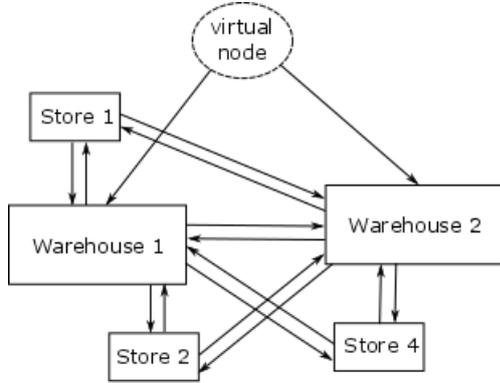


Figure 1 Example network

The number of capacity units stored in node u at any time is a variable called $f_{uct} \in \mathbb{R}^+$. It is constrained by a lower and upper limit namely a_u and b_u . The cost of storing one capacity unit at node u for one time step is called α_u .

Each mode m has a certain maximum capacity z_m . The capacity is to be distributed over all edges that specify that mode. Not all of the capacity needs to be assigned. In fact it will be a goal to minimize the employed capacities. So for each edge and each time step the decision has to be made how much capacity is to be assigned. That is what the continuous variable $x_{ect} \in \mathbb{R}^+$ is used for. The modes are also associated with a certain cost per unit of capacity transported. These

transport costs together with the handling costs that occur in the source and target nodes of an edge are called β_e the total cost of transporting one unit of any commodity over edge e .

The amount of each commodity c initially stored in each node u is given by $i_{uc} \in \mathbb{R}^+$. The demand $d_{uct} \in \mathbb{R}^+$ is given for each node, commodity and time step. When to buy how much of which commodity is to be determined by the model and described by the variable $s_{ct} \in \mathbb{R}^+$.

The price of one unit of commodity c is given as γ_c . To be able to account for fixed cost per order there needs to be an additional variable $o_{ct} \in \{0, 1\}$ which indicates whether an order was placed. The value of the fixed costs per order is given by δ_c .

The factor ε_c is used to convert between pieces of a commodity which is the unit used by demand, supply and initial stock and transport units which are used to describe how much is stored or transported

The goal is to minimize the sum of all acquisition, storage, handling, and transportation costs:

$$\sum_c \sum_t \left(\sum_e x_{ect} \beta_e + \sum_u f_{uct} \alpha_u + s_{ct} \gamma_c + o_{ct} \delta_c \right)$$

subject to:

$$+ \sum_{e \in E^+(u)} x_{ect} - \sum_{e \in E^-(u)} x_{ect} = f_{uct} \quad \forall u, c, t \quad (1)$$

$$\sum_{e \in E^-(q)} x_{ect} = s_{ct} \varepsilon_c \quad \forall c, t \quad (2)$$

$$\sum_c \sum_{e \in E_m} x_{ect} \leq z_m \quad \forall m, t \quad (3)$$

$$\sum_c \sum_{e \in E^+(u)} x_{ect} \leq p_u \quad \forall u, t \quad (4)$$

$$a_u \leq \sum_c f_{uct} \leq b_u \quad \forall u, c, t \quad (5)$$

$$s_{ct} \varepsilon_c \leq o_{ct} M \quad \forall c, t \quad (6)$$

$$f_{uc0} = i_{uc} \varepsilon_c \quad \forall u, c \quad (7)$$

Constraint (1) guarantees that commodities do not simply disappear from the network by requiring that the storage capacity used at each node and time step for each commodity depends on the previous time step as well as the inflow, outflow and demand at that time step. The supply is not mentioned in the first constraint instead there is constraint (2) which states that the sum of all the flows out of the virtual node named q in the model must equal the supply for each commodity and time step. This is useful because otherwise the supply variable would need another index for the node where it has to be delivered which would increase the complexity of the model considerably. Additionally a

new constraint equivalent to (4) would be needed to limit the processing rate for the supply.

The constraints (3, 4 and 5) all limit the amount of commodities both stored in as well as transported between nodes. The next constraint (6) guarantees that there is an order whenever the supply is greater than zero by requiring the supply to be smaller than the order times the maximal possible order size. The maximum possible order size M is defined to be the sum of the processing rates of all nodes connected to node q . The final constraint deals with the initially stored commodities.

3. SPLITTING THE MODEL

As shown in Table 1 the effort required to solve this model rises very rapidly. In order to overcome this problem it is possible to split the model into two models, one model that makes decision on the acquisition of commodities, i.e. a purchasing problem and another model that decides on the movement of commodities, i.e. a network flow model. The purchasing model ignores the network aspects but decides when to buy how much of which commodity. Figure 2 illustrates the two different approaches. On the left is the two step approach described above and on the right is the combined model described in Section 2.

3.1. Purchasing Model

The network flow model gets the supplies calculated by the other model as input and optimizes the transportation and storage in the network. The purchasing model has the goal to minimize the storage costs and the purchasing costs:

$$\sum_c \sum_t (f_{ct} \alpha + s_{ct} \gamma_c + o_{ct} \delta_c)$$

subject to:

$$f_{ct-1} - d_{ct} \varepsilon_c + s_{ct} \varepsilon_c = f_{ct} \quad \forall c, t \quad (1)$$

$$\sum_c s_{ct} \varepsilon_c \leq M \quad \forall t \quad (2)$$

$$a \leq \sum_c f_{ct} \leq b \quad \forall t \quad (3)$$

$$s_{ct} \varepsilon_c \leq o_{ct} M \quad \forall c, t \quad (4)$$

$$f_{c0} = i_c \quad \forall c \quad (5)$$

The variables and constants have the same meaning but because the network aspect is ignored the variables that originally contained a value per node are now summed over all nodes. The only constraint that does not have a direct equivalent in the combined model is constraint (2). It limits the supply at each time step to the combined processing capacity of all nodes connected to the virtual source node q . This is not needed in the combined model because constraint (4) includes this.

3.2. Network flow model

The second model is very similar to the combined model but since s_{ct} and o_{ct} are now constant instead of

variable the purchasing costs in the objective also become a constant. The objective function that needs to be minimized then looks like this:

$$\sum_c \sum_t \left(\sum_e x_{ect} \beta_e + \sum_u f_{uct} \alpha_u \right) + C$$

Where C stands for the total purchasing costs as defined by the supply matrix s_{ct} the commodity prices γ_c and the fixed cost per order δ_c . All the constraints of the original model remain unchanged except for constraint (6) which can be removed entirely.

This new model even though it looks very similar is much easier to solve because there are a lot less variables. Additionally are all of the remaining variables continuous which means that the problem is no longer a MILP but is now simply a Linear Program.

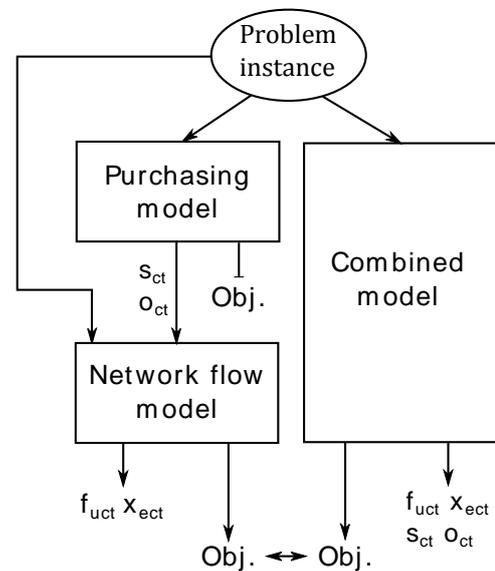


Figure 2 Comparing the two approaches

4. EXPERIMENTS

A generator for random problem instances was written. The problem instances model a retailer with several stores, warehouses, commodities and modes of transport. The warehouses are the sources where commodities are delivered to by the suppliers and the stores are sinks where the commodities sold to the customers disappear from the network. Each problem instance is completely specified by the number of stores, warehouses, commodities days and a seed value for the random number generator. A problem instance is generated using the following steps:

- Store and warehouse nodes are generated and the storage capacities are randomly chosen.
- Transport modes are generated and the nodes are connected by edges.
- Commodities are generated along with prices and conversion factors.

- The nodes are filled with an initial stock of commodities.
- Demands are generated for each store proportional to the stores storage.
- The capacity of the transport modes and the processing capacity of the nodes are chosen so that a feasible solution exists.

4.1. Comparing the two approaches

The first important question is how difficult is it to solve each of the models presented earlier. The models were set up like illustrated in Figure 2 then the time and memory consumption needed to obtain the solution using the two different approaches was recorded for problem instances of varying size.

As a next step the solution quality was compared between the two alternative ways of solving the problem. The idea is that if the solution is good enough it is not even necessary to run the bulky combined model. Finally it was examined if the solution of the two step approach could be improved by varying parameters of the purchasing model used as a first step. If this is the case it should be possible to run an algorithm that solves the problem by repeatedly solving the optimization network and varying these parameters.

5. RESULTS

All three models were implemented and tested using IBM ILOG CPLEX (IBM 2015). All tests were run on a Dell Latitude E6540 with an Intel i7 4810MQ CPU @2.80GHz and 16 GB RAM running Windows 7.

Table 1 Time and memory consumption for differently problem instances

C	T	t_{sep}	m_{sep}	t_{com}	m_{net}
10	10	0.41	15.61	0.91	87.61
10	20	0.66	31.09	511.96	290.78
10	30	1.12	28.93	111.29	99.98
100	10	5.02	105.59	404.65	529.21
100	20	38.38	200.98	>12h	
100	30	126.74	281.84	>12h	

Table 1 compares the time and memory needed to find the solution using the two sub models with the integrated model. The column labeled C and T shows the number of commodities and time steps in the problem instance. The size of the logistics network was for all tests set to ten nodes. The values under t_{sep} and t_{com} are the total time in seconds, it took to find the solution using the two separate models and the combined model respectively. Finally m stands for the peak memory consumption in megabytes.

Solving both of the separate models is as expected much faster. For tiny problem instances the runtimes are

comparable, but as the problem instances grow larger is the differences are so big that the integrated model does not even find a single solution candidate before the solution of the two models is found. Additionally the solving time of the combined model varies greatly depending on the problem instance.

5.1. Comparing the solution quality

As mentioned before is the two step solution likely to be worse than the one from the combined model. To find out how much worse the objective values of the combined model and the two step approach were compared while varying problem parameters. The problem instances used in this test all have 3 warehouses 7 stores 10 commodities and 30 days and fixed cost per order δ_c was varied between 0 and 100 with step size 10 and the storage cost multiplier α'_u was varied between 0 and 2 with step size 0.2. Each problem instance was solved to optimality with both approaches and the gap between the optimal solution and the approximation is shown in Figure 2. It can be seen that using two separate models can provide solutions that are close to the optimum. In five cases the solution of the optimization network was exactly the same as the one of the combined model. Considering how much faster it is to solve the two models than to solve the integrated model this is quite remarkable.

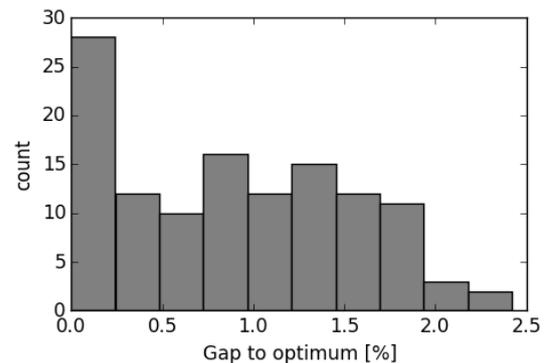


Figure 3 Differences in solution quality between the integrated and the two step model.

5.2. Parameter Variation

For the problem instance with the parameters $\delta_c = 50$ and $\alpha'_u = 1$ the objective value of the network flow model was 1.163% bigger than that of the combined model. The goal of the next test is to determine if a better solution can be found if the parameters of the purchasing model are changed. So the δ_c was again varied between 0 and 100 this time using a step size of 5 and α'_u was varied between 0 and 2 with a step size of 0.1. The purchasing model was parametrized with all the combinations and the resulting supply matrix s_{ct} was passed on to the network flow model. This way we obtain a number of different solutions which can be compared to the solution to the integrated model that was calculated for the previous test.

Figure 3 shows the gap in percent between the optimal and the approximated solution for all the combinations of parameters the purchasing model was set to. While the gap was 1.163% using the original parameters the lowest gap found in this case was 0.119% and was found at $\delta_c = 95$ and $\alpha'_u = 2$. With this parameter grid variation of input parameters we can thus obtain much better results while still having a model that can be solved in reasonable time and close to optimality.

Additionally, Figure 4 shows a rough sketch of the fitness landscape of the optimization problem if we would use a meta-solver to vary these input parameters. The landscape is not as nice to optimize as we would wish as the best solutions can be obtained with only a very small range of parameters that are surrounded by much worse solutions.

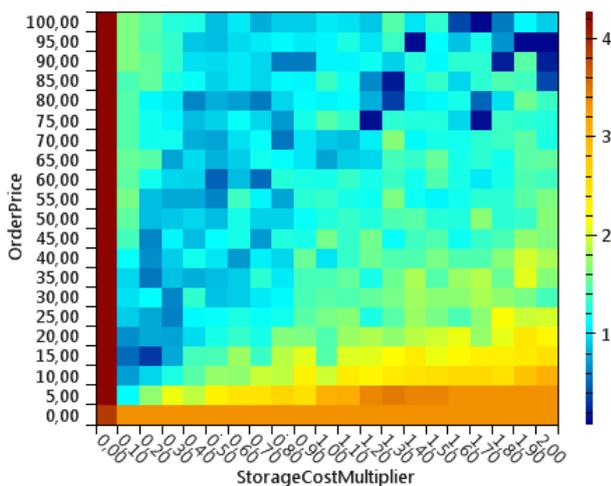


Figure 4 Results of the parameter variation for one instance

6. MODEL APPLICATION

The model describes relevant costs arising due to logistics operations such as handling costs when accepting or commissioning items to or from a warehouse, costs for storing an item for a certain number of time steps and costs that arise out of purchase and transportation. The optimization decides on the transportation and thus determines all movements of commodities in the network. The constraints enforce that demand has to be supplied in time and that commodities that enter the network are being moved should the maximum storage capacity be reached.

One application of the model that is interesting for the real-world is to act as a scenario evaluator for logistics operations. The inputs to this model, i.e. the nodes, edges, when and where items appear and disappear, capacities, cost factors, etc. may be subject to a master level optimization algorithm to which the optimization of this model may appear as an evaluation function. The scenario evaluation aspect is useful in critical business situations such as when new warehouses are to be opened, logistics capacities are expanded or reduced, or

also to evaluate environmental effects such as a change in demand. The model can then be used to evaluate the effects on the network and determines costs and capacities.

7. DISCUSSION

Using the two separate models is clearly not equivalent to solving the integrated problem, because in this case the layout of the logistics network does not influence the purchasing decision made by the first model. The network flow model only has to find the best solution for one given supply vector while the combined model optimizes both the flow through the network as well as the supply. So the two step approach is likely to deliver a solution that is suboptimal. What makes the two step solution interesting is that it can be obtained much faster while still providing good solutions. The biggest benefit gets apparent when considering real world problem instances where the number of commodities is likely to be at least an order of magnitude larger and the logistics networks may be bigger. In this case the optimization network can still provide solutions while the combined model becomes utterly unsolvable.

It was shown that by varying the parameters of the purchasing model it is possible to arrive at solutions that are better than the one found with the original parameters. This means that it is possible to optimize the combined problem even in cases where the combined MIP is too large to be solved at all. All that is needed is some kind of strategy of how to modify the parameters of the purchasing models. Whatever strategy is used to choose the parameters while optimizing when only the fixed cost per order δ_c and the storage cost multiplier α'_u are used the number of different supply matrices that can be generated is very limited. This can be remedied by adding a dimension to the price array γ_c and turn it into a price matrix γ_{ct} . Using the price matrix the purchasing model can be rewarded for buying a commodity at a certain point in time. It should therefore be possible to generate every valid supply matrix.

8. OUTLOOK

The next step is to implement a meta-solver for the optimization network approach. While changing the parameters randomly might work it is certainly not very efficient. The goal therefore is to find ways to intelligently adapt the parameters of the purchasing model using the solutions of the network flow model.

CPLEX offers a way to add a starting point for a MIP problem. In theory it should be possible to combine the two solutions of the sub problems into a starting point for the combined model. This effectively turns the two smaller models into a construction heuristic for the combined model. Using this approach it should be possible to cut down the runtime considerably without having to let go of the advantages of an exact solution.

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

- Ahuja, R., Magnanti, T. & Orlin, J., 1993. *Network Flows: Theory, Algorithms and Applications*. Upper Saddle River: Prentice-Hall, Inc..
- Assad, A. A., 1978. Multicommodity network flows— A survey. *Networks*, 8(1), pp. 37-91.
- Castro, J. & Nabona, N., 1996. An implementation of linear and nonlinear multicommodity network flows. *European Journal of Operational Research*, 92(1), pp. 37-53.
- Fonoberova, M., 2010. Algorithms for Finding Optimal Flows in Dynamic Networks. In: *Handbook of Power Systems II*. Berlin: Springer, pp. 31-54.
- Ford, L. R. & Fulkerson, D. R., 1958. Constructing maximal dynamic flows from static flows. *Operations Research* 6, pp. 419-433.
- Ford, L. R. & Fulkerson, D. R., 1962. *Flows in Networks*. s.l.:Princeton University Press.
- Hall, A., Hippler, S. & Skutella, M., 2007. Multicommodity flows over time: Efficient algorithms and complexity. *Theoretical Computer Science*, pp. 387-404.
- IBM, 2015. *IBM ILOG CPLEX Optimizer*. [Online] Available at: <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/> [Accessed 1 5 2015].
- Karakostas, G., 2008 . Faster approximation schemes for fractional multicommodity flow problems. *ACM Transactions on Algorithms*, 4(1), pp. 13:1-13:17.
- Klinz, B. & Woeginger, G., 2004. Minimum-cost dynamic flows: The series-parallel case. *Networks*, 43(3), pp. 153-162.
- Kotnyec, B., 2003. *An annotated overview of dynamic network flows*, Paris: HAL - Inria.
- McBride, R., 1998. Progress Made in Solving the Multicommodity Flow Problem. *SIAM J. on Optimization*, 8(4), pp. 947-955.

AUTHORS BIOGRAPHY

SEBASTIAN RAGGL received his Master of Science in Bioinformatics in 2014 from the University of Applied Sciences Upper Austria, and is a research associate at the Research Center Hagenberg. His research interests include metaheuristic methods applied to combinatorial and simulation-based problems.

JUDITH FECHTER received her Master in Mathematics in 2014 from TU Vienna, Austria, and is a research associate at the Research Center Hagenberg. Her research interests include metaheuristic methods applied to simulation-based optimization problems.

ANDREAS BEHAM received his Master in computer science in 2007 from JKU Linz, Austria, and is a research associate at the Research Center Hagenberg. His research interests include metaheuristic methods applied to combinatorial and simulation-based problems. He is a member of the HeuristicLab architects team.

MODELLING A CLUSTERED GENERALIZED QUADRATIC ASSIGNMENT PROBLEM

Judith Fechter^(a), Andreas Beham^(b), Stefan Wagner^(c), Michael Affenzeller^(d)

^{(a),(b),(c),(d)} Heuristic and Evolutionary Algorithms Laboratory, School of Informatics, Communications and Media,
University of Applied Sciences Upper Austria, Hagenberg Campus
Softwarepark 11, 4232 Hagenberg, Austria

^{(a),(b),(d)} Institute for Formal Models and Verification, Johannes Kepler University Linz,
Altenbergerstr. 69, 4040 Linz, Austria

^(a)judith.fechter@fh-hagenberg.at, ^(b)andreas.beham@fh-hagenberg.at, ^(c)stefan.wagner@fh-ooe.at
^(d)michael.affenzeller@fh-hagenberg.at

ABSTRACT

This paper is about the modelling of an assignment problem motivated by a real world problem instance. We consider multiple pieces of equipment which need to be assigned to several locations taking into account capacities as well as relations between equipment and distances between locations. Additionally, a clustering of locations is taken into account that groups locations into areas or fields. It is forbidden to assign the same equipment to locations in different fields. The problem arises in many real world applications such as facility layout or location problems. We discuss the complexity of the problem and prove its NP-hardness. Further two linearization approaches are presented as well as computational studies of the original and the linearized models are conducted. Experimental tests are carried out using CPLEX.

Keywords: generalized quadratic assignment problem, location problem, logistics, linearization

1. INTRODUCTION

This research work is motivated by a real world problem instance. A production company should store multiple equipment to several storage locations which are grouped into areas, here groups. It is necessary to assign every equipment, but allow a single equipment to be assigned to more than one location, since the required storage can exceed the capacity of single locations. Thus, the overall number of locations must exceed the amount of equipment to be assigned. One location may only contain one equipment due to the nature of the storage order and transport vehicles.

The transportation costs which arise when equipment are demanded for further processing should be minimized. Those costs capture the costs of a vehicle moving between groups or locations collecting equipment that are required at the same time. The costs rely on a matrix which assigns a probability of collaborative further processing, as well as on a matrix which indicates the distances between locations.

1.1. Related Problems

This problem class is strongly related to the quadratic assignment problem (QAP) (Koopman and Beckmann 1957) as well as to the generalized quadratic assignment problem (GQAP) (Lee and Ma 2004). The QAP usually assigns one equipment to one location. However, some researchers considered the QAP in a more general way (Garfinkel and Nemhauser 1972) (Sahni and Gonzalez 1976), where the number of equipment is less than the number of locations but without regarding demanded storage of equipment or capacities of locations. The GQAP assigns several equipment to one location but one equipment only to a single location, while considering distances, relations, storage demands and capacities. A generalization of the GQAP is the multi-resource generalized quadratic assignment problem (MRGQAP) (Yagiura et. al. 2007) that addresses an assignment of tasks to agents whereas an agent can handle several tasks but one task may only be handled by one agent. Further, a task may require various resources and one agent has a capacity of each resource, which represents the difference to the GQAP.

1.2. Computational Research and Complexity

Methods for solving the QAP have been researched extensively. (Burkard, Çela, Pardalos, and Pitsoulis 1998) present an overview of studies of the research community. Approaches for solving the GQAP are very limited. The authors of the GQAP (Lee and Ma 2004) presented three linearization approaches of the problem and a branch and bound algorithm for solving those. They employ a single-assignment method as a branching rule. By solving a heuristic greedy algorithm an upper bound is provided, whereas the lower bound at each node of the branch and bound tree is found by solving the generalized linear assignment problem (GLAP) (Ross and Soland 1975). They present an empirical study of 27 test instances varying from 5–16 equipment for 5–30 locations for all three linearization approaches. (Hahn, Kim, Guignard, Smith, and Zhu 2008) propose another exact solution method for the GQAP. A branch and bound algorithm is used whose bound is based on a Lagrangean dual. The dual is derived using the

Reformulation Linearization Technique (RLT) (Sherali and Adams 1996). The basis of the algorithm is a dual ascent procedure using first-level or Level-1 RLT. (Kim 2006) also propose an exact algorithm partially based on a reformulation linearization technique enhanced by adding a subgradient optimization step.

(Kim 2006) also developed a heuristic technique using new evaluations of neighboring solutions for various strategies for exploring the neighborhood. A memetic heuristic approach is presented by (Cordeau, Gaudioso, Laporte, and Moccia 2006) obtaining good results in reasonable computational time. In (Ahlatcioglu, et al. 2012) a combined method of convex hull relaxation (CHR) (Ahlatcioglu and Guignard 2007), which is a special case of the primal relaxation, and quadratic convex reformulation (QCR) (Plateau 2006), which aims to convert non-convex quadratic functions into convex quadratic functions, is proposed. The QAP and generalizations of it remain one of the hardest optimization problems. It is a NP-hard problem. Even finding an approximate solution cannot be done in polynomial time unless P=NP (Sahni and Gonzalez 1976). Due to this fact heuristic methods have become widely used. Most efficient heuristics are construction methods, enumeration or improvement methods or metaheuristics, such as simulated annealing, tabu search, genetic algorithms or ant systems (Burkard, Çela, Pardalos, and Pitsoulis 1998).

1.3. Our Contribution

So far the research community has done a lot of essential work regarding the development of various solution techniques for already existing models, which shall not be within the scope of this project. The focus of this work is the modelling of a new problem class. So far, problem classes considering a 1:1 or a n:1 assignment have been developed, some regarding further constraints like limited resources or capacities. Even so, no problem class has been defined regarding the splitting of equipment to several locations, speaking of a 1:n assignment. Further we take into account that locations belong to a certain area, while regarding distances between locations and relations between equipment. This problem class is important when the storage demands of equipment exceed the storage capacities of single locations as well as when collateral storage is needed due to collaborative further processing. Due to considering locations as part of a group, we name the developed model a clustered generalized quadratic assignment problem. The aim of this paper is to formulate the model and to perform a comprehensive computational study on it.

In section 2 the problem formulation is outlined in a descriptive and mathematical way. Properties like complexity and linearizations will be discussed. Computational studies are conducted in section 3. Tests are carried out on academic instances and show results regarding the solution quality and runtime and draw a comparison between the linearization approaches and the original problem.

2. PROBLEM DESCRIPTION

The clustered generalized quadratic assignment problem (CGQAP) considers the assignment of multiple and various equipment to one or several storage locations depending on the storage demands and location capacities. Storage locations are clustered and hence belong to a single group. When assigning equipment to locations we consider a probability of further collaborative processing between equipment as well as distances between locations inducing that two pieces of equipment having a high common probability shall be stored close to each other. Additionally, the problem takes into account that a single equipment may not be placed in more than one group. The objective is to minimize the transportation costs arising when articles are demanded for further processing.

The space requirement of the equipment has to be met by the space capacities of the locations. The CGQAP is a generalization of the GQAP in that a 1:n assignment is considered, meaning 1 equipment may be assigned to n locations which is an inverted assumption of the GQAP. Additionally, each location belongs to exactly one group. For the purpose of an optimal assignment probabilities of further collaborative processing as well as distances between locations are regarded. The objective is to optimally assign equipment to storage locations whereas an equipment may be located on several locations but only within one group, while regarding space limitations.

2.1. The Model

For the formulation of the mathematical model we use the following notation:

$M = \{1, \dots, m\}$: Set of equipment
 $N = \{1, \dots, n\}$: Set of locations
 $G = \{1, \dots, g\}$: Set of groups
 $r_{i \in M} \in \mathbb{R}^+$: Space requirements per equipment
 $c_{k \in N} \in \mathbb{R}^+$: Space capacity per location

Let α be a surjective mapping, such that each location is mapped to a group and G_l be the set of locations belonging to a group:

$$\alpha: N \rightarrow G: \alpha(k) \in G, \forall k \in N$$

$$G_l = \{n | \alpha(n) = l\} \text{ with } G_l \cap G_o = \emptyset \quad \forall l \neq o.$$

The function $a: N \times N \rightarrow \{1,0\}$ is equal to the identity in the case that two locations belong to the same group.

$$a(k, h) = \begin{cases} 1 & \alpha(k) = \alpha(h) \\ 0 & \text{otherwise} \end{cases} \quad k, h \in N$$

Between equipment there is a certain probability of how likely two pieces of equipment have a further collaborative processing. Function $w: M \times M \rightarrow \mathbb{R}_0^+$ assigns a probability, also called weights, to each possible pair of equipment.

$$w_{ij} \in \mathbb{R}_0^+ \quad \forall i, j \in M.$$

Between locations, function $d: N \times N \rightarrow \mathbb{R}_0^+$ determines the distance between each pair of locations.

$$d_{kh} \in \mathbb{R}_0^+ \quad \forall k, h \in N.$$

The costs arising when equipment are demanded for further processing shall be minimized. Those costs capture the distance the transport vehicles have to travel in order to pick up the demanded articles.

The binary decision variable determines the assignment of equipment to locations and can be written as

$$x_{ik} \rightarrow \begin{cases} 1 & \text{equipment } i \text{ is assigned to location } k \\ 0 & \text{otherwise} \end{cases}$$

We also define a binary decision expression denoting if at least one location in a group contains an article

$$z_l \rightarrow \begin{cases} 1 & \text{if } \sum_i^m \sum_{k \in G_l} x_{ik} \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

We can finally write the objective

$$\min \delta * \sum_l^g z_l + \gamma * \sum_i^m \sum_k^n \sum_j^m \sum_h^n x_{ik} * x_{jh} * w_{ij} * d_{kh} \quad (\text{CGQAP})$$

subject to

$$\sum_i^m x_{ik} \leq 1 \quad \forall k \in N \quad (1)$$

$$\sum_k^n x_{ik} * c_k \geq r_i \quad \forall i \in M \quad (2)$$

$$x_{ik} x_{ih} \leq a(k, h) \quad \forall i \in M, k, h \in N \quad (3)$$

$$x_{ik} \in \{0, 1\} \quad \forall i \in M, k \in N \quad (4)$$

$$\delta, \gamma \in \mathbb{R}^+$$

Constraint (1) ensures that each location is assigned only once. Constraint (2) ensures that the space requirements are met by the assigned locations. Constraint (3) ensures that the equipment is not assigned to locations in different groups.

2.1.1. Complexity

The formulation above is a generalization of the formulation of Koopman and Beckman (Koopman and Beckmann 1957) with further constraints. Assuming that one piece of equipment can be split into several pieces, we can prove that the CGQAP is NP-hard by reducing it to a QAP, which is known to be NP-hard (Sahni and Gonzalez 1976).

Theorem 1: The CGQAP is strongly NP-hard.

Proof: Assuming that equipment can be split into several pieces of the same equipment with lower demand, solving a CGQAP can be considered as solving a QAP, with additional constraints on clustered areas, for each possible way of splitting an equipment. Hence the

CGQAP is at least as complex as the QAP which is known to be strongly NP-hard (Sahni and Gonzalez 1976). ■

Following the thought of the proof, a CGQAP becomes more complex as the storage capacity decreases since the equipment need to be split more often which results in solving more QAPs. The maximum number of QAPs to be solved is the number of multiples of the minimum storage unit within the e demand.

2.1.2. Linearization

Due to the high complexity of the QAP, there have been many linearization approaches (Xia and Yuan 2006) (Burkard, Çela, Pardalos, and Pitsoulis 1998) (Erdogan, 2006) (Punnen and Kabadi 2013). (Lee and Ma 2004) have conducted three approaches for the GQAP, namely following Frieze and Yadegar, Kaufman and Broeckx and a new linearization. In this paper, following (Lee and Ma 2004), we introduce Frieze and Yadegar linearization (Frieze and Yadegar 1983) and Kaufman and Broeckx (Kaufman and Broeckx 1978) linearization for the CGQAP.

Frieze and Yadegar

The products $x_{ik}x_{jh}$ of the binary variables are replaced by continuous variables $y_{ikjh} := x_{ik}x_{jh}$. Applying this approach to the CGQAP formulation we achieve the following mixed integer linear programming formulation.

$$\min \delta * \sum_l^g z_l + \gamma * \sum_i^m \sum_k^n \sum_j^m \sum_h^n y_{ikjh} * w_{ij} * d_{kh} \quad (\text{FYL})$$

$$\sum_{k=1}^N y_{ikjh} = x_{jh} \quad \forall i, j \in M, h \in N \quad (1a)$$

$$\sum_{h=1}^N y_{ikjh} = x_{ik} \quad \forall i, j \in M, k \in N \quad (2a)$$

$$\sum_i^m x_{ik} \leq 1 \quad \forall k \in N \quad (3a)$$

$$\sum_k^n x_{ik} * c_k \geq r_i \quad \forall i \in M \quad (4a)$$

$$x_{ik} + x_{ih} - 2 * a(k, h) \leq 1 \quad \forall i \in M, k, h \in N \quad (5a)$$

$$x_{ik} \in \{0, 1\} \quad \forall i \in M, k \in N \quad (6a)$$

$$y_{ikjh} \in [0; 1] \quad \forall i, j \in M, k, h \in N \quad (7a)$$

$$\delta, \gamma \in \mathbb{R}^+$$

The problem (FYL) has m^2n^2 continuous variables, mn binary variables and $n^2m + 2m^2n + mn + m + 3n$ constraints.

Theorem 2: The FYL is equivalent to the CGQAP.

Proof: We will only show one direction of the proof namely that a feasible solution of FYL is a feasible

solution to the CGQAP since the other direction is trivial to show.

Let (x, y) be a feasible solution to FYL. First we show:

$$x_{ik} = 0 \Rightarrow y_{ikjh} = 0, \exists i \in M, k \in N, \forall j \in M, h \in N$$

Considering (1a), we obtain:

$$\sum_{i=1}^M \sum_{k=1}^N y_{ikjh} = \sum_{i=1}^M x_{jh} = Mx_{jh}, \forall j \in M, h \in N$$

In case $x_{jh} = 0$ we have:

$$\sum_{i=1}^M \sum_{k=1}^N y_{ikjh} = 0 \Rightarrow y_{ikjh} = 0, \forall i \in M, k \in N.$$

Considering (2a), we obtain:

$$\sum_{j=1}^M \sum_{h=1}^N y_{ikjh} = \sum_{j=1}^M x_{ik} = Mx_{ik}, \forall i \in M, k \in N.$$

In case $x_{ik} = 0$ we have:

$$\sum_{j=1}^M \sum_{h=1}^N y_{ikjh} = 0 \Rightarrow y_{ikjh} = 0, \forall j \in M, h \in N$$

Further we show:

$$(x_{ik} = 1 \wedge x_{jh} = 1) \Rightarrow (y_{ikjh} = 1).$$

For this purpose we define a function ϕ , denoting the assignment of equipment to locations. In our case, it is an injective function (without proof).

$$\phi: M \rightarrow N: \phi(i) = k, \text{ such that } x_{i\phi(i)} = 1, \forall i \in M.$$

We need to show:

$$y_{i\phi(i)j\phi(j)} = 1, \forall i, j \in M.$$

We write

$$\sum_{k=1}^N y_{ikjh} = y_{i\phi(i)jh}, \forall i, j \in M, h \in N.$$

We have

$$\sum_{i=1}^M \sum_{k=1}^N y_{ikjh} = \sum_{i=1}^M y_{i\phi(i)jh} = Mx_{jh}, \forall j \in M, h \in N$$

For $\phi(j) = h \Leftrightarrow x_{jh} = 1$, we achieve

$$\sum_{i=1}^M y_{i\phi(i)j\phi(j)} = M, \forall j \in M$$

$$\text{Hence } (x_{ik} = 1 \wedge x_{jh} = 1) \Rightarrow (y_{ikjh} = 1). \quad \blacksquare$$

Kaufman and Broeckx

The products $x_{ik}x_{jh}$ of the binary variables are replaced by continuous variables y_{ik} determining the product of relations and distances to all located equipment of a fixed located equipment. Applying this approach to the CGQAP formulation we achieve the following mixed integer linear programming formulation.

$$\min \delta * \sum_i^g z_i + \gamma * \sum_i^m \sum_k^n y_{ik} \quad (\text{KB})$$

$$\sum_i^m x_{ik} \leq 1 \quad \forall k \in N \quad (1b)$$

$$\sum_k^n x_{ik} * c_k \geq r_i \quad \forall i \in M \quad (2b)$$

$$x_{ik} + x_{ih} - 2 * a(k, h) \leq 1 \quad \forall i \in M, k, h \in N \quad (3b)$$

$$\sum_{j=1}^M \sum_{h=1}^N w_{ij} * d_{kh} * x_{jh} + v_{ik} * x_{ik} - y_{ik} \leq v_{ik}, \forall i \in M, k \in N \quad (4b)$$

$$x_{ik} \in \{0, 1\} \quad \forall i \in M, k \in N \quad (5b)$$

$$y_{ik} \geq 0 \quad \forall i \in M, k \in N \quad (6b)$$

$$\delta, \gamma \in \mathbb{R}^+$$

We define $\forall i \in M, k \in N$:

$$y_{ik} := x_{ik} * \sum_{j=1}^m \sum_{h=1}^n w_{ij} * d_{kh} * x_{jh}$$

$$v_{ik} := \sum_{j=1}^m \sum_{h=1}^n w_{ij} d_{kh}$$

The problem (KB) has mn continuous variables, mn binary variables and $m + mn + 3n$ constraints.

Theorem 3: The KB is equivalent to the CGQAP.

Proof: We will only show one direction of the proof namely that a feasible solution (x, y) of KB is a feasible solution to CGQAP since the other direction is trivial to show.

Let (x, y) be a feasible solution to FYL. First we show that (4b) is satisfied for $x_{jh} = 0$ and/or $x_{ik} = 0$.

$$(x_{jh} = 0) \Rightarrow v_{ik} * (x_{ik} - 1) \leq y_{ik}, \forall j \in M, k \in N$$

$$(x_{ik} = 0) \Rightarrow -w_{ik} + \sum_{j=1}^m \sum_{h=1}^n w_{ij} * d_{kh} * x_{jh} \leq y_{ik}, \forall i \in M, k \in N$$

Obviously $v_{ik} * (x_{ik} - 1)$ and

$$-\sum_{j=1}^m \sum_{h=1}^n w_{ij} d_{kh} + \sum_{j=1}^m \sum_{h=1}^n w_{ij} * d_{kh} * x_{jh}$$

are both non-positive. Since $y_{ik} \geq 0$ and y_{ik} shall be minimized, y_{ik} must be zero.

Further we need to prove (7b) for all $i \in M$ and $k \in N$. For this purpose we use the function ϕ , defined in 2.1.2: Frieze and Yadegar.

Since $w_{ik} * (x_{ik} - 1)$ is non-positive, we know:

$$\begin{aligned} y_{ik} &= y_{i\phi(i)} \\ &\geq (\sum_{j=1}^m \sum_{h=1}^n w_{ij} * d_{\phi(i)h} * x_{jh}) + v_{i\phi(i)} * (x_{i\phi(i)} - 1) \end{aligned}$$

$$= \sum_{j=1}^m w_{ij} * d_{\phi(i)\phi(j)} * x_{j\phi(j)} + v_{i\phi(i)} * (x_{i\phi(i)} - 1)$$

$$= \sum_{j=1}^m w_{ij} * d_{\phi(i)\phi(j)}$$

In order to minimize the sum of y_{ik} , we can conclude:

$$y_{ik} = x_{ik} * \sum_{j=1}^m \sum_{h=1}^n w_{ij} * d_{kh} * x_{jh}, \forall i \in M, k \in N.$$

3. EXPERIMENTAL STUDIES

In this section computational studies on both linearization approaches and on the original problem are presented. The studies are conducted on a variety of GQAP problem instances (see 3.1). We draw a comparison of results achieved by CPLEX (IBM C. , 2015) regarding solution quality and runtime.

All tests were calculated on a laptop with an Intel(R) Core(TM) i7-4600U CPU @2.10GHz 2.70GHz processor.

3.1. Problem Instances

The problem instances used are described in (GitHub, 2015). However, they had to be adapted to this problem formulation due to the inversed assignment order (1:n) and the constraint on locations belonging to certain groups. The instances are turned around such that the original locations become equipment and the original equipment become locations and also distances and weights have been interchanged. The capacities of locations are varied in order to obtain instances of various complexity. The instances are named in the way “N-M-p” meaning that N equipment shall be assigned to M locations. The total storage requirement of equipment is p percentage of the total capacity of locations. Obviously the higher the percentage the more complex is the problem instance. The problem instances are published on the HeuristicLab website (HEAL, 2015).

3.2. Experimental Results

Due to the complexity of the problem we set a timelimit of 10800 seconds in CPLEX. The reason for choosing three hours is that preliminary tests show good results within that time.

Table 1 summarizes results regarding the best found solution and runtime to that solution. What can be seen is that the KB-formulation obtains the best solution quality for most instances. The FYL-linearization shows best results for the instances of lowest complexity (10-50-38, 15-35-45, 20-30-45) and also for all instances of the 10-50-class.

Note that two of the most complex instances (15-35-91, 20-30-91) could not be solved by the FYL-model within the timelimit (marked with an “x”), which in turn is due to the high number of variables and constraints. Here it is worth to mention that the original problem formulation could be solved to its best found solution within 257 seconds for one of those problems (15-35-91), which is way better than the solving time of the KB-model for that instance.

Having a closer look at the runtime until the best solution is found, again the KB-linearization performs best for most problem instances. This is a reasonable fact due to the little number of variables and constraints. Nevertheless, it is noticeable that the runtime to the best found solution of KB is significantly higher for the instances of highest complexity (10-50-77, 15-35-91, 20-30-91). This characteristic is not that distinct for FYL and the original formulation. However, in table 2 we take a look at the time until a first solution is found, where the KB-linearization is the most efficient one for each problem instance. At this point we can conclude that the KB-model is very efficient in finding a first solution which is also further improved during the solving process. It takes considerable time to find the optimal solution for those instances.

Table1: Best solution found within 10800 sec. and time to this solution

Instance	Original		FYL		KB	
	Best Solution	Time to Best	Best Solution	Time to Best	Best Solution	Time to Best
10-50-77	1,0037	6820	1,0006	10561	1,003	10379
10-50-51	0,669	7398	0,6673	1601	0,6681	250
10-50-38	0,6675	5202	0,6671	947	0,6674	214
15-35-91	1,0036	257	x	x	1,0026	10777
15-35-61	0,6678	6827	1,0013	3322	0,6677	16
15-35-45	0,6675	105	0,6673	3092	0,6673	78
20-30-91	1,0022	1611	x	x	1,0019	1065
20-30-61	1,0011	579	1,001	4403	1,0009	60
20-30-45	1,0009	4502	1,0009	10569	1,0009	17

Table 2: First solution found

Instance	Original		FYL		KB	
	First Solution	Time to First	First Solution	Time to First	First Solution	Time to First
10-50-77	1,0053	121	1,001	1200	1,0038	16
10-50-51	1,0024	143	0,6673	1601	0,6687	12
10-50-38	1,0013	66	0,6671	947	0,6676	11
15-35-91	1,2934	144	x	x	1,0043	19
15-35-61	1,0013	166	1,0013	3322	0,6679	9
15-35-45	0,6675	105	0,6674	1811	0,6675	12
20-30-91	1,2916	1144	x	x	1,0023	62

The FYL-formulation also obtains the best solution quality for some instances, the runtime to the first solution found is significantly longer though. This fact can also be seen in the performance of all three formulations for the problem instances with lowest complexity. Having a look at figure 1, it can be seen that the FYL-model has a long initialization time for each instance while the KB model reaches a first feasible solution very fast. However, it is worth to mention that no definite preference can be identified between the original model and FYL regarding the solution quality.

equivalence to the original problem. Computational studies are conducted regarding runtime and solution quality on adapted problem instances of the GQAP. The presented problem class has numerous applications whenever multiple equipment need to be assigned to several locations which again can be grouped into certain areas where also capacities and storage amount need to be taken into account. The problem is also very interesting from a theoretical point of view since it is based on the QAP which is one of the most challenging combinatorial optimization problems.

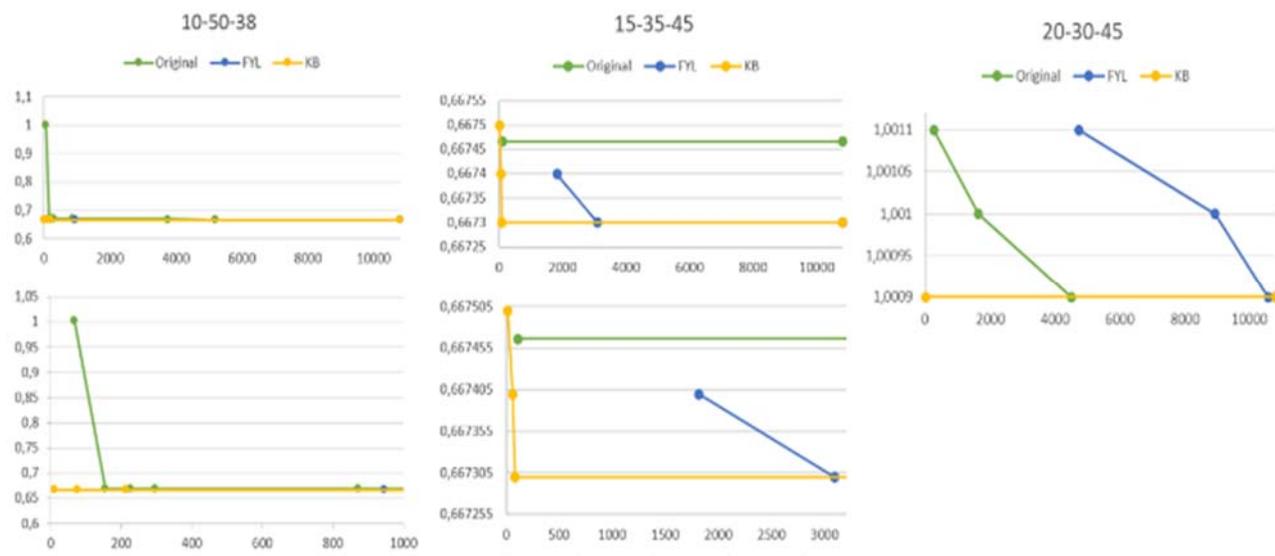
4. CONCLUSION

In this paper we define a new problem class namely the clustered generalized quadratic assignment problem (CGQAP). It is a generalization of the quadratic assignment problem (QAP) and the generalized quadratic assignment problem (GQAP). Based on the complexity of the QAP, which is known to be NP-hard, we show the NP-hardness of the CGQAP. Further we present two linearization approaches and prove its

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).

Figure 1: Performance of the original model, FYL and KB



REFERENCES

- Ahlatcioglu, A., Guignard, M. 2007. The convex hull relaxation for nonlinear integer programs with linear constraints. OPIM Department Report, revised 2008.
- Ahlatcioglu, A., Bussieck, M., Esen, M., Guignard, M., J.-H., J., Meeraus, A. 2012. Combining QCR and CHR for convex quadratic pure 0-1 programming problems with linear constraints. *Annals of Operations Research* 199(1), 1-17.
- Burkard, R., Çela, E., Pardalos, P., Pitsoulis, L. 1998. The Quadratic Assignment Problem. In P. Pardalos, P., Resende, M. Eds., *Handbook of Combinatorial Optimization Vol.3*, 241-337. Dordrecht: Kluwer Academic Publishers.
- Cordeau, J., Gaudioso, M., Laporte, G., Moccia, L. 2006. A memetic heuristic for the generalized assignment problem. *INFORMS Journal on Computing* 18, 433-443.
- Frieze, A., Yadegar, J. 1983. On the quadratic assignment problem. *Discrete Applied Mathematics* 5, 89-98.
- Garfinkel, R., Nemhauser, L. 1972. *Integer Programming*. New York: John Wiley & Sons.
- Gavish, B., Prikul, H. 1991. Algorithms for the Multi-Resource Generalized Assignment Problem. *Management Science* 37(6), 695-713.
- Hahn, P., Kim, B.-J., Guignard, M., Smith, J., Zhu, Y.-R. 2008. An algorithm for the generalized quadratic assignment problem. *Computational Optimization and Applications* 40(3), 351-372.
- HEAL. 2014 Additional Material. Retrieved from HeuristicLab: <http://dev.heuristiclab.com/AdditionalMaterial> [accessed: 15 May 2015].
- IBM, C. IBM Software 2015. CPLEX Optimizer: <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/> [accessed: 30. April 2015].
- Kaufman, L., Broeckx, F. 1978. An algorithm for the quadratic assignment problem using benders' decomposition. *European Journal of Operational Research* 2(3), 204-211.
- Kim, B. 2006. Investigation of methods for solving new classes of quadratic assignment problems (QAPs). Philadelphia, PA 19104: (Ph.D). University of Pennsylvania.
- Koopman, T., Beckmann, M. 1957. Assignment problems and the location of economic activities. *Econometrica* 25, 53-76.
- Lee, C.-G., Ma, Z. 2004. The generalized quadratic assignment problem. Toronto, Canada: Department of Mechanical and Industrial Engineering, University of Toronto.
- Plateau, M. 2006. Reformulations quadratiques convexes pour la programmation quadratique en variables 0-1. (PhD) Laboratoire Cedric, CNAM, France.
- Punnen, A., Kabadi, S. 2013. A linear time algorithm for the Koopmans-Beckmann QAP linearization and related problems. *Discrete Optimization* 10(3), 200-209.
- Ross, G., Soland, R. 1975. A branch-and-bound algorithm for the generalized. *Mathematical Programming* 8, 91-103.
- Sahni, S., Gonzalez, T. 1976. P-complete approximation problems. *Journal of the ACM* 23, 555-565.
- Sherali, H., Adams, W. 1994. A hierarchy of relaxations and convex hull characterizations for mixed-integer zero-one programming problems. *Discrete Applied Mathematics* 52, 83-106.
- Wagner, S., Kronberge, G., Beham, A., Kofler, M., Scheibenpflug, A., Pitzer, E., Affenzeller, M. 2014. Architecture and Design of the HeuristicLab Optimization Environment. *Topics in Intelligent Engineering and Informatics* 6, 197-261.
- Xia, Y., Yuan, Y.-X. 2006. A new linearization method for quadratic assignment problems. *Optimization Methods and Software* 21(5), 805-818.
- Yagiura, M., Komiya, A., Kojima, K., Nonobe, K., Nagamochi, H., T., I., & Glover, F. (2007). A path relinking approach for the multi-resource generalized quadratic assignment problem. *Proc. Engineering Stochastic Local Search Algorithms—Designing, Implementing and Analyzing Effective Heuristics (SLS 2007)* (pp. 121-135). Brussels, Belgium: Lecture Notes in Computer Science 4638.

AUTHORS BIOGRAPHY

JUDITH FECHTER received her Master in Mathematics in 2014 from TU Vienna, Austria, and is a research associate at the Research Center Hagenberg. Her research interests include metaheuristic methods applied to simulation-based optimization problems.

ANDREAS BEHAM received his Master in computer science in 2007 from JKU Linz, Austria, and is a research associate at the Research Center Hagenberg. His research interests include metaheuristic methods applied to combinatorial and simulation-based problems. 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, Campus Hagenberg. 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 UAS, Campus Hagenberg.

COMBINATION OF LEARNING APPROACHES IN MODELLING AND SIMULATION

Maja Atanasijević-Kunc, Marko Papić, Borut Zupančič, Janez Bešter

University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

maja.atanasijevic@fe.uni-lj.si, marko.papic@fe.uni-lj.si, borut.zupancic@fe.uni-lj.si, janez.bester@fe.uni-lj.si

ABSTRACT

Modelling and simulation are disciplines which have an important influence to the learning processes in technical fields, but are well accepted also in many other scientific areas. So is also at the Faculty of Electrical Engineering, University of Ljubljana where such lectures have long-lasting tradition. Through the last year we have introduced a new program called E-CHO with which additional e-learning approaches have been opened for teaching staff as well as for students. The question arises: how to take the advantage of these new possibilities and how to combine them with the established approaches in an efficient way? Some of the solutions are presented in the paper.

Keywords: modelling, simulation, education, E-CHO program

1. INTRODUCTION

Modelling and simulation are disciplines which have an important influence to research and learning processes in technical fields (Löscher et al. 2005, Karer et al. 2007, Sodja and Zupančič 2009, Bicher et al. 2012, Logar, Dovžan, and Škrjanc 2012a, 2012b, Glavan et al. 2013, Mušič et al. 2013, Štampar, Sokolić, and Karer 2013, Fathi et al. 2015), but are well accepted also in many other scientific areas (Atanasijević-Kunc, Drinovec, and Mrhar 2008, Čorović et al. 2015). At the Faculty of Electrical Engineering, University of Ljubljana (FE-UL), Slovenia, several subjects are devoted to mentioned contents, but even more are using actively the knowledge learned through these lectures. The history of the subjects is rather long, but the last major reorganization has been realized through the so called bologna study in 2009-10. As can be observed in Figure 1 the study has three levels. The first and the third one are lasting three years, while for the second one two years are needed. At the first level students can choose between University program and Professional higher education program, the latter being more oriented to the practice, while the University program gives strong theoretical starting points. Graduates of both first-level programs can, when passing the corresponding evaluation, enroll the second – master level, while the best of them can later on continue at Doctoral Study.

The majority of the first and the second level subjects are organized in the form of lectures and laboratory exercises, while at the end of the semester first written and then oral exam are carried out. In some subjects, especially those in the first year, during the semester teaching staff sometimes prepares also colloquia to intensify students' work. Very high colloquia's exam mark sometimes means that student has also passed written examinations.

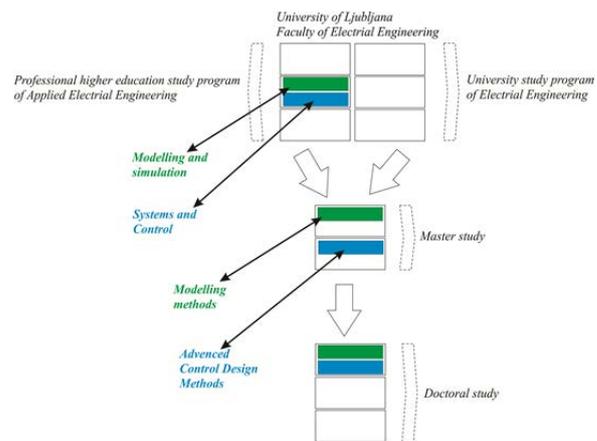


Figure 1: Organization of education process at the Faculty of Electrical Engineering, University of Ljubljana

In 2014/15, we have, in addition to the aforementioned forms, introduced an e-learning approach. It was decided to use program E-CHO (Čorović, Bešter and, Miklavčič 2009, Papić et al. 2011, Atanasijević-Kunc et al., 2011, Logar et al. 2011), as the development of this program has long been conducted by the Laboratory for Telecommunications at FE-UL. The efficacy of this program was tested and complemented for the purposes of studies at FE-UL already in previous years (Čorović, Bešter and, Miklavčič 2009, Atanasijević-Kunc et al., 2011). Because of many positive experiences we have decided to try now even more intensively to develop the so-called combined or blended learning taking into account new possibilities. The goals are numerous, but some of the most important are the following:

- gradualness,
- flexibility and multi-functionality,
- enhancement of motivation,

- step by step improvement of program E-CHO regarding specific experiences and arising needs,
- explicit subjects' connections, where possible (some possibilities are indicated in Figure 1).

The paper is organized in the following way. In the next section some of the most important features of the program E-CHO are described where the emphasis is given to its new properties and to those features which enable the extensions and/or improvement of education process. The subject Modelling and simulation is presented in the third section regarding lectures and laboratory exercises education flow. Here also Toolbox LABI is introduced which was developed in program MATLAB as a Laboratory of mathematical models, where also efficient analysis in the form of toolbox is prepared and can support education process at different levels of the study. In the fourth section the realization of blended learning approach is presented taking into account lectures, laboratory exercises, homework, online testing and exams, motivation seminars and projects, and the possibilities to extend the work to a diploma thesis. The paper ends with the concluding remarks and some ideas for future investigations.

2. E-CHO PROGRAM

E-CHO is an e-learning internet-based platform. It is the result of a long lasting development of experts from Laboratory for Telecommunications at the FE-UL (Čorović, Bešter and, Miklavčič 2009, Papić et al. 2011, Atanasijević-Kunc et al., 2011, Logar et al. 2011) It is therefore not surprising that it associates and combines numerous e-learning functionalities. It can also be integrated into other web-based applications.

E-CHO provides e-learning management (LMS – Learning Management System), e-learning content managing (LCMS – Learning Content Management System), tracking of the teacher's activity and progress, knowledge assessment/testing, e-learning standards' support (SCORM and QTI) and multilingualism.

During the last few years improved version of E-CHO was developed where special attention was devoted to the following new and/or extended program properties:

- Simplification of content development procedures:
Rapid and user-friendly content development is now possible. This means that each user with sufficient access rights can add text, pictures, videos, questions or other learning objects within the developed interface. No special program knowledge is needed. This content can be delivered to selected users or groups directly after the development. E-CHO enables also fully automatic video lectures content development process. Arbitrary multimedia element can be added to the video and played synchronously.
- Improvement of progress tracking tools and learning statistics reports:
Several different filters were added to the progress tracking mechanisms. The users with

sufficient access rights can analyze in depth data regarding learning statistics, they can assess quality of developed learning objects on the large scale (whole courses) or on the individual learning objects level (e.g. assessing quality of questions and tests).

- Adaptive access – rights management:
E-CHO program maps system features with user roles. That means that desired user role within E-CHO can be defined and the system use can be generalized. Users and group management can also be simply integrated with Active Directory (AD) or other Human Resources Management (HRM) systems.
- Video and multi – media support:
E-CHO enables simple uploading of different video formats with corresponding multimedia learning objects. Video is converted to the format which is the most appropriate for web or mobile viewing and corresponding multimedia is added. It supports catalogue of all video and other multimedia objects, which can be reused in any way for learning purposes.
- Simplified third party integration through open interfaces:
E-CHO provides open interfaces, enabling third party developers to integrate their application for various e-learning purposes. As an example, third party remote experiment can be simply integrated into E-CHO. Open web interfaces enable exchange of data between the remote experiment and E-CHO LMS in order that remote experiment is delivered through E-CHO system and results of experiment are saved within the E-CHO system progress tracking tools.
- New mobile interfaces (tablets, smart-phones):
E-CHO program provides mobile interfaces for Android and iOS operating systems and corresponding terminals. This enables users to access e-learning from tablets or smart phones.
- Improved knowledge assessment (new question types, simplified tests and questionnaires development and delivery):
E-CHO supports more than 20 different question types for knowledge assessment. Video and other multimedia can be added to questions. Questions can be sorted into folders (directories) and E-CHO test player can deliver random numbers of questions from folders, based on their difficulty level. Tests development is simplified and time to delivery is shortened. Each test or question can also be delivered to the users or groups directly, without first integrating it into courses or groups of learning objects.

As already mentioned in the school-year 2014/15 we have chosen E-CHO to be an official e-learning support

for the whole FE-UL. The consequence was that special adaptations were introduced for this purpose. Each professor and his/hers assistants, as well as all students, can enter E-CHO through corresponding user name and password information. Starting window is illustrated in Figure2.

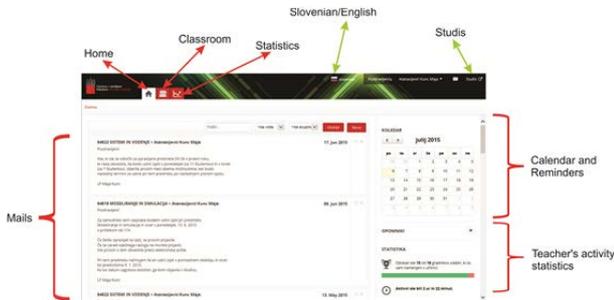


Figure 2: E-CHO starting window

E-CHO has been connected to program *Studis* (see upper part of Figure 2) which is the environment for managing all the information of education processes at FE-UL. In corresponding data basis information of students, teaching staff, lectures, exams, diplomas, etc. is stored. For entering *Studis* also digital certificate and additional password is needed due to security reasons. To each person involved into education process corresponding rights are assigned regarding the information transferred from *Studis* data base. For example, each professor can develop e-learning contents for the subjects he/she is responsible for by clicking the button *Classroom* (see upper part of Figure 2). This opens the window of the form as is presented in Figure 3.

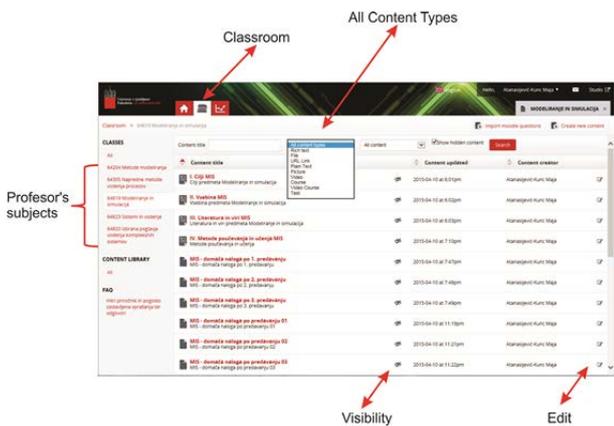


Figure 3: E-classroom with professor's subjects and developed content elements

In the left part of the window the titles of the subjects are displayed. Below also the access to the short manual is available and some answers to frequently asked questions. Titles of already developed contents are displayed in the next column. It is possible to observe all developed

contents or only those which are connected to the certain subject.

The user can develop contents in the following forms: file, figure, questions, folder with questions, web-page, url-address, video, different tests, group of contents, course, video course.

Content types *video*, *figure*, and *file* enable file transfer from the user's computer or definition of url-address where the file is available.

When developing content type *web-page* it is possible to edit text, to add figures and video clips and to use prepared UL-FE templates.

Content type *url-address* enables a definition of corresponding address which is not inside E-CHO application.

Content type *course* enables aggregation of developed content types into corresponding sequence. In the same time the user can generate multi-level table of contents. In addition also *video-course* can be prepared where the user can combine video clips with corresponding transparencies and/or tests or other prepared graphic material.

All prepared contents can be connected to one or more subjects in corresponding way. During development or during education process the chosen contents can be made visible only for developers. When desired this status can simply be changed and the chosen content is then delivered to the chosen students' group or groups.

It is important to mention that teaching staff and students can develop contents in the same way. Of course if students are uploading their homework, projects, or exams these materials can be accessed by teaching staff only, but if their project results can represent interesting information to the whole group it can also be made visible for everyone from corresponding group. These properties are important because in this way developed contents can also be combined into different course forms and flexibility of education can be increased.

3. MODELLING AND SIMULATION

As presented in Figure 1, students from Professional higher education program of Applied Electrical Engineering Program have in the third semester subject entitled Modelling and simulation if they choose to study Control Engineering (other options are Electronics, Power Engineering Technology and System Automation, Telecommunications, and Quality Engineering). It is organized into lectures (3 hours per week) and laboratory exercises (2 hours per week). Main chapters, which are presented through lectures (Atanasijević-Kunc 2013), (similar courses are rather frequent: Cellier 1991, Matko, Zupančič, and Karba 1992, Close and Frederick 1993, Cellier and Kofman 2006) are the following:

1. Introduction
2. Dynamic systems presentations
3. Introduction to simulation
4. Experimental modelling

5. Theoretical modelling

Due to time-limitations the main focus is given to continuous systems.

It is important to mention that through systems' presentations also computer presentations are taken into account. Different programs can be chosen for demonstration purposes. We have decided to use Matlab as it is very suitable for modelling and simulation problems but also for control design of dynamic systems and so the time needed for program teaching can be minimized also regarding several other subjects. During introduction to simulation also basic information regarding Simulink and Control System Toolbox is given.

In laboratory we have installed a great number of pilot plants (see Figure 4) which mimic different dynamic processes.

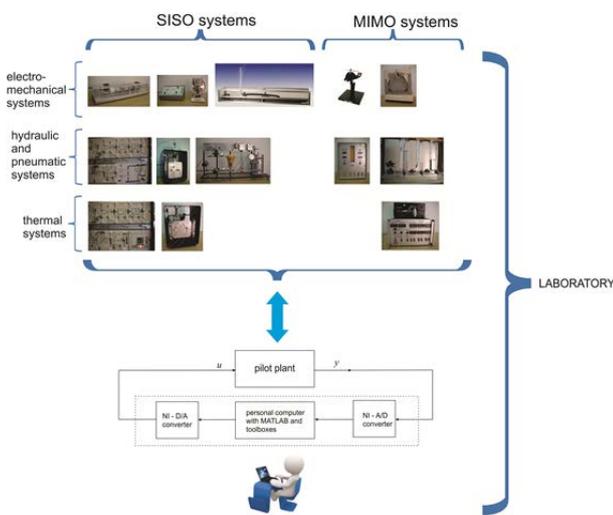


Figure 4: Laboratory work organization

They can be divided regarding different criteria. Some of them are for example single-input single-output (SISO) systems, some can operate as multivariable (MIMO) devices, some are stable and some not. When time for experimentation is very limited (and this aspect is important during laboratory exercises) also systems' time constants are important. Regarding presented groups of devices in Figure 4 it can be expected that the fastest are electro-mechanical pilot plants, while the slowest are thermal, where another problem can be expected and that is very slow cooling. This is problematic when we want to repeat experiments with the same initial conditions in short time intervals.

Each pilot plant is through A/D and D/A converters connected to the personal computer, where also Matlab with corresponding toolboxes is installed. In Simulink library one block was added which enables the communication with mentioned A/D and D/A converters. In this way experimentation with pilot plants can be realized in the same way as with mathematical models. So the gap between theory and practice was minimized.

During lectures also a great number of examples are explained and for some of them also computerized interpretation is given (Atanasijević-Kunc 2008). This has motivated the development of the so called LABI (Atanasijević-Kunc, Karba, and Zupančič 2011) which stands for the Laboratory of mathematical models of dynamic systems and consists of a group of files which are organized through graphical user interface (GUI) in Matlab. The development started in Slovene language but now we are trying to realize it bilingually. This is important because the number of foreign students is constantly growing, but also our students can in this way learn some important expressions.

The starting window of LABI is illustrated in Figure 5.

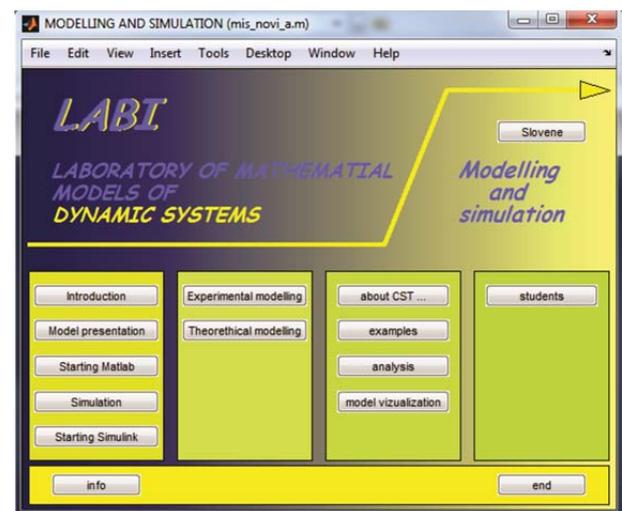


Figure 5: Starting window of LABI

This and also all other windows of the GUI are organized in a similar way. In the upper part the title or short description of the current level is given with the possibility to switch between English and Slovene language. In the frame of the bottom part push-buttons enable to finish the work (*end*), to obtain some information about the current level (*info*) and to return to the previous level (*back*) or to the very beginning (*beginning*) if possible.

The push-buttons of the central part enable to observe prepared information or files, or to execute developed files if buttons are active. If certain button is not active there are two possibilities: this part of LABI is still under development or needed data for file execution is not available in Matlab's workspace.

In the name of each graphical window (top of Figure 5) also the name of the m-file which generated the window is given. So it is very simple to trace the information needed for observed calculations. This aspect is very important as all the users are invited to participate in extensions and improvement of this software.

In Figure 5 presented buttons of the first and the second column open windows where information is available regarding each chapter of the lectures (pdf-files). In addition Matlab files with illustrative examples can be observed and executed.

The role of the buttons of the third column is the following. Button entitled *about CST ...* displays very short manual of Control System Toolbox and presents some of its functions, button *examples* enables to observe modelling and simulation results of simple, but also more complex problems (case studies) where in many cases (if of interest) also animation of the problem is illustrated. One such example is presented in Figure 6 for TQ CE9 Ball & Hoop pilot plant.

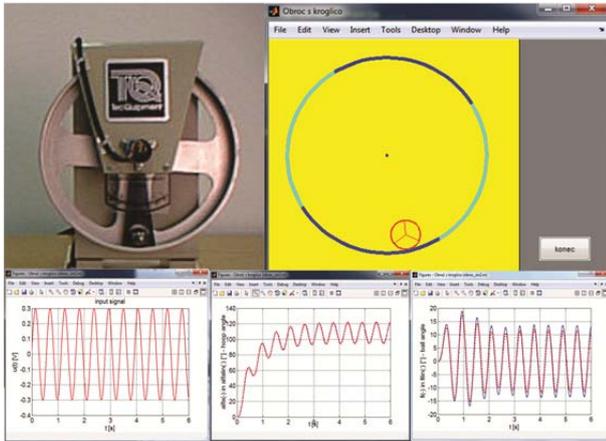


Figure 6: Simulation with animation window of mathematical model of TQ CE9 Ball & Hoop pilot plant

Very precise instructions for graphical results presentation in Matlab can be observed by pushing the button *model visualization*.

For analysis purposes we have developed a toolbox which can be started by pushing the button *analysis*. This opens the window as is illustrated in Figure 7.



Figure 7: Starting window of analysis toolbox

Analysis toolbox consists of a great number of functions which represent an extension and enlargement of the functions available in Matlab and Control System Toolbox in the way that several functions have been added, they are organized through graphical windows

and can be called also only by pushing the button if all needed information is in workspace. In all functions explanation and where possible graphical representation of results is added. The amount of additional explanations can be controlled by the so called communication vector.

Mentioned functions can be divided in several different ways. Some of them can be used with SISO systems (see Figure 8), some with MIMO and some are suitable for both groups of problems.

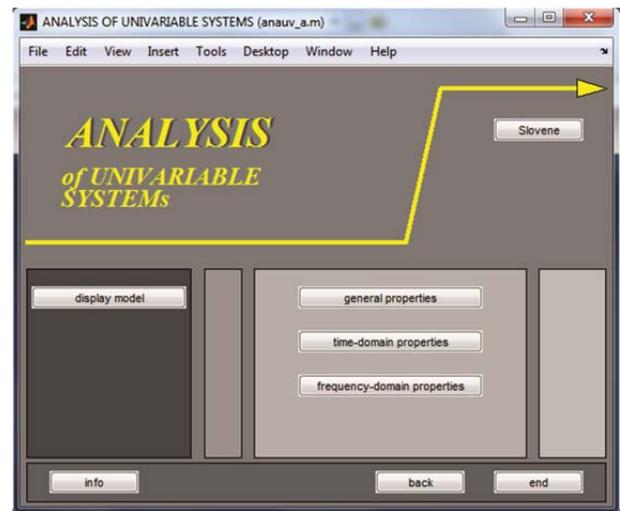


Figure 8: Analysis of SISO systems

Regarding different design situations both groups are organized into four levels.

Open-loop analysis functions are using only the information of linear system model. They are grouped into *general, time-domain, and frequency-domain properties*, as is for SISO systems illustrated in Figures 9, 10, and 11.

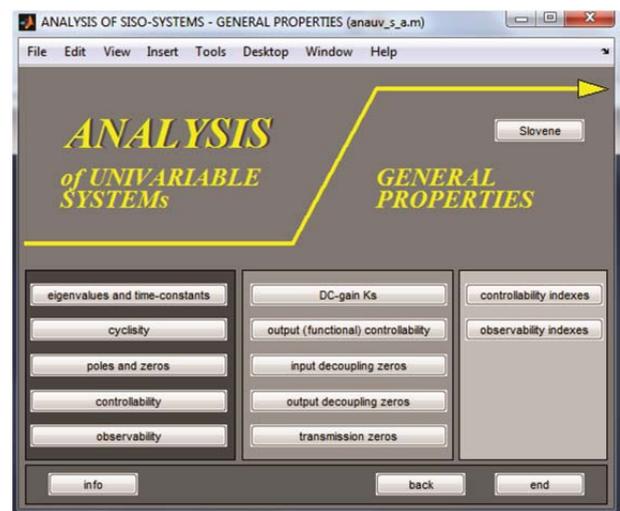


Figure 9: Functions available in the group of general properties

Closed-loop analysis functions are organized identically however in this case also the information of used controller is needed.

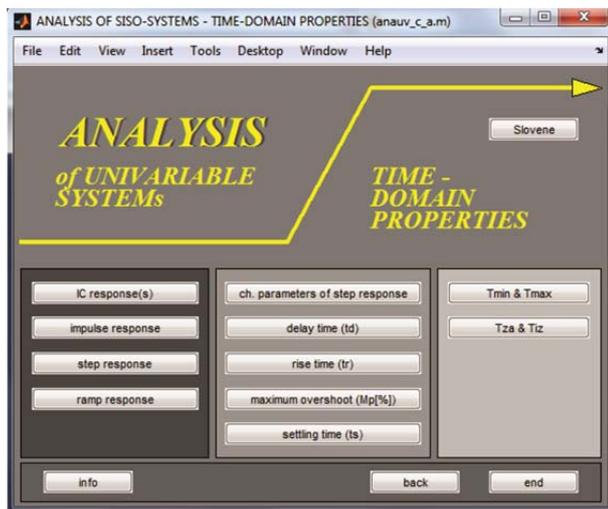


Figure 10: Functions available in the group of time-domain properties

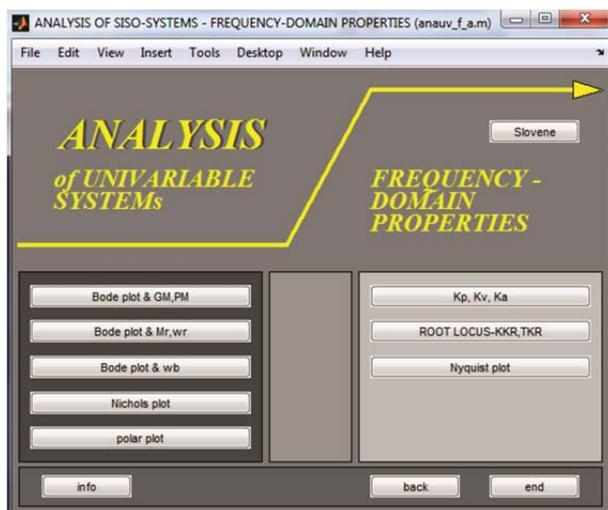


Figure 11: Functions available in the group of frequency-domain properties

Absolute validation functions are used for evaluation of matching desired design goals. In comparison with closed-loop analysis here also design goals have to be specified. Suitable and enough general specification of design goals can of course be very difficult especially in earlier design stages where needed information is not available or in situations where the user can define contradictory design goals. To avoid somehow these problems we have introduced the possibilities with which the user can define the importance of each specified design goal. In this way design goals can be specified either very precise or in a very approximate manner. The result of absolute validation is in the range between 0 and 1. If the result is 0 design solution is

unacceptable as one or more design goals are violated more than allowed. If the result is 1 this means that all design goals are completely satisfied. If the result is between 0 and 1 the solution is acceptable, but all design goals are not completely fulfilled. Better are of course solutions which are closer to 1. These results can also be used for some kind of relative validation. But it can occur that several solutions have the same validation result. In this case the next level can be used. Relative validation functions tend to help the user to compare the efficacy of different design solutions and to prepare him/her to the real situation where also different kind of non-linearities have to be expected. One, which is in practice always present, is for example limitation of control signals. Relative validation functions can simultaneously compare up to five design solutions. The validation result is calculated regarding the first solution. Result greater than 1 therefore means that solution is better than the first one. Higher validation values mean better or more efficient solutions.

It is important to mention also the button *students* which opens the graphical window for students work. Students are during laboratory exercises organized into groups of two persons and each group organizes their resulting files into a folder and they connect their files with the buttons of GUI. In this way documentation of their work is transparent and also potential problems are easier to be debugged.

The work flow through the semester is usually organized so, that first few weeks (5-6) all hours are used for the lectures and work in the classroom. When the first two chapters are explained lectures and exercises are running in parallel and at the end of semester lectures end earlier and students have time to finish all laboratory work and to prepare themselves to written and oral exams which are organized at the end of the semester.

Our experiences have indicated the following problems:

- Some students are not motivated enough because they have problems with combining abstract and practical knowledge.
- Some students like to postpone their study. This is problematic because they come to laboratory exercises unprepared and desired studying goals cannot be reached.
- Addition problem is that the knowledge of this subject is necessary already in the fourth semester (for example for the subject Systems and Control, see Figure 1) and online study is therefore very important.
- These students are relatively young and sometimes not capable to organize their work effectively.

So we have decided to slightly reorganize our work and have at the end of the semester introduced the so called seminars. These are small projects which are different for each student. If the student choose to solve one such seminar this work is evaluated as written part of exam. These seminars have improved the situation but we are still not satisfied. So, further improvements are

designed taking into account new possibilities which have been opened by introduced e-learning approach.

4. BLENDED LEARNING

As already mentioned we have introduced E-CHO into the studying process systematically in the year 2014/15 but some features of this program have been tested at UL-FE also earlier (Čorovič, Bešter and, Miklavčič 2009, Atanasijević-Kunc et al 2011, Logar et al 2011). From these testing experiments and of course from previous education experiences the following was learned:

- New teaching approaches are usually interesting for students and can attract their attention to certain degree, but young people are critical and their suggestions are important.
- E-learning is opening new and/or additional and/or complementary possibilities in education processes and it is important to investigate and to develop efficient forms of these new opportunities. It can be expected that all novelties are not equally efficient so it is important to evaluate them and to be prepared for further modifications if all effects are not positive.
- Established, efficient, and well accepted forms should not be eliminated (at least completely) from education process because of (especially unverified) novelties.
- Project work, especially if organized in small groups of students, is intensifying cooperation between the students and consequently also slower workers are forced to become more active.
- From well prepared projects also the reasons why certain knowledge is important can be become more convincing.
- Additional positive effect can be expected with the organization of work in a form of a game, perhaps even competition game. Natural reaction to competition is to try to be better (Matko, Blažič, and Belič 2001, Atanasijević-Kunc et al 2011).
- E-learning contents are in general available 24 hours per day and as such they introduce a certain degree of freedom into learning processes which is always an important advantage.
- It can be expected that a lot of effort and time is needed for teaching staff to prepare all new technical equipment and to become familiar with these new possibilities. New equipment also demands corresponding adaptation of materials and education flow.

Taking into account mentioned experiences we have started with the adaptations of lectures and exercises in the form as is illustrated in Figure 12. Each week the file (ppt-file) of corresponding lectures was added. This file was complemented by the file describing the homework. It is important to mention that it was not

checked if the given material was studied online and if homework was done. Later on (after the lecture no. 6) also instructions for laboratory exercises were uploaded. Matlab files were installed at the computers in the laboratory. At the end of semester also files describing project work were added. All the materials were generated in Slovene.



Figure 12: E-classroom of the subject Modelling and simulation: content development

It was observed that after few starting problems students have adapted to this new extension very promptly. It is interesting that practically all decided to download the prepared files in spite of the fact that they can be equivalently used also online.

In the last year different projects were prepared for each student. All students decide for this form of written part of exam.

After the first year the following was observed and/or suggested and/or already modified:

- The level of online study should be improved. To achieve this goal some kind of online knowledge verification is necessary. This solution was suggested even by the students (of course after the end of semester). This seems to be very important because learning efficacy of laboratory exercises depends crucially on this knowledge. As a consequence we have started with the preparation of short tests, which can now be relatively simply prepared and evaluated in E-CHO. Each lecture and corresponding homework is now followed by a short test verifying understanding of the presented material as illustrated in Figure 13.
- Some students needed quite long time to prepare results of project work and so also oral exams were postponed in these cases. This problem is not simple to solve, but perhaps positive results can be expected by the reorganization of project work (as will be described) in combination with previously mentioned short tests.
- Final average level of presented knowledge should and could be improved, especially regarding more ambitious students. It is usually a stimulating aspect if the achieved level (and with this also the current achieved mark) can be observed where at the same time

further steps are well defined and lead to more demanding final results.

- Students' questionnaires further revealed that even more practical work is desired.

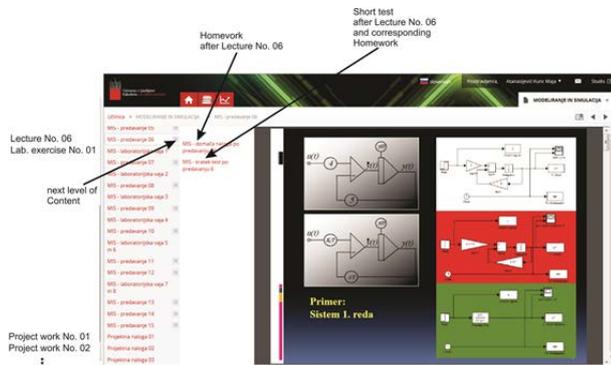


Figure 13: E-classroom of the subject Modelling and simulation: multi-level content development including lectures, exercises, homework, testing and project work

In the previous years we have developed remote experimentation using E-CHO learning environment, corresponding pilot plants and computers, web-cameras and Matlab with Simulink as described in details in (Logar 2011, Atanasijević-Kunc et al. 2011). In these cases the situation, as presented in Figure 4, was extended in a way that the computer connected to the pilot plant was as a slave computer through the LAN connected to the main server where E-CHO was installed. Students can start prepared experiments by uploading corresponding parameters. During experiment it is possible to observe video-stream of the pilot plant and after the experiment the responses can be observed and measurement data can be uploaded from the main server. In the same way also virtual experiments were realized.

It is important to notice the following:

- The guided work in the laboratory where the students obtain important basic information is very valuable and should not be replaced with other forms of education.
- Remote experimentation is convenient for additional projects, games and/or competition games, and demonstration purposes.
- Remote real experiments are efficient only when using suitable pilot plants. That means that they are robust, have corresponding (not too long) time constant and, if possible, the observation of the experiment can be informative for the user.
- Virtual experiments can be of great help, especially in the case of larger group of students. In such cases projects can be less time-consuming, different and yet similar enough, that they represent comparable burden for all students.

Regarding the importance and potential efficacy of project work we are adapting described technology also

to a new version of E-CHO where prepared experiments (virtual and real) can be accessed now through defined url-address which can be developed as one of the contents form inside E-CHO as was mentioned in the second section. This configuration is illustrated in figure 14.

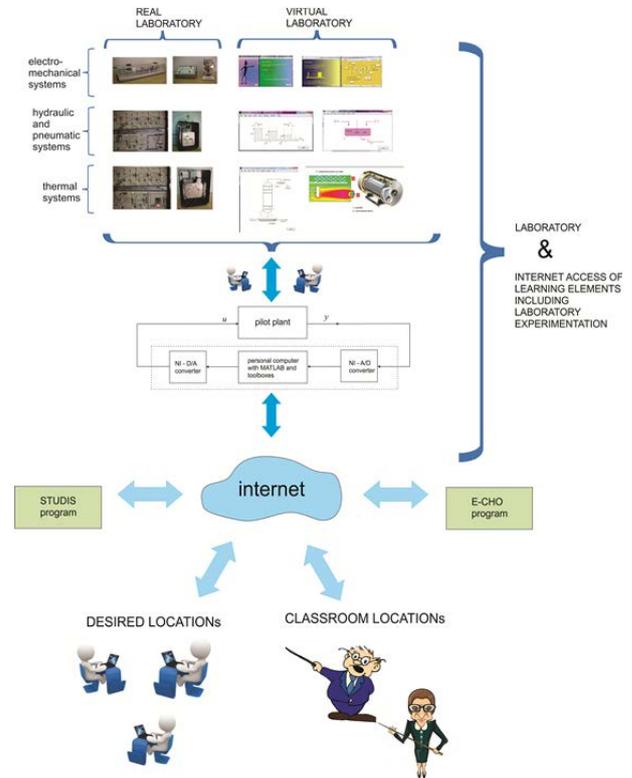


Figure 14: Extended laboratory work organization

For the next year the projects are prepared where the groups of students (each team will consist of two to three students) are instructed to:

- develop theoretical and/or experimental model of given problem,
- observe model responses in the whole operating range,
- choose five working points and develop corresponding linear approximations,
- analyze models' properties in all working conditions,
- and realize qualitative evaluation of models with measurements.
- Written reports and corresponding files are going to be uploaded into E-CHO regarding pre-defined timing.

Each project task realized correctly will be evaluated with corresponding number of points contributing to final mark, but also the evaluation of short tests will be taken into account.

On the basis of the evaluated efficacy of modified education and regarding numerous possibilities which are indicated in figure 14 further improvements will be designed.

One very important which will be realized as soon as possible is also realization of all content elements in English as each year we have a greater number of foreign students.

5. CONCLUSIONS

Through the last year we have introduced at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia, a new program called E-CHO with which additional e-learning approaches have been opened for teaching staff as well as for students. The important question arises from this situation: how to take the optimal advantage of this new program and how to combine its capabilities with the established approaches in an efficient way? Some of the solutions are presented in the paper for the subject Modelling and simulation.

In the future we are planning

- to investigate the options which would enable the introduction of interdisciplinary students' projects,
- and projects which would represent the bridge to innovative enterprises,
- we will further develop remote experimentation with pilot plans,
- and with additional complex mathematical models presenting real processes, like is described in (Fathi et al. 2015),
- more attention will also be given to video contents where it is planned to activate especially students.

It is important to mention that some questions and problems are still not solved completely, like for example:

- which lectures or contents should be visible for the staff and/or students of the faculty,
- which contents should be made visible to everyone,
- should teachers have the possibility to develop additional courses, or only their lectures,
- can this additional equipment stimulate life-long learning,
- or activities which are not part of the regular study,
- etc..

These problems represent the challenge for our future endeavors.

REFERENCES

- Atanasijević-Kunc M., Drinovec J., Mrhar A., 2008. Usage of modelling and simulation in medicine and pharmacy. *Slovenian Medical Journal*, 77(1), 57–71.
- Atanasijević-Kunc M., 2008. Process Modelling: Examples with illustrations in Matlab-Simulink. Faculty of Electrical Engineering, University of Ljubljana (in Slovene).
- Atanasijević-Kunc M., Logar V., Karba R., Papić M., Kos A., 2011. Remote multivariable control design using a competition game. *IEEE Transactions on Education*, Vol. 54, No. 1, 97–103.
- Atanasijević-Kunc M., Karba R., Zupančič B., 2011. LABI - Laboratory of Mathematical Models. *EUROSIM Simulation News Europe*, Vol. 22, No. 2, 38–42.
- Atanasijević-Kunc M., 2013. Modelling and simulation, Lectures. Faculty of Electrical Engineering, University of Ljubljana (in Slovene).
- Bicher M., Fitsch U., Atanasijević-Kunc M., Nagele N., Wild W., Breitenecker F., 2012. Modelling and simulation e-learning set of hydraulic models. 9th International Multidisciplinary Modelling & Simulation Multiconference, Vienna, Austria, 507–512.
- Cellier F. E., 1991. Continuous System Modeling. Springer-Verlag, New York.
- Cellier, F. E., Kofman, E. 2006. Continuous System Simulation. Springer Science + Business Media, New York.
- Close C. M., and Frederick D. K., 1993. Modeling and Analysis of Dynamic Systems. Houghton Mifflin Company, Boston.
- Čorović S., Bešter J., Miklavčič D., 2009. An e-learning application on electrochemotherapy. *BioMedical engineering online*, Vol. 8, No. 26.
- Čorović S., Markelc B., Dolinar M., Čemažar M., Jarm T., 2015. Modeling of microvascular permeability changes after electroporation. *PLoS one*, <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0121370>, doi: 10.1371/journal.pone.0121370.
- Fathi A., Saboohi Y., Škrjanc I., Logar V., 2015. Low computational-complexity model of EAF arc-heat distribution. *ISIJ international*, Vol. 55, No. 7, 1353–1360.
- Glavan M., Gradišar D., Strmčnik S., Mušič G., 2013. Production modelling for holistic production control. *Simulation modelling practice and theory*, Vol. 30, 1–20, doi: 10.1016/j.simpat.2012.07.010.
- Karer G., Mušič G., Škrjanc I., Zupančič B., 2007. Hybrid fuzzy modelling for model predictive control. *Journal of intelligent & robotic systems*, Vol. 50, No. 3, 297–319.
- Logar V., Karba R., Papić M., Atanasijević-Kunc M., 2011. Artificial and real laboratory environment in an e-learning competition. *Mathematics and computers in simulation*, Vol. 82, No. 3, 517–524.
- Logar V., Dovžan D., Škrjanc I., 2012a. Modeling and validation of an electric arc furnace. Part 1, Heat and mass transfer. *ISIJ international*, Vol. 52, No. 3, 402–413.
- Logar V., Dovžan D., Škrjanc I., 2012b. Modeling and validation of an electric arc furnace. Part 2, Thermo-chemistry. *ISIJ international*, Vol. 52, No. 3, 414–424.
- Löscher T., Breitenecker F., Mušič G., Gradišar D., 2005. Petri net modelling and simulation in MATLAB - a Petri net toolbox. *EUROSIM Simulation News Europe*, iss. 43, 20–21.
- Matko, D., Zupančič, B., Karba, R., 1992. Simulation and Modelling of Continuous Systems: A Case

- Study approach, Prentice Hall International Series in Systems and Control Engineering, New York.
- Matko, D., Blažič, S. Belič, A., 2001. Virtual race as an examination test : models, solutions, experiences. IEEE transactions on education, Vol. 44, No. 4, 342–346.
- Mušič G., Hafner I., Breitenecker F., Körner A., 2013. Learning Petri net dynamics through a Matlab web interface. 8th EUROSIM Congress on Modelling and Simulation, Cardiff, Wales.
- Papić M., Zebec L., Pogačnik M., Bešter J., Atanasijević-Kunc M., Logar V., 2011. Personalized learning environment E-CHO. EUROSIM Simulation News Europe, Vol. 22, No. 2, 17–24.
- Sodja A., Zupančič B., 2009. Modelling thermal processes in buildings using an object-oriented approach and Modelica. Simulation modelling practice and theory, Vol. 17, No. 6, 1143–1159.
- Štampar S., Sokolić S., Karer G., 2013. Predictive functional control of temperature in a pharmaceutical hybrid nonlinear batch reactor. Chemical industry & chemical engineering quarterly, Vol. 19, No. 4, 573–582.

SUSTAINABLE PERFORMANCE MANAGEMENT MODEL OF MANUFACTURING COMPANIES ACCORDING TO SAMPLED METHODS FROM PERFORMANCE MANAGEMENT SYSTEM

Veronika Burešová^(a), Lilia Dvořáková^(b)

^{(a),(b)}University of West Bohemia, Faculty of Economics, Department of Finance and Accounting

^(a)buresovv@kfu.zcu.cz, ^(b)ldvorako@kfu.zcu.cz

ABSTRACT

The paper focuses on the topic of research and development of a complex sustainable performance management model for medium-sized manufacturing companies operating in the Central Europe economies. The form of the model is determined by the most significant tendencies and phenomena with a worldwide reach influencing manufacturing companies (sustainable entrepreneurship, environment protection, social responsibility of a company, gender equality, national non-discrimination etc.) and by the results acquired by the performed primary research in the area of sustainable performance of manufacturing companies. The model is designed as a four-element one. The reason is that the competitive level of an enterprise in a long-term period is the reflection of the management ability to manage the company successfully, to guide its performance in the right direction and at a needed speed and to evaluate the intercompany results effectively by controlling methods and tools – a system of feedback and mutual causalities and dependencies.

Keywords: Sustainable entrepreneurship, performance management model, manufacturing company, empirical research

1. INTRODUCTION

Since the 20th century the behaviour and reactions of companies have dramatically changed in the area of industrial production and in the context of the world economy development. In the Central European countries the major part of the 20th century was marked by the protectionist policies, communication and geographic barriers which was connected to the fact that there was no significant pressure on the increase of productivity, economization and business activity efficiency, there was low pressure on improvement of managerial methods and processes or generally enterprise costs minimization (including manufacturing companies), to the contrary of what happens now. The increasing production costs were transferred to the customer as a result of the excess of production demand over production supply.

The continuing internationalization of environment and the progressing integration tendencies lead to blurring of boundaries, removing of barriers and therefore to the growing dynamism of information, capital, raw materials and workforce flow. As a result of increasing dependence of markets and national economies caused by high dynamics of trade with goods, capital, services and by transfer of technologies and know-how, the competition significantly increases, sometimes to such an extent that we often speak of so-called super- or hyper-competition (Drucker 1980 and 2011). In the second half of the 20th century (e.g. in the Czech Republic) these were the main reasons which caused the restructuring wave of production companies which were learning to do business in much larger and more intensive competitive environment than which they had been a part of so far. Technical and technological development, boom of information technologies and communication means (informatization) lead to processes acceleration. In the paper by Aziri, Zeqiri and Ibraimi (2014) Gary Dessler, Professor of Business at Florida International University's College of Business Administration in Miami, Florida (2013) summarizes the tendencies which the companies are forced to face over the framework of economic and financial crisis (Koshovets and Frolov 2014) and the areas which they mainly have to focus on in business management – see Figure 1.



Figure 1: Reflection of tendencies and phenomena with a worldwide reach in company's targets and in

management of the most valuable capital of a company – human sources

Source: Aziri, Zeqiri and Ibraimi (2014), adapted from Dessler (2013)

Also the performed operations have registered some changes. The work share of labourers and other company workers (and also the level of direct staff costs) has generally shown a decreasing trend in a long-term period. On the contrary the share of automated ongoing processes grows which leads to the growth of indirectly attributable (overhead) costs at the expense of directly attributable costs in the overall scale. A major part of the overhead costs is formed by the fixed costs. Compared to the past the costs of service, assistance, information, planning, inspection and strategically oriented activities are growing (e.g. research and development costs, production planning, organization and management costs, costs connected to quality checking, purchase, sale and advertising – marketing activities costs, accompanying services and after-sales service costs, ecological destruction of production after its lifetime expiration costs, costs of environment revitalization, area after mining etc.).

The customers' behaviour is also changing – not only the final consumers but also the buyers of intermediate products in the whole complex value-creating chain. The company's customers as a result of their higher awareness, thanks to the easier availability of information and communication technologies, as e.g. Poláková and Petrová (2013), Rolínek, Holátová, Březinová and Kantnerová (2011), Vollman, Berry and Whybark (1988) and others emphasize, „dictate“ to the company what, for whom and when to produce. The customers (purchasers) still more often assume that the goods will be delivered to them in the right (required) quality, for the best price, at the shortest delivery time, with a maximal after-sales service... Brychta, Benda, Knápková (2010) summarize the mentioned customers' expectations in three letters – C – Q – T (illustrating three crucial parameters of success – Costs – Quality – Time). Customers expect and require quality from their suppliers. This term no more means only the quality product or service, thus such product which is non-defective, working the whole lifetime period. Currently the terms quality and quality product are perceived in a much wider and more complex way. As Šoljaková (2009) states, „*quality is the ability to satisfy customers' expectations and requirements and is not related only to technical parameters but also to safety, health and environment friendliness.*“ The product which takes into account the quality requirements in the early 21st century should achieve following features beyond respecting the product character:

- ability to be „tailor-made“ to the wishes, needs, requirements and preferences of customers – The operations are still more individualized.;
- wide range (variety in offer) – It leads to a gradual expansion of structures of performed operations,

or higher specialization. The importance of outsourcing grows.;

- reliability, durability – probability that the product will be functional and will keep its declared features for all the guaranteed lifetime/use period;
- high quality grade including user friendliness, customer service, post-warranty services;
- quick response – Products, which enable it, are gradually started to be realized in still smaller but more often delivered quantities. The faster satisfying of customers' needs is reflected and leads to still clearer trends to shorten the lifetime period of offered products (Rudzki 2004).;
- practicality – easy operation and maintenance, or repairs, easy communication in a case of accompanying services necessity;
- perfect after-sales service;
- safety – damage risk which the product or its use can cause including negative health effects on the user and the environment (Veber 2002); (the fact whether during the production/providing there are no children employed – generally the issue of responsible behaviour to employees);
- correct pricing;
- aesthetic look of the production and the space where the business is done.

The following Table 1 summarizes the changes in behaviour of business entities towards customers as results of extensive changes which resulted in the change of the economic development paradigm itself.

Table 1: Comparison of companies' behaviour to customers in the 20th and the 21st century

20th century	21st century
The target of the business activities is to make the customer buy the offered product. The target is a single satisfaction of a customer.	The target of the business activities is to make the customer buy the offered product repeatedly. The target is a repeated satisfaction of a customer.
The contact with a customer is limited only to the time when the customer decides to buy the product. By making the sale the relationship business-customer ends.	The contact with a customer is direct and repeated. The relation with a customer continues even after the sale is made. The after-sale service is available to the customer and after-warranty services are offered to the customer.
During the product presentation the most attention is devoted to the advantages of the product – price, technical parameters, useful properties etc.	During the product presentation the most attention is devoted not only to the advantages of the product – price, technical parameters, useful properties etc. but

20th century	21st century
	especially to the value which the offered product can bring to the customer.
The level of customer services is limited.	The level of customer services is developed.
The production section is responsible for the quality grade.	All company is responsible for the quality grade.

Source: own processing, 2015

The changes of the market environment induce changes in the structure of business costs, influence business processes. Therefore it is necessary to adjust the whole complex system of business activities management and to modify the methods and tools of companies' performance measurement and management. Every company wants to achieve and keep long-term competitiveness, tries to stabilize its performance on an acceptable level and to stay attractive for investors, generally stakeholders. These targets can be achieved by innovations in the settings of company performance measurement and management system.

Although it can still be said that globalization is today's phenomenon, there are already authors who question the validity and the effect of this phenomenon and of the events accompanying it or caused by it. For example prof. Zelený (2014) states that *„the movement for which it was typical to substitute local companies by national, international and global political units has reached its limit and is transforming into the reverse movement. After globalization a long-term process of re-localization follows, we witness its acceleration today.“* As Odkladal (2014) or prof. Zelený (2014) state, there were thousands of re-localization technologies till 2015. It will not be necessary any more to transport material goods globally but only to transfer/flow digital information for local knowledge creation. The main motto of re-localization is: *„Think globally, act locally.“* The re-localization process is based on three pillars:

1. Self-service (outsourcing to customer). Self-service can take place only locally, individually and in a personalized way.
2. Disintermediation (removing useless interlinks) which means direct contact and communication between the provider and the user (producer and consumer). It is retreated from traditional dealers, agents, middlemen etc. The consumers merge with the producers.
3. Massive customization (individualization, personalization). Each product and service is „tailor-made“, using the internet in the local context: at first sell, then produce – not conversely. Customization is not related only to style, material, colour and quality but also to price (so-called Customized pricing).“

2. RESEARCH FOCUS, TARGETS AND METHODOLOGY OF THE PAPER PROCESSING

On the level of companies the adequate response and certainly the necessary reaction to the macroeconomic changes is to adjust the style – the settings (priorities) of business management. It is current and still more needed to deal with the topic of enterprise performance measurement and management. No progressive management systems can work without systems for enterprise performance measurement and feedback and efficient evaluation. The same old „what I do not measure – I do not manage“ is still true.

The business practice results of production companies e.g. in the Czech Republic instead of this usually show that they are partial problems of a less serious character which are often solved, solutions of a more non-strategic areas are searched which are not important for long-term success of a company. The company managements of successful companies seem to differ from those managements which did not survive the economic crisis of the last few years in a few areas. They digress from monitoring and tools for measurement (management) of short-term performance which were preferred to strategic development and were the easier solutions with short-term impact and lower efficiency in the financial management area, and now they prefer implementation of „just a few“ financial management methods and tools in the company performance measurement and management system. The result and the effect of this approach and these preferences is a more sensitive implementation of methods and tools of financial (economic) performance measurement and management system often with the target to fulfil the long-term ideas of the stakeholders, expectations of the owners (shareholders) of the company.

In the enterprise performance management context on one hand it is necessary to adjust the known and developed theoretical apparatus to the new conditions but on the other hand it is appropriate to enrich the used methods and implemented tools by brand new approaches. This is confirmed by e.g. prof. Zelený or by the original form of Carl Gustav Jung who wrote: „a man keeps forgetting that what used to be good once, does not stay good forever. He takes the old paths which proved themselves good a long time after they became bad ways; only with the highest sacrifices and with immeasurable efforts he manages to accept that what used to be good has got old and it is not good any more. It happens to him in small and large scale too.“ In other words: Let's do what is necessary and not what we have always done so far.

Only individual persons entrepreneurs, business corporations, public institutions – subjects of the profit and non-profit sector with a well stated, implemented and regularly corrected vision and thence derived targets, strategies and system of company performance management supported by human capital and information system, will not weaken their market

position, will keep their prosperity and will still be successful with their core business. And they will also be competitive in comparison with their local and international competitors (Souček 2010) „in the period of the fastest and the deepest transformation of production, consumption, values and behaviour – lifestyle of individuals, groups and nations – the age of entrepreneurship“ (Zelený 2015).

The so far known concepts, methods and tools of company performance measurement and management get a new form, also new ones are developed. Generally there is the need to create a new complex model of company performance management. Although there has been an intensive demand for a new company performance management model for medium-sized production companies in financial theory and also company practice for several years, the beginnings of efforts to create it are of a much later date. Therefore the target of this paper is to present a system solution of sustainable development in the form of a sustainable performance management business model proposal for medium-sized production companies operating in open economies of the Central Europe which would integrate the three basic pillars – economic, environmental (ecological) and social dimension of entrepreneurship performance. And to define its assumptions.

2.1. Methodology of the paper processing

The authors' work is based on critical literary search, especially scientific studies and papers to the topics associated with sustainable performance management of a company, social responsibility of firms which were published in professional magazines, monographs and in international scientific conferences proceedings. The theoretical approaches evaluation within the qualitative research by the desk research method in the area of enterprise performance management systems models is enriched by the results of the empirical – quantitative research.

The questionnaire survey method was used for the data collection within the quantitative research. The data were collected in the period autumn – winter 2014. The key criteria for business entities, their sorting, consequent selection, addressing and starting of cooperation at the research were:

- size in terms of employees number (according to European Commission, 2005 and Commission Recommendation 2003/361/EC),
- headquarters and place of their operation – running the business (production),
- object of operation and
- legal form of business.

The specifically targeted enterprises were those of medium size (enterprises of 50 to 250 employees including), with place of operating in the Czech Republic (specifically in two regions - Plzeň and Karlovy Vary region), focused on production and the legal form of business in the form of a limited liability

company or a joint stock company. The data of thoroughly completed questionnaires were collected in total from 128 business entities in the Czech Republic, the country which is one of the countries in the Central Europe which is intensively pro-export oriented and has so far less liquid capital markets in comparison with the West European countries. Considering the manner of the researched issue in the area of the business sector in the Czech Republic, the financial department workers were questioned. The questionnaires were filled in by mainly controlling section workers, higher management (financial directors), executive managers or financial accountants.

3. SUMMARY OF THE QUALITATIVE RESEARCH RESULTS IN THE AREA OF COMPANY PERFORMANCE MANAGEMENT SYSTEM MODELLING

For the production companies in the Central Europe it is critical to manage to create such environment with convenient conditions for dynamic reaction to the market environment changes, starting a high speed of company growth and achieving stable competitiveness. It is necessary for the business entities to manage not only their economic side of business (although it still remains essential) but also to perform their social function and to implement the principles of environmentally responsible behaviour (environmental friendly production) – not just formally. The social and environmental dimension of entrepreneurship (social and environmental aspect of performance and sustainable entrepreneurship) becomes still more important for keeping and strengthening performance, prosperity, success or in other terms also the competitiveness of a company.

Although company performance and competitiveness are common topics of discussions and debates on or out of the academic ground, it is not possible to find their unified, generally respected definition. The reason for the variety of the company performance and competitiveness concept can be found in the different perception of different groups of interest which comes from their expectations, existing experience, set targets etc. Furthermore these terms are not usually defined by the authors; they just implicitly assume that the meaning of these terms is generally known. This is one of the reasons for misunderstanding, errors and mistakes in a case that the same terms are used at different understanding of their meaning, which however is not a unique phenomenon. Often, but not completely correctly the terms performance and competitiveness of a company are interchanged with the terms like efficiency, success, competitiveness, economy, stability, prosperity, productivity etc. In this paper competitiveness is perceived as Jirásek (2001) states, as „*market potential of a company in the clash about the market position with other companies – the ability to make offensive pressure on the competitors and at the same time to face the competitors onslaught.*“ The result of such clash is considered to be

„the ability to achieve good market position and to ensure corresponding prosperity on that level for at least a short or medium period“. Competitiveness is understood as a wider concept which covers the term for company success and short-term performance of a company. The company performance is then one of the basic assumptions for company's long-term existence. (Jirásek 2001). The level of company's success is the reflection of the ability to react to external influences; the company's performance reflects the potential and economic results of the company's internal environment. Competitive ability and advantage are changing in character, they have no static characteristics which premiums that they, as a «consequence» of numerous relations and correlations of internal and external factors, have to be approached in a highly responsible manner, with a lot of forethought and necessity to balance economic, as well as social, ethical, and ecological dimensions, which undoubtedly affects the increase in competitiveness (Kutnjak 2014).

The company's performance will further be understood as the ability of a company to use the available sources at the best so as the clearly defined targets are achieved in a short period. These can be the maximal valorisation of input sources, the maximization of produced profit, company's values – generally building firm basis of the company's success on the market and ensuring the ability to survive in the future in view of global competition. An interesting concept which emphasizes the significance of human capital is presented by e.g. Pitra (2001) who defines the company's performance as „the result of the organized activity of specific people who joined their strength and means to achieve clearly defined targets“. Further on he adds: „Business performance is not only an economic demand but it is a demand of all society, and the people who make decisions about its level – employees, managers, owners, are personally responsible for its achieving.“

According to the authors of the paper, the main target of a company should be the maximization of the company's goodwill. The company's goodwill is seen here not as an often stated sum which is the difference between the accounting valuation of a company and the market valuation of a company, but as an overall result of the company's long-term operation on the market. It reflects the position of the company on the market, the tradition and reputation in the eyes of the main customers of the company. The company's goodwill is determined by these factors:

- the quality of created production (the products are expected to have a perfect quality grade, appropriate technical and artistic level, external look and other qualities which lead to gaining the popularity with the customers on the market – HPCG 2015),
- the quality of workforce of the company,
- convenient location of the company or the sales network of the company,

- level of relations with public, with stakeholders, (given by long-term social responsibility, environmental friendly production, enforcing equal opportunities – in gender, nationality areas),
- level of marketing activities.

Since the second half of the 20th century the principles of sustainable development and sustainable entrepreneurship have been enforced especially by international organizations measures of which the loudest sound the interventions of the United Nations, Organisation for Economic Co-Operation and Development, European Union and World Bank. As Isada F. and Y. (2014) state, sustainability is (according to Hart 2007) a key word „in shareholder value, which leads to a sustainable society.“ The problem is that nor this term is sufficiently firmly anchored in the financial theory. „The umbrella term sustainability covers many different ways of thinking, issues and concepts, and modes of practice.“ For the purpose of this study the term sustainability is defined as „the long term capability of well being by encompassing the responsible management of all resources, which results in meeting the needs of the present without compromising the ability of future generations to meet their own needs.“ The companies in their own interest should use the tools of sustainable development. For now most of them are on a voluntary basis. One of the ways how not to lose, or even get the trust of customers, is for the companies (especially production companies) to have a responsible approach to sustainable development (corporate social responsibility, further just CSR). This concept includes environmental, social and economic areas, by accepting this concept the company can increase its credibility. There so-called 3Ps, three spheres – profit (economic sphere), people (social sphere) and planet (environmental sphere). In practice it means that companies set high ethic standards, look after their employees and minimize negative impacts on the environment. „According to the Meadows⁴ and Randers (1995, p. 211–212) there are three ways how the human world can react to the signals that exploiting resources and harmful emissions have crossed their permanently sustainable limits. The first reaction is to hide these signals, refuse or obfuscate. The second reaction is to ease the pressures from the limits by technical or economic measures. The third possibility is to step back and realize that the human socioeconomic system is beyond control, that it crossed its limits, that it is going to collapse and therefore it is necessary to change the system's structure.“ (Loužek 2013, p. 395) In the corporate practice of the Central Europe economies reaction number 2 prevails. The problems associated with the impacts of industrial production on the environment are not taboo any more, they are spoken of more and possible solutions are searched, which would help to „ensure the needs of current generations without threatening to meet the needs of future generations and

without acting at the expense of other nations.“ (Ministry of the Environment 2014).

4. RESULTS OF THE QUANTITATIVE RESEARCH IN THE COMPANY SUSTAINABLE PERFORMANCE MANAGEMENT SYSTEMS AREA

The target of the performed empirical research was to find out about the awareness and level of use of the financial theories known concepts, methods and tools for company sustainable performance management. The data acquired by the empirical research and associated results reflect the positive attitude of the researched sample to company sustainable performance management. As a substantial component of company performance management they perceive the methods, tools and indicators which can be used for environmental and social performance management. But still the medium-sized production companies in the Czech Republic use the most the approaches based on financial scales which indicate the level of financial performance of a company. The aspects of sustainable entrepreneurship and sustainable performance are not systematically monitored and evaluated. It was found out that medium-sized production companies in the Czech Republic use the most the traditional methods for evaluation of the company's financial health – i.e. financial analysis in its narrower concept (in structure – see Table 1) which includes especially horizontal and vertical analysis and quantification of ratio indicators. It happens although the companies are aware, at least partly, of the limitations, pitfalls and disadvantages of this approach to company performance measurement.

Table 1: Summary of answers to the question: „In the context of company's financial performance measurement do you use:

(It was possible to cross more than one item.)

Item	Number of respondents in a sample
Traditional indicators of financial analysis (indicators like profitability, activity, liquidity etc.)	128
Traditional indicators (ratios) based on determinants of working capital and cash flows (e.g. net working capital)	79
Horizontal and vertical analysis of reports, especially of balance sheet and profit and loss financial statement of a company	59
Pyramidal analysis of profitability indicators (DuPont analysis of ROE or ROA indicator)	43
Pyramidal analysis of Economic Value Added indicator (EVA)	10
Indicator like Economic Value Added and Market Value Added	48
Economic Value Added	44

Item	Number of respondents in a sample
(EVA)	
Market Value Added (MVA)	4
Profitability based on cash flows (e.g. CF ROI)	5
Bankruptcy and credibility models quantifying the probability of decline/termination of entrepreneurship	55

Source: own processing, 2015

A large contradiction between financial theory and corporate practice in the Czech Republic can be observed e.g. in the case of the pyramidal decomposition of Return on Equity (or Return on Assets) ratio or the Economic Value Added indicator. Pyramidal system belonging to the modern methods of analysis comes from the retrospective point of view. Nevertheless it also provides management staff with the information allowing to the prospective view. In sum approximately in only one third of companies from the research sample it finds its use as a tool for managers in the field of internal business analysis within comparison between the actual and the desired values of the parameters. Nevertheless in the Czech Republic there e.g. Neumaier's (2008) believe that "the concept of a system of indicators with the character of the pyramid is just the most convenient for the present, which is as a result of globalisation characterised by a high degree of dynamic complexity."

The research results prove the general awareness of the possibilities which the modern (advanced) approaches to the company performance management bring, such as Balanced Scorecard (BSC), Economic Value Added (EVA), Market Value Added (MVA), Cash Flow Return on Investment (CFROI), Activity Based Management (ABM/ABC) etc. However they are perceived as too complicated and demanding for implementation into the company environment and processes.

Table 2: Summary of answers to the question: Do you use the below listed concepts for the measurement and management of performance of your company? (It was possible to choose more than one item.)

Item	Number of respondents in a sample
Value Based Management (VBM)	59
Total Quality Management (TQM)	76
Management based on activities, operations and (ABM/ABC)	43
Investment management	88
Strategic controlling	41
Operative controlling	99
Environmental managerial accounting	89

Item	Number of respondents in a sample
Balanced Scorecard (BSC)	59
Strategic maps	11
European Foundation Excellence Model (EFQM)	44
Six Sigma	9
Performance prism	1
Continuous improvement	47
Lean management	41
Tableau de Bord	0
Others (please specify.....)	0

Source: own processing, 2015

5. PROPOSAL OF SUSTAINABLE PERFORMANCE MANAGEMENT MODEL OF MANUFACTURING COMPANIES IN SMALL OPEN ECONOMY ENVIRONMENT

Considering the great interest in new performance management systems, especially because the traditional systems for company performance measurement and management often focus just on financial indicators which however fail as a result of the highly dynamic and turbulent environment, the target of the paper is to propose a model for management of sustainable performance of a company. In terms of the theory and practice needs it seems needed to research the issue of performance measurement and management systems development (Kennerley and Neely 2003) and to propose a model for company performance management which would be easily implemented, efficient and effective for medium-sized companies operating in the environment of a small, open, pro-export oriented economy. Companies defined by European Commission as large (according to European Commission 2005; Commission Recommendation 2003/361/EC) already have their efficient methodology, whereas innovation introduction in the area of company performance measurement and management (new methodologies of implementation, new model) would further burden small and micro companies in terms of the range of activities.

There have been lots of discussions (Henri 2003, Behn 2003) and a few attempts to create a suitable performance model (e.g. Carroll 1979; Bartoš). Bartoš states, based on purposes identification which is necessary for measurement (management) of performance and ideal qualities of integrated system for performance measurement, that „none of the existing models or approaches completely meets the requirements“.

The model presented in this paper (it will be improved within authors' further research and its functionality will be verified in Czech enterprise praxis by a case study research form) was proposed based on the synthesis of knowledge acquired by studying and analysis of secondary sources and data and also of the performed primary research results. The target of the proposed model of sustainable performance

management of a company is to direct a medium-sized company towards strengthening and consequent keeping of sustainable performance and therefore its long-term competitiveness by the model's implementation in the company practice.

The model proposal was inspired by the strategic maps of authors Kaplan and Norton (2000), by the proposal of integrated model for company performance measurement and management by authors Knápková, Pavelková and Chodúr (2011) and by Pollak's methodology, who based it on the experience with unhealthy companies rehabilitation in West Europe and proposed a new method of companies' viability evaluation based on evaluation of the ten following significant parameters:

- cash flow sufficient for adequate coverage of interests for invested capital,
- sufficient satisfaction of customers, shareholders, creditors, employees, suppliers and the area where the company operates,
- stable base of satisfied customers,
- products/services corresponding to the market requirements,
- continuous market research whose results the company implements into new products/services corresponding to the changing requirements,
- training of its qualified and motivated staff,
- optimal capital structure,
- capable, reliable and innovative suppliers of raw materials, products and services,
- operate at a convenient place,
- positive attitude to the environment.

Pollak noted that "the future viability of an enterprise is determined by mainly exogenous (non-firm) factors which in the company accounts do not come into account". Referring to the monograph of Measuring Performance in the Organisation of the Future, Nolan Institute, USA, 1990, Pollak with monograph's authors argues that "the corporate accounting rather obscures than supports the future of the company." That's why the authors' model proposal (within model's assumptions) was inspired by mentioned Pollak's significant parameters.

The proposed reference model of sustainable performance management of a medium-sized production company is built on four main components. These are:

- management built on the idea of Sustainable Balanced Scorecard (SBSC),
- controlling activities based on so-called sustainable controlling,
- sustainable performance management of a company,
- benchmarking.

The model is based on the idea that the level of competitiveness of a company in long term is the reflection of the management's ability to manage the company successfully, to guide its performance in the right direction and at an appropriate speed and to evaluate effectively the company's internal results by controlling methods and tools – feedback and mutual causalities and dependencies system.

It is generally known that what cannot be measured, cannot be managed. This managerial axiom is sometimes used in a modified version, i.e. „What we do not measure, we cannot improve.“ The meaning is still the same. A significant part of the future success of a company is the human capital and innovative abilities of the workers, the level of management – managerial processes and the manager's personality in the company's performance management model. The main tasks of the company management are primarily to create such environment where individuals together carry on the effectively defined targets – to achieve maximal outputs and use with optimal structure and composition of sources. Šilha (2006) adds to this that „company management had to start looking for methods ensuring that company's survival and profit will not be a matter of chance or one person's skill but it will be possible to control it“.

Figge, Hahn et al point out that „environmental and social aspects can be subsumed under the four existing BSC perspectives like all other potential strategically relevant aspects (Deegen 2001; Epstein 1996; Figge et al 2002). This means that environmental and social aspects are integrated in the four perspectives through respective strategic core elements or performance drivers for which lagging and leading indicators as well as targets and measures are formulated (Kaplan and Norton 2000)

5.1. Model assumptions

The model of sustainable performance of a medium-sized production company is based on the following assumptions:

- Assumption of continuous lasting of the accounting unit (going concern).
- The company representatives are aware of the costs associated with respecting environmental and social perspectives of entrepreneurship. They act systematically not randomly in the environmental and social areas. They expect that the revenues of such activities will be taken into account by the company stakeholders to such extent that the benefits outweigh the costs connected to the environmental protection and activities in the social area.
- The corporate culture reflects in everyday reality the company activities associated with environment protection and activities in the social area, which assumes that the company workers were introduced to the sustainable entrepreneurship rules and behave according to

them. (Sustainable entrepreneurship is a journey, not a goal. (Pascual et al 2011))

- The company climate is pro-active to the changes in the sustainable entrepreneurship area.
- If there are possibilities of new activities or improvement of the current ones in the area of sustainable entrepreneurship of a company, a senior worker deals with the presented proposal, checks the possibilities to apply it and the benefits to improve the company's results and finally he rejects the proposal, suggests its completing or reworking or its approval.
- The systems of environmental protection and social area activities are included and based on the strategic documents of a company.
- The company's target is maximalisation of long-term economic result (so-called sustainable profit).
- A company can offer something useful to the others and to fill them with enthusiasm about it. (It's true that no core endures forever (Zook 2007).)

Based on the study of professional sources and knowledge from production companies' practice, it can be proposed the following composition of steps which can be used for implementation of the model in a production company practice (Bartoš 2006):

1. Evaluation of the company's readiness, formulation and approval of changes.
2. Creation of organization framework, personnel shield and information providing of changes.
3. Analysis and completing the basic strategy documentation.
4. Deriving the key strategic targets and proposing the measurement quantities - indicators.
5. Setting the target indicators values and proposing of strategic operations set.
6. Unification of system measurements for ISO standards and BSC.
7. Extending BSC implementation to all organization units and management levels – creating a consistent and internally interconnected system of criteria for company performance measurement and management.
8. Ensuring fluent integration of BSC into the company's management system.

6. CONCLUSION

If a company wants to „keep up with“ the high dynamics of the economy, an essential assumption is to continuously work on making the company's performance management system more effective. It means the right setting of priorities, functionalities and respecting the causal dependencies in the whole system, precise checking, regulating in the desired direction. For business entities at the beginning of the 21st century it is not enough to be a unit whose only target is to satisfy the needs of the entrepreneur/its owners by maximalisation of the short-term accounting profit. To

make a company successful in the growing competition fight, it has to fulfil more or less also its social function and follow the environmental rules of responsible behaviour which must be evaluated in terms of effectiveness, efficiency and economy. In spite of all the displeasure and protests brought by the environmental summits, conferences, aroused by published news and government recommendations and directives on environmental topics, the social and environmental dimensions of entrepreneurship play a still more significant role in company management in the early 21st century, still more significance is laid on them for keeping and strengthening of company's performance, success and long-term competitiveness in the global context.

The paper reacts to the need of financial theory and business practice to propose a model of sustainable performance management of a company.

„People don't need big cars, they just need respect. They don't need closets full of clothes, they need to feel attractive and they need excitement, variety and beauty. People don't need electronic entertainment, they need to do something worthwhile with their lives...People need identity, society, motivation, appreciation, love and joy. Trying to meet these needs by material things means to develop an unquenchable appetite for false solutions of real and never consolable problems“ (Meadows et al 1995, p. 234).

ACKNOWLEDGMENTS

The paper was made within the project SGS-2015-021 Development of financial management approaches as a tool of company value growth.

REFERENCES

- Aziri B., Zeqiri I., Ibraimi S., 2014. Human Resource Management in Contemporary Business Organizations: A Literature Review. *Economy & Business Journal of International Scientific Publications* 8: 815-821.
- Bartoš V. Model integrovaného systému měření výkonnosti, [online]. [accessed 30 June 2015] Available from: <http://www.google.cz/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCEQFjAA&url=http%3A%2F%2Fwww.vse.cz%2Fpolek%2Fdownload.php%3Fjnl%3Deam%26pdf%3D45.pdf&ei=EN2TVY7iF4G5UbvUveAN&usg=AFQjCNGJ-2ncw4wPQMGIJK5kSnx70KeXGA&bvm=bv.96952980,d.d24>.
- Bartoš V., 2006. Principiální postup implementace systému měření výkonnosti v českých podnicích, Sborník z 3. mezinárodní konference Řízení a modelování finančních rizik Ostrava, VŠB-TU Ostrava, Ekonomická fakulta, 6 – 7 September 2006, 24-33.
- Behn R.D., 2003. Why Measure Performance? Different Purposes Require Different Measures. *Public Administration Review*: 63(5). pp. 586-606.
- Brychta K., Benda R., Knápková A., 2010. Creating a Model for Business Performance Assessment - a background. *Economics and Management*: 3.
- Carroll A.B., 1979. A Three-Dimensional Conceptual Model of Corporate Performance. *Academy of Management Review* 4(4): 495-505.
- Commission Recommendation 2003/361/EC, the Official Journal of the European Union L 124, p. 36 of 20 May 2003
- Deegen T., 2001. Ansatzpunkte zur Integration von Umweltaspekten in die „Balanced Scorecard“. Lüneburg: Center for Sustainability Management.
- Dessler G., 2013. *Human Resource Management*, 13th edition, New Jersey: Pearson.
- Drucker P.F., 2011. *THE AGE OF DISCONTINUITY: Guidelines to Our Changing Society*, 9th printing, New Jersey: Harper&Row.
- Drucker P.F., 1980. *Managing in Turbulent Times*, Routledge: New York.
- Epstein M. 1996. *Measuring corporate environmental performance: best practices for costing and managing an effective environmental strategy*. Chicago: Irwin Professional Publ.
- European Commission, 2005. *The new SME definition. User guide and model declaration*, Enterprise and industry publications, [online]. [accessed 20 June 2015] Available from: http://ec.europa.eu/enterprise/policies/sme/files/sme_definition/sme_user_guide_en.pdf
- Figure F., Hahn T., Schaltegger S., Wagner M., 2002. *The Sustainability Balanced Scorecard – Theory and Application of a Tool for Value-Based Sustainability Management*, [online] Available from: <http://www.ecnc.org/uploads/documents/the-sustainably-balanced-scorecard-theory-and-application-of-a-tool-for-value-based-sustainability-management.pdf>
- Hart S.L., 2007. *Capitalism at the Crossroads: Aligning Business, Earth, and Humanity*. New Jersey: Wharton School Pub.
- Henri J.-F., 2003. *Performance Measurement and Organizational Effectiveness: Bridging The Gap*. [online]. [accessed 20 June 2015] Available from: https://www.google.cz/search?q=Measuring+Performance+in+the+Organisation+of+the+Future+pdf&ie=utf-8&oe=utf-8&gws_rd=cr&ei=PJKvVcuPFcbUbnutqgL.
- HPCG, 2015. *HPCG: tvoříme goodwill. Definice goodwillu*. [online]. [accessed 20 June 2015] Available from: <http://www.hpcg.cz/stranka/cz/28/goodwill/>.
- Isada F., Isada Y., 2014. An Empirical Study Of A Sustainable Strategy And Profitability In The Electrical-Manufacturing Industry, *Journal of International Scientific Publications* 8, *Economy & Business*: 282-295.
- Jirásek J., 2001. *Konkurenčnost – Vítězství a porážky na kolbišti trhu*, Praha: Professional Publishing.

- Kaplan R.S., Norton, D.P., 2000. *Balanced Scorecard. Strategický systém měření výkonnosti podniku*. Praha: Management Press.
- Kennerley M., Neely A., 2003. Measuring performance in a changing business environment, *International Journal of Operations & Production Management*: 23(2), pp. 213-229.
- Knápková A., Pavelková, D., Chodúr, M., 2011. *Měření a řízení výkonnosti podniku*. Praha: LINDE.
- Koshovets O.B., Frolov I.E., 2014. The Current Financial And Economic Crisis As A New Stage Of Transformation Of The Global Economy. *Economy & Business Journal of International Scientific Publications* 8, ISSN 1314-7242, pp. 399-412.
- Kutnjak G., 2014. Reflection of Socially Responsible Operations on the Increase In Competitiveness, *Journal of International Scientific Publications, Economy & Business*:8.
- Loužek M., 2013. Global Problems as Seen by Environmental Economics. *Political Economics* 3, 393-410.
- Meadows D.H., Meadows D.L., Randers J., 1995. *Překročení mezí – konfrontace globálního kolapsu s představou trvale udržitelné budoucnosti*. Praha: Argo.
- Ministry of the Environment, 2014. Sustainable development. [online]. [accessed 20 June 2015] Available from: http://www.mzp.cz/cz/udrzitelny_rozvoj.
- Neumaierová I., Neumaier I., 2008. Why index IN was accepted and pyramidal metrics system INFA not?. *Economics and management*:4.
- Pascual O., van Klink A., Grisales J.A.R., 2011. *Create Impact! Handbook for Sustainable Entrepreneurship*. Enviu
- Pitra Z., 2001. *Zvyšování podnikatelské výkonnosti firmy: strategický obrat v podnikatelském chování*. Praha: Ekopress.
- Poláková I., Petrová M., 2013. CSR – trend dnešní doby. *CSR forum* 2: 30-31.
- Pollak J., the publishing year unknown. How to assess the vitality of business. [online]. [accessed 20 June 2015] Available from: http://www.google.cz/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=0CCsQFjADahUKEwjaxK73wZPGAhWFPBQKHVneAFM&url=http%3A%2F%2Fnb.vse.cz%2F~synek%2FMethodika_Pollak.doc&ei=oLhVdr0EIX5UNm8g5gF&usg=AFQjCNGYGDHtpguqUeVCmdqcwJDWbZWo1Q&bvm=bv.96041959,d.d24
- Rolínek L., Holátová D., Březinová M., Kantnerová L., 2011. Behaviour of small and medium-sized enterprises in terms of their strategic management and their goals, *Mathematical Methods in Engineering and Economics*: 92-99.
- Rudzki R.A., 2004. The Advantages of Partnering Well. *Supply Chain Management Review* 8(2): 44–51.
- Souček Z., 2010. *Firma 21. století (Předstihneme nejlepší!!!)*. 2nd edition. Professional Publishing.
- Šoljaková L., 2009. *Strategicky zaměřené manažerské účetnictví*. Praha: Management Press.
- Veber J. a kol., 2002. *Řízení jakosti a ochrana spotřebitele*. Praha: Grada Publishing.
- Vollman T.E., Berry W.L., Whybark, D.C., 1988. *Manufacturing Planning and Control Systems*, Illinois: Irwin.
- Zelený M., 2011. *High Technology and Barriers to Innovation: From Globalization to Relocalization. Finance and the Performance of Firms in Science, Education, and Practice*, 629-641. 28 April 2011, Zlín, Czech Republic.
- Zelený M., 2014. P.F. 2014: A new world of technology – from globalization to re-localization. *Aktuálně.cz*. Available from: <http://blog.aktualne.cz/blogy/milan-zeleny.php?itemid=21950> [accessed 01 January 2014].
- Zook C., 2007. *Finding Your Next Core Business*. Harvard Business Review: April 2007.

AUTHORS BIOGRAPHY

Veronika Burešová has worked as an assistant at the Department of Finance and Accounting of the Faculty of Economics at the University of West Bohemia in Pilsen since 2014. Since 2012 she has actively published the results of her research activity for the national and international professional community. The author's main research area is the problems associated with sustainable performance management of companies.

Lilia Dvořáková works as the Head of the Department of Finance and Accounting at the University of West Bohemia in Pilsen in the Czech Republic. She works as a tutor of students of the doctoral studies in the Company economy and management area. Her research activity is focused on the area of measurement and management of the financial position and performance of a company.

USE OF DELPHI METHOD IN OBTAINING REQUIREMENTS FOR THE GAME STRUCTURAL MASONRY

Ailton Soares Freire ^(a), Leiliane Santana Souza ^(b), Daniela Matschulat Ely ^(c), Antônio Edésio Jungles ^(d)

^(a)Instituto Federal de Educação, Ciência e Tecnologia do Piauí

^(b)Instituto Federal de Educação, Ciência e Tecnologia da Bahia

^(c) Centro Federal de Educação Tecnológica de Minas Gerais

^(d)Universidade Federal de Santa Catarina

^(a)ailton.freire@ifpi.edu.br, ^(b)santana.leiliane@gmail.com, ^(c)daniela.ely@gmail.com

^(d)ajungles@ceped-ufsc.com

ABSTRACT

The article presents the achievement of the project requirements for the development of learning material from the formulation of a game which simulates the management of a work in the building system in structural masonry. The methodology for obtaining these requirements goes through literature survey and application of the Delphi method with a group of 12 specialists (Professor, Researcher and Game Designer Professional) who develop their professional activities in this building system. The game will serve to qualify the students of civil engineering undergraduate courses, production engineering and architecture that will act as work managers in this system.

Keywords: simulation, education, Delphi method, requirements, serious games

1. INTRODUCTION

The quality of instruction of the civil engineer will have a positive impact on the success of his/her professional career, for this, he should use learning tools that combine theoretical training and professional practice.

One of the possible ways to get to this combination - theory and practice- is to use Serious Games in student training.

The term "Serious Game" is used to identify the games that go beyond entertainment and reach the learning and training level.

This paper seeks to develop a tool for qualifying undergraduate students in Civil Engineering as part of building system subject in structural masonry, serving to establish a link between the theories taught in the classroom and the practice of management.

The game will be developed to serve as a complement to the student training in the subject of work management in undergraduate courses of Civil Engineering, Architecture, Production Engineering and it will focus on constructive system in Structural Masonry.

The game setting is a simulation of a construction containing several buildings of two decks each, with four apartments per floor. The players are divided into teams containing one construction manager and the general supervisors in charge of operate the procurement of materials, plan the different stages of execution, and organize the building site.

The team which has a better cost/benefit development, planning and management wins the game.

The development of the research follows the steps shown in Figure 1, performing the stages to its conclusion with the implementation of the instructional design proposed, the development of the game Structural Masonry.

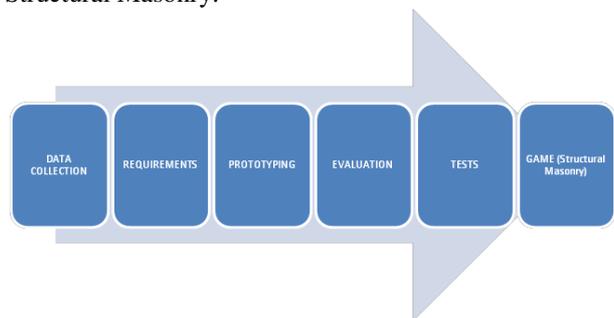


Figure 1: Logical scheme of the game development

Source: Adapted from Preece, Rogers and Sharp, 2007.

This article will present the steps taken to develop the first two stages, data collection and obtainment of the requirements.

The purpose of the data collection stage is to gather sufficient, relevant and necessary information for building a set of stable requirements for the game Structural Masonry.

"A requirement consists in a statement about an intended product that specifies what it should do or how it should perform" (Preece, Rogers and Sharp, 2007, p. 224).

To identify the requirements we used the Delphi method to experts, professionals working in the building system which is the focus of this work.

2. SIMULATION IN EDUCATION

Soldiers were the first to realize the great advantage of using simulation in the teaching - learning process. Besides them, this technique is also widely used in training in health and business areas.

Games and simulations are present in society in the last two centuries. It has been realized that the use of these techniques can go beyond simple fun and can add value to technical, scientific and intellectual education (Magee 2006).

Studies have shown several manners to use simulation in the educational environment. They range from the search for solutions to specific problems to the formulation of an environment that allows qualification for the accomplishment of more activities (Larréché 1987; Wolfe 1993; Curland and Fawcett 2001; Khaled 2001; Santos 2002; Van Der Zee and Slomp 2009; Pasin and Giroux 2011).

Simulation is understood as a technique that allows studying, predicting and understanding the behavior of a system in a simplified manner, enabling a "permanent representation of reality" (Ellington et al. 1981).

According to Salas *et al.* (2009) and Mahboubian (2010), learning from simulation takes the learner to build knowledge in a safe environment where he can make mistakes, understand them and get the correct results without the need of damaging equipment or changing the production flow of the activity to be learned.

Learning from the experimentation and simulation, along with the application of theoretical lessons results in a solid foundation for the students who need knowledge for decision making in real scenarios (Latorre and Jiménez 2012).

Developing competence in disciplines of undergraduate courses, from teaching methods that go beyond the traditional method already established of teaching, can be an alternative to achieve better results in education of the student. Among these alternatives we can mention; Latorre and Jiménez (2012); Bruzzone et al. (2013); Kazimoglu et al. (2012), Koops and Hoevenaar (2013), Cavalheiro, Freire and Jungles (2014) and Bellotti et al. (2012).

In Latorre and Jiménez (2012) it is possible to see the modeling, simulation and optimization application in a didactic tool which will help the training and construction of competences in students, so they can become decision makers in a business environment.

Bruzzone et al. (2013), contributes in the training process in the health area inserting the MARIA (Model for Advanced and Realistic patient simulation driven by Intelligent Agents) a virtual model of a human being used as a patient in the game simulating the continual cares in the hospital environment to teach issues, such as the evolution of pathology, preventive actions and identification of clinical symptoms. Thus, the solution proposed becomes an innovation in the student training and requalification of professionals from new technologies.

In Kazimoglu et al. (2012), the Serious Game "Program

your robot" is a didactic tool projected to enable the student in the subject of Introduction to Programming, seeking to develop mostly the proficiency related to the development of algorithms, data depuration and computer simulation. Divided into levels, the "Program your robot" gradually builds the necessary competencies to the student, expanding the knowledge required in each of the six levels of the game.

In the articles by Latorre and Jiménez 2012; Bruzzone et al. 2013; Kazimoglu et al. 2012, the authors seek the training of the students and the development of the professional competencies, starting from the application of the serious games in the training environment.

Koops and Hoevenaar (2013) verifies the acquisition of concepts of Newtonian mechanics from the comparison of two students groups, the first one exposed to a traditional class and the second one using the simulation of learning games.

Cavalheiro, Freire and Jungles (2014), from the use of a table-top simulation with the application of serious game to a group (named A) of Civil Engineering students and application of the traditional teaching method to another group (named B), it is measured the acquisition of the competencies to the subject of programming techniques and construction work planning, coming to the results: (a) for the students, the acquisition of these competencies is more effective with the application of the serious games which simulates the reality of the building sites, (b) evasion is lower when applied this teaching method. Pode-se observar nestes trabalhos a utilização de serious game para a medição de aquisição de competências utilizando dois métodos de ensino diferentes.

Bellotti et al. (2012), part of the obtainment of the requirements thru the literature survey and consultation with entrepreneurs, students and professors to create educational games which bring to students basic concepts of entrepreneurship and business management from this game practices.

According to Magee (2006), an approach to learning, from the use of simulation and games in adult education has worked well for the training paradigms, in which the educational objective has been to create a more homogenous workforce with consistent and predictable levels of competence.

3. METHODOLOGY

Preece, Rogers, and Sharp (2007) present some forms of data collection and determination of requirements for the construction of an instructional design, among these are: questionnaires, interviews, study group, natural observation and analysis of the documentation.

To Drenth (1984), the techniques of diagnostic of the requirements found in the literatura are in the figure 2 below:

Instrument or source	Level of method structuring	
	Structured	Unstructured
Instrument	Questionnaires standardized scale Structured interview Objective test	Unstructured interview Projective test
Researcher	Sistematic observation Measuring scales	Register of impressions Participant observation
Specialist	Evaluations Structured interview	Brainstorm Delphi Method
Recording equipment	Indirect indications Experimental apparatus	Video, film, tape
Archive	Saved data	Personal documents Official documents

Figure 2: Main data collection methods
Source: Drenth, 1984.

Guimarães, Bruno-Faria and Brandão (2006) highlight that the utilization of techniques complementary to the principal one chosen auxiliary in obtaining the desired information, since it maintains the methodological rigor. At the time of data collection, the starting point is the survey on the existing literature on the skills necessary to the professional manager of the building process on structural masonry and the main requirements for the development of learning materials for those skills.

The Delphi method will be used to determine the requirements creation of the game Structural Masonry as well as questionnaire and interview.

It is a method that serves to identify the objects requirement of this work, and it is related with the literature as seen in the paragraphs above.

The method was developed by the Rand Corporation to study the impact of technology on war (Dalkey and Helmer 1962). With this method, it is possible to obtain a set of information on a specific topic from the consultation and consensus from specialists.

The research presents a qualitative, systematic and interactive method, for which specialists are selected according to pre-defined guidelines and are invited to participate anonymously in two or more rounds. The rounds are for collecting their impressions and providing impressions from the other specialists, in order to analyze, reflect and revise the points of view on the subject discussed.

The technique ends reaching the defined criteria which can be the rounds number or the reaching of the consensus (Hallowell e Gambatese 2010) and a statistical work provides the final result.

The procedure used to obtain the following requirements is shown through steps in Figure 3.

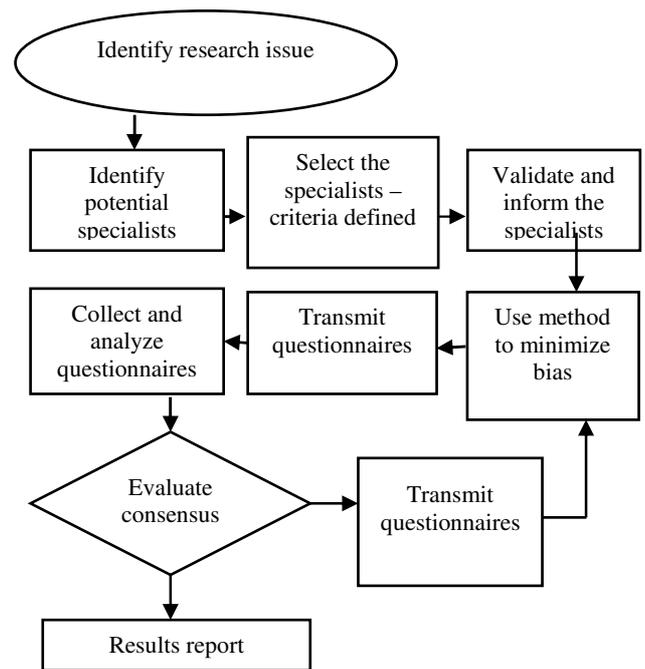


Figure 3: Procedure used in the execution of the Delphi method

Source: Adapted from Hallowell and Gambatese, 2010

For the identification and selection of specialists, there was an observation of Roger and Lopez's guidelines (2002), where the specialist must be or have been reached at least two of the following requirements:

- Primary or secondary author of two or more publications in the study area;
- Presenter in conferences of three or more publications on the subject covered in this work;
- Member and president of an institution or committee related to the topic of this work;
- Employee with at least 5 years of experience working with the practice of matters related to the subject;
- Professor at an institution of higher education within the field of study covered in this work.

Authors such as Brockhoff (1975), Boje and Murnighan (1982), Rowe and Wright (1999) and Hallowell (2008) recommend that the number of specialists to be selected for the application of Delphi method should be in a range of 9 to 15 specialists.

In this work the adopted number of specialists was 12, following Hallowell's recommendations (2008). According to him, this amount meets the study characteristics and the available number of potential specialists.

The choice of the specialists selected was for professionals who work on building sites, professors and researchers who use the building system in structural masonry in their work, divided into equal parts. It resulted in the professional staff as Figure 4.

Specialist	Experience (years)	International Experience	Professor	Researcher	Work Site Professional
01	09	x		x	
02	06				x
03	23	x	x	x	
04	12				
05	09	x	x	x	
06	18	x	x		x
07	05				x
08	06		x		
09	20	x	x		
10	11	x		x	
11	13				x
12	08	x		x	

Figure 4 – Characteristics of the participating specialists

Along with the definition of the specialists, a literature analysis for construction of the initial set of requirements was held.

To the specialists who expressed an interest in participating in this research, it was highlighted the importance of providing information, and the importance of working for the scientific and productive environment.

In the first round, there was an exposition of the theoretical conception on what would be the conditions required for the professional focus of the work and the skills these professionals needed to have in order to manage the production of a building in structural masonry.

Then, the following question was asked:

- "What are the requirements for the work manager in building system in structural masonry?"

The result of the first round was an ordinal enumeration of the requirements raised by the specialists, elimination of repetitions or similarities and the incorporation of the requirements observed from the literature, generating the matrix of initial requirements.

In the second round, the array of initial requirements was sent to the specialists, and they were asked about the approval or disapproval of each of these as a fundamental element for the professional of this area. In addition, there was the definition of values on the importance of each requirement reported. In this second round a consensus began to be drawn. It was possible to notice some changes of opinion compared to the responses from the first round. The feedback makes the specialists analyze and reassess the responses provided in the previous round. Thus, the views can be maintained, modified and changed according to the accomplishment of the following interactions.

To determine the required number of rounds, literature was consulted. According to studies by Dalkey et al, 1970, Brooks, 1979, Ludwig, 1994, Custer, Scarcella and Stewart, 1999, three rounds are enough to reach consensus. However, Pill, 1971, Delbecq, Van de Ven

and Gustafson, 1975, Gupta and Clarke, 1996, indicate about three to six rounds as necessary to reach consensus.

For checking if there is consensus, literature recommends using a benchmark of $Ca \geq 60\%$ agreement to be considered acceptable and validated, thus forming the requirements matrix.

The responses of the second round were used to determine the level of agreement. It was made from the expression below, creating a new matrix, named requirements matrix with agreement level.

$$Ca = \left(1 - \frac{Vn}{Vt}\right) * 100 \quad (1)$$

Where:

Ca = coefficient of agreement expressed in percentage;

Vn = amount of specialists in disagreement with the prevailing criteria;

Vt = total amount of specialists

Empirically when $Ca \geq 60\%$ the agreement is considered acceptable (Santos 2001).

4. RESULTS

Running the Delphi method, from the steps suggested by Hallowell and Gambatese (2010), has resulted in this work a total of three rounds. The compilation of the results is presented in Figure 5.

TYPES OF REQUIREMENTS	REQUIREMENTS
FUNCTIONAL	- The simulation will help in the development of planning and managing skills on execution and acquisition of materials for building in structural masonry.
OF DATA	- The system must have access to the database of the student skills required at different levels; - Data must be current and accurate.
ENVIRONMENTAL	- The sharing of data must be synchronous; - Players (students) should be in the same intranet system;
OF USER	- Targeted to undergraduate students in Civil Engineering, Industrial Engineering or Architecture; - Having knowledge in using technology; - Age should be less than 30 years old.
OF USABILITY	- The game should have complexity divided into levels (beginner, intermediate and advanced)

Figure 5 : Necessary requirements for formulation of instructional design to be proposed

The requirements presented were the basis for the new stage of the creation of instructional design Structural Masonry, which is prototyping.

5. FINAL REMARKS

With this article, an absence of connection between classroom theory and professional practice could be identified. Thus, it is necessary to rethink the pedagogical practice of professors, trying to integrate

the classroom with the reality of the job market that awaits future engineers.

The proposal to obtain the requirements from the Delphi method makes the next steps a consistent work when gathering the information coming from specialists in the field.

This study contributes both for academia and for retraining by searching the theoretical knowledge, from the simulation techniques and associate it with the management process in undergraduate courses related to the civil construction.

It is worth to mention that the obtainment of requirements from the use of the Delphi method is a new element part of the practices and principles in the student training in the existing literature..

It is hoped that this article can contribute to expand the theoretical framework regarding the requirements for training of students who will act as experts and professionals in the construction system in structural masonry.

REFERENCES

- Bellotti, F.; Berta, R.; De Gloria, A.; Lavagnino, E.; Dagnino, F.; Ott, M.; Romero, M.; Usart, M.; Mayer, I. S. 2012. Designing a Course for Stimulating Entrepreneurship in Higher Education through Serious Games. *Procedia Computer Science*, 15, p. 174 – 186.
- Boje, D. M.; Murnighan, J. K., 1982. Group confidence pressures decisions. *Manage. Sci.*, 28 (1), 1187–1196.
- Brockhoff, K., 1975. The performance of forecasting groups in computer dialogue and face-to-face discussion. *The Delphi method: Techniques and applications*, Addison-Wesley, Reading, Mass., 291– 321.
- Brooks, K. W., 1979. Delphi technique: Expanding applications. *North Central Association Quarterly*, 54 (3), 377-385.
- Bruzzzone, A. G.; Frascio, M.; Longo, F.; Massei, M.; Nicoletti, L. 2013. An Innovative Serious Game for education and Training in Health Care. *International Journal Privacy and Health Information Management*. 1 (2), p.132 – 146.
- Cavalheiro, C; Freire, A. S.; Jungles, A. E. 2014. Table Top for scheduling and controlling constructions course. In: *Proceedings of the European Modeling and Simulation Symposium*, 2014. pp. 24 – 29.
- Curland, S. R.; S. L. Fawcett. 2001. Using simulation and gaming to develop financial skills in undergraduates. *International Journal of Contemporary Hospitality Management*, 13 (3), pp. 116 – 119.
- Custer, R. L., Scarcella, J. A., & Stewart, B. R., 1999. The modified Delphi technique: A rotational modification. *Journal of Vocational and Technical Education*, 15 (2), pp. 1-10.
- Dalkey, N; Helmer, O., 1962. Na experimental application of the Delphi method to the use of experts. Memorandum RM – 727/1 – ABRIDGED. United States Air Force Project Rand. Santa Monica, California.
- Dalkey, N., Brown, B., Cochran, S. 1970. Use of self-ratings to improve group estimates. *Technological Forecasting*, 1 (3), pp.283–291.
- Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H., 1975. *Group techniques for program planning*. Glenview, IL: Scott, Foresman, and Co.
- Drenth, P. J. D. 1984. *Research in work and organizational psychology: principles and methods*. In: PIETER, J. D. et al (Ed.) *Handbook of work and organizational psychology*. John Wiley & Sons, p. 13 – 50.
- Ellington, H.; Addinall, E.; Percival, F.; 1981. *Games and simulations in science education*. London: Kogan Page Limited.
- Guimarães, T. de A.; Bruno-Faria, M. de F.; Brandão, H. P. 2006. Aspectos metodológicos do diagnóstico de competências em organizações. In: Borges-Andrade, J.E.; Abbad, G. da S; Mourão, L. *Treinamento, Desenvolvimento e Educação em organizações e trabalho: Fundamento para gestão de pessoas*. Porto Alegre: Artmed.
- Gupta, U. G.; Clarke R. E., 1996. Theory and applications of the Delphi technique: A bibliography 1975–1994. *Technol. Forecast. Soc. Change*, 53(2), pp. 185–211.
- Hallowel, M.; 2008. *A formal model of construction safety and health risk management*. Ph.D. dissertation, Oregon State University, Corvallis, OR.
- Hallowel, M. R; Gambatese, J. A., 2010. *Qualitative research: Application of the Delphi method to CEM research*. *J. Constr. Eng. Manage.*, 136 (1), p. 99–107.
- Kazimoglu, C.; Kiernan, M.; Bacon, L.; Mackinnon, L. 2012. A serious game for developing computational thinking and learning introductory computer programming. *Procedia - Social and Behavioral Sciences*, 47, pp. 1991 – 1999.
- Khaled, N., 2001. *Managing construction equipment buy and sell decisions replacement: A simulation game*. In: *ASC proceedings of the 37th annual conference*, pp. 187–198.
- Koops, M. Hoevenaar, M. 2013. *Conceptual Change During a Serious Game: Using a Lemniscate Model to Compare Strategies in a Physics Game*. *Simulation & Gaming*, V. 44, Issue 4, pp. 544 – 561.
- Larréché, J.C., 1987. *On Simulations in Business Education and Research*. *Journal of Business Research*, 15 (6), pp 559 – 571.
- Latorre, J.I., Jiménez, E., 2012. *Simulation for Education in Business decision-Making*. *Magazine of the Society for Modeling & Simulation International (SCS M&S Magazine)*, (2), 59–65.
- Ludwig, B. G., 1994. *Internationalizing Extension: An exploration of the characteristics evident in a state university Extension system that achieves*

- internationalization. Unpublished doctoral dissertation, The Ohio State University, Columbus.
- Magee, M., 2006. State of the Field Review: Simulation in Education. Internal report. Alberta Online Learning Consortium. Calgary AB.
- Mahboubian, M., 2010. Educational aspects of business simulation softwares. *Procedia Social and Behavioral Sciences*, v.2, pp. 5403–5407.
- Pasin, F.; Giroux, H., 2011. The impact of a simulation game on operations management education. *Computers & Education* 57: 1240–1254.
- Pill, J., 1971. The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Economic Planning Science*, 5, 57-71.
- Preece, J.; Rogers, Y.; Sharp, H., 2005. *Design de Interação: Além da interação homem-computador*, Porto Alegre: Bookman.
- Rogers, M. R.; Lopez, E. C., 2002. Identifying critical crosscultural school psychology competencies. *J. Soc. Psychol.*, 40 (2), p. 115–141.
- Rowe, G.; Wright, G., 1999. The Delphi technique as a forecasting tool: Issue and analysis. *Int. J. Forecast.*, 15, p. 353 – 375.
- Salas, E.; Wildman, J. L.; Piccolo, R. F., 2009. Using simulation-based training to enhance management education. *Academy of Management Learning and Education* 8: 559–573.
- Santos, A. C. 2001. O uso do método Delphi na criação de um modelo de competências. *Revista de Administração*, São Paulo, 36 (2), p.25 – 32, april-june.
- Santos, J., 2002. Development and implementing an internet-based financial system simulation game. *Journal of Economic Education* 33: 31–39.
- Van Der Zee, D. J.; Slomp, J. 2009. Simulation as a tool for gaming and training in operations management case study. *Journal of Simulation* 3: 17–28.
- Wolfe, J., 1993. A history of business teaching games in English-speaking and post-socialist countries: the origination and diffusion of a management education and development technology. *Simulation & Gaming* 24: 446–463.

AUTHORS BIOGRAPHY

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

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

Daniela Matchulat Ely – Professor at *Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)* in Brazil, Master in Civil Engineering by *Universidade Federal de Santa Catarina* in Brazil and doctoral student in Civil Engineering by *Universidade Federal de Santa Catarina*.

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

TOWARDS AN EXTENSION OF PROMELA FOR THE MODELING, SIMULATION AND VERIFICATION OF DISCRETE-EVENT SYSTEMS

Aznan YACOUB^(a), Maamar HAMRI^(a), Claudia FRYDMAN^(a), Chungman SEO^(b), Bernard P. ZEIGLER^(b)

^(a) Aix-Marseille Université, CNRS, ENSAM, Université de Toulon, LSIS UMR 7296, 13397, Marseille, FRANCE

^(b) RTSync Corp. and Arizona Center for Integrative Modeling and Simulation, AZ

^(a)aznam.yacoub@lsis.org, amine.hamri@lsis.org, claudia.frydman@lsis.org

^(b)cseo@rtsync.com, zeigler@rtsync.com

ABSTRACT

PROMELA is a well-known formalism for the modeling and the verification of concurrent systems. PROMELA deals with high-level specifications. As a result, PROMELA models are expressed in a high-level abstraction which not considers explicit representation of time or events for example. But, the efficiency of the processes of Verification and Validation relies on the accuracy of the models. That is why we propose in this paper preliminary work to a new extension of PROMELA for the modeling of discrete-event systems. The verification of these models is then done by combining formal verification and simulation-based verification.

Keywords: DEv-PROMELA, Simulation, Formal Verification, Verification and Validation

1. INTRODUCTION

Process Meta Language (PROMELA) is a well-known formalism proposed by Holzmann (1991, 2004) for the modeling and the verification of concurrent systems by model-checking. Model-Checking (Clarke and Emerson 1982, Baier and Katoen 2008), and more generally formal verification techniques (Huth and Ryan 2005), represent promising methods of verification. Because they are potentially able to explore all the states of a model, these methods appear as comprehensive and efficient methods. A verification model is focusing on the conceptual aspects of a design that are relevant to the properties one wants to verify. The meaning of that is that a verification model must be as simpler as possible, and PROMELA is designed to encourage the user to make abstraction of the computational aspects of the system under study. This also guarantees the efficiency of the model-checking algorithms. But, by enforcing these restrictions, the expressiveness of PROMELA is reduced. For example, the interactions between the components of a system can strongly depends on the values of data or on their timely coordination. Making strong assumptions about that or a too high-level abstraction can lead to the development of a non-efficient model.

In the opposite, Discrete-Event system Specification (DEVS) formalism (Zeigler 1976) is well-suited for low-level modeling and analysis of real systems. By providing a clean operational semantics, the DEVS formalism allows a clean interpretation of the model elements in the real world. The few restrictions enforced by DEVS increase its expressiveness. The other side of the coin is that it becomes impossible to apply formal proofs on these models. The challenge to introduce the DEVS semantics into a formalism of formal verification like PROMELA then seems to provide some benefits. First, it allows accurate modeling of discrete-event systems (DES) by using a formalism which supports formal verification. In other words, the purpose of such an approach is increasing the expressiveness of PROMELA without breaking its formal verification capabilities. Second, specifying a DES model into a syntactic formalism like PROMELA can make easier the translation from the conceptual model to a computerized simulation model. Indeed, transformation rules can easily be defined and verified between two syntactic formalisms. Third, the verification and validation of the conceptual DES models can be done through two combined methods, formal verification and discrete-event simulation, without needing to use multiple formalisms.

The goal of this paper is to present preliminary work about an extension of PROMELA for modeling of DES. The section 2 is a quick overview of PROMELA. The section 3 presents our extension. Two example is given in section 4 to illustrate the verification process using the MS4Me environment (Seo et al. 2013) and the SPIN (Simple PROMELA Interpreter) model-checker (Holzmann 1991).

2. QUICK OVERVIEW OF PROMELA

PROMELA is a specification language with a semantics of executability. Its syntax is influenced by the Dijkstra's Guarded Command Language (Dijkstra 1975) and the C programming language defined by Kernighan and Ritchie [1978]. This fact makes that its use is relatively easy compared with other formal methods because its syntax is really close to any implementation languages. We are focusing in this section only on concepts

interesting to the scope of this paper. The section 2.1 is about how specifying components of a system in PROMELA. The section 2.2 expose the semantics of PROMELA. We don't deliberately talk about the specification of the properties that one wants to verify on the model in this paper, because we are focusing on how increasing the accuracy of conceptual models.

2.1. System Specifications

A PROMELA system relies on three main types of objects: processes, data objects and messages.

The components of the system are modeled by a finite set of instances of processes. The processes can communicate with each other thanks to different mechanisms: buffered messages, shared global variables or rendez-vous handshakes. Each process is a finite set of guarded commands called instructions. At any time t , only one instruction is executed without any assumptions about its duration. Note that a set of instructions can be labeled as atomic: in this case, these instructions are considered as a unique instruction, like any common atomic operations. Processes can be also prioritized, meaning that a process with a higher priority will always execute its instructions before other processes. Syntactically, a process is defined by a **proctype** block of instructions, as given in Program 1.

Program 1 A simple example of PROMELA program.

```

1: int z = 1;
2:
3: active proctype A {
4:   int x = 2, y = 2;
5:   if
6:   :: ( x == 2 ) → x = 3;
7:   :: ( y == 2 ) → y = 4;
8:   fi;
9: }
10:
11: active proctype B priority 1 {
12:   int x = 2, y;
13:   do
14:   :: ( z == 1 ) → x = 2;
15:   :: ( x == 2 ) → y = 4;
16:   :: ( y == 4 ) → z = 0;
17:   od;
18: }
19:
20: !tl {[](z == 1); }

```

Instructions are divided into two categories: statements that modify the state of the system on the one hand, and control-flow instructions on the other hand. Assignments involve local and global variables, whereas communication statements involve global buffered channels. Control flow instructions are classical conditional and loop structures. They allow the selection of the next statement among different branches regarding a guard. Because PROMELA processes are non-deterministic, if several guards are satisfied, the next instruction is randomly selected. If none of them is satisfied, the control flow structure is blocked. PROMELA provides two special guards called **else** and **timeout**: if these guards are present in a control flow

structure, and if this one is blocked, these instructions are then executed.

Data in PROMELA is represented by local and shared variables. Local variables are relative to the process only in which they are declared, whereas global variables are shared by all the processes. A variable is characterized by its value and its type, among either all the PROMELA scalar basic datatypes as given in Table 1, or any finite combinations (structures) or finite arrays of these types.

Table 1 : Basic PROMELA Datatypes

Type	Size (bit)	Range of Values
bit, bool	1	[0;1]
byte	8	[0;255]
mtype (constants)	8	[0;255]
short	16	$[-2^{16}; 2^{16}-1]$
int	32	$[-2^{32}; 2^{32}-1]$

2.2. PROMELA Semantics

PROMELA centers on a semantics of executability (Natarajan and Holzmann 1997, Holzmann 2004). We can study it at two levels: on the one hand, at the process level, and on the other hand, at the whole program level.

2.2.1. Executability of a Process

A PROMELA process with a set L of statements is a finite state machine $P = (Q, T, q_0, F)$ where

- $Q = \{ q_i = (i, l_1, \dots, l_m, g_1, \dots, g_n, c_1, \dots, c_o) \in \mathbb{N} \times \prod_{i=1}^m L_i \times \prod_{j=1}^n G_j \times \prod_{k=1}^o C_k \}$ is the finite set of states. Each state is defined by its id, the value of each local and global variable, and the value of each channel;
- $T \subset Q \times L \times Q$ is the set of labeled transitions;
- q_0 is the initial state;
- F is the set of final states.

Denote $(q_i, q_j) \in Q^2$ and $l \in L$ (l is an instruction in the process). Then, $t = (q_i, l, q_j) \in T$ iff the process can change its state from q_i to q_j by only executing l . This means that an instruction syntactically denotes two consecutive states. A transition can be executed only if it is enabled:

- l is a non-blocking instruction: an assignment, a conditional instruction with a satisfied guard, or any control-flow atomic instruction (**else**, **skip**, **break**, etc);
- l is an asynchronous message sent over a non-full channel;
- l is an asynchronous message received from a non-empty channel;
- l is an unblocking rendez-vous message.

If a transition is enabled and executed, the process changes its state from the source state to the target state by applying an appropriate action function. This action function modifies the values of the variables, the content of the channels, etc. If more of one instruction is enabled,

the semantics engine randomly selects one of them and executes it.

2.2.2. Executability of a Program

A PROMELA program can be simply seen as the asynchronous product of the automata of each process. Then, a PROMELA program is defined by:

- $S \subseteq \prod_{i=1}^n Q_i$ where Q_i is the set of states of the process P_i ;
- T is the set of transitions; $t = (s_i, l, s_j) \in T$ if it exists a transition that changes the state of any of the processes by applying l ;
- s_0 is the initial state;

A transition t is enabled if its equivalent t_i is enabled in the process p_i . If two or more transitions are enabled, the semantics engine selects one of them at random and executes it (unless there is a defined priority between processes).

3. DISCRETE-EVENT PROMELA

The previous section shortly described the PROMELA semantics. As seen, PROMELA does not define either explicit time representation or events. But as stated by Tripakis (1996), the systems which can be modeled with PROMELA can be characterized as real-time systems. The representation of time becomes thus important to develop a more efficient design. Many timed extensions of PROMELA (Tripakis and Courcoubetis 1996; Bosnacki and Dams 1998; Nabialek et al. 2008) and the algorithms for their verification were studied in the literature. In opposite of these approaches which try to deal with the algorithms of formal verification and with the PROMELA semantics, we propose to integrate the DEVS semantics into PROMELA. This approach has two main advantages. On the one hand, it gives to PROMELA the capability of modeling DES with a lower level of abstraction. The model is more accurate and formal verification can be applied only on interesting paths. Formal verification can also be used to check structural properties on the model. Moreover, while formal verification deals with finite models, discrete-event simulation can be used for simulation-based verification of behavioral properties on infinite models, validation and analysis. On the other hand, modeling DES with a syntactic language close to an implementation language make easier the translation from the conceptual model to the computerized simulation model.

The sections 3.1 and 3.2 will recall concepts around discrete-event systems and the DEVS operational semantics. Then, the section 3.3 will introduce our new extension called Discrete-Event PROMELA (DEV-PROMELA).

3.1. Discrete-Event Concepts

DES are a specific class of timed systems (Zeigler 1976, Zeigler 1984). A DES can be seen as an extension of a Moore Machine, in the sense that the outputs of a DES

can depend only on its current state. Zeigler (1976) introduced two important concepts: the association of a lifespan to each state and the hierarchical composition. A DES evolves along the events that it emits or consumes, the distribution of events can be non-linear and time can be represented (through the lifespan concept) by any value of the \mathbb{R}^+ -space. A DES thus relies on the following notions:

1. Each state has a lifespan whose the value is a real. When the lifetime of state expired, the current state changes according to the transition table. An output is then emitted;
2. When an input is consumed by a DES, its current state changes according to the transition table, regardless the current lifetime of the current state;
3. Each event are well-dated. If e_1 and e_2 are two events, thus e_1 and e_2 can always be ordered;
4. If two events overcome at the same time, then either they are probably equivalent events ($e_1 = e_2$) or they are prioritized; the priority between two events is well-defined (non-deterministic behavior are not permitted);
5. The state trajectory is stable between two successful events;
6. As a result of the previous points, transitions are characterized by their nature: internal transitions correspond to autonomous behaviors while external transitions allow the system to react to any external events.

3.2. Overview of the DEVS Formalism

A DEVS system is built upon DEVS atomic models which can be coupled to get a DEVS coupled model. A DEVS atomic model is then a small simulation unit defined by $A = (X, Y, S, \delta_e, \lambda, ta)$ where

- X is the set of inputs;
- Y is the set of outputs;
- S is the set of sequential states;
- $\delta_e: Q \times X \rightarrow S$ is the external transition function;
- $\delta_i: S \rightarrow S$ is the internal transition function;
- $\lambda: S \rightarrow Y$ is the output function;
- $ta: S \rightarrow \mathbb{R}^+$ is the time advance function;
- $Q = \{(s, e) \mid s \in S, 0 \leq dt \leq ta(s)\}$ is the set of total states.

Taking a DEVS atomic model A in a current state q . If the lifespan of q is ended (i.e. $q = (s, ta(s))$), the model changes its state to $s' = \delta_i(s)$. It emits the output $y = \lambda(s)$. If A receives an event x , it change its state according to $s' = \delta_e(q, x)$. Otherwise, it stays in q , letting the time progress.

A DEVS coupled model is defined by $C = (X, Y, D, \{M\}, EIC, EOC, IC, Select)$ where

- X is the set of inputs;

- Y is the set of outputs;
- D is the name set of components;
- M is the set of atomic components;
- EOC is the set of external output coupling;
- EIC is the set of internal input coupling;
- IC is the set of internal coupling;
- Select is the tie-breaking function to choose the next event from the set of simultaneous events.

Informally, a DEVS coupled model is composed by atomic models. The date of the next event is computed by getting the minimum value among all the dates of all the next events. By processing an event, each component is updated according to its transition functions. If multiple events occur at the same date, the Select function chooses the next event to proceed.

3.3. Building Discrete-Event PROMELA

Introducing a new semantics in PROMELA needs to define new syntactic elements to model the concepts we are adding. The section 3.3.1 introduces these new syntactic elements. The section 3.3.2 and 3.3.3 define the meaning a DEv-PROMELA model.

3.3.1. Syntactic Rules

A new datatype called **real** is introduced. The purpose of real variables is to represent infinite and unbounded real values. Real variables can be local or global, and their declaration does not differ from any scalar variables:

```
real i,j,k;
```

Real variables can be combined in complex structures or in arrays, without any restriction.

An implicit local clock is associated to each process. Consequently, like in Timed PROMELA, introducing a specific type for clocks is not needed. Each process is ensured to have a clock. The clock valuation can be accessed through a specific **getCurrentDate** instruction. Clocks can be used to define exactly the date of next events. The purpose of clocks is only to determine the next event by taking the minimum value of the dates of all next events.

Missing concepts of events, state lifespan and type of transitions as viewed in the previous section are then introduced. Syntactically, a state is defined between two instructions in PROMELA. This means we can easily define the lifespan associated to each state and the type (internal or external) of the transitions associated to each instruction. We extend each statement with the following grammar described in the Backus-Naur Form:

```
<proctype decl> ::= "[" priority "=" <int> "]" <proctype>
<event stmt> ::= "[" <timed trans> "]" <stmt> | <stmt>
<timed trans> ::= <clt expr> | <evt expr> |
                <clt expr> "[" <evt expr>
<clt expr> ::= "clt:" <real expr> "->emit:" <evt val>
<evt expr> ::= "evt:" <evt val> [ <op> <evt expr> ]
<op> ::= " | "
<evt val> ::= <mtype> | "silent"
```

```
<real expr> ::= <real> | "infinity" | /* Any C-function
returning a real value */
```

Each statement is prefixed by an event descriptor (in brackets) which defines the lifespan of the source state of the considering transitions and how transitions are triggered. For instance, look at these instructions:

```
[clt:3.0->emit:newa] a = a + 3; (1)
```

```
[evt:newc] d = a + 9; (2)
```

```
[clt:lifespan(a)->emit:newa | evt:newb ] b = a * 7; (3)
```

(1) defines a state with a lifetime equal to 3 units of time, and an internal transition will emit the event "newa". By executing this transition, the assignment is executed. (2) defines a passive state (with an infinite lifetime) and an external transition which will be enabled by receiving the event "newc". (3) defines a state with a lifetime equal to "lifespan(a)" and two transitions: an internal one which will emit the event "newa" and an external one which will be enabled by consuming the event "newb". This third example shows how event descriptors can define transitions either as internal (with the *clt* command), or as external (with the *evt* command), or as a combination of multiple external events with probably one internal transition. A PROMELA statement can now label several transitions. Note that the default lifespan of each state is equal to infinity if an *evt* descriptor is defined and if there is no *clt* command. The *clt* command is optional and needed only to denote internal transitions. For convenient purposes, if a statement is not prefixed by an event descriptor, that is interpreted like it is prefixed by *[clt:0 ->emit:silent]*. The statement is executed without any delay by emitting the silent event. Priority descriptors on processes are also added. Indeed, the concept of DEv-PROMELA priority allows the ordering the events coming (resp. generating) from (resp. by) each process. This is a different meaning than the PROMELA initial concept of priority.

3.3.2. Semantics of a DEv-PROMELA Process

A DEv-PROMELA process P with a set L of statements is an automaton $T = (S_\tau, E, \delta_i, \delta_e, s_0, F)$ where

- $S_\tau = \{ s_i = (t_s, i, l_1, \dots, l_m, g_1, \dots, g_n, c_1, \dots, c_o) \in \mathbb{N} \times \prod_{i=1}^m L_i \times \prod_{j=1}^n G_j \times \prod_{k=1}^o C_k \}$ is the finite set of states;
- E is the finite set of events; E contains at least the silent event denoted ϵ ;
- $\delta_i : Q_f \rightarrow Q_0 \times E$ is the partial function that defines transitions labeled by a statement $l \in L$; note that it is a partial function because no internal transition can be defined for passive state.
- $\delta_e : Q \times E \rightarrow Q$ is the partial function that defines external transitions labeled by a statement $l \in L$; note that it is a partial function

because for a state q , there is not necessarily an external transition defined for each event e ;

- s_0 is the initial state;
- F is the set of final states.

We also define:

- $ta: \begin{cases} S_\tau \rightarrow \mathbb{R}^+ \\ ta(s) \mapsto t_s \end{cases}$ is the lifespan function;
- $Q = \{ q = (s, dt), s \in S_\tau \text{ and } 0 \leq dt \leq ta(s) \}$ is the set of total states;
- $Q_0 = \{ q = (s, 0), s \in S_\tau \}$
- $Q_f = \{ q = (s, ta(s)), s \in S_\tau \}$

Take s a state, and l a statement with an event descriptor. If l denotes an internal transition, l is enabled and $\delta_l(s)$ is triggered only if the value of the current clock of the process is equal to the lifetime $ta(s)$. The event associated to the transition is emitted and the next event date for the process is computed by:

$$d_e = \text{getCurrentDate} + ta(s') \text{ if } (s'; e) = \delta_l(s).$$

If l denotes an external transition on an event e , l is enabled if the process receives the event e . If the next event of the current process is at the same date than another external event, the priority statement resolves the conflict. If $ta(s) = \infty$, all of the statements related to s are blocked. Only external transitions can be enabled in this case. Because there is at most one internal transition per state and because there is a priority statement, a DEV-PROMELA process is deterministic. What about the non-determinism expressed in the PROMELA control-flow structures ?

Program 2 Simple condition structure in Classic PROMELA.

```

1: int x, y = 2;
2:
3: if
4: :: ( x == 2 ) → x = 3;
5: :: ( y == 2 ) → y = 4;
6: fi;
```

Take a look at the Program 2. In the meaning of PROMELA, both guards are satisfied, meaning one of them is non-deterministically chosen. From another point of view, this means the model checker will verify the both paths, and this is the purpose of a good low-level verification model. But, if we take a look at the state space (Figure 3), we can clearly note that there is three cases. Two of them are unambiguous; concerning the intersection, the behavior is not well-defined. Transposed to an event system, this model would describe two situations:

1. Denotes e_1 and e_2 two different events that overcome at the same date. The verification model does not care about which event is

consumed at first; the two orders will be checked;

2. The lifespan of the state $(x=2, y=2)$ is ended, but two behaviors are possible. In this case, the verification model is not well-defined and effort are put into the verification of paths that are maybe meaningless for the real system.

We mean that the non-determinism involved by the PROMELA control-flow structures doesn't come necessarily from the non-deterministic nature of the considered program, or from a high-level abstraction, but can be involved by incomplete specifications.

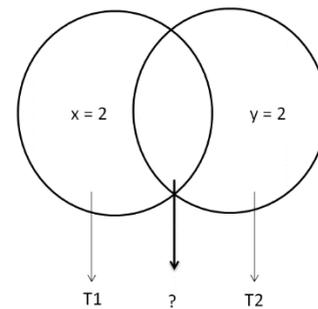


Figure 1 : State space of the Algorithm 2. T1 denotes the transition labelled by the statement line 4, and T2 the transition labelled by the statement line 5.

Enforcing determinism allows a unique, clear and unambiguous interpretation of the model in the real world, even non-determinism is really convenient for the verification of high-level models.

Taking into account these considerations, the Program 3 introduces a DEV-PROMELA version of the Program 2.

Program 3 Simple condition structure in DE-PROMELA.

```

1: int x, y = 2;
2: real t1, t2;
3:
4: if
5: :: [clt : t1 → emit : changex] ( x == 2 ) → x = 3;
6: :: [clt : t2 → emit : changey] ( y == 2 ) → y = 4;
7: fi;
```

This program can be interpreted in three manner:

1. if $t_1 \neq t_2$, there is not ambiguity: the state concerned by the branches are not the same. Indeed, if t_1 is lesser than t_2 , T1 will be always triggered in first. If t_2 is lesser than t_1 , T2 will be always triggered in first;
2. if $t_1 = t_2$, this model is non-deterministic. Determinism is then enforced by always executing the first enabled branch T1 in the case $(x=2, y=2)$.

This property is applicable to any branching structures.

3.3.3. Semantics of a DEv-PROMELA Program

A DEv-PROMELA program is a product of the automata of each process that composed the program, as in PROMELA. The difference is that, while in PROMELA the sequences of the states are interleaved (i.e. the final graph is composed by all the possible permutations of the statements), DEv-PROMELA takes into account only the execution paths given by the order of the events. Denote e_1 and e_2 , two events of the DEv-PROMELA system composed by two processes P1 and P2. If $\text{date}(e_1) < \text{date}(e_2)$, then the statement associated to e_1 will be executed in first. Both processes update their state according to the action function and computes their local next event e_1' and e_2' . The next event date is then given by $\min(e_1', e_2')$.

If $\text{date}(e_1) = \text{date}(e_2)$, priority as defined in the previous section determines the next event between both events. Non-determinism is implicitly resolved for a precise instance: this does not mean that non-determinism doesn't exist in the global system. Because order of events can depend on the execution context, multiple branches could exist for a given global state. A verification algorithm would then have to check any possible orders of events. But in opposite of PROMELA, for a given order of events, one and only one path can be walked through.

It is also important to note that if the value of a global variable is changed, an implicit event is emitted. Other processes change their respective state. Their respective clock is however preserved, meaning their respective next event remains unchanged.

4. APPLICATIONS

We used our extension to model two examples of systems. First is the well-known Fischer's Mutual Exclusion Protocol. Second concerns the Alternate Bit Protocol with Finite Queue and Infinite Queue. Verification of properties were done through combined formal verification using the classic SPIN model-checker, and simulation-based verification using the MS4Me Environment. Because the DEv-PROMELA semantics is the same than the one of DEVS, the translation from DEv-PROMELA specifications to DNL format were easy.

4.1. Fischer's Mutual Exclusion Protocol

The Fischer's Mutual Exclusion Protocol (Abadi and Lamport 1994) looks like a very simple algorithm (Algorithm 4) for handling mutual exclusion: firstly, the process p checks whether another process either is already or wants to enter the critical section (line 4). If it is the case, the process stays in an active wait. Then, the process p declares his willing to enter the critical section before entering a sleep mode. When it wakes up, it checks whether another process is entered the critical section during its sleeping. If it is the case, the protocol restarts from the beginning, else the process can enter the critical section. Fisher's protocol can be seen as a timed system, and also a timed event system.

Using DEv-PROMELA, it is easy to obtain a verification and a simulation model (Program 6) of this algorithm. Because of the system can be considered as closed, almost of the transitions correspond to an autonomous behaviors which depend only on the current state. Priority ensures that each process will have the hand in turn. When a process is sleeping, it is blocked until the delay is exactly expired or until another process changes the value of the id variable. This is modeled in l.16: the model stays in this state during $delay$ units of time or until it receives the $changeid$ event. In other words, the next event for a process will be at $\text{getCurrentDate} + delay$ after the execution of line 14.

Algorithm 4 Fischer's Mutual Exclusion Protocol (1984).

```
1: while true do
2: begin
3: /* non-critical section */
4: L: if id ≠ 0 then goto L;
5: id := i;
6: pause(delay);
7: if id ≠ i then goto L;
8: /* critical section */
9: id := 0;
10: end
```

If both processes wake up at the same time, priority will decide of the next event to proceed. Remember that the execution of this DEv-PROMELA program is deterministic.

The structural preservation property of DEv-PROMELA is out-of-scope of this paper. This property allows the generation of an untimed PROMELA verification model which has the same structural properties than the DEv-PROMELA model (Program 5). Thanks to that model, two properties can be easily verified:

1. "Two processes cannot be in critical section in same time." ($\square \neg (c_i \wedge c_j)$)
2. "Is there cases in which two processes cannot be in critical section in same time?" ($\diamond \neg (c_i \wedge c_j)$)

While the property 2 is well-verified by the model, the SPIN model-checker returns a violation error as expected for the property 1. Indeed, a process P can execute the l.11 and l.14, then a process Q can execute the l.11 and l.14 and enters the critical section while P is always in it. This is well-known that the timed constraint $delay > C$ (where C is the longest time that a process may take to perform a step while trying to enter the critical section) must be met to guarantee the mutual exclusion by this protocol. This only result could enforce the designers to design solutions to prevent such an error.

But if the system is well specified, for example if the duration taken to execute each instruction is quantified, a simulation-based verification can bring another interpretation. Remember that the DEv-PROMELA semantics is the same than the DEVS semantics, meaning a DEv-PROMELA model is a DEVS model. As a consequent, a DEv-PROMELA model is also a simulation model. We encode the Program 6 into a DNL

format given in Appendix A by translation; then, we simulate it into the MS4 Me Environment (Figure 2). A simulation-based verification is then performed. In the context in which two processes exactly acts in the same manner and with the semantics of Classic DEVS, the analysis reveals that whatever the value of the delay, two processes can't be in the critical section at the same time. Property 1 is thus verified by this DEv-PROMELA model. How interpreting the difference between the results of formal verification and simulation? On the one hand, the formal verification gives first information about the correctness of the model, on a high-level abstraction. For example, the formal verification ensures the absence of deadlocks. It can help designer to fix the design before considering other forms of testing. On the other hand, the simulation-based verification gives more precise information to the designer to understand what really happened. Because the DEv-PROMELA model is at lower level of abstraction than the PROMELA model, the designers can decide that the error found by the model checker is outside the scope of the model or not. Another advantage is that discrete-event simulation can be also used for validation at the same time. Designer can then get information about the validity of this model against its intended use.

4.2. Alternating Bit Protocol

We also use DEv-PROMELA to model the Alternating Bit Protocol (APB) (Bartlett et al. 1969), which has been used to illustrate the analysis capabilities of PROMELA (Holzmann 2004). The only things we have to do is to model the time T to transmit a bit, and the time T_1 up to that we can consider a message is lost. The obtained model does not really differ from the PROMELA one. But what happened if we want take into account a queue which stores messages before sending them? As shown by Zeigler and Nutaro (2014), the capacity of the queue has an impact on the behavior of the system. Modeling the queue in PROMELA certainly increases the size of the state space, and reduces the efficiency of the formal verification algorithms. But DEv-PROMELA can be used in two manner: as a verification model, assumptions can be verified in idealized conditions (in this case, by abstracting the capacity of the queue or by considering there is no queue), and as a simulation model. Then, the DEv-PROMELA model of the APB can be coupled with any simulation model of queues. The simulation-based verification then extends the formal verification by exploring scenarios outside the scope of the formal verification.

Program 5 Fischer's Mutual Exclusion Protocol in Classic PROMELA.

```

1: int pid = 0;
2:
3: active proctype P ( int id ) {
4: do ::
5: /* non-critical section */
6: wait_L;
7: if
8: :: pid != 0 → goto wait_L;
9: :: else → skip;
10: fi;
11: pid = id;
12: if
13: :: pid != id → goto wait_L;
14: :: else → skip;
15: fi;
16: /* critical section */
17: pid = 0;
18: od;
19: }
20:

```

Program 6 Fischer's Mutual Exclusion Protocol in DE-PROMELA.

```

1: int pid = 0;
2: mtype = { changepid }
3:
4: [ priority = id ]
5: active proctype P ( int id ) {
6: real delay = 2.0 * id;
7: do ::
8: /* non-critical section */
9: wait_L;
10: if
11: :: [clt:0.1 → emit:silent] pid != 0 → goto wait_L;
12: :: [clt:0.1 → emit:silent] else → skip;
13: fi;
14: [clt:0.1 → emit:changepid] pid = id;
15: if
16: :: [clt:delay → emit:silent | evt: changepid]
17:   pid != id → goto wait_L;
18: :: [clt:delay → emit:silent] else → skip;
19: fi;
20: /* critical section */
21: [clt:0.1 → emit:changepid] pid = 0;
22: od;
23: }
24:

```

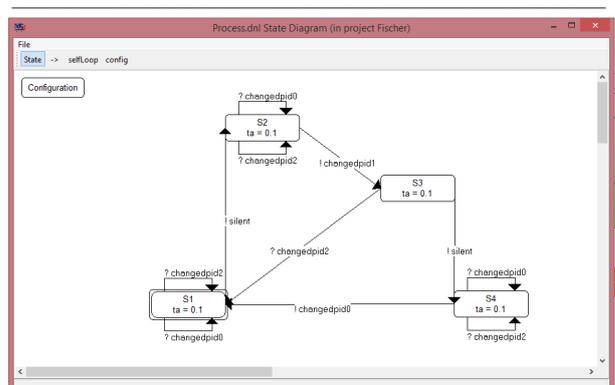


Figure 2 : State Diagram of the DEv-PROMELA model of Fischer's Algorithm in MS4 Me Environment.

5. CONCLUSION AND FUTURE WORK

We described in this paper some preliminary work to develop a new extension of PROMELA for the modeling and simulation of DES. This is done by introducing the DEVS operational semantics into PROMELA, without breaking the verification capabilities of PROMELA. A DEv-PROMELA model is then a verification model and a simulation model which can be verified by combining formal verification for structural properties and simulation-based verification for behavioral properties. The other advantage is that we are providing a syntactic formalism for DEVS modeling. This can help to implement computerized models from conceptual models. This maybe also enables the formal verification of the simulation models by using rewriting rules (for example from DEv-PROMELA to DNL).

The major drawbacks of our methodology is that DEv-PROMELA is limited to the Classic DEVS and its subclasses. Parallel DEVS cannot be modelled with this approach due to the limitations of PROMELA. But it can be interested to study how introducing the semantics of Parallel DEVS into PROMELA. Future work also concerns providing a formal proof that a DEv-PROMELA model is well a DEVS model, and working on limiting the formal verification algorithms only to the paths which are really expressed by the DEv-PROMELA models, without generating all the state space of the PROMELA equivalent.

ACKNOWLEDGMENTS

This work is a part of the R&D project “MAGE”, from French “Investing for the Future” national program.

APPENDIX A: DEV-PROMELA MODEL OF FISCHER’S ALGORITHM ENCODING IN THE DNL FORMAT

use delay with type double and default “0.1”!
use id with type int and default "1"!
use pid with type int and default "0"!

generates output on silent!
generates output on changedpid1!
generates output on changedpid0!
accepts input on changedpid0!
accepts input on changedpid2!

to start hold in S1 for time 0.1!
after S1 output silent!
from S1 go to S2!

when in S1 and receive changedpid0 go to S1!
when in S1 and receive changedpid2 go to S1!
external event for S1 with changedpid0
<%pid = 0;%>!
external event for S1 with changedpid2
<%pid = 2;%>!

internal event for S1
<%if(pid==0) holdIn("S2",0.1);

else holdIn("S1",0.1);%>!
hold in S2 for time 0.1!
after S2 output changedpid1!
from S2 go to S3!

when in S2 and receive changedpid0 go to S2!
when in S2 and receive changedpid2 go to S2!
external event for S2 with changedpid0
<%pid = 0;%>!
external event for S2 with changedpid2
<%pid = 2;%>!

internal event for S2
<%pid = id;%>!
hold in S3 for time delay!
when in S3 and receive changedpid0 go to S1!
when in S3 and receive changedpid2 go to S1!

external event for S3 with changedpid0
<%pid = 0; holdIn("S1",0.1);%>!
external event for S3 with changedpid2
<%pid = 2; holdIn("S1",0.1);%>!
after S3 output silent!
from S3 go to S4!

internal event for S3
<%if(pid==id) holdIn("S4",0.1);
else holdIn("S1",0.1);%>!
hold in S4 for time 0.1!
after S4 output changedpid0!
from S4 go to S1!

when in S4 and receive changedpid0 go to S4!
when in S4 and receive changedpid2 go to S4!
external event for S4 with changedpid0
<%pid = 0;%>!
external event for S4 with changedpid2
<%pid = 2;%>!

internal event for S4
<%pid = 0;%>!

REFERENCES

- Abadi M. and Lamport L., 1994. An old-fashioned Recipe for Real-Time. ACM Transactions on Programming Languages and Systems. Vol. 16, pp1543-1571.
- Baier C. and Katoen J.P., 2008. Principles of Model Checking. MIT Press.
- Bartlett K.A., Scantlebury R.A., and Wilkinson P.T. 1969. A note on reliable full-duplex transmission over half-duplex lines, Comm. of the ACM, Vol.12, No. 5, pp260-265.
- Bosnacki D. and Dams D., 1998. Discrete-Time PROMELA and SPIN. Formal Techniques in Real-Time and Fault-Tolerant Systems, Vol.1486, 307-310. Berlin, Springer Berlin Heidelberg.
- Clarke E.M. and Emerson E.A., 1982. Design and Synthesis of Synchronization Skeletons Using

- Branching-Time Temporal Logic. Logic of Programs, Workshop, pp52-71. London, UK.
- Dijkstra E.W., 1975. Guarded commands, nondeterminacy and formal derivation of programs. *Comm. of the ACM*. Aug. 1975. Vol. 18, No. 8, pp. 453-457.
- Holzmann G.J., 1991. *Design and Validation of Computer Protocols*. Englewood Cliffs, NJ, Prentice Hall.
- Holzmann G.J., 2004. *The SPIN Model Checker: Primer and Reference Manual*. Readings Massachusetts, Addison-Wesley.
- Huth M. and Ryan M., 2005. *Logic in Computer Sciences: Modelling and Reasoning about Systems*. Cambridge University Press.
- Kernighan B.W. and Ritchie D.M., 1978. *The C Programming Language*. First Edition. Englewood Cliffs, NJ, Prentice Hall.
- Nabialek W., Janowska A. and Janowski P., 2008. Translation of Timed PROMELA to Timed Automata with Discrete Data. *Journal Fundamata Informaticae*, Vol.85, pp409-424. Amsterdam, IOS Press.
- Natarajan V. and Holzmann G.J., 1997. Outline for an Operational Semantics of PROMELA. *The SPIN Verification System*. DIMACS Series in Discrete Mathematics and Theoretical Computer Sciences. AMS, 1997. Vol. 32, pp.133-152.
- Seo C., Zeigler B.P, Coop R. and Kim D., 2013. DEVS Modeling and Simulation Methodology with MS4 Me Software. *Theory of Modeling and Simulation Symposium, SpringSim MultiConference*. San Diego, CA.
- Tripakis S. and Courcoubetis C., 1996. Extending PROMELA and SPIN for Real-Time. *Proceedings of the Second International Workshop on Tools and Algorithms for Construction and Analysis of Systems, TACAs*, pp329-348. 1996, London, UK.
- Zeigler B.P., 1976. *Theory of Modeling and Simulation*. John Wiley.
- Zeigler B.P., 1984. *Multifaceted Modelling and Discrete-Event Simulation*. San Diego, CA, USA. Academic Press Professional Inc.
- Zeigler B.P. and Nutaro J., 2014. Combining DEVS and Model-Checking: Using System Morphisms for Integrating Simulation and Analysis in Model Engineering. *Proceedings of the 26th European Modeling and Simulation Symposium*, pp350-356. 2014.

VALIDATION AND APPROVAL OF PN CIPHERED SUBNETS: SECURITY IN PN SHARING AND STORAGE USING PNML, XMLSECURITY AND XMLSIGNATURE

Iñigo León Samaniego ^(a), Mercedes Perez-Parte ^(b), Juan-Ignacio Latorre-Biel ^(c), Emilio Jiménez-Macías ^(d)

^(a,d) University of La Rioja. Department of Electrical Engineering

^(b,c) University of La Rioja. Department of Mechanical Engineering

^(a)inigo.leon@gmail.com, ^(b)mercedes.perez@unirioja.es, ^(b)juan-ignacio.latorre@unirioja.es, ^(b)emilio.jimenez@unirioja.es

ABSTRACT

In this work I board the study of Petri nets from the point of view of the security. There several goals. First of all, I will take advantage from my previous work in order to complete a subnetting process. Then, the creation of a PNML extension that allows the representation of subnets structures.

One application of this subnetting and PNML representation is the possibility of hiding part of a Petri net, facing a possible distribution, maintaining the privacy of the critical, secret, or complex parts of the system. However this hidden information is not eliminated from the net, but encrypted inside.

Other application explained is the possibility of digital signature of subnets, providing security services to the net and/or subnets.

My contributions to knowledge are:

1. Comprehensive study of subnets, abstracting their internal structure from the exterior by using front-ends. A method to build these subnets from the complete Petri net is explained and analyzed matrixed.
2. PNML has no way to represent subnets, so I approach a possible PNML extension to do it.
3. Subnetting and PNML extension to represents subnets allow to apply several security technics that offers encryption, data integrity, authentication and non repudiation

Keywords: Petri net, Petri subnet, subnetting PNML, XMLEncryption, XMLSignatures

1. INTRODUCTION

1.1. Background of the research

Petri nets are widespread for modeling many classes of systems, such as manufacturing logistics processes and services, concurrent systems, etc. However, all these nets are described in a comprehensive way and must have the information of the entire net to determine its evolution.

Furthermore, these nets can be modified with no control of integrity or authoring, for example.

1.2. Research problem

The problem occurs when somebody doesn't want to describe the whole subnet. Or, maybe, is wanted one part of the process to be only accessible for one specific person or entity.

The first approach to solve this problem is to take two Petri nets:

- one Petri Net with only the public information, extracting the private data. This is an incomplete model of the process
- another Petri Net with the whole information for the interested person or entity.

As you can notice, this is not an efficient way to publish this kind of Petri nets.

Other problem appears when I want to protect parts of the net from undesired modifications or ensure the authoring of some parts (or the whole net).

1.3. Justification of the research

It would be interesting to provide security to a Petri net:

- hiding a part of it. This can be useful, for example, distributing a process we want to be secret (León 2011), or simply to be a part of the net to be complex and do not interest handle for any reason (León 2011).
- avoiding not allowed changes in it.
- authenticating it (or a part of it). Useful to ensure who has developed a Petri net or subnet.
- avoiding the possibility of supplant other people in the authority of the Petri net or some of its parts.

So here is my contribution. I have researched the possibilities of hiding a part of a Petri Net so that everybody can access the public information, maintaining the secret of the private data. This private data is accessible only for authorized people. And not only that: I ensure data integrity, authentication and non-repudiation to Petri nets or subnets.

Some authors study the possibilities of Petri nets reduction (Valette 1979; Suzuki and Murata 1983; Fahmy 1990; Druzhinin and Yuditskii 1992; Fahmy 1993; Xia 2011), grouping in one place or transition a subnet, so that what happens on this subnet, is encapsulated in a single point of execution. However, we want to go further by defining parts of the net that are hidden (not clustered) and what are the implications, studied within network properties.

The main objective of this work is to extract parts (subnets) of a Petri net and provide them of wide security (privacy, integrity, authentication and non-repudiation).

1.4. Methodology

In order to achieve this goal, I have defined three milestones:

- Extend Petri Nets in order to define subnets, abstracting the internal structure from the rest of the net using front-ends, focusing on hiding information.
- Choose a lossless and extendible representation of this kind of Petri Nets
- Define both hiding and signing methods for this representation

For the first milestone, I complete my previous works for the creation of the theoretical basis for further study of Petri nets in which certain parts are hidden. So we setup a generic framework of definitions and notations that allow us to deepen in the study of the characteristics and properties of Petri nets and their subnets (Murata 1989; Silva 1985).

Also mention work already carried by other researchers in which we rely for our goal (Silva 1993; David and Alla 2010; Jensen and Kristensen 2009; Peterson 1981). All of this will be necessary to create the framework that allows us to study occultation in Petri nets. We will expand the vision of Petri nets, providing them with greater functionality, such as attachable subnet (León 2013).

The next step in this work is to choose (or define) a flexible representation of Petri nets that allows us to translate the previous extended nets. This representation has to be really extendible and flexible in order to be able to show actual and future characteristics of Petri nets.

I can advance you that the selected representation is the standard PNML and I have to define and extension for it in order to represent subnets that are going to be secured.

Once selected this representation, the last step is the hiding and signing method (digital signature provide integrity, authentication and non-repudiation services). Once more, I bet for standard protocols like XMLEncryption and XMLSignature.

This is a very basic investigation because I extend the very early definitions of Petri nets. Because of it, the results of this work are very probably extensible to any other development whose basis are the classic Petri

nets. For example, I am not going to study colored Petri nets, neither timed Petri nets, etc. But it is very easy to see that the results achieved in this thesis can be applied to them with little problems.

2. SUBNETTING

Let's take a Petri Net $R = \langle P, T, \alpha, \beta \rangle$ where P is the set of places, T is the set of transitions, α is the pre-incidence function and β is the post-incidence function. We define $R' = \langle P', T', \alpha', \beta' \rangle$ such that $P' \sqcap P$ and $T' \sqcap T$, α' and β' are restrictions of α y β over $P' \times T'$ (P' and T' are not empty).

Then we have split the original Petri Net in two disjoint subnets, R' and the rest of the net.

The next step is to analyze the inputs and the outputs of R' . For example, starting from the Figure 1, I want to extract R_1 and its interface:

$$C = \begin{matrix} & & t_1 & \cdots & t_s & t_{s+1} & \cdots & t_r \\ \begin{matrix} p_1 \\ \vdots \\ p_m \\ p_{m+1} \\ \vdots \\ p_n \end{matrix} & \left(\begin{array}{cccc|cccc} a_{11} & \cdots & a_{1s} & & a_{1(s+1)} & \cdots & a_{1r} & \\ \vdots & & R' & \vdots & \vdots & HN & \vdots & \\ a_{m1} & \cdots & a_{ms} & & a_{m(s+1)} & \cdots & a_{mr} & \\ \hline a_{(m+1)1} & \cdots & a_{(m+1)s} & & a_{(m+1)(s+1)} & \cdots & a_{(m+1)r} & \\ \vdots & & HT & \vdots & \vdots & \ddots & \vdots & \\ a_{n1} & \cdots & a_{ns} & & a_{n(s+1)} & \cdots & a_{nr} & \end{array} \right) \end{matrix}$$

Figure 1: Petri Net divided into two subnets

The resultant extracted interface is:

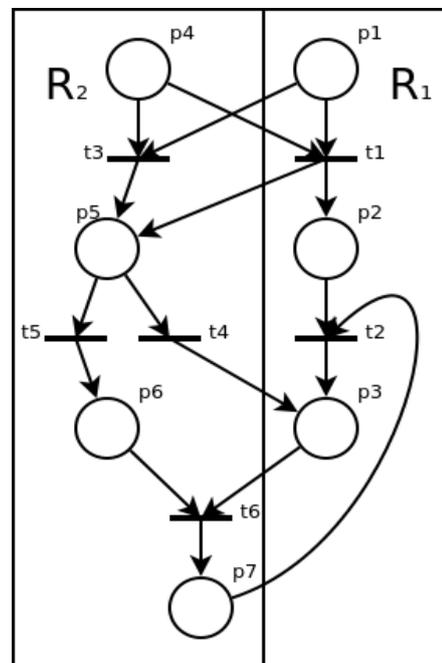
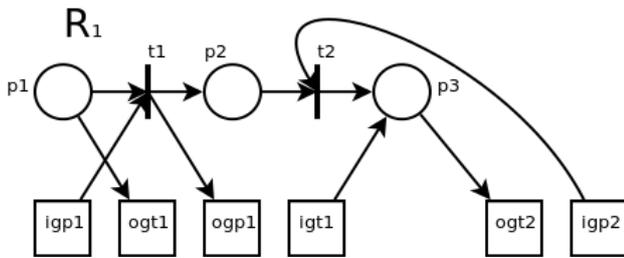


Figure 2: R1 with its interface

So we can define the input interface and the output interface in this way.



or the same:

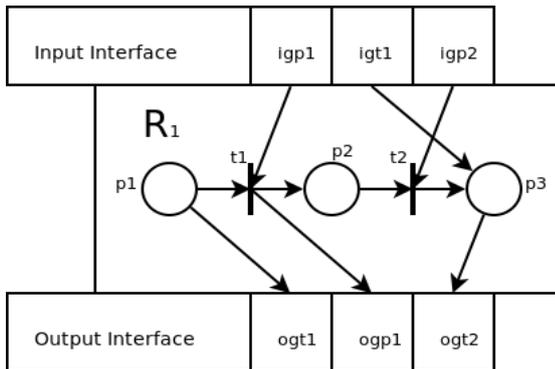


Figure 3: R1 with its input and output interface.

Once this is done, the next step is to implement this information in a Petri Net representation that will allow security implementation. For this objective, I am going to use PNML.

3. PETRI NET REPRESENTATION FOR SUBNETS SUPPORT.

3.1. Petri net representations

There are four standard ways to represent Petri nets. Each one of them have their properties, advantages and disadvantages. But I want to select one that I am able to represent any kind of Petri net, its subnets and allow to hide information without erasing it.

3.1.1. Graphic representation

This is the clearest and extended way to represent Petri nets. It has a very important advantage and it is that a picture is worth a thousand words.

Subnets can be defined simply drawing a vertical line. The right part is one subnet and the left part is other subnet. Places and transitions can be moved from one location to another depending on the subnet they are situated. This is only an example. Other way would be to use colors for the nodes (same color indicates same subnet) or use rectangles, etc.

So this representation is useful in order to show at one sight the Petri net structure, but I can't choose it for my goals.

I have not been able to discover a way to show some people the hidden information (the hidden subnet). However I will continue using it where a clear idea of the Petri structure if necessary.

3.1.2. Matrix representation

This representation is very useful to study properties and evolution of a Petri net, independently of its graphic representation. We can reorder rows and columns and define subnets in the matrix.

For example, let's take the Petri net represented by the matrix of Figure 4

$$\begin{array}{c}
 \begin{array}{ccc|ccc}
 & t_1 & t_2 & t_3 & t_4 & t_5 & t_6 \\
 p_1 & \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & -1 & 0 \end{pmatrix} & & \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ -1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \\
 p_2 & & & & & & \\
 p_3 & & & & & & \\
 p_4 & & & & & & \\
 p_5 & & & & & & \\
 p_6 & & & & & & \\
 p_7 & & & & & & \\
 p_8 & & & & & &
 \end{array}
 \end{array}$$

Figure 4: Matrix representation of a Petri net

We could group the places and transitions of the hidden subnet in this way, maintaining those elements other than zero:

$$\begin{array}{c}
 \begin{array}{c|ccc}
 & t_4 & t_5 & t_6 \\
 \begin{pmatrix} \blacksquare \\ 1 \\ (3, -1) \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} (1, -1, 2) \\ 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} & \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\
 p_5 & & & \\
 p_6 & & & \\
 p_7 & & & \\
 p_8 & & &
 \end{array}
 \end{array}$$

Figure 5: Matrix representation of a Petri net with hidden subnet

With this representation it is possible to study properties and it may be really important as a complement to graphic mode. With both representations together, everyone has a clear idea of the Petri net structure and properties. But I haven't found a way to store information inside the black box.

3.1.3. Equation representation

The third representation way for Petri nets is the equation representation.

Basically, transitions are selected and, for each one, the tokens of the places connected to that transition are modified. This is very useful to compute the evolution of a Petri net, choosing the transition fired.

However it is difficult to find a way for representing subnets with this notation. And, of course, if a subnet cannot be represented, it cannot be hidden.

In the Figure 6 we have an example of equation representation of a Petri net.

```

if (p1>0) then
  p1 <- p1 - 1
  p2 <- p2 + 1
if (p2>0) then
  p2 <- p2 - 1
  p1 <- p1 + 1
  p3 <- p3 + 1
if (p3>0) then
  p3 <- p3 - 1
  p2 <- p2 + 1
  p4 <- p4 + 1
if (p4>0) then
  p4 <- p4 - 1
  p1 <- p1 + 1
  p5 <- p5 + 1

```

Figure 6: Equation representation of a Petri net

I have tried to think about these ways of representation but, in my opinion, no one of them is suitable enough to represent subnets in a clear way than can be occulted. Because of that I have chosen the fourth representation, which is PNML, and is explained in the next section.

3.1.4. PNML Petri Net Marked Language

PNML is a way to represent Petri nets as xml content. The advantages over the other three representation ways described before are clear. By one side, XML is a widely extended format to represent almost everything. In the other side, XML is a robust technology free of errors and it is really flexible. Its flexibility comes from the possibility of adding any kind of labels and functionality with a very little amount of work. Its robustness come from the strict specification of the schemas declared to define completely the XML files that support. Once the schema is defined, the associated files have no way to get out of this definition, so we can validate a XML given its schema. And this is not all. If you have an XML file, you can extract information and complete it to create a schema for that file.

Originally, with basic Petri nets, the structure of a Petri net was fully provided. The only thing that is not supported in comparison with graphic mode is the graphical appearance of Petri nets: the position of nodes and transitions was not important, but with the arrival of High level Petri nets and Petri nets design software, it is necessary to store this kind of information.

In this work, I am going to study only basic Petri nets, but that concept and the method is easily exportable to other kind of nets, such as Symmetric nets and High Level Petri nets (representable in PNML format too).

4. PNML

4.1. Description

Petri Net Marked Language is an xml language created to represent Petri Nets. With it, we can take a Petri Net and store it into an xml file without loss of information. One of the best properties of PNML is that, as it is an xml based schema, it can be extended with more functionality extending the grammar. Virtually, any extension over Petri nets can be translated into PNML in a logical and natural way.

Moreover, this extension is defined by Petri net type definition (Billington et al 2003; Iso/iec 15909-2:2011}. In this case, PNML hasn't got a way to represent subnets. There is something named <page> that is used to represent several nets in the same PNML file. But, by default, a node inside a page cannot connect with a node of other page. So it cannot be used as "subnets". So I will extend the language in order to get several goals:

1. Represent subnets of a Petri Net.
2. Include input and output interfaces for every subnet.

As we can think, definition of several subnets of a Petri net is possible and the connection over them is always through their respective interfaces.

4.2. PNML grammar

As PNML is an xml based language it has to be described by a schema that define the creation rules of the PNML representation of a Petri net.

The grammar is defined since 2009 and updated until 2012, which is the most recent revision. I am not going to do an extensive explanation of all the possibilities of the grammar, but the most important. As we can see later, anything we think is useful can be added to the process with little effort.

So I am going to study only the most basic elements of a Petri net. The rest of the element can be attached later with facility.

4.2.1. PNML basics

In this section I am going to explain several characteristics of PNML files. With these explanations it is going to be easier the understanding of PNML structure.

First of all, as PNML files are xml files, there are several things to comply:

1. A xml file normally starts with a line defining some characteristics of the file, like the version and the encoding type. It has an aspect like this (Note: for clarity, in the following examples, this line can be deleted.):
`<?xml version="1.0" encoding="utf-8"?>`
2. A root node must exist. In this case, the root node is <pnml>. So every PNML files has to start with the tag <pnml> and end with the tag </pnml>. Below this tag, there is a new tag <net> that can contain:
 - (a) Type: the type of the Petri net as an attribute. In this case, as I am going to study only Place/Transition nets, it will be `ptnet`.
 - (b) Name of the net: New tag `<name><text>...</text></name>`
 - (c) Pages: one page is an invention to store several Petri nets inside an unique PNML file, but usually there is only one page for file. It is nested inside a <page> tag.

Furthermore, we cannot link elements from different pages.

- Each element in PNML has to have a unique id inside the net to be identified unambiguously. So there cannot be two elements with the same id.

With these three observations, we can have an idea about how a PNML file is structured:

```
<?xml version="1.0" encoding="utf-8"?>
<pnml>
  <net id="myNet"
    type="http://www.pnml.org/version-2009/grammar/ptnet">
    <name>
      <text> My new net </text>
    </name>
    <page id="page1">
      .....
    </page>
  </net>
</pnml>
```

Figure 7: Example of general PNML file

Once this structure is defined, I am going to explain the next stage, which is the most important one in this work.

4.2.2. Places, transitions and arcs in PNML

As Petri nets have three main elements (places, transitions and arcs), PNML has them too. These three elements have several things in common in a PNML file:

- They are all nested in a tag `<page>`.
- Places and transitions (not arcs) can contain a tag `<name>` with its name. This tag has been defined before for the net's name. It can store information about the text of the name and the graphical position of this label in this way:

```
<name>
  <text> Element Name </text>
  <graphics>
    <offset x="22" y="-10"/>
  </graphics>
</name>
```

- They can contain information about its position and dimension in a tag `<graphics>`:

```
<graphics>
  <position x="100" y="200"/>
  <dimension x="40" y="40"/>
</graphics>
```

These are the common properties of places, transitions and arcs. Now let's go on the particular characteristics of each one of them.

Places are represented with the tag `<place>` and they can have a marking with the tag `<initialMarking>`.

Here we have an example of a place with two tokens in PNML:

```
<place id="p1">
  <name>
    <text> Place number one </text>
    <graphics>
      <offset x="130" y="130"/>
    </graphics>
  </name>
  <graphics>
    <position x="130" y="90"/>
    <dimension x="40" y="40"/>
  </graphics>
  <initialMarking>
    <text> 2 </text>
  </initialMarking>
</place>
```

Figure 8: PNML representation for places

Transitions are represented with the tag `<transition>`. Except the initial marking, it is really similar to a place. This is an example of a transition in PNML:

```
<transition id="t1">
  <name>
    <text> Transition number one </text>
    <graphics>
      <offset x="270" y="140"/>
    </graphics>
  </name>
  <graphics>
    <position x="270" y="100"/>
    <dimension x="40" y="40"/>
  </graphics>
</transition>
```

Figure 9: PNML representation for transitions

Arcs are represented with the tag `<arc>`. Arcs must have a source and a target, which are defined by the attributes `source` and `target` that have to point to a transition and a place, identified by their `id`. Furthermore, the arc weight can be fixed by the tag `<inscription>`. If the weight is one, the tag `inscription` is not necessary because this is the default value. This is an example of the arc with weight 3 that connects the place and the transition of the previous examples in PNML:

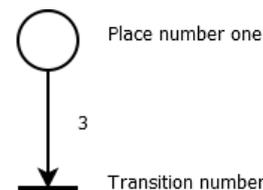
```
<arc id="a1" source="p1" target="t1">
  <inscription> 3 </inscription>
</arc>
```

Figure 10: General PNML representation for arcs with arbitrary weight

And this is the same example but with weight 1, that is obviated:

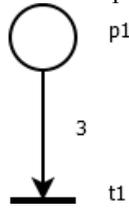
```
<arc id="a1" source="p1" target="t1"/>
```

If we take the last examples all together, we can represent the following Petri net:



For clarity and space reasons, I am going to obviate several options. I am not going to draw the graphic information and the names, but the id. And moreover,

the tags `<?xml>`, `<pnml>`, `<net>` and `<page>` are going to be obviated too. In many of the later examples they will not be present. So this last example is as follows:



```
<place id="p1">
  <initialMarking>
    <text> 2 </text>
  </initialMarking>
</place>
<transition id="t1"/>
<arc id="a1" source="p1" target="t1">
  <inscription> 3 </inscription>
</arc>
```

Figure 11: Simplified PNML representation for a basic Petri net

4.3. PNML extension for representing subnets

In this section I am going to define new tags and structures in PNML. At this point, I have developed all the necessary to extend PNML in order to represent subnets inside a concrete Petri net.

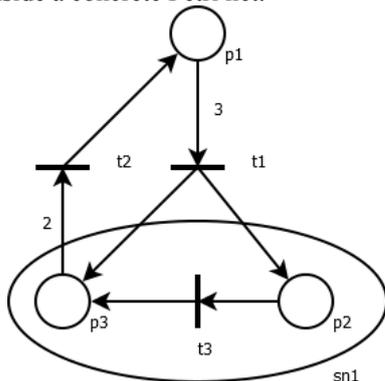


Figure 12: Subnet to represent in PNML

Let's take simple Petri net of the figure 12. It will serve us to explain the method to achieve a subnet representation and the PNML extension associated to it parting from a determinate Petri net:

The PNML code for this net is:

```
<place id="p1"/>
<place id="p2"/>
<place id="p3"/>
<transition id="t1"/>
<transition id="t2"/>
<transition id="t3"/>
<arc id="a1" source="p1" target="t1">
  <inscription>
    <text> 3 </text>
  </inscription>
</arc>
<arc id="a2" source="t1" target="p2"/>
<arc id="a3" source="t1" target="p3"/>
<arc id="a4" source="p3" target="t2">
  <inscription>
    <text>2</text>
  </inscription>
</arc>
<arc id="a5" source="t3" target="p3"/>
<arc id="a6" source="p2" target="t3"/>
<arc id="a7" source="t2" target="p1"/>
```

I want the ellipse region to be a subnet, so I have to specify a subnet with the elements inside the ellipse.

The first step is to define a new tag `<subnet>`. This tag will have an id, as the rest of PNML elements. And now we proceed in this way:

1. The places and transitions inside the subnet are moved into the tag `<subnet>`.
2. The arcs linking two elements that are both inside the subnet will be moved to into this new tag too.
3. The arcs entering or leaving the subnet will be copied inside the tag. This means that there are arcs duplicated inside and outside the tag.

If we apply these rules to the example:

1. p2, p3 and t3 are moved into the tag `<subnet>`.
2. a5 and a6 are put inside the tag.
3. a2, a3 and a4 are copied inside the tag.

And we have this other PNML extended code:

```
<subnet id="sn1">
  <place id="p2"/>
  <place id="p3"/>
  <transition id="t3"/>
  <arc id="a2" source="t1" target="p2"/>
  <arc id="a3" source="t1" target="p3"/>
  <arc id="a4" source="p3" target="t2">
    <inscription>
      <text> 2 </text>
    </inscription>
  </arc>
  <arc id="a5" source="t3" target="p3"/>
  <arc id="a6" source="p2" target="t3"/>
</subnet>
<place id="p1"/>
<transition id="t1"/>
<transition id="t2"/>
<arc id="a1" source="p1" target="t1">
  <inscription>
    <text> 3 </text>
  </inscription>
</arc>
<arc id="a2" source="t1" target="p2"/>
<arc id="a3" source="t1" target="p3"/>
<arc id="a4" source="p3" target="t2">
  <inscription>
    <text> 2 </text>
  </inscription>
</arc>
<arc id="a7" source="t2" target="p1"/>
```

Now I will separate the inside and the outside of the subnet completely. Taking advantage of the process described in section 2, I can extract its front-end.

In this case I have two igp (input gate to a place) and an ogt (output gate to a transition) with weight 2.

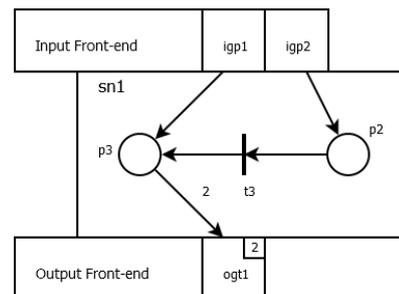


Figure 13: Subnet with its front-end

This is the complete net including subnet and front-end.

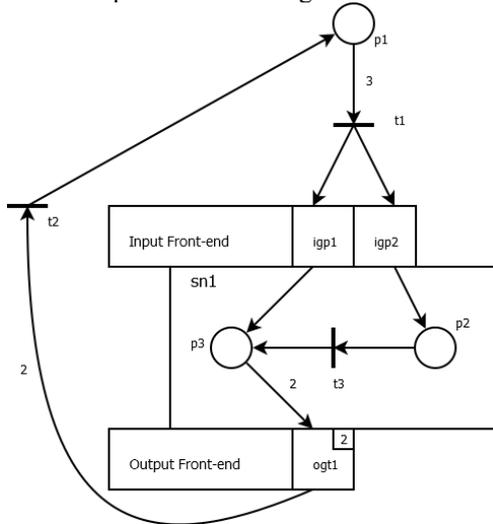


Figure 14: Petri net with subnet

And now that I have the graphic, how can I represent it in PNML? To answer this question I will define four new tags: `<interface>`, `<gate>`, `<inscription>` and `<content>`. Let's explain them.

As its name says, `<interface>` is the tag name for encapsulate the front-end. This tag has no attributes (just the id, of course) but it has embedded the gates inside of it. These gates are represented by `<gate>`. This tag has two new attributes: `action` and `type`. These two attributes have information about the gates. The attribute `action` can take two different values: `input` and `output`. It indicates whether the gate is an input or an output gate. The other attribute, `type`, can take other two values: `place` and `transition`.

As arcs have weight, gates have it too. For being in accordance, I define the tag `<inscription>` embedded in the tag `<gate>`. It has the weight of the arc associated. If the weight is 1 this tag can be obviated.

There is one other tag `<content>` that probably at this moment seems useless, but it is necessary for the rest of the process, as we will see in next sections. So I am going to introduce it now. This tag is used to encapsulate the rest of the subnet outside the interface. That is, `<subnet>` has two children: `<interface>` and `<content>`, that have the input/output place/transition gates and the rest of the elements, respectively.

At this moment I have to do only one thing more. The last step is to modify the arcs that are repeated inside and outside the net changing their id and source or target, depending on where is it:

- Change the id of one of the copies of the arc.
- If the arc is entering the subnet
 - For the `<arc>` tag inside the tag `<subnet>`, the `source` attribute of the arc is changed by the `input gate` associated
 - For the `<arc>` tag outside the tag `<subnet>`, the `target` attribute of the arc is changed by the `output gate` associated
- If the arc is leaving the subnet

- For the `<arc>` tag inside the tag `<subnet>`, the `target` attribute of the arc is changed by the `input gate` associated
- For the `<arc>` tag outside the tag `<subnet>`, the `source` attribute of the arc is changed by the `output gate` associated

Applying again all these rules to the example we have the definitive code for this Petri net:

```
<subnet id="sn1">
  <interface id="sn1-interface">
    <gate id="igp1"
      action="input" type="place"/>
    <gate id="igp2"
      action="input" type="place"/>
    <gate id="ogt1"
      action="output" type="transition">
      <inscription>
        <text> 2 </text>
      </inscription>
    </gate>
  </interface>
  <content id="sn1-content">
    <place id="p2"/>
    <place id="p3"/>
    <transition id="t3"/>
    <arc id="sn1-a2" source="igp2" target="p2"/>
    <arc id="sn1-a3" source="igp1" target="p3"/>
    <arc id="sn1-a4" source="p3" target="ogt1">
      <inscription>
        <text> 2 </text>
      </inscription>
    </arc>
    <arc id="a5" source="t3" target="p3"/>
    <arc id="a6" source="p2" target="t3"/>
  </content>
</subnet>
<place id="p1"/>
<transition id="t1"/>
<transition id="t2"/>
<arc id="a1" source="p1" target="t1">
  <inscription>
    <text> 3 </text>
  </inscription>
</arc>
<arc id="a2" source="t1" target="igp2"/>
<arc id="a3" source="t1" target="igp1"/>
<arc id="a4" source="ogt1" target="t2">
  <inscription>
    <text> 2 </text>
  </inscription>
</arc>
<arc id="a7" source="t2" target="p1"/>
```

Figure 15: Final PNML subnet representation

Once this is done, the only way to enter or leave the subnet is crossing the front-end.

This is a comprehensive definition of how to represent subnets in PNML. Now there are several ways to create a grammar extension that frame this structure of xml. For example, we can define a dtd file, a xsd file or, by coherence with the original grammar of PNML, a Relax NG file. I have defined a way to represent subnets, but the formal grammar is outside the scope of my work because of the wide casuistry of these Petri net types. However, the method is explained enough in order to each one of these types to define their own extension.

5. SECURITY

This is the second main goal, after subnetting. Once the possible subnets are defined it is the turn of securing them. It is possible to secure subnets or the entire net.

With secure, I mean four goals:

- Privacy. Concrete parts of the net must be occulted: the content is secret, so not everybody should be able to know it.
- Integrity. Any change in the secured parts has to be detected. If any of these parts suffers any kind of modification, the information may have been compromised, and perhaps it is not valid or correct. But I cannot know what has been modified: I can only detect that the original content has been changed.
- Authentication. I can authenticate the source of that net/subnet (signer, author or guarantor).
- Non repudiation. With this characteristic, the possibility of supplant other people is avoided. So the person that signs that part can't say that he hadn't done it. The signer cannot deny it.

But this information should be accessible to authorized people without necessity of supplying any other kind of data. So the whole information may be stored in the same file.

In the same way, there can be reasons for the rest of the security characteristics. For example, suppose that we have a Petri net that several people can access and:

- Some parts of that Petri net have been validated and accepted, so I want nobody to change them. In this case *integrity* is needed.
- I want to know who has developed a concrete chunk of the net. *Authentication* is required.
- There is a part of the Petri net that is bad defined and goes wrong. The person responsible of this part says that he hasn't made it and somebody has supplanted him. Then, *non-repudiation* is needed.

The best way to reach these goals is using standard and proved technologies. In this case, the selected technologies are:

- XMLEncryption (Xml encryption syntax and processing version 1.1, 2013) for privacy.
- XMLSignature (Xml signature syntax and processing version 1.1, 2013) for integrity, authentication and non-repudiation.

5.1. XMLEncryption

XMLEncryption is a World Wide Web Consortium (W3C) Recommendation for encrypting xml or non xml content. It is a standard xml file cipher. Both symmetric and asymmetric ciphering can be used. The main idea of this encryption is to replace the xml element or elements we want to be ciphered by other xml code that contains the ciphered data, in addition to information of the algorithms and keys used for the encryption process. When a non xml file is ciphered, the only option is to encrypt it completely. But, when it is applied to xml content, this technology allows us to define concrete

fragments of the document we want to hide. Moreover, the xml document can be transformed before applying the encryption, for example, in order to normalize the xml content.

In this work, the pieces of xml content susceptible to be ciphered are, obviously, the subnets represented in PNML format.

Regardless of the data source (xml or non xml) the result is always a xml element. Normal is that this xml encrypted chunk has the whole necessary information to be decrypted. Among that information we can find:

- Ciphering algorithm: it is the name of chosen method to encrypt the data. It can be not included. In this case, both ciphering and deciphering agents have to know which is the exact this algorithm.
- The ciphered data: obviously this part is mandatory and has always to be present.
- Name of the chosen key: it is optional. It is used when a set of keys is known by both ciphering and deciphering agents.
- Key: it is optional. In this case there is a symmetric key in order to encrypt the data and an additional pair of keys: one (known by the cipher agent) to encrypt the symmetric key and the other (known by the decipher agent) to decrypt it.

Actually, there are several options to apply XMLEncryption, such as the algorithm or the key. The exact election of those option values is responsibility of the Petri net sender.

This section does not want to be an extensive explanation about XMLEncryption but a general idea about its functionality. So I am not going to deepen the whole characteristics of XMLEncryption. The final decision about which options use is responsibility of those people that want to apply this work, basing their decision on the requirements of their own Petri net.

Petri subnets are represented by a `<subnet>` tag that contains `<interface>` and `<content>`. This last tag contains the xml content that is going to be ciphered. Obviously, if we encrypt the interface we will have no way to connect the subnet with the rest of the net. So here is the utility of `<content>` tag.

Let's take the example of the figure 14 and its PNML representation. The goal is to hide the internal content of the subnet. If we apply XMLEncryption to the data contained inside the `<content>` tag, we will get something like this, depending on the algorithm and key selected for the ciphering:

```
<subnet id="sn1">
  <interface id="sn1-interface">
    <gate id="igp1" action="input" type="place"/>
    <gate id="igp2" action="input" type="place"/>
    <gate id="ogt1" action="output" type="transition">
      <inscription>
        <text> 2 </text>
      </inscription>
    </gate>
  </interface>
  <content id="sn1-content">
    <xenc:EncryptedData
```

```

xmlns:xenc=http://www.w3.org/2001/04/xmenc#
Type="http://www.w3.org/2001/04/xmenc#Element">
<xenc:EncryptionMethod
  Algorithm=http://www.w3.org/2001/04/xmenc#aes128"
  xmlns:xenc="http://www.w3.org/2001/04/xmenc#" />
<xenc:CipherData
  xmlns:xenc="http://www.w3.org/2001/04/xmenc#">
  <xenc:CipherValue
    xmlns:xenc="http://www.w3.org/2001/04/xmenc#">
Wrl1njyJ1YOM91AYqcwGCWkw2L4pUjQD2GGVoU91VZ0wKqHY8y31Gy8FY4i5K
3GY8FY4i5K3G8grIe1HRFqe7RtkFiXZgGMeYnQp6oB6ckKp3KFKHVqtucc9rA
VzOgC7XAwe61HRFqe6RRVzXjNM9h1VZ0wKqHY8y31Gy8FY4i5K3G8grIe2xN
4u7x7fRtkFiXZgGMeYnQp6oB6ckKp3KFKRRVzXjNatVzOgC7XAwoe61HRFqe6
RRVzXjNMLU5ZgGMeYny8NVFQmUSDX7NRtnR6YnQp6oB6GY8F=
  </xenc:CipherValue>
  </xenc:CipherData>
</xenc:EncryptedData>
</content>
</subnet>
<place id="p1"/>
<transition id="t1"/>
<transition id="t2"/>
<arc id="a1" source="p1" target="t1">
  <inscription>
    <text> 3 </text>
  </inscription>
</arc>
<arc id="a2" source="t1" target="igp2"/>
<arc id="a3" source="t1" target="igp1"/>
<arc id="a4" source="ogt1" target="t2">
  <inscription>
    <text> 2 </text>
  </inscription>
</arc>
<arc id="a7" source="t2" target="p1"/>

```

Figure 18: Ciphered Petri subnet content

The subnet content has vanished and replaced by xml code, corresponding to the ciphered data.

5.2. XMLSignature

Although privacy is a solved question with XMLEncryption, there are other aspects of the security that XMLEncryption can't cover that are: integrity, authentication and non-repudiation.

In this case I want to sign the whole Petri net or concrete parts of it. Obviously, if I sign only a fragment of a Petri net, this part keeps integrity, authentication and non-repudiation, but the rest of net doesn't.

As with XMLEncryption, the best way to achieve these goals is using standard technologies. For signing, the method chosen is XMLSignature.

With XMLSignature we can sign any kind of content but the result is XML content. It requires the use of digital certificates and a set of public/private keys, using asymmetrical ciphering algorithms for the process.

This signature can be enveloped, enveloping or detached, but they are very similar. I am going to use enveloped signature

A XML enveloped signature consists of <Signature> tag that is defined in the namespace: <http://www.w3.org/2000/09/xmldsig#> and attached to the original xml file before closing the root node.

XMLSignature forces a digital signature to have:

- Canonicalization method: All the equivalent xml files are transformed in the same representative.
- Reference: Each reference indicates a part of the document that has to be signed. The set of

all of these references are signed together and generates only one signature.

- Key information: Optionally, the signature can include necessary information to be validated, for example, with X.509 digital certificates.
- Transforms: It is a ordered list of processing steps that has to be applied to the content before being signed.

Many times we will need to sign the whole Petri net, but it in this case I want to sign certain parts of a Petri net, for example a critical subprocess. The modus operandi here is similar to XMLEncryption. First of all, the content to be signed should be grouped in a subnet and then, this subnet is signed.

The standard way to indicate a subnet to sign in XMLSignature is through a XPath expression. In XMLSignature, the way to specify XPath addresses is using XMLSignature XPath Filter . XPathFilter returns the node set that is going to be signed and it is placed into /Signature/SignedInfo/Reference/Transforms as a new <Transform>.

I am not going to explain all the possibilities of XPath Filter. I will explain only those main configurations useful to my objective.

The exact configuration depends on the particular necessities of each case.

The goal in this work is the signing of a concrete subnet. In this case, it is a little different as in XMLEncryption. Remember that in XMLEncryption, if I want to mask a subnet I don't process the <subnet> tag but the subnet/content. This is because the interface has to be visible. But in a signature I want to sign the complete subnet, including the interface. Suppose that this subnet has id="sn1". The XPath expression that represents it is: /pnml/net/page/subnet[@id="sn1"]. And the result of signing the subnet of the figure 14 is:

```

<?xml version="1.0" encoding="UTF-8"?>
<pnml>
  <net id="myNet" type="http://www.pnml.org/version-2009/grammar/ptnet">
    <name>
      <text> My new net </text>
    </name>
    <page id="page1">
      .....
      <!--This is the same content as in the figure 15 -->
      .....
    </page>
  </net>
  <ds:Signature
    xmlns:ds="http://www.w3.org/2000/09/xmldsig#">
    <ds:SignedInfo>
      <ds:CanonicalizationMethod
        Algorithm="http://www.w3.org/TR/2001/REC-xm1-c14n-20010315"/>
      <ds:SignatureMethod
        Algorithm="http://www.w3.org/2000/09/xmldsig#rsa-sha1"/>
      <ds:Reference URI="">
        <ds:Transforms>
          <ds:Transform
            Algorithm="http://www.w3.org/2000/09/xmldsig#enveloped-signature"/>
          <ds:Transform
            Algorithm="http://www.w3.org/TR/2001/REC-xm1-c14n-20010315#WithComments"/>
          <ds:Transform
            Algorithm="http://www.w3.org/2002/06/xmldsig-filter2">

```

```

<dsig-xpath:XPath xmlns:dsig-
xpath="http://www.w3.org/2002/06/xmldsig-filter2"
Filter="intersect">
  /pnml/net/page/subnet[@id="sn1"]
</dsig-xpath:XPath>
</ds:Transform>
</ds:Transforms>
<ds:DigestMethod
Algorithm="http://www.w3.org/2000/09/xmldsig#sha1"/>
<ds:DigestValue>
prCzhLgTCZlck6MjQnFy6cASCZw=
</ds:DigestValue>
</ds:Reference>
</ds:SignedInfo>
<ds:SignatureValue>
Q0o7mQmGBFTg2UxgiZnz1snKi8V477JC0v12JPItL53zIOCPjhOwLoyxEN16v81C
r3GdgrgZimNXMJjwR4zkd9FVNCIrn85DurjHA/zDwSuPMg9w0N5A07c0xJ24uvn9
+2pbQxfblyTbkIy08+S0pqcZU/bv5+g=
</ds:SignatureValue>
<ds:KeyInfo>
<ds:X509Data>
<ds:X509Certificate>
MIICgTCCAeaggAwIBAgIETfh4CTANBgkqhkiG9w0BAQUFADCBDELMAkGA1UEBhMC
BAGTCEXBIJFTpBMREWdYDVQQHDAhMT0dST8K1TzEgMB4GA1UEChMXV5JkV5
TEEGUk1PSkEXDDAKBgNVBAsTAlBGQzEEMBOGA1UEAwWSck1SudPIExFw6BOIFNB
Fw0xMTA2MTUwOTE0ND1aFw0xMTA5MTMwOTE0ND1aMIGEMQswCQYDVQQGEwJFUzER
TEEGUk1PSkEXETAPBgNVBACMEExPR1JPwqVPMsAwHgYDVQQKEXdVTKLWRVJTSURB
SU9KOTEMMAoGALUECxDUEZDMR8wHQYDVQDDBZJwqVJUR08gTEEDoE4gU0FNQ0U5J
CSqGSIB3DQEBAAQAA4GNADCBiQKBgQCHePFNVCI fphFlyXQ9Bysir5BfXIuv3AnA
nVUjjeGnkUYQ032ouU+eEBK8Wsqjeh8A7zrHTRQjFYZnyuGWwM8gJX0a/POMROP
1/+zLwR0tYkqLI2xqDOFI2RwK5L2yGeV4T4y8i3h1U00FTSEWIDAQBA0GCSqG
A4GBAIDovAA4OCaTpy+83bGB2KmgMJrNxxWDPa15LGFrN8iCSHmbTpIeIbYBUAA
wD5Q0ENSFipQcH5GEpPM9Rquy6xMwfda9EU5UF0SEmbk4fK2vaIOVjynpQsJ9P9
9en02smQlyvw=
</ds:X509Certificate>
</ds:X509Data>
<ds:KeyValue>
<ds:RSAKeyValue>
<ds:Modulus>
oXjxTVQih6YRzcl0PQcrIkeQXlyLr9wJwCvNBbsOLUxcAplVIYXhp5FGEDT9gFLV
/AO86x00UI32GVsrhlqzPICVzmvz9DETj5v3Px/G+Wujedf/sy8EdLWJKIyNSagz
hSCNkcCuS9shnleE+MvIt4dVNDhU0Hm=
</ds:Modulus>
<ds:Exponent>AQAB</ds:Exponent>
</ds:RSAKeyValue>
</ds:KeyValue>
</ds:KeyInfo>
</ds:Signature>
</pnml>

```

6. CONCLUSIONS

I have explained a comprehensive method to define subnets with interfaces. The interaction between these subnets and the rest of the net is always through this front-end. The selected representation method selected has been PNML. Once this is done, the conclusion of this work is to show that it is possible to apply security measures to Petri nets in order to fulfill four characteristics to Petri nets, or parts of it:

1. Privacy: not everybody can access some data
2. Integrity: not allowed changes are detected
3. Authentication: ensure who guarantee some information
4. Non-repudiation: one person cannot supplant other one

With XMLEncryption we achieve privacy. Integrity, authentication and non-repudiation are achieved with XMLSignature

REFERENCES

Silva, M., 1985. Las Redes de Petri: en la Automática y en la Informática. Madrid: Editorial AC
 León, I., 2011. Security and Protection in Petri Nets sending and storage. Signing and Encryption.

Proceedings of EMSS, September 12-15, Rome (Italy)
 León, I., 2013. Analysis of information partial encryption options for exchanging Petri Nets systems. Proceedings of EMSS , September 25-27, Athens (Greece)
 León, I., 2014. Petri net representation with ciphered subnets: definition of PNML extensions for subnets representation and use of XMLEncryption for ciphering. Proceedings of EMSS, September 10-12, Bordeaux (France)
 Valette, R., 1979. Analysis of petri nets by stepwise refinements. Journal of Computer and System Sciences, 18(1):35–46.
 Suzuki, I and Murata, T., 1983. A method for stepwise refinement and abstraction of petri nets. Journal of Computer and System Sciences, 27(1):51–76.
 Fahmy, H.M.A., 1990. Analysis of petri nets by partitioning: Splitting transitions. Theoretical Computer Science, 77(3):321–330.
 Druzhinin, Va and Yuditskii, 1992. Construction of well-formed petri nets from standard subnets. Automation and Remote Control, 53(12):1922–1927.
 Fahmy,HMA,1993.Analysi of petri nets by partitioning: splitting places or transitions. International Journal of Computer Mathematics, 48(3-4): 127–148.
 Xia, C., 2011. Analysis and application of petri subnet reduction. Procs. of the IEEE, 6 (8):1662–1669.
 Murata, T., 1989. Petri nets: properties, analysis and applications. Proceedings of the IEEE, 77(4):541–580.
 Silva, M., 1993. In Practice of Petri Nets in Manufacturing. Chapman and Hall, London, UK.
 David, R. and Alla, H., 2010.. Discrete, Continuous and Hybrid Petri Nets. Springer, Berlin, Germany, 1st ed., 2004 edition.
 Peterson, JL., 1981. Petri Net Theory and the Modeling of Systems. Prentice Hall, Englewood Cliffs, NJ.
 Jensen, K. and Kristensen L.M., 2009. Coloured Petri Nets: Modelling and validation of concurrent systems.
 Billington, J., Christensen,S., Van Hee, K., Kindler, E., Kummer, O., Petrucci, L., Post, R., Stehno, C. and Weber, M., 2003. The petri net markup language: Concepts, technology, and tools. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2679: 483–505.
 Iso/iec 15909-2:2011 - systems and software engineering – high-level petri nets – part 2: Transfer format. http://www.iso.org/iso/catalogue_detail.htm?csnumber=43538. [Online; accessed 21-May-2015].
 Xml encryption syntax and processing version 1.1, 2013. URL <http://www.w3.org/TR/xmlenc-core1/>. [Online; accessed 20-Jun-2015].
 Xml signature syntax and processing version 1.1, 2013. URL <http://www.w3.org/TR/xmldsig-core1/>. [Online; accessed 20-Jun-2015].

LCA SOFTWARE FOR ENVIRONMENTAL IMPACT ASSESMENT OF INJECTED MOULDED PLASTIC PARTS

Ana Elduque^(a), Daniel Elduque^(b), Carmelo Pina^(c), Carlos Javierre^(d), Ángel Fernández^(e)

^{(a),(b),(d),(e)} Department of Mechanical Engineering, University of Zaragoza, C/María de Luna, 3, 50018 Zaragoza, Spain
^(c) BSH Electrodomésticos España, S.A., Avda de la Industria, 49, Zaragoza (Spain)

^(a) aelduque@unizar.es, ^(b) delduque@unizar.es, ^(c) carmelo.pina@bshg.com ^(d) carlos.javierre@unizar.es, ^(e) angel.fernandez@unizar.es

ABSTRACT

A software for the simulation of the environmental impact of injected moulded plastic parts is presented in this paper. The LCA model and assumptions made for calculations are explained, and the software's interface is shown. A case study is carried out in order to provide a clearer explanation of the applicability of this environmental impact simulation software. Results show how the raw material is the most influential factor. In second position is the injection moulding process, which has been modelled in detail, depending on many parameters.

Keywords: environmental impact simulation, software, life cycle assessment, injection moulding

1. INTRODUCTION

A more environmental benign manufacturing is an important goal for all industrial companies nowadays. As pointed out by (Dufloy et al. 2012), a more demanding environmental legislation (European Parliament 2009), achieving cost savings thanks to more efficient technologies, or competitive advantages, should be important reasons to motivate a movement towards sustainability.

During the last decades, simulation tools have been developed in order to evaluate the environmental performance of products and processes. One of the most important methodologies to achieve that is the Life Cycle Assessment. There are numerous examples of this methodology where the environmental performance of different kind of products or processes is simulated. Studies like the ones performed by (Martínez et al. 2009), (Cellura et al. 2012), (Ribeiro et al. 2013a), (Jiménez et al. 2014), (Elduque et al. 2014b), (Tsiropoulos et al. 2015), (Martínez et al. 2015), (Wäger and Hischer 2015) and (Simon et al. 2015) are some examples of them.

Plastics manufacturing is one of the most important industrial manufacturing processes worldwide.

Several authors have analyzed this process, trying to improve its efficiency by means of design of experiments (Packianather et al. 2013) or software

simulation. Also calculating its environmental impact in detail is an issue that is being addressed in other researches (Elduque et al. 2015).

In this paper a software simulation tool to perform the LCA of an injected moulded part is presented.

When a LCA is carried out, all the life stages of the product have to be taken into account in the model. Thereby, the production of the polymer, distribution of the raw material to the manufacturing company, the usage of auxiliary equipment and the injection moulding process itself should be included in the simulation. Also packaging materials used to deliver the product to the client, its distribution and its end-of life should be considered, as well as the waste generated throughout the life cycle, (Figure 1).

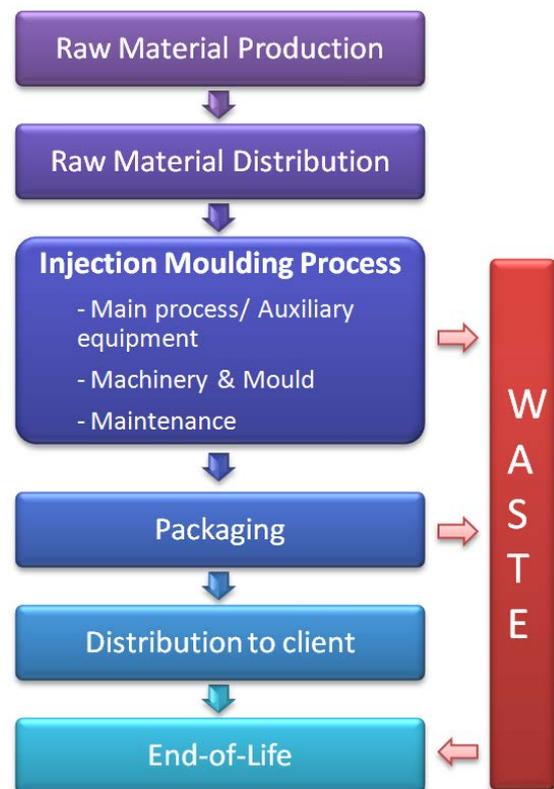


Figure 1: Life Cycle Stages of an Injected Moulded Part

This is a relevant research topic as a wide range of authors have analyzed the energy efficiency of this process to reduce it (Gutowski et al. 2006), (Borchardt et al. 2011), (Huang and Yang 2012), (Madan et al. 2014), (Müller et al. 2014), (Spiering et al. 2015) whereas others have analyzed the Life Cycle Cost of injection moulding (Ribeiro et al. 2013b). Also, several researchers have tried to reduce the impact of plastic parts (Lucchetta and Bariani 2010), (Park et al. 2013). Simulations tools have been widely used to improve injection moulding (Gerber et al. 2006), (Aisa et al. 2006), (Jimenez et al. 2009), (Fernandez et al. 2013). A streamlined environmental impact simulation tool based on an LCA model, that takes into account all the relevant factors related to the process of manufacturing an injected moulded plastic part, is presented in this paper.

2. SOFTWARE DEVELOPMENT AND LCA MODELLING IMPLEMENTATION

This LCA software has been developed using Visual Basic .NET having a very clear structure. It has three main blocks: databases used as inputs to the environmental impact simulation model, user interface, where input data are introduced by the user throughout several screens to characterize the studied part and results are displayed and, at last, the software's code which acts as a black box for the user. The input data are processed, saving it in variables, which interoperate and return the simulation results that are then displayed in the interface.

In subsection 2.2, the program's screens are going to be described as well as the internal assumptions that are behind them to simulate the environmental impact of the considered blocks.

In order to help the user during the data introduction, multiple guides and suggestions are included in the software's interface.

2.1. Databases

In order to simulate the environmental impact results of an injected moulded plastic part, several databases have been developed. One of them contains the inventory datasets and their environmental impact calculated in mPt (ReCiPe Methodology) and kilograms of CO₂ eq. Aspects like the electric mix, lubricating oil, materials such as polymers for the injection moulding, metals for the mould, packaging materials...etc. are included. These values have been obtained from EcoInvent database.

Other databases contain properties of materials, or data related to required equipment for the process that the program uses, for example, to simulate the energy consumption.

2.2. Software Methodology

Software interface is divided into different screens where required data can be introduced by the user or selected between different modelling options provided by the software.

A summarized flux diagram of the software is shown in Figure 2.

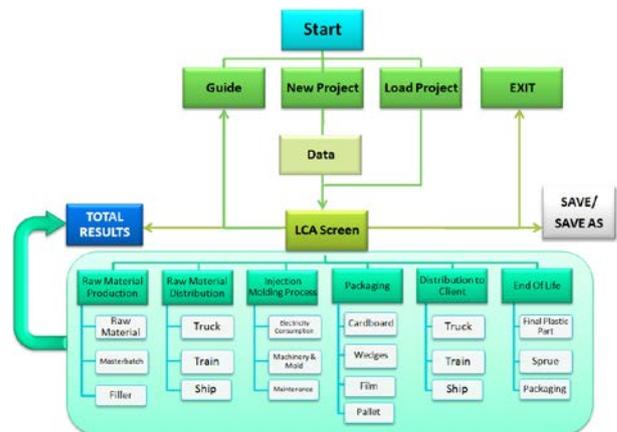


Figure 2: Software's Flux Diagram

Data regarding the part's net weight, the mould's feeding system and the number of the mould's cavities are required at the start of the modelling case study, in order to determine the quantity of sprue associated with the final part. This sprue could be assessed separately at the end-of-life phase. For example a mould without cold runners and one cavity hasn't got any sprue, but a plastic part obtained from a multi-cavity mould with a cold runner feeding system, will have associated the proportion of the runner also as used material. In addition, the total of kilograms injected to obtain the moulded part are calculated, as the section of the injection molding process simulates the environmental impact per processed kilogram.

Figure 3 shows the main LCA screen of the program which allows to go to the input data screen of each phase of the plastic part's life cycle.

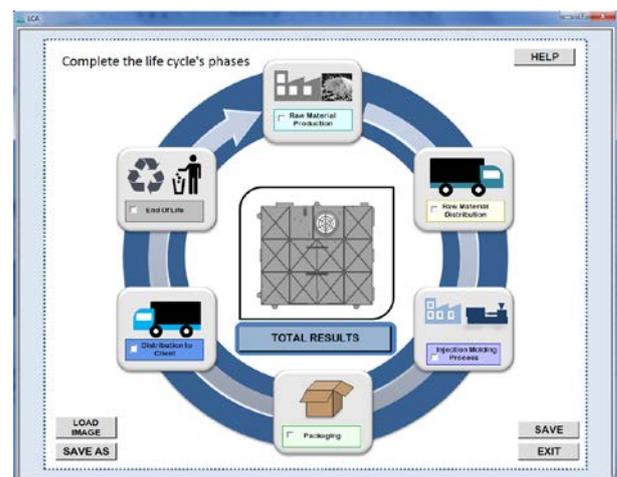


Figure 3: Software's LCA Screen

Environmental impact results are displayed in each section to check how the input data affect the results but also a "Total Results" screen has been developed to compare which life cycle phases have more importance in the ReCiPe Methodology and the carbon footprint (kg CO₂ equivalent).

2.2.1. Raw Material Production

The first phase of the LCA model for an injected moulded part is the raw material production. The database provides the environmental impact of materials obtained from professional Life Cycle Inventory databases such as EcoInvent. In addition, if the user wants to incorporate a new raw material, a mix of thermoplastic polymer, masterbatch and fillers can be modelled to better evaluate the environmental impact of the part's material.

2.2.2. Raw Material Distribution

The raw material is then distributed to the conversion's factory. A transport route can be configured using either road, rail or marine transport. The environmental impact's unit for this section is kg.km and the transported weight corresponds to the part's gross weight. An extra weight can be added if sacks or boxes are considered for the modelling of this phase.

2.2.3. Injection Moulding Process

The most important part of the program is the section where the environmental impact of the process is modelled.

Energy consumption is simulated as a sum of individual electric consumptions like the heating of the barrel, the hot runners if they are used, pneumatic or conveyor systems, dryers, coolers, and the drivers of the injection moulding machine. If known, it is possible to introduce experimental power measurements of the machine, in order to have a more precise model and therefore more adequate simulation results (Figure 4).

If it is not possible to introduce experimental measurements, a simulation of the energy consumption can be done. Energy consumption is modelled as a sum of six blocks:

- **Plastification:** the required energy is approximated as the specific heat of the thermoplastic material multiplied by the difference between injection temperature and factory's temperature, and divided by an assumed barrel's yield value.
- **Hot runner:** two options have been considered for the model. First one is to consider its energy consumption as a percentage of the plastification phase. The other one provides a power value (in kW) of typical equipment and a usage factor that usually is 50% (Thiriez 2006).
- **Conveying system:** manufacturer's specifications from several equipments of different power's level are provided. The user can select between different modelling options and indicate if the equipment's working point is optimum, in average load or over-sized. A value of consumption is returned depending on the selection. This value was obtained as an arithmetic mean of the different working points.

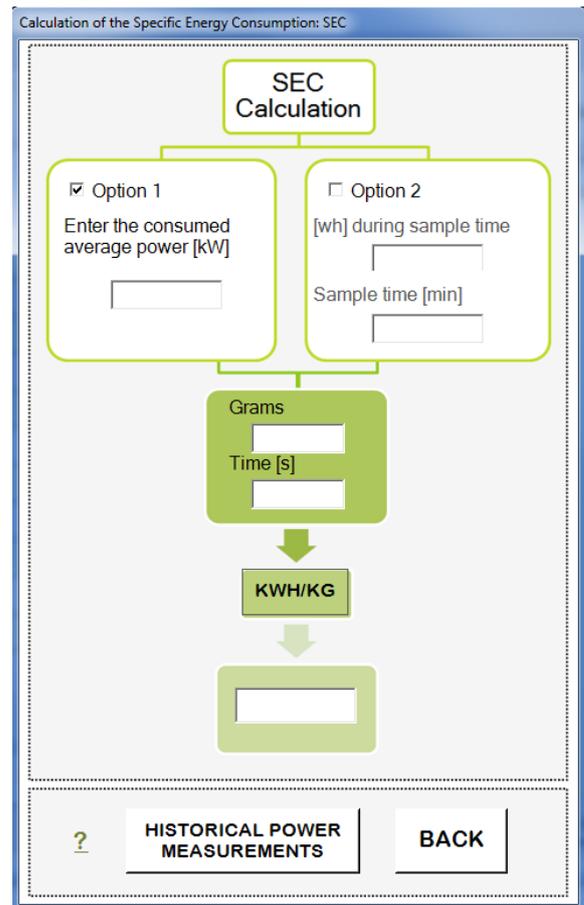


Figure 4: Specific Energy Consumption Calculator

- **Dryer:** as in other phases two alternatives can be selected for the model. Values of kWh/kg needed by material are provided (Thiriez 2006). Also the energy consumption in kWh/kg can be determined by introducing the dryer's power and the kilograms to be dried. A typical drying time for the material is suggested by the program's database.
- **Refrigeration:** cooling of the mould is simulated as the specific heat of the part's material multiplied by the difference between the injection and ejection temperature, and divided by a coefficient of performance. The oil hydraulic's cooling is of great importance as well, due to the oil's temperature has to be maintained stable to assure a good performance. The kWh/kg required is modelled with a lineal correlation obtained from specifications of a chiller's manufacturer.
- **Drivers:** the energy requirements for mould clamping (opening and closure) and the extruder are incorporated to the model.

In addition to the electricity consumption, the machinery and the mould, needed to manufacture the plastic part, have to be included as well in the model. Therefore the environmental impacts of metals like steel, aluminium or copper, which are part of the mould,

are also assessed. They contribute to the final results proportionally, as a function of the quantity of parts produced with each one (Figure 5 and Figure 6). Its end of life scenario is also included, providing suggestions of percentage for recycling and landfilling.

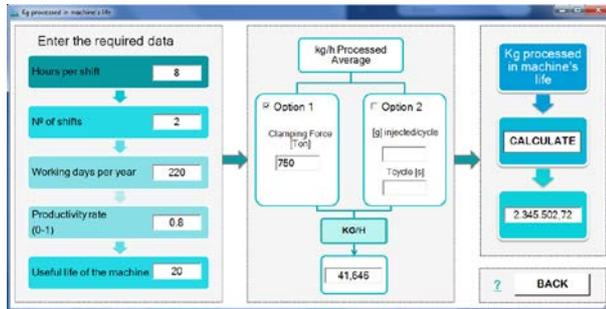


Figure 5: Production in the Injection Machine's Life

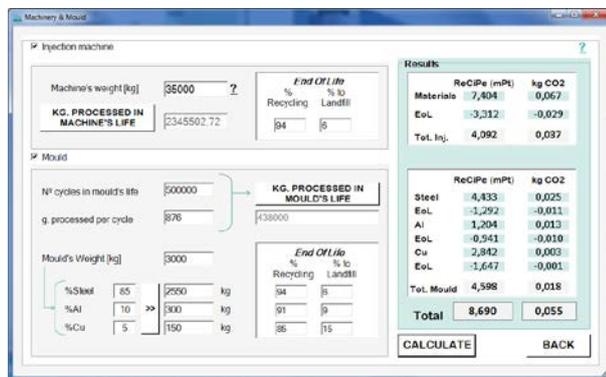


Figure 6: Environmental Impact Calculation of the Mould and the Injection Machine

The maintenance of the machine is also considered in the LCA model. The capacity of the oil tank has to be introduced along with the work hours between changes of the hydraulic oil in order to evaluate the environmental impact of the machine's maintenance. The environmental burden of the tank's oil has to be divided between all the kilograms injected during the time that has been used.

2.2.4. Packaging

Packaging materials like corrugated cardboard, expanded polystyrene, or plastic film, made for example of polyethylene, can be considered in this section. These three materials constitute a box in which one or several parts can be packaged. Also pallets can be introduced as a part of the packaging, dividing its environmental impact between the number of parts per pallet. As some of these packaging materials might be reused, a number of uses can also be assigned for pallets, wedges and cardboard to properly assess its environmental impact. Giving the main dimensions of the part, an estimation of the packaging material needed can be done in a help's screen, taking into account the volume of the part.

2.2.5. Distribution to Client

Similarly to the raw material distribution, the transportation of the manufactured plastic part is taken

into account in our model. Two different routes can be configured in order to give flexibility to the introduction of data.

Different scenarios can be simulated, for example if the plastic part is manufactured in a different factory than the assembly's factory. In both routes the user has to indicate the weight to be considered, net part's weight with or without the packaging weight.

2.2.6. End-of-Life

In this last phase, an end of life scenario is assigned for both the part and its sprue and the packaging materials, considering percentages of recycling, incineration or landfilling.

The percentage values suggested are from the technical report IEC TR62635 (IEC 2012), but they can be changed by the user if it is convenient to do so.

The sprue from the injection moulding process can be treated as the final part, or regrinded if it is reused in the same factory to inject another plastic part. If this alternative is selected, the environmental impact of this material will be subtracted from the total results. An electricity consumption value for the regrind process will be considered instead.

The end-of-life scenario for the packaging materials is also included in the LCA model, using values from Eurostat (Eurostat 2015).

3. CASE STUDY

In order to provide a clearer explanation of how the program works and give some quantitative results, an application example is going to be shown.

3.1. Part's Data

The plastic part that is going to be studied is a housing of an induction hob as the one analyzed by (Elduque et al. 2014a). Its function is to fix the electronic boards of the induction hob as well as closing the assembly. This plastic part is made of polyamide reinforced with glass fibre (PA66 GF30). It has a weight of 876 grams and its main dimensions are 460x415x40 mm (Figure 7).

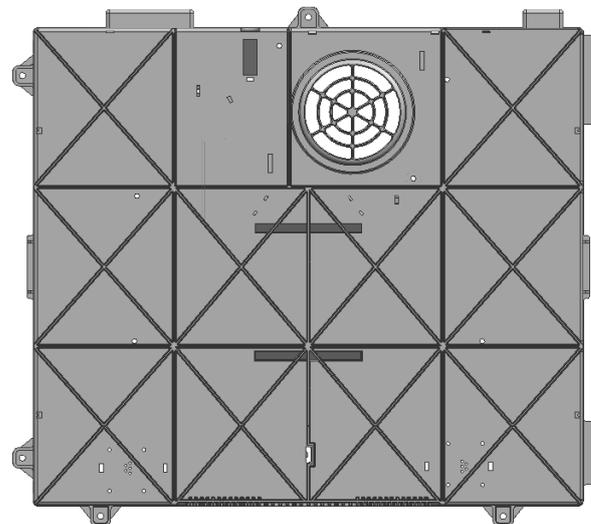


Figure 7: 3D Model of the Case Study Part

Its mould has only one cavity, and it has not got cold runners (Figure 8). This plastic part is injected in an injection moulding machine of 750 tons of clamping force.

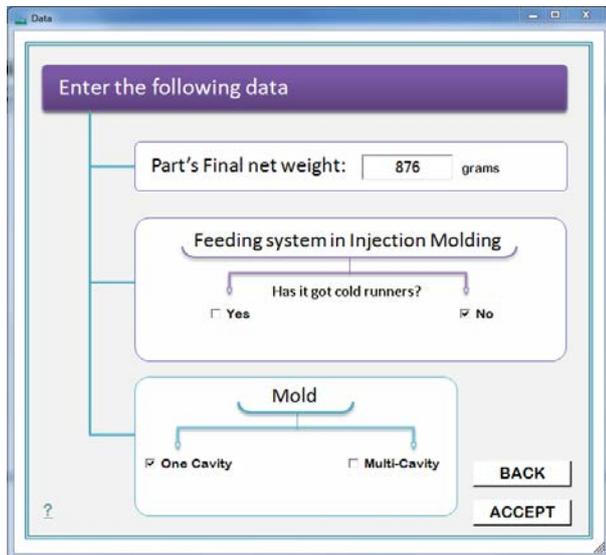


Figure 8: Data Required at the Start of the Modelling Case Study

3.2. Raw Material

Figure 9 shows the software's screen where the raw material is selected. For this case the "Nylon 6-6, glass-filled (RER) production" EcoInvent dataset was selected as the 100% of raw material. This thermoplastic has a high environmental impact in both studied categories. (522,6 mPt/part and 6,17 kg of CO₂ eq./part).

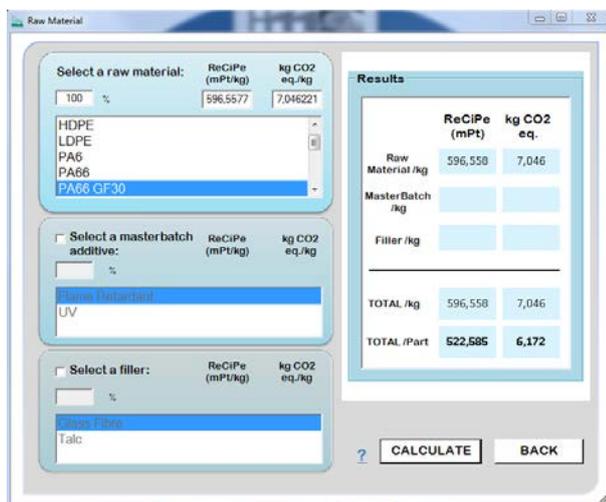


Figure 9: Selection of the Raw Material

3.3. Raw Material Distribution

A distance of three hundred kilometres is introduced to distribute the raw material to the factory by a >32 ton EURO 5 truck where the part is going to be manufactured. (Figure 10).

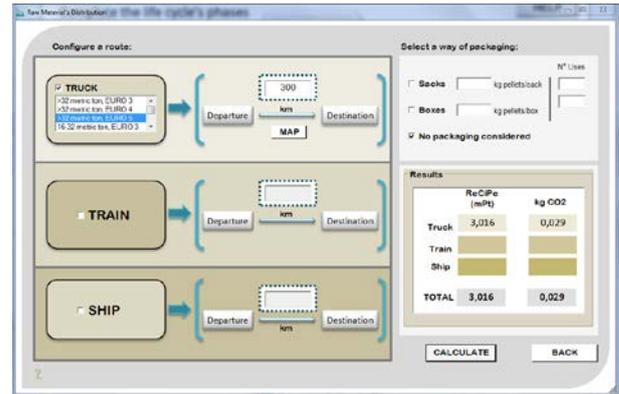


Figure 10: Raw Material's Distribution

3.4. Injection Moulding Process

To simulate the environmental impact of this phase, three subsections have to be completed. The most important one, as the results will show, is the electricity consumption. Figure 11 indicates the introduced data for the plastification phase, where the injection temperature is 300 °C.

Also the consumption of the hot runner is estimated considering an equipment of 5.35 kW and a usage factor of 50%.

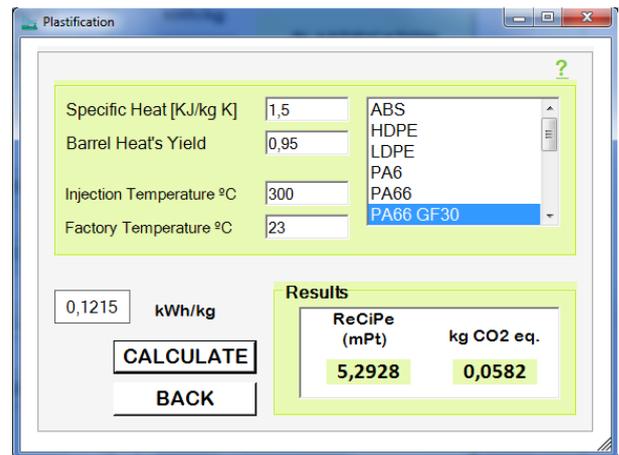


Figure 11: Modelling of the Plastification Phase

Using the program's help, a 2.2 kW power for the conveying system is selected, which for this case is an optimum working point taking into account the kg/h required by the manufacturing process (Figure 12).

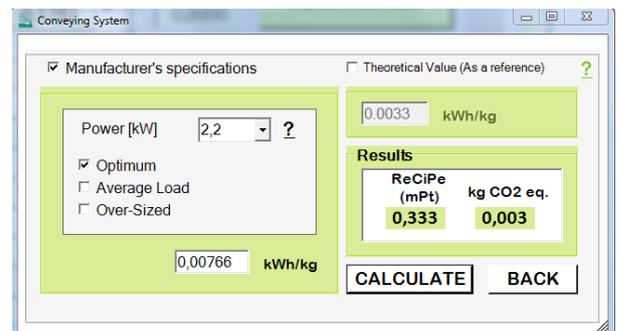


Figure 12: Modelling of the Conveying System

The raw material of this part (PA66) requires a drying process. A reference value of electricity consumption of 0.07 kWh/kg extracted from the literature is selected. The oil's cooling is considered in the refrigeration block considering the size of the injection machine. Given that this component is manufactured in Spain, a Spanish electric mix is selected for this model. In Figure 13, the main screen for the electricity consumption collects the results of this section.



Figure 13: Modelling of the Electricity Consumption

An electricity consumption of 0,786 kWh/kg is obtained which is considered as an acceptable estimation, as this value is only a five percent higher than the one registered by the power measurement equipment (Elduque et al. 2014a).

It is worth noting that the plastification phase and the drivers of the injection moulding machine represents almost the 70% of the electricity consumption.

The second subsection to evaluate this manufacturing process is the Machinery & Mould. The injection moulding machine weights 35000 kilograms. As previously showed in Figure 5, a calculation of the kilograms produced during the useful life of the machine is used to evaluate its impact.

In addition the environmental impact of the 3000 kg-mould is divided between the total of kilograms processed in the mould's life, considering 500000 cycles (Figure 6). An end of life scenario is assigned for the steel, copper, and aluminium belonging to the injection machine and the mould. Also the hydraulic oil consumption is calculated (Figure 14).

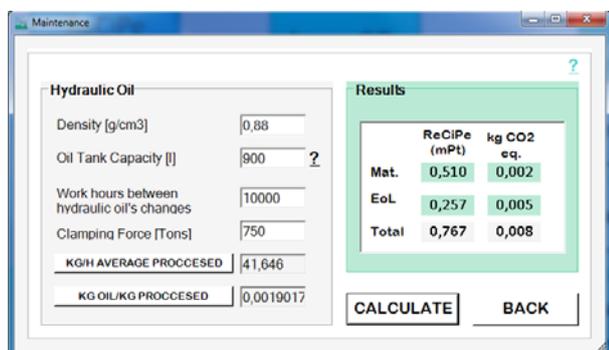


Figure 14: Maintenance for the Injection Moulding Process

The final results of the injection moulding process phase, are obtained by processed kilogram, therefore, in the total results this value will be multiplied by the part's gross weight (0.876 kg).

As it can be seen in Figure 15, the electricity consumption is the most influential factor in the environmental impact of the injection moulding process (nearly 80% of the impact simulation results for the ReCiPe methodology and almost 86% for the carbon footprint).



Figure 15: Results for the Injection Moulding Process

3.5. Packaging

The packaging required to distribute the manufactured part to the assembly plant is estimated as the next figure shows (Figure 16).

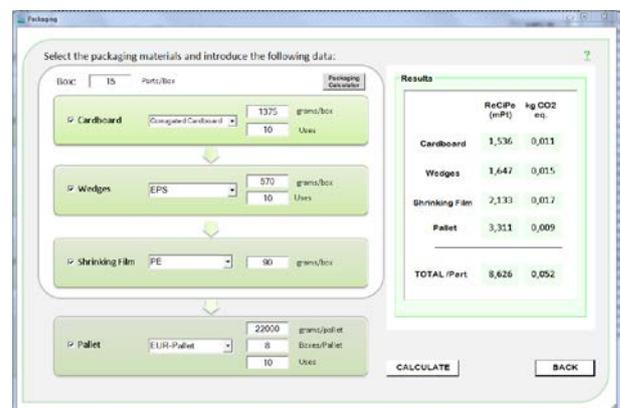


Figure 16: Packaging Materials Selected for the Case Study

3.6. Distribution

For this case study, two different routes are configured. The first one of 35 kilometres covers the distance between the injection factory and assembly's factory where this part will be assembled to the induction hob. The weight of the packaging materials is added to the part's weight for this calculation.

In addition, it is going to be considered the final distribution of this part to client, in order to observe the influence in the distribution's phase of the weight of this part. To achieve this, a total of 1800 kilometres by truck are considered (Pina et al. 2015).

For both routes a >32 ton EURO 5 truck is considered, using EcoInvent's dataset (Figure 17).

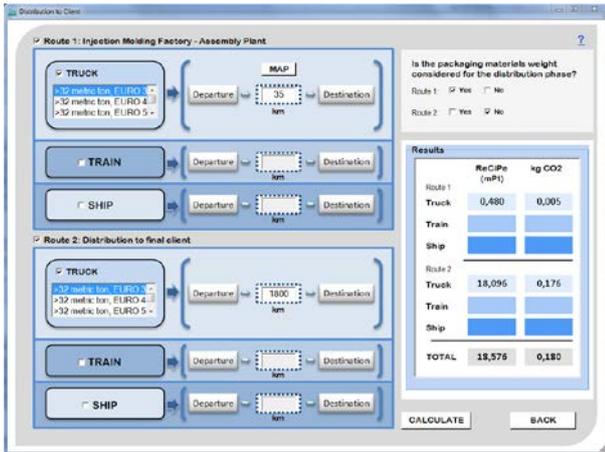


Figure 17: Distribution to Client

3.7. End of Life

As the material of this part is a reinforced thermoplastic, it is considered as not suitable for direct recycling and it is all sent to landfill, increasing the environmental impact of the material in 8,4 mPt/part, and 0,08 kg CO₂ eq./part. The end of life scenario for the packaging materials is included as well, as is indicated in Figure 18.

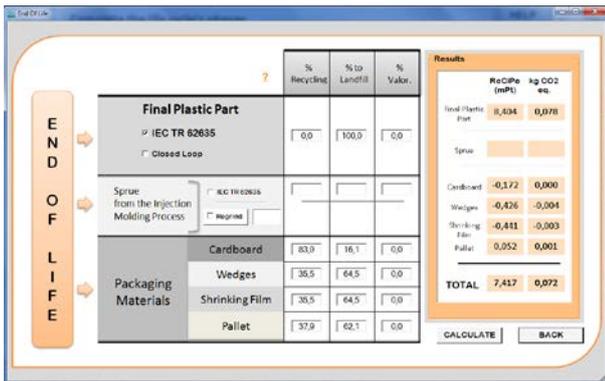


Figure 18: End-of-Life Scenarios

3.8. Results

Figure 19 summarizes the final results of this case study. The environmental impact of the polyamide is quite high in comparison with other thermoplastics. This causes that more than the 85% of the environmental impact for both categories is due to the raw material production. Nevertheless the manufacturing process contributes with more than 6% to the total results.

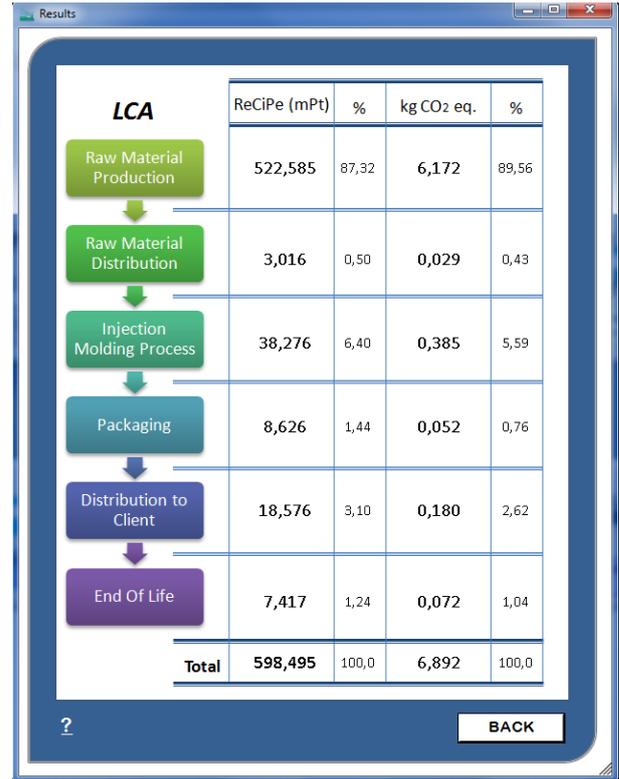


Figure 19: Total Results of the Case Study

4. CONCLUSIONS

The software shown in this paper simulates the environmental impact of an injected moulded plastic part by means of a LCA model. The results provided by this program allows designers to evaluate in detail the environmental impact of a specific plastic part and thereby simulate different alternatives, as the model is sensitive to many different parameters such as the raw material of the studied part, routes and means of transport used for distribution, the characteristics of the mould or the injection moulding machine that define the manufacturing process, or several end of life scenarios to complete the life cycle analysis simulation.

Results from the case study and other studied parts by the authors have shown that usually the most important phase, due to its higher environmental impact, is the production of the raw material but also the second one which has a remarkable importance is the manufacturing process, depending on many factors on which designers or engineers have influence.

Having this simulation tool implemented, our future work would be to relate these simulations of the harm caused to the environment, with the economic cost of the part's manufacturing. The modelling and simulation performed by this software would allow companies to evaluate in a two dimensional way the sustainability of their products. In addition, several alternatives could be compared, letting the engineer choose the most favourable case, achieving impact's reductions from an economic and environmental standpoint.

REFERENCES

- Aisa, J., Javierre, C., De la Serna, J., 2006. An example of simulation tools use for large injection moulds design: The CONTENUR™ 2400 l solid waste container. *Journal of Materials Processing Technology*, 175(1-3), pp. 15-19.
- Borchardt M., Wendt M.H., Pereira G.M., Sellitto M.A., 2011. Redesign of a component based on ecodesign practices: environmental impact and cost reduction achievements. *Journal of Cleaner Production*, 19, pp. 49-57.
- Cellura, M., Longo, S., Mistretta, M., 2012. Life Cycle Assessment (LCA) of protected crops: an Italian case study. *Journal of Cleaner Production*, 28, pp. 56-62.
- Duflou, J.R., Kellens, K., Renaldi, R., Guo, Y., Dewulf, W. 2012. Critical comparison of methods to determine the energy input for discrete manufacturing processes. *CIRP Annals - Manufacturing Technology*, 61, pp. 63-66.
- Elduque, A., Javierre, C., Elduque, D., Fernández, A., 2015. LCI Databases Sensitivity Analysis of the Environmental Impact of the Injection Molding Process. *Sustainability*, 7 (4), pp. 3792-3800.
- Elduque D., Clavería I., Fernández, Á., Javierre, C., Pina, C., Santolaria, J. 2014a. Analysis of the Influence of Microcellular Injection Molding on the Environmental Impact of an Industrial Component. *Advances in Mechanical Engineering*, 6, pp 7.
- Elduque, D., Javierre, C., Pina, C., Martínez, E., Jiménez, E. 2014b. Life cycle assessment of a domestic induction hob: Electronic boards. *Journal of Cleaner Production*, 76, pp. 74-84.
- Eurostat, 2015, Recycling rates for packaging waste, Available from: <http://ec.europa.eu/eurostat/data/database> [April 2015]
- European Parliament, 2009. Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009 Establishing a Framework for the Setting of Ecodesign Requirements for Energy-Related Products, EU Publications Office, Luxembourg: Official Journal of the European Union.
- 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), pp. 176-183.
- Gerber A.G., Dubay, R., Healy, A., 2006. CFD-based predictive control of melt temperature in plastic injection molding. *Applied Mathematical Modelling*, 30, pp. 884-903.
- Gutowski T., Dahmus J., Thiriez A., 2006. Electrical Energy Requirements for Manufacturing Processes. 13th CIRP International Conference of Life Cycle Engineering, May 31st – June 2nd, Lueven.
- Huang Y., Yang G., 2012. The Applied Research of Green Production Technologies Based on the Production in Plastic Molding Factories. *Energy Procedia*, 14, pp.247-254.
- IEC, IEC/TR 62635 Guidelines for end-of-life information provided by manufacturers and recyclers and for recyclability rate calculation of electrical and electronic equipment. Geneva, International Electrotechnical Commission.
- 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), pp. 23-30.
- Jiménez, E., Martínez, E., Blanco, J., Pérez, M., Graciano, C., 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.
- Lucchetta, G., Bariani, P.F., 2010. Sustainable design of injection moulded parts by material intensity reduction. *CIRP Annals - Manufacturing Technology*, 59, pp. 33-36.
- Madan J., Mani, M., Lee, J.H., Lyons K.W., 2014. Energy performance evaluation and improvement of unit-manufacturing processes: injection molding case study. *Journal of Cleaner Production*, pp. 1-14.
- Martinez, E., Sanz, F., Pellegrini, S., Jiménez, E., Blanco, J., 2009. Life cycle assessment of a multi-megawatt wind turbine. *Renewable Energy*, 34(3), pp. 667-673.
- Martínez, E., Blanco J., Jiménez, E., Saenz-Díez J.C., Sanz F., 2015. Comparative evaluation of life cycle impact assessment software tools through a wind turbine case study. *Renewable Energy*, 74, pp- 237-246.
- Müller E., Schillig R., Stock T., Schmeiler M., 2014. Improvement of injection moulding processes by using dual energy signatures. *Procedia CIRP* 17, pp.704-709.
- Packianather, M., Chan, F., Griffiths, S., Dimov, S., Pham, D.T., 2013. Optimisation of micro injection moulding process through design of experiments. *Procedia CIRP* 12, pp. 300-305.
- Park, H.S., Dang, X.P., Roderburg, A., Nau, B., 2013. Development of plastic front side panels for green cars. *CIRP Journal of Manufacturing Science and Technology*, 6, pp. 44-52.
- Pina C., Elduque D., Javierre C., Martínez E., Jiménez E., 2015. Influence of mechanical design on the evolution of the environmental impact of an induction hob. *International Journal of Life Cycle Assessment*, 20(7), pp. 937-946.
- Ribeiro, I., Peças, P., Henriques, E., 2013a. A life cycle framework to support materials selection for Ecodesign: A case study on biodegradable polymers. *Materials & Design*, 51, pp. 300-308.

- Ribeiro, I., Peças, P., Henriques, E., 2013b. Incorporating tool design into a comprehensive life cycle cost framework using the case of injection molding. *Journal of Cleaner Production*, 53, pp. 297-309.
- Simon, B. Amor, M.B., Földényi, R., 2015. Life cycle impact assessment of beverage packaging systems: focus on the collection of post-consumer bottles. *Journal of Cleaner Production*, pp. 1-11.
- Spiering T., Kohlitz S., Sundmaeker H., Herrmann C., 2015. Energy efficiency benchmarking for injection moulding processes. *Robotics and Computer-Integrated Manufacturing*, 36, pp. 45-59.
- Thiriez, A., 2006, *An Environmental Analysis of Injection Molding*, Massachusetts Institute of Technology.
- Tsiropoulos, I., Faaij, A.P.C., Lundquist, L., Schenker, U., Briois, J.F., Patel, M.K., 2015. Life cycle impact assessment of bio-based plastics from sugarcane ethanol. *Journal of Cleaner Production*, 90, pp. 114-127.
- Wäger, P.A., Hischer, R., 2015. Life cycle assessment of post-consumer plastics production from waste electrical and electronic equipment (WEEE) treatment residues in a Central European plastics recycling plant. *Science of the Total Environment*, 529, pp. 158-167.

SIMULATION-BASED ASSESSMENT OF THE INFLUENCE OF REQUESTS FOR INFORMATION ON INDUSTRIAL MODULE CONSTRUCTION PIPING WORK

Amin Amini Khafri ^(a), Wenying Ji ^(b), Seungjun Ahn ^(c), Simaan AbouRizk ^(c)

^{(a),(b),(c),(d)}University of Alberta
Edmonton, Alberta, Canada

^(a)amin.amini@ualberta.ca, ^(b)wenying1@ualberta.ca, ^(c)seungjun@ualberta.ca, ^(d)abourizk@ualberta.ca

ABSTRACT

Engineering errors (e.g., missing information and ambiguities) in isometric drawings result in significant labor productivity loss and schedule delays in industrial modular construction projects. Given this background, this paper aims to assess the influence of requests for information (RFIs) on module construction duration using computer simulation. In particular, pipe installation tasks are of interest in this paper due to the large number of RFIs related to pipe designs occurring in many modular construction projects. A hybrid modeling approach combining discrete event simulation (DES) and continuous simulation (CS) was used to model the piping work of a module construction project in Alberta, Canada, considering the impacts of RFIs on construction duration. Through simulation, it was assessed that RFIs will increase the piping work duration by 8% on average on the most-likely scenario of the project.

Keywords: simulation, request for information (RFI), module construction

1. INTRODUCTION

Over the past few decades, modularization has been implemented widely in lieu of the conventional stick-built construction in Alberta oil sands projects. Modularization has contributed to not only reducing the time and cost of on-site construction under northern Alberta's harsh weather conditions, but also improving the safety and quality performance of the projects.

Industrial modular construction is a construction method involving large-scale use of offsite prefabrication and preassembly (Taghaddos et al. 2014). The offsite, open-space locations where modules are assembled are called module assembly yards, or module yards. A typical industrial module construction process includes erection of structural steel, installation of pipe spools, equipment and electrical units, insulation, hydro testing, and fireproofing work. Due to the unique design and internal components of modules, each module in a project can be seen as a unique sub-project.

In a module construction project, requests for information (RFIs) are commonly issued by the constructor when they require further clarification or

information on the module design from the engineer (Hanna 2012). Responses to RFIs should be provided in a timely manner to mitigate the impact of design uncertainties on construction schedules. The number of RFIs and the response duration have been cited as significant factors influencing productivity loss and schedule delays in projects (Song 2009).

Given this background, this paper aims to assess the influence of RFIs on module construction duration using computer simulation. It is expected that assessing the schedule impacts of RFIs can help practitioners develop more accurate estimations of module construction durations.

In particular, pipe installation tasks are of interest in this paper due to the large number of RFIs related to pipe designs occurring in many module construction projects. Piping is usually the most complex work of industrial module construction due to the complicated design of pipes, the various types of work involved with different pipe sizes and shapes, and the various resources required. Figure 1 shows the distribution of RFIs on different types of work in an industrial module construction project in Alberta, Canada. In this particular project, piping-related RFIs made up 53% of the total number of RFIs. Such a large number of piping-related RFIs can significantly decrease labor productivity and delay the entire project schedule.

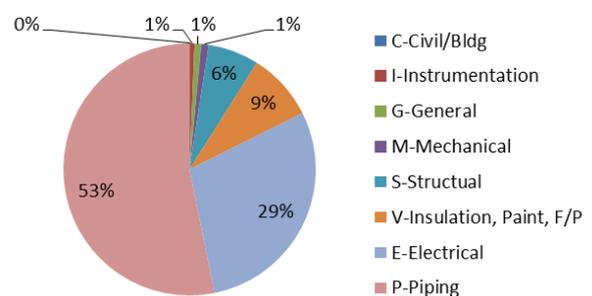


Figure 1: Pie Chart of RFIs

In the next section, an overview of simulation applications in construction, as well as a review of previous research efforts for simulating industrial module construction processes and assessing the

schedule impact of RFIs is presented. In the methodology section, a hybrid modeling approach combining DES and CS is introduced. Then, the piping work process is described and a simulation model for estimating piping work duration and piping work man-hours is presented. Lastly, the simulation results from the model are analyzed and the impact of RFIs on piping work duration is discussed.

2. LITERATURE REVIEW

As construction projects involve complex and uncertain processes, it is difficult to use conventional techniques to manage them (AbouRizk 2010, 2011). Construction simulation is “the science of developing and experimenting with computer-based representations of construction systems to understand their underlying behavior” (AbouRizk 2011). A simulation model is a composition of objects that represent the abstraction of a construction system. Once a simulation model is developed, the model output can be used to predict the process cycle time, productivity, resource utilization, etc. (AbouRizk 2010). Typically, simulation research is conducted in four phases: 1) product abstraction phase; 2) process abstraction and modeling phase; 3) experimentation phase; and 4) decision-making phase (AbouRizk 2010).

Mohamed et al. (2007) developed a discrete event simulation model to schedule resource-constrained module assembly processes. Taghaddos et al. (2009) tackled the scheduling complexity issue of a module assembly yard using the high level architecture technique. In addition, Taghaddos et al. (2014) attempted to schedule a module assembly yard using a simulation-based auction protocol to optimize module assembly yard layout.

Piping work is usually the most complex work in a module assembly yard, and therefore, it has a high rate of RFI occurrence. If an RFI is issued, workers are usually assigned to another workface until a response to the RFI is made, but such an interruption significantly decreases labor productivity (Hanna 2012). Many researchers have attempted to study the impacts of such interruptions on project duration. Christian and Hachey (1995) found that crews without interruptions improve their working productivity as time passes. Han et al. (2007) studied the effect of non-value adding activities, such as RFIs, on the project schedule and budget using a system dynamics simulation approach. However, there have been few research efforts to assess the schedule impact of RFIs in module construction.

3. METHODOLOGY

Assembling pipe spools in module construction involves multiple tasks, such as support welding, spool erection and pipe welding. The duration and man-hours required to conduct these tasks are dependent on the complexity of work. Since each module has unique specifications and combinations of pipes and supports, there is no normal or average size and quantity of pipes that can represent the typical module.

To deal with the different levels of work complexity involved in piping work, spool erection tasks are usually grouped into several work categories by the pipe size (diameter) in module construction projects. Then, based on the historical data, production rate is calculated for different work categories, and the rate is used to estimate construction durations. An assumption made here is that pipes under the same category are assembled with the same production rate. This kind of work rate information that is available in module construction projects is compatible with continuous simulation modeling (CS). Discrete event simulation (DES) modeling is usually more compatible with other construction tasks. Therefore, a hybrid simulation approach combining CS and DES was used for modeling piping work processes, considering RFI impacts.

3.1. Discrete Event Simulation (DES)

DES is an approach for modeling and simulating a dynamic system. DES models are simulated by advancing time in discrete steps based on the events that take place in the model. Since most of the operations in construction could be discretized to activities with specified start and finish times, DES is a suitable approach for modeling and simulating construction operations (Puri and Martinez 2012). Furthermore, DES is useful for representing repetitive processes, which are very common in construction tasks. In DES, activities can be modeled either stochastically or deterministically. Once a model is developed, detailed resource interactions and utilization in different scenarios can be observed through simulations. In addition, interruptions, such as RFIs, equipment breakdown, and shifts of work can be easily modeled in DES due to its discrete-time nature.

3.2. Continuous Simulation (CS)

CS is another approach for modeling and simulating a dynamic system. In CS, dynamic models are processed by incrementing time in equal steps (i.e., time steps). Therefore, changes in the model variables can occur at each of the time steps in CS. In certain circumstances, the simulationist may consider a construction operation as having a continuous nature (e.g., concrete pouring). To model a continuous phenomenon using CS, two types of variables, stocks and flow rates, are commonly used. A stock can be thought of as a container whose contents are changed by inflow and outflow, and a flow rate is a variable that determines the rate of the inflow or outflow. An illustrative example of this is a water tank; the tank is the stock, and the valve changes the inflow and outflow rate, and as a result, the stock value can change gradually.

3.3. A Hybrid Approach Combining CS and DES

The need to combine CS and DES models arises when we need to include both continuous variables and discrete events that can occur in one system (Klingener 1996). If the piping work in module construction is

modeled in DES, it would be challenging to define the exact start and finish time due to the continuous nature of pipes. On the other hand, CS is compatible with modeling a continuous nature process (i.e., work that can be best modeled by work rates). However, CS has limitations in modeling discrete events, such as interruptions. Therefore, a hybrid approach combining DES and CS could be a suitable approach to simulating the piping work in module construction.

To build a combined DES and CS model, the simulationist first needs to define the discrete events that can occur in the system and identify the variables that can be best modeled as a continuous variable. In the context of our research, the work sequence and possible interruptions are modeled in DES, while the production rate is modeled as a continuous variable.

4. MODEL DEVELOPMENT

In this section, a DES-CS hybrid simulation model developed to represent the piping work of a module construction project in Alberta is presented.

4.1. Model Description

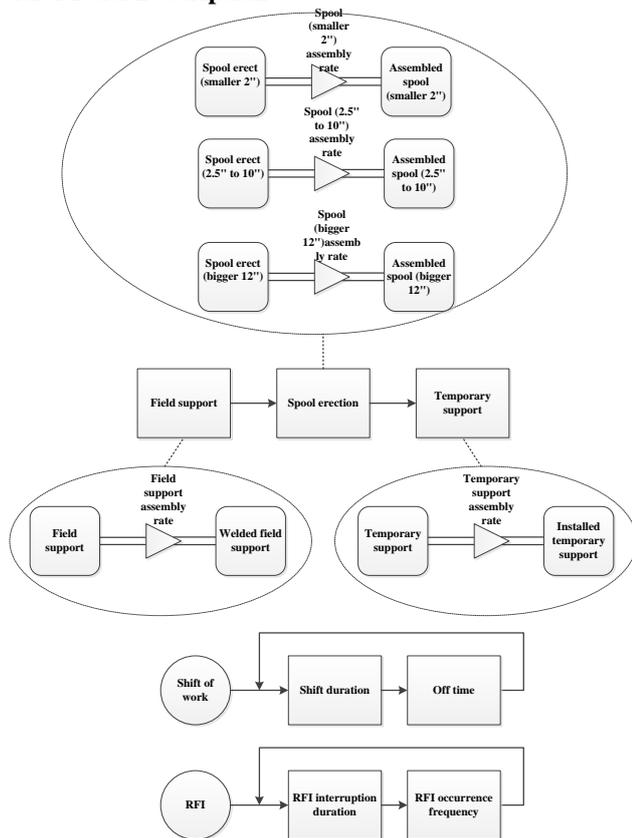


Figure 2: Conceptual Model of the Pipe Assembly Process

Figure 2 illustrates the conceptual model of the pipe assembly process. As shown in the figure, the high-level processes were modeled as discrete events, while the sub-processes—shown in the ovals in Figure 2—were modeled as continuous processes. Each piping task was also modeled as a continuous variable in the

model. A quantity was assigned to each piping task as the stock value, and a production rate was assigned to each variable as the outflow rate of each piping work task (i.e., the outflow of products). Each task was regarded as completed when the stock value became zero (i.e., there was no more work to do).

Table 1 shows the description of different piping tasks for simulating the piping work of the module construction project. In this particular project, three different categories of pipe spools (P1, P2 and P3) based on the pipe spool diameter, were used, as shown in Table 1.

Table 1: Tasks of Piping Work in the Industrial Module Construction Project

Task	Description
P1	Spool Erect - 2" (50 mm) & Less
P2	Spool Erect - 2-1/2 -10" (63 - 250 mm)
P3	Spool Erect - 12 -34" (300 - 850 mm)
P4	Field Supports
P5	Temporary Supports (Shipping)

The shifts of work and RFI interruptions are modeled as discrete events in the proposed model. The routine eight hour shifts of work take place every sixteen hours in the model. RFI interruptions are randomly generated according to an occurrence interval preset in the model. The piping work progress stops when an RFI is issued or during the non-working hours.

4.2. Model Inputs and Settings

In the simulation model, all inputs were set based on the real data collected from the case project. The real data were collected from the projects through their stored historical data or interviews with experienced managers/foremen. Table 2 shows the input settings for variables.

Based on the observations at the module assembly yard, the size of piping crews was determined to be nine to twelve, and two cranes were assigned to be shared among three modules. When an RFI is issued, it is likely to take three to five hours to start work on another work face. This duration includes checking the drawings, notifying the foreman and preparation work for the new task. Based on the interviews conducted with on-site experts involved in the studied case, it was found that RFI interruptions happen on average once a week to each of the modules. Literature shows that exponential distribution is suitable for modeling the RFI inter-arrival time (Nasrallah and Bou-Matar 2008).

Take-off quantities for each task for three particular modules in the project were collected from the module design package, and were utilized as the input value for the required amount of work. Table 2 shows the take-off quantities in linear meters for all the piping tasks for the selected three real modules.

Table 2: Take-off Quantity of the Tasks to be Executed on Each Module (P1 – P5 are Different Types of Piping Tasks)

Task	Quantity of Each Task (Linear Meter)		
	Module 1	Module 2	Module 3
P1	194.75	86.63	127.54
P2	38.81	14.32	62.42
P3	27.51	65.74	128.13
P4	261.07	166.69	293.72
P5	17.18	16.17	25.91

As per the Construction Industry Institute (2011), estimated productivity can be calculated for each task of the piping work using the following equation.

$$\text{Estimated Productivity} = \frac{\text{Estimated Man-hour}}{\text{Quantity}} \quad (1)$$

Take-off quantity and man-hour estimation data were utilized to determine the distribution of production rate for each task. Normal distribution was selected to represent the data set for the tasks P1, P2, and P3, and a constant production rate was used for P4 and P5 (i.e., temporary and field pipe support tasks), because of the pattern in the distribution of production rates. All fitted data sets passed the Kolmogorov-Smirnov test, which means the fitted curve was sufficient to represent the distribution of real data. The production rates for all five tasks of piping work are shown in Table 3 and Table 4.

Table 3: Production Rate Distribution for P1, P2, and P3 Based on 100% Labor Effectiveness (n: Sample Size; L: Low; H: High)

Task	Data Description	Productivity (P_i) Distribution
P1	n = 129 L = 1.11 H = 1.69	Normal(1.533,0.096) K-S = 0.097
P2	n = 128 L = 1.69 H = 2.97	Normal(2.395,0.367) K-S = 0.079
P3	n = 121 L = 2.85 H = 6.92	Normal(4.927,1.091) K-S = 0.115

Table 4: Constant Production Rate for P4 and P5 Based on 100% Labor Effectiveness

Task	Productivity (P_i)
P4	0.35
P5	1.10

It should be noted that the production rates shown in Table 3 and Table 4 are for each laborer. Therefore, the production rate needs to be multiplied by the number of workers in each crew to get the entire crew production rate.

In addition, a few assumptions, which are based on the result of observations and interviews, were made for setting the simulation model, as follows:

- 1) A piping crew foreman is assigned to manage a ten-laborer crew for all three modules.
- 2) The piping crew works eight hours per day.
- 3) RFIs occur once per week on average, and follows an exponential distribution.
- 4) The number of workers is determined at the beginning of each working day, and varies from three to five.
- 5) 50% of labor effectiveness is assumed based on a study of the Alberta labor force (Hewage et al. 2011). Through simulation, the total piping work duration and man-hours are measured for all three selected modules.

4.3. Model Validation

In order to validate the proposed model, the simulation parameters were first adjusted for the scenario of 100% production rate and no RFI occurrence, and the man-hours estimated by simulation were compared to the existing man-hour estimate. The results showed that the difference between the model output and the company's man-hour estimation was within 15% error, and this was regarded as acceptable, considering the simulation model was not intended to capture every factor that can influence the production rate of piping work. Since the base model is validated with the company's estimation data, the end user can now add more labor productivity factors to the model to get more realistic results with simulation.

5. SIMULATION RESULTS

Resource utilization is one of the things that can be observed from simulation, and it can be an indicator of how efficient the entire operation system is. Figure 3 shows the crane utilization during the piping work for the three modules in simulation. The crane utilization was estimated to be 72% on average during the pipe assembly process for all three modules. For tasks P4 and P5, cranes are not required, and thus the crane utilization for these tasks is zero. In contrast, during pipe spool erection tasks, P1, P2, and P3, cranes are shared between three modules. Practitioners can further optimize crane utilization based on this kind of simulation result. For example, during the period of tasks P4 and P5, the cranes can be shared with other groups of modules to increase the utilization of the cranes.

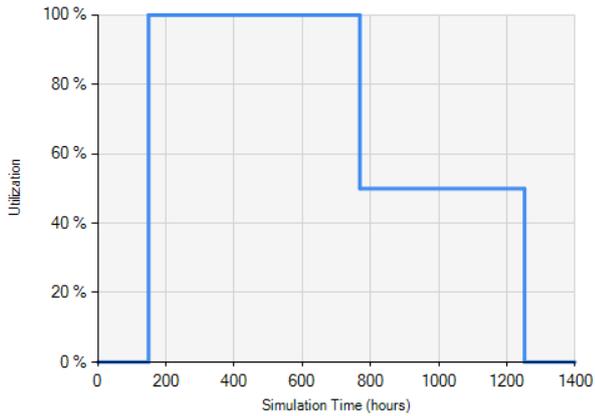


Figure 3: Crane Utilization

Another thing that can be observed from simulation is the pattern of work progress for each module. As an example, Figure 4 illustrates the piping work progress over time for Module 3. As can be inferred from Figure 4, task P1 could not begin until task P4 was completed. The different slope of each task progress is determined by the pipe assembly production rate, which depends on the randomly assigned production rate (according to the probability distribution) and the number of workers each working day. Horizontal segments in the graphs in the figure result for two reasons: 1) working shifts after each working day, and 2) work delays due to the RFI interruption.

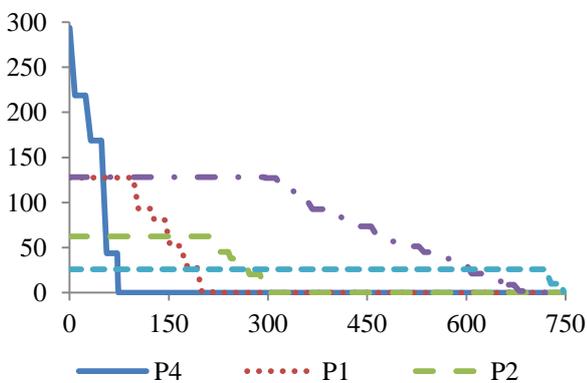


Figure 4: An Example of Work Progress Graph Generated from Simulation (Module 3)

As a next step, a sensitivity analysis was performed to see the impact of the RFI occurrence rate on the module construction durations/man-hours. Table 5 shows the results of running the model 100 iterations with different RFI occurrence rates. As expected, the total piping work duration is increased as the number of RFIs increases. In the most likely scenario, the duration was increased by four and half days (around 8%) of the total duration for constructing all three modules.

Table 1: Simulated Duration for Finishing Piping Work on Three Modules

	RFIs Occurrence per Week			
	0	1 (Most Likely)	2	3
Total Piping Duration (Days)	54.5	59.0	63.6	68.5

Similarly, simulation results also show that total man-hours required for piping work also increased as the number of RFIs increased. In the most likely scenario, total man-hours required to finish the piping work for all three modules increased by 7.7%. Table 6 shows the comparison of man-hours required to finish the piping work for different numbers of RFIs.

Table 6: Simulated Man-hours for Finishing Piping Work on Three Modules

	RFIs Occurrence per Week			
	0	1 (Most Likely)	2	3
Total Piping Man-hours (Labor-hours)	2047	2204	2391	2560

As a result of this sensitivity analysis, for the most likely RFI issuance frequency scenario (once a week), it was estimated that RFIs will increase total piping duration by 8%, and increase the total man-hours for the piping work by 7.8%.

6. CONCLUSION

In this research, a simulation model was developed using a hybrid simulation approach combining DES and CS, and the simulation was implemented using Symphony.NET 4.0 (Hajjar and AbouRizk 1999) to assess the RFI influence on piping work duration and man-hours in an industrial module construction project. Historical data from a real industrial module construction project in Alberta, Canada was used to build the model and to validate the simulation results. Simulation results showed that RFIs have an average of 8% impact on piping work duration.

It is expected that the simulation model for estimating piping work man-hours, taking into account the schedule impact of RFIs, can help estimators with assessing the impacts of different numbers of RFIs and achieving more accurate estimations. Work progress and resource utilization could also be observed from the simulation. The results show the proposed hybrid model performed well to analyze the pipe assembly process and to assess the schedule impacts of productivity factors.

Future work may include: 1) expanding the model for the entire module construction process; 2) collecting more data from real projects to train the models; 3) developing a more user-friendly estimation tool for

industry use; 4) using different labor production rates for the beginning and end of each shift of work.

REFERENCES

- AbouRizk, S., 2010. Role of simulation in construction engineering and management. *Journal of Construction Engineering and Management*, 136(10), 1140–1153.
- AbouRizk, S., Halpin, D., Mohamed, Y. and Hermann, U., 2011. Research in modeling and simulation for improving construction engineering operations. *Journal of Construction Engineering and Management* 137, 843–852.
- Christian, J. and Hachey, D., 1995. Effects of delay times on production rates in construction. *Journal of Construction Engineering and Management*, 121(1), 20-26.
- Hajjar, D., & AbouRizk, S., 1999. Symphony: an environment for building special purpose construction simulation tools. In *Proceedings of the 31st conference on Winter simulation: Simulation--a bridge to the future-Volume 2* (pp. 998-1006). ACM.
- Han, S., Lee, S., Fard, M. G. and Peña-Mora, F., 2007. Modeling and representation of non-value adding activities due to errors and changes in design and construction projects. *Proceedings of the 39th Conference on Winter Simulation*, 2082-2089. December 9-12, 2007, Washington, DC, USA.
- Hanna, A., Tadt, E. and Whited, G., 2012. Request for information: Benchmarks and metrics for major highway projects. *Journal of Construction Engineering and Management*, 138(12), 1347–1352.
- Hewage, K. N., Gannoruwa, A. and Ruwanpura, J. Y., 2011. Current status of factors leading to team performance of on-site construction professionals in Alberta building construction projects. *Canadian Journal of Civil Engineering*, 38(6), 679-689.
- Klingener, J. F., 1996. Programming combined discrete-continuous simulation models for performance. In *Proceedings of the 28th conference on Winter simulation* (pp. 833-839). IEEE Computer Society.
- Mohamed, Y., Borrego, D., Francisco, L., Al-Hussein, M., AbouRizk, S. and Hermann, U., 2007. Simulation-based scheduling of module assembly yards: Case study. *Engineering, Construction and Architectural Management*, 14(3), 293-311.
- Nasrallah, W. F. and Bou-Matar, R., 2008. Exponential, gamma, and power law distributions in information flow on a construction site. *Journal of Construction Engineering and Management*, 134(6), 442-450.
- Puri, V., & Martinez, J. C. (2012). Modeling of simultaneously continuous and stochastic construction activities for simulation. *Journal of Construction Engineering and Management*, 139(8), 1037-1045.
- Song, L., Mohamed, Y. and AbouRizk, S. M., 2009. Early contractor involvement in design and its impact on construction schedule performance. *Journal of Management in Engineering*, 25(1), 12–20.
- Taghaddos, H., AbouRizk, S. M., Mohamed, Y. and Hermann, U., 2009. Integrated simulation-based scheduling for module assembly yard. *Proceedings of the 2009 ASCE Construction Research Congress* 1270-1279. April 5-7, 2009, Seattle, WA, USA.
- Taghaddos, H., Hermann, U., AbouRizk, S. and Mohamed, Y., 2014. Simulation-based multiagent approach for scheduling modular construction. *Journal of Computing in Civil Engineering*, 28(2), 263–274.

ENVIRONMENTAL ASSESSMENT TOOL TO ANALYZE THE PRESENCE OF CRITICAL AND VALUABLE RAW MATERIALS

Patricia Gómez^(a), Daniel Elduque^(b), Carmelo Pina^(c), Judith Sarasa^(d), Isabel Clavería^(e), Carlos Javierre^(f)

^{(a),(c)} BSH Electrodomésticos España, S.A., Avda de la Industria, 49, Zaragoza (Spain)

^{(b),(e),(f)} i+, University of Zaragoza, C/ María de Luna 3, Zaragoza (Spain)

^(d) Department of Chemical Engineering and Environmental Technologies, University of Zaragoza, C/ María de Luna 3, Zaragoza (Spain)

^(a)patricia.gomez@bshg.com, ^(b)daniel.elduque@gmail.com, ^(c)carmelo.pina@bshg.com

^(d)jsarasa@unizar.es, ^(e)isabel.claveria@unizar.es, ^(f)carlos.javierre@unizar.es

ABSTRACT

The aim of this paper is to show the Software "Sustainable Electronics", developed in the University of Zaragoza as an environmental impact assessment tool, specially developed to design components taking into account the presence of critical and valuable raw materials consumption; simulating environmental impact and measuring the overall raw material consumption. It considers raw material obtaining, manufacturing processes, transports and end of life. This Software allow us to easily update and use the datasets provided by Life Cycle Inventory databases, like for example EcoInvent, developed by the Swiss Center for Life Cycle Inventories. The methodology has been tested through the software in an electronic board of a Touch Control of an induction hob, obtaining that there is a high consumption of materials such as copper, tin or aluminum.

Keywords: life cycle assessment, environmental impact simulation, methodology, critical materials.

1. INTRODUCTION

Companies and the society want to reduce the environmental impact of products and services. The concept of ecodesign began in the 1990s in order to produce more sustainable products.

In order to reducing the environmental impact of a product, methodologies and techniques such as Life Cycle Assessment (LCA) are the key, as they allow researchers to assess and reduce the impacts to the ecological environment. This tool has been used to model a wide range of products: from wind turbines (Martinez et al. 2009) (Martinez et al. 2015), electronic boards (Elduque et al. 2014) or induction hobs (Pina et al. 2015) to compost production (Leiva-Lazaro et al. 2014), food packaging (Fernández et al. 2013) or wine production (Jiménez et al. 2014).

Currently there is a large concern about the materials that affect environmental impact. Due to environmental risk, economic importance and supply risk, materials can be considered as critical (European Commission

2014). The concept of Critical Material emerged firstly in 1939 by the US Administration.

In 2010, the restriction on the exportation of neodymium in China caused a supply chain crisis, as result, prices increased by an order of magnitude (Sprecher et al. 2015)

Although nowadays, the methodology used to determine the criticality of a material is based in the combination of three main indicators (Chapman et al. 2013) (Binnemans et al. 2013) (Graedel & Nuss 2014):

- Economic vulnerability: the end of life recycling has to be taken into account and also the economic benefit that these raw materials have at the sectors in which them arise.
- Supply risk: this value arises from a combination of the stability in the production of the material in a specific country, the substitutability of the material and end-of-life recycling rates of the studied material.
- Ecological risk: this value is estimated taking into account similar criteria than supply risk, raw material country concentration, the ability to be substituted and the recyclability of the material

When calculating the criticality of a material, it is necessary to take into account that an increase in the demand of some materials, in a specific moment, involves a decrease in the demand of others, due to technological change; creating changes in risk indicators of these materials (European Commission 2010) (Achzet & Helbig 2013).

As Peck (Peck et al. 2015) points out, critical materials are "invisible", as they are normally alloyed with other materials. For this reason researchers are using LCA and Life Cycle Sustainability Assessment to systematically compile inventories of the consumption of resources (Mancini et al. 2015) (Sonnemann. et al. 2015). Environmental impact indicators for criticality are still currently being developed, as authors are developing several perspectives (Dewulf et al. 2015) (Glöser et al. 2015) (Rorbech et al. 2014) (Adibi et al. 2014)

The consumption of critical materials have been studied for products such as solar photovoltaics (Goe and

Gaustad 2014), bulbs (Lim et al. 2013) or iron alloys (Nuss et al. 2014).

Several authors have studied ways to reduce the overall consumption of critical materials, focusing specially on recycling (Rademaker et al. 2013) (Dhammika et al. 2014) (Eckelman et al. 2014), recovery (Gutierrez-Gutierrez et al. 2015) (Funari et al. 2014) (Hennebel et al. 2015) and also on reducing the consumption of raw critical materials in new products, such as permanent magnets (Mcguinness et al. 2015).

European Union laws have focused on reducing the environmental impact, by means of ecodesign (EuP 2005/32/CE) (European Parliament, 2005) (ErP 2009/125/CE) (European Parliament, 2009), chemical control and restriction of hazardous substances (REACH 1907/2006) (European Parliament, 2006) (RoHS 2002/95/CE) (European Parliament, 2003).

Applying a suitable ecodesign methodology is very interesting from an environmental point of view in electrical and electronic industry, improving all phases of electrical and electronic life cycle and analyzing the influence of material composition on the environmental impact (Gómez et al. 2015).

The aim of this paper is to show the Software "Sustainable Electronics", developed in the University of Zaragoza as an environmental impact assessment tool, specially designed to simulate environmental impact and to measure the overall raw material consumption, reducing the consumption of critical and valuable materials.

2. "SUSTAINABLE ELECTRONICS" ENVIRONMENTAL ASSESSMENT METHODOLOGY

Nowadays, most LCA models are carried out with professional databases such as EcoInvent one of the most used Life Cycle Inventory databases, developed by Swiss Centre for Life Cycle Inventories. However, these databases provide generic data that is not always adequate for specific products. Our methodology is based on a LCA model, which uses customized datasets to simulate the environmental impact and also quantifies the critical materials consumption.

2.1. General approach

The main goal of the methodology, shown in Figure 1, is to calculate the critical material consumption of an electrical or electronic component, simulating the environmental impact. For this reason, it is necessary to know all material compositions of all the parts in a component or product.

Once analyzed and compared the critical material composition of the components and also, of the products, the user decides the design of the component or the product depending on the life cycle of the component.

This methodology allows the user to compare different designs of the same component depending on the quantity of critical materials in the composition of the component. The user could choose the component with

less critical or valuable raw material and also with less environmental impact.

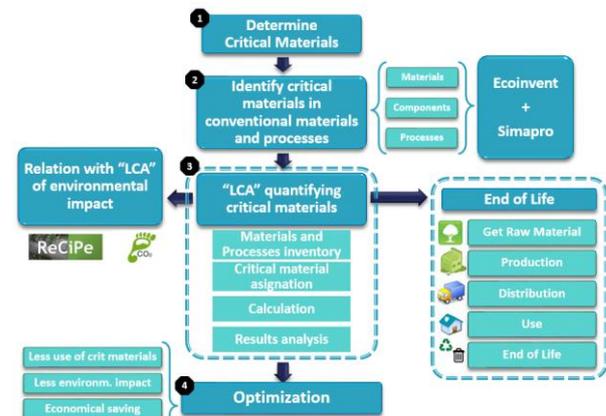


Figure 1: Methodology Diagram

2.1.1. Life Cycle stages

The software carries out environmental impact simulations by means of a LCA model that takes into account all the critical and valuable materials consumed in the life cycle. All the life cycle phases (Figure 2) of the component have to be taken into account, from getting raw materials to the end of life of the component. There will be processes that provide more critical materials consumption than others, affecting also the environmental impact simulation results.



Figure 2: Life Cycle Stages

2.2. Environmental impact

The calculation methodology consist on an improved LCA adapted to critical materials, where the total amount of critical materials is obtained from the life cycle using the critical materials in raw materials of the components, critical and strategic materials in production and in distribution processes. Furthermore, the end of life phase is considered in the calculations of critical materials (Figure 3).

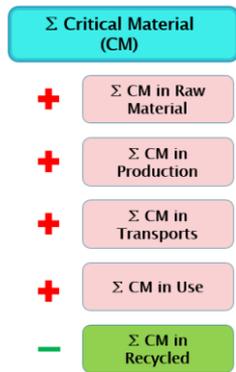


Figure 3: Critical Material Calculation

Also, the methodology considers different end of life scenarios to calculate the amount of critical materials and the environmental impact.

2.3. Results

The methodology can carry out the LCA model calculations and also take into account all the critical and valuable materials consumed in the life cycle.

Results should be clear and concise, in order to the user understand them. So that, the results would be:

- List of critical and strategic materials.
- Environmental impact simulation results.
- Summary of critical and strategic materials percentages.

In order to obtain proper results with this methodology, it is necessary to build a customized database structure that helps calculations. This methodology will be applied to the electrical and electronic field, so these are the components that should be considered for the particularization of the methodology in that field: materials, boards, components, connections and processes.

3. “SUSTAINABLE ELECTRONICS” SOFTWARE

This methodology has been implemented by means of a software tool, named “Sustainable Electronics”. This tool has the aim of calculate the critical materials content and simulate the environmental impact of electrical and electronics devices.

The structure of this software is divided in three blocks, which are: New Project/ Load Project, Databases and also Results (Figure 4), where the user could evaluate quantity of critical materials and environmental impact.

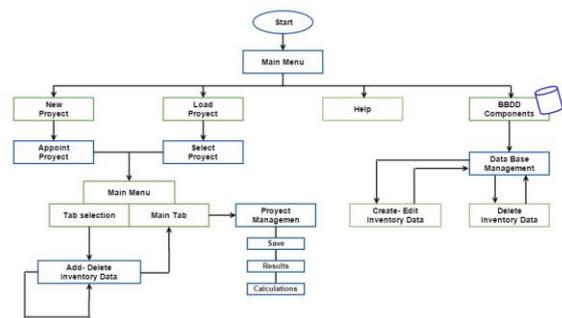


Figure 4: Software Structure

This structure of the software is represented in the main screen of the program (Figure 5).

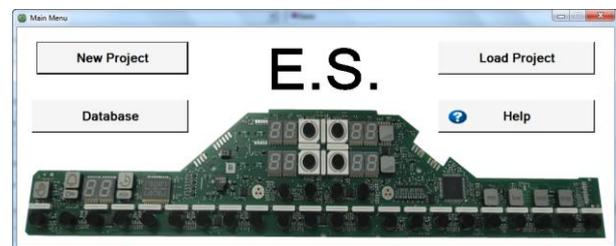


Figure 5: Main Screen of the Software

3.1. Database (LCA model inputs)

Selecting “Database” button from the main screen, the software shows the complete database (Figure 6). Navigating in this screen the user can filter and also manage the database, creating, editing and deleting inventory data.

Nombre	Tipo	Cod_Proj	Piso	Comentarios	E11	E12	E13	E14	E15
Aluminio	Básico	71	1.000000	Sheet, electrolyzed, hot r	0	0	0	0	0
Aluminio	Básico	1	1.000000	Aluminum, wrought al	0	0	0	0	0
Antimonio	Básico	2	1.000000	Antimony (GLC) mark	0	0	0	0	0
Aréne	Básico	70	1.000000	Silica sand (GLC) me	0	0	0	0	0
Aréne	Básico	42	1.000000	Silica sand (GLC) m	0	0	0	0	0
Bario	Básico	3	1.000000	Barite (GLC) market f	0	0	0	0	0
Berilio	Básico	4	1.000000	Copper (GLC) market f	0	0	0	0	0
Borato	Básico	5	1.000000	Boric oxide (GLC) me	0	0	0	0	0
Caliza y ca	Básico	6	1.000000	Lime, hydrated, packe	0	0	0	0	0
Carbon bla	Básico	66	1.000000	Carbon black (GLC) m	0	0	0	0	0
Cellulosa	Básico	67	1.000000	Kraft paper, unbleache	0	0	0	0	0
Cobalto	Básico	7	1.000000	Cobalt (GLC) market f	0	0	0	0	0
Cobre	Básico	8	1.000000	Copper (GLC) market f	0	0	0	0	0
Cristal	Básico	64	1.000000	Flat glass, uncoated (0	0	0	0	0
Cristal	Básico	47	1.000000	Vacio	0	0	0	0	0
Cromo	Básico	9	1.000000	Chromium (GLC) mark	0	0	0	0	0
Electrodo	Básico	44	1.000000	(null)	0	0	0	0	0
Epoxy resi	Básico	94	1.000000	Epoxy resin, liquid (L	0	0	0	0	0
Estallo	Básico	14	1.000000	Ti (GLC) market for	0	0	0	0	0
Etileno g	Básico	61	1.000000	Ethylene glycol (GLC)	0	0	0	0	0
Fenole	Básico	15	1.000000	Phenol (GLC) market	0	0	0	0	0
Ferita	Básico	41	1.000000	Ferite (GLC) market f	0	0	0	0	0
Fierro lita	Básico	59	1.000000	Phosphorus, white, liq	0	0	0	0	0
Fluorita	Básico	16	1.000000	Phosphor, 99% purity	0	0	0	0	0
Gaio	Básico	17	1.000000	Gallium, semiconductor	0	0	0	0	0
Germanio	Básico	18	1.000000	Gallium, semiconductor	0	0	0	0	0
Grafito	Básico	53	1.000000	Graphite, battery grade	0	0	0	0	0
Grafito	Básico	19	1.000000	Graphite, battery grade	0	0	0	0	0

Figure 6: Software Database

After customizing the database, the software allows to select the required components, in order to create the inventory (Figure 7). For an electronic board, the researcher can select between components, boards, production processes and also connections and other elements thanks to the previous work and research that have been performed in order to create the database.

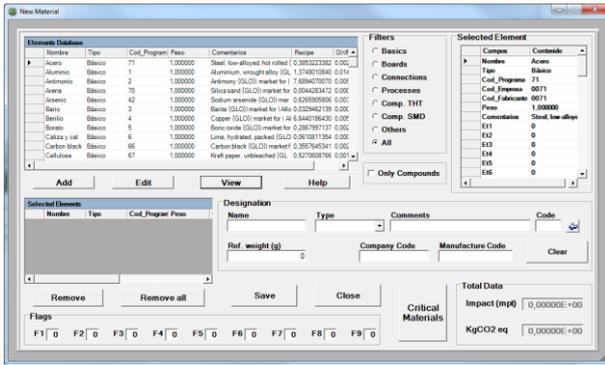


Figure 7: Database Personalization

3.2. Software development

“Sustainable Electronics” software allows researchers and engineers to analyzing different design alternatives in order to reduce the impact to the ecological environment and diminish the use of critical raw materials, showing the overall quantity of critical and valuable raw materials in each component, process, board or connection.

Users can take design decisions taking into account critical materials and according to the environmental impact.

Figure 8 shows “New Project” and “Load Project” options, where there are several tabs such as Project, components, Processes, Boards and Connections and others, the user can navigate through them creating its own project.



Figure 8: Project Screen

All of these screens are prepared for add easily data from the inventory to the project. Once data is added, in Project screen the results could be calculated and viewed by users. In panel “Total data” appears environmental impact and also, pressing “Critical Materials” bottom, critical materials window is shown (Figure 9), where the user can see the quantities of critical and strategic materials.

These screen shows 43 critical and strategic materials, where 22, marked in bold letters, are considered by the latest EU report as critical (Commission 2014).

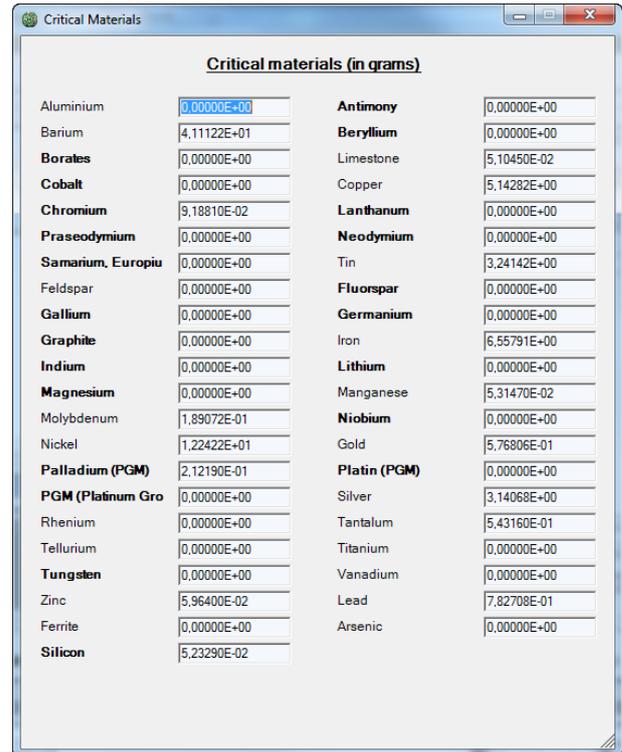


Figure 9: Critical Materials Screen

4. SOFTWARE APPLICATION

The main goal of the software application is to show the methodology and the software, and at the same time, verify the performance of the tool.

4.1. Touch Control

As previously mention, this implementation will be carried out in the field of electronics, specifically the tool is used to calculate critical materials content and simulate the environmental impact in a Touch Control device, shown in Figure 10.



Figure 10: Touch Control

It is an electronic control board, called “Touch Control”, used in induction hobs.

4.1.1. Touch Control Inventory

The inventory for this application consists on data related to electronics, such as capacitors, resistors, diodes, transistors and so on, boards and processes as welding SMD technology.

All of the inventory used in this software application, shown in table 1, has been obtained from the manufacturer of the touch control and desoldering the

electronic components. Also manufacturer’s datasheets of the components have been used in order to consider the exact composition of the components of the studied case, the touch control of an induction hob, following the same methodology shown in (Gómez, et al., 2015).

TOUCH CONTROL INDUCTION HOB			
Name	Material	Units	Weight per unit (g)
Small plastic parts	Nylon 6	7	0,199
Foam cylinders	Polyurethane, flexible foam	13	0,311
Ceramic capacitors SMD	Capacitor, for surface-mounting	52	0,0053
SMD0603 resistor	Resistor, surface-mounted	79	0,0019
Diode BAV 99	Diode, glass-, for surface-mounting	6	0,009
Tantalum capacitor	Capacitor, tantalum	1	0,237
Logical IC	Integrated circuit, logic type	5	0,102
Memory IC	Integrated circuit, memory type	2	0,057
Resonator	Resonator CPM SMD	1	0,013
7-segmentos displays	7 Segment display	6	0,75
LEDs	Light emitting diode	9	0,0015
PCB	Printed wiring board	1	37,995
Welding SMD Technology	Mounting, surface mount technology	1	0,9355

Table 1: Summarized Touch Control Inventory

Once the inventory is completed, elements that are not in the database of the software should be introduced in the tool from “New/Edit Database” option (Figure 11). Weight of each component has to be introduced in the software every time that user wants to introduce a new component in database

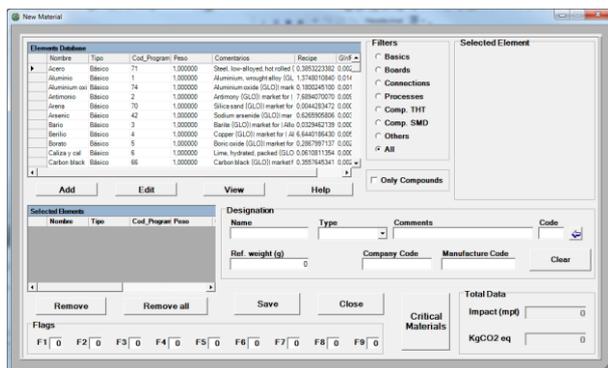


Figure 11: Add or Create Data Screen

After all the inventory of the project has been created, the next step is create the application case. For that, it is necessary to access to “New Project” (Figure 12), where appears the main menu and it is compulsory named the project.

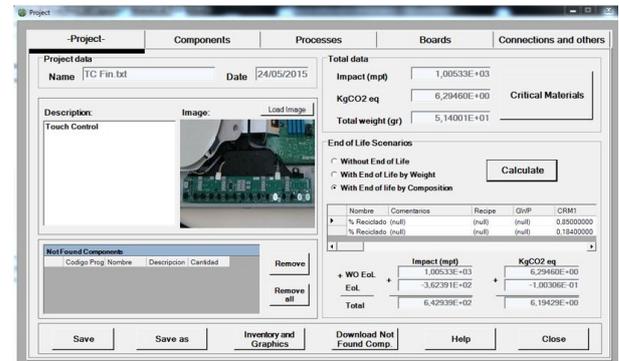


Figure 12: Project Screen

Browsing in tabs Project, Components, Processes, Boards, Connections and Others, data from the inventory should be introduced. Depending on the type of data it will be added in ones or in others (Figure 13).

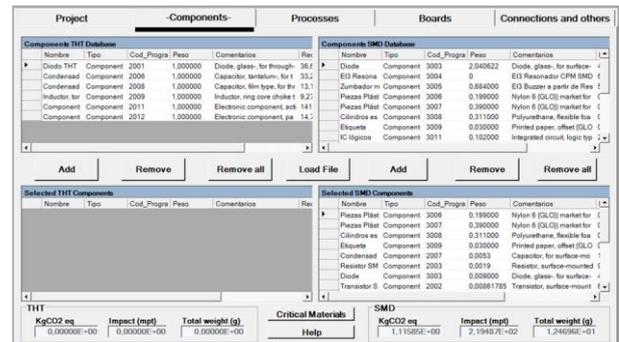


Figure 13: Components Screen

4.1.2. Touch Control Case Results

Afterwards, once the project is completed, “Project” screen gives tool’s options for calculate both critical materials and environmental impact.

“Sustainable Electronics” enables apply several end of life scenarios, such as end of life by composition or by weight. The results simulated by the software show an environmental impact measurement in Recipe of 642.9 mPt and of 6.19 Kg eq. CO₂ in Carbon Footprint.

The PWB creates most of the environmental impact, followed by the SMD components. Although these components have a low overall weight, they generate a significant amount of impact and critical material consumption.

Table 2 shows the consumption of critical and strategic materials in the studied Touch Control. The quantity of copper is the highest, mostly due to its use in the Printing Wiring Board, followed by Tin, used in components soldering, and Aluminium, which is used mainly in the buzzer.

Material	Consumption (g)
Copper	34,93585
Tin	0,981889
Aluminium	0,526671
Barium	0,218226
Tantalum	0,128728
Silver	0,089304
Nickel	0,086837
Iron	0,065517
Lead	0,025029
Manganese	0,012595
Molybdenum	0,01047
Gold	0,009901
Zinc	0,005100
Feldspar	0,004296
Limestone	0,001060
Silicon	0,000934
Palladium (PGM)	0,000666
Chromium	0,000192

Table 2: Overall Material Consumption

Furthermore, there is an option in the software that allows to stand out materials with a high consumption, either global consumption or in percentage. It can be really useful for future regulations, also could be implemented to adapt the software to the legislation.

Table 3 shows critical and strategic material consumption of SMD components. These are essential because, although SMD components have slight weight, the total consumption of critical and strategic material and the environmental impact is really important.

The value of Copper consumption is the highest in SMD components, followed by Aluminium, Barium and in fourth place Tantalum.

Material	Consumption (g)
Copper	1,427462
Aluminium	0,526671
Barium	0,218226
Tantalum	0,128728
Nickel	0,077808
Iron	0,065517
Tin	0,055989
Silver	0,046820
Lead	0,025029
Manganese	0,012595
Molybdenum	0,010472
Gold	0,008097
Zinc	0,005100
Feldspar	0,004296
Limestone	0,001060
Silicon	0,000934
Palladium (PGM)	0,000666
Chromium	0,000192

Table 3: SMD Components Material Consumption

After analyzing the environmental impact and the overall critical and strategic material consumption, several conclusions can be reached. For example, there are several materials with economic importance, such as gold and silver. Gold, with a consumption of 0,0099 grams, was used mostly in integrated circuits and transistors whereas Silver, with 0,0089 grams, was used in soldering processes. On the other hand, there are also materials that present supply risk, such as Palladium and Chromium, which are used in SMD resistors.

5. CONCLUSIONS

The software SE (Sustainable Electronics), developed in Visual Basic .NET, allows the user to quantify the critical and strategic materials associated with the design of a component, also the user can simulate the environmental impact. The user interface makes easy for the user to compare between different components design, making material selection easier.

Although it has been developed for electrical and electronic components, it could be adapted to other sectors easily.

The simulation of the environmental impact of an induction hob delivers an environmental impact result of 642.9 mPt in Recipe and 6.19 Kg eq. CO₂ in Carbon Footprint. Furthermore, the consumption of Copper, Tin and Aluminium supposes a significant amount of environmental impact and material consumption.

This analysis can be used to generate new ecodesign proposals, changing critical components by others with less critical material consumption. For example, the substitution of SMD transistors for components without gold content, such as the ones offered by several electronic components suppliers, would decrease the environmental impact, and also reduce cost and supply risk.

All these results can be achieved thanks to the modeling and simulation of the environmental impact carried out in this study. This approach will allow companies to reduce supply risk, environmental impact and costs.

ACKNOWLEDGMENTS

We would like to thank Ignacio Alecha for his valuable help.

REFERENCES

- Achzet B., Helbig C., 2013. How to evaluate raw material supply risks – an overview. *Resources Policy*, 38(4), pp. 435-447.
- Adibi N., Lafhaj Z., Gemechu E.D., Sonnemann G., Payet J., 2014. Introducing a multi-criteria indicator to better evaluate impacts of rare earth materials production and consumption in life cycle assessment. *Journal of Rare Earths*, 32(3), pp.288-292
- Binnemans K., Jones P.T., Blanpain B., Van Gervenc T., Yang Y., Walton A. & Buchert M., 2013. Recycling of rare earths: a critical review. *Journal of Cleaner Production*, Volumen 51, pp. 1-22
- Chapman A., Arendorf J., Castella T., Thompson P. & Willis P., 2013. Study on Critical Raw Materials at EU Level, Buckinghamshire: OAKDENE HOLLINS
- Dewulf J., Benini L., Mancini L., Sala S. Blengini G.A., Ardenete F., Recchioni M., Maes J., Pant R., Pennington D., 2015. *Environmental Science & Technology*, 49, pp. 5310-5317
- Dhammika Bandara H.M., Darcy J.W., Apelian D., Emmert M.H., 2014. Value Analysis of Neodymium Content in Shredder Feed: Toward Enabling the Feasibility of Rare Earth Magnet Recycling. *Environmental Science & Technology*, 48, pp. 6553-6560
- Eckelman M.J., Ciacci L., Kavlak G., Nuss P., Reck B.K., Graedel T.E., 2014. Life cycle carbon benefits of aerospace alloy recycling. *Journal of Cleaner Production*, 80, pp. 38-45
- Elduque D., Javierre C., Pina C., Martínez E. & Jiménez E., 2014. Life cycle assessment of a domestic induction hob: electronic boards. *Journal of Cleaner Production*, 76, pp. 74-84
- European Commission, 2010. Report of the Ad-hoc Working Group on defining critical raw materials. Enterprise and Industry Directorate General website.
- European Commission, 2014. Report on Critical Raw Materials for the EU. Critical Raw Materials Profiles, European Union: DG Enterprise and Industry.
- European Parliament, 2003. Directive 2002/95/EC of the European Parliament and of the Council of 27 January 2003 on the restriction of the use of certain hazardous substances in electrical and electronic equipment (RoHS). EU Publications Office, Luxembourg. Official Journal of the European Union.
- European Parliament, 2005. Directive 2005/32/EC of the European Parliament and of the Council of 6 July 2005 Establishing a Framework for the Setting of Ecodesign Requirements for Energy-Using Products, EU Publications Office, Luxembourg: Official Journal of the European Union.
- European Parliament, 2006. Regulation (Ec) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency. Official Journal of the European Union. EU Publications Office, Luxembourg.
- European Parliament, 2009. Directive 2009/125/EC of the European Parliament and of the Council of 21 October 2009 Establishing a Framework for the Setting of Ecodesign Requirements for Energy-Related Products, EU Publications Office, Luxembourg: Official Journal of the European Union.
- 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), pp. 176-183.
- Funari V., Braga R., Nadeem S., Bokhari H., Dinelli E., Meisel T., 2014. Solid residues from Italian municipal solid waste incinerators: A source for “critical” raw materials. *Waste Management* (article in press)
- Glöser S., Espinoza L.T., Gandenberger C., Faulstich M., 2015. Raw material criticality in the context of classical risk assessment. *Resources Policy*, 44, pp. 35-46
- Goe M and Gaustad G., 2014. Identifying critical materials for photovoltaics in the US: A multi-metric approach. *Applied Energy* 123, pp. 387-396
- Gómez P., Elduque D., Sarasa J., Pina C., Javierre C., 2015. Influence of the material composition on the environmental impact of surface-mount device (SMD) transistors. *Journal of Cleaner Production*.
- Graedel T.E., Nuss P., 2014. Employing Considerations of Criticality in Product Design. *JOM*, 66(11), pp. 2360-2366.
- Gutierrez-Gutierrez S., Coulon F., Jiang Y., Wagland S., 2015. Rare earth elements and critical metal content of extracted landfilled material and potential recovery opportunities. *Waste Management*, 42, pp.128-136
- Hennebel T., Boon N., Maes S., Lenz M., 2015. Biotechnologies for critical raw material recovery from primary and secondary sources: R&D priorities and future perspectives. *New Biotechnology*, 32 (1), pp. 121-127
- Jiménez E., Martínez E., Blanco J., Pérez M., Graciano C., 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.
- Leiva-Lázaro F.J., Blanco J., Martínez, E., Jiménez E., 2014. Production of compost for mushroom cultivation: A life cycle assessment study 26th European Modeling and Simulation Symposium, EMSS 2014; Bordeaux; France; pp 620-625.
- Lim S.R., Kang D., Ogunseitan O.A., Schoenung J.M., 2013. Potential Environmental Impacts from the

- Metals in Incandescent, Compact Fluorescent Lamp (CFL), and Light-Emitting Diode (LED) Bulbs. *Environmental Science & Technology*, 47, pp. 1040-1047
- Mancini L., Sala S., Recchioni M., Benini L., Goralczyk M., Pennington D., 2015. Potential of life cycle assessment for supporting the management of critical raw materials. *International Journal of Life Cycle Assessment* (20), pp.100-116
- Martínez, E., Blanco J., Jiménez, E., Saenz-Díez J.C., Sanz F., 2015. Comparative evaluation of life cycle impact assessment software tools through a wind turbine case study. *Renewable Energy*, 74, pp- 237-246.
- Martínez E., Sanz F., Pellegrini S., Jiménez E., Blanco J., 2009. Life cycle assessment of a multi-megawatt wind turbine. *Renewable Energy*, 34(3), pp. 667-673.
- Mcguinness P., Akdogan O., Asali A., Bance S., Bittner F., Coey J. M. D., Dempsey N. M, Fidler J., Givord D Gutfleisch., O., Katter M., Le Roy D., Sanvito S., Schrefl T., Schultz L., Schwöbl C., Soderžnik M., Šturm S., Tozman P., Üstüner K., Venkatesan M., Woodcock T. G., Žagar K., Kobe S., 2015. Replacement and Original Magnet Engineering Options (ROMEOS): A European Seventh Framework Project to Develop Advanced Permanent Magnets Without, or with Reduced Use of, Critical Raw Materials. *The Journal of The Minerals, Metals & Materials Society (TMS)*, 67 (6), pp. 1306-1317
- Nuss P., Harper E.M., Nassar N.T., Reck B.K., Graedel T.E., 2014. Criticality of Iron and Its Principal Alloying Elements. *Environmental Science & Technology*, 48, pp. 4171-4177
- Pina C., Elduque D., Javierre C., Martínez E., Jiménez E., 2015. Influence of mechanical design on the evolution of the environmental impact of an induction hob. *International Journal of Life Cycle Assessment*, 20(7), pp. 937-946.
- Peck D., Kandachar P., Tempelman E., 2015. Critical materials from a product design perspective. *Materials and Design*, 65, pp.147-153.
- Rademaker J.H., Kleijn R., Yang Y., 2013. Recycling as a Strategy against Rare Earth Element Criticality: A Systemic Evaluation of the Potential Yield of NdFeB Magnet Recycling. *Environmental Science & Technology*, 47, pp. 10129-10136
- Rorbech J., Vadenbo C., Hellweg S., Astrup T.F., 2014. Impact Assessment of Abiotic Resources in LCA: Quantitative Comparison of Selected Characterization Models. *Environmental Science & Technology*, 48, pp. 11072-11081
- Sonnemann G., Gemechu E.D., Adibi N., Bruille V.D., Bulle C., 2015. From a critical review to a conceptual framework for integrating the criticality of resources into Life Cycle Sustainability Assessment. *Journal of Cleaner Production*, 94, pp 20-34
- Sprecher B., Daigo I., Murakami S., Kleijn R., Vos M., Kramer G.J., 2015. Framework for Resilience in Material Supply Chains, With a Case Study from the 2010 Rare Earth Crisis. *Environmental Science & Technology*, 49, pp. 6740-6750

SUPERVISORY FCM SIMULATION IN TUNNEL CONSTRUCTION

Limao Zhang^(a), A. J. Antony Chettupuzha^(b), Seungjun Ahn^(c), Stephen Hague^(d), Chrysostomos D. Stylios^(e),
Simaan M. AbouRizk^(f)

^{(a),(b),(c),(d),(f)} Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada

^(e) Department of Computer Engineering, Technological Educational Institute of Epirus, Arta, Greece

^(a) limao@ualberta.ca, ^(b) chettupu@ualberta.ca, ^(c) seungjun@ualberta.ca,
^(d) steve.hague@ualberta.ca, ^(e) cdstylios@gmail.com, ^(f) abourizk@ualberta.ca

ABSTRACT

This paper presents a Fuzzy Cognitive Map (FCM) approach for controlling the Tunnel Boring Machine (TBM) operation in tunnel construction. The fuzzy logic based approach is used to capture the construction experience and knowledge from domain experts in creating FCMs. A supervisory FCM model is developed to represent the cause-effect relationships among the variables involved in the operation of a TBM. The developed model can be used to test what-if scenarios and to perform a real-time control of operational variables in response to the changing environment. A tunneling project in China is used as an example case study to demonstrate the applicability of the developed approach. This example case study shows that the TBM performance is influenced by a variety of geological and operational variables, and that FCMs can provide guidance on adjusting operational parameters in tunnel construction. The proposed approach can be used as a decision support tool for ensuring satisfactory performance of TBMs, and thus increases efficiency of tunnel construction projects.

Keywords: Simulation; Fuzzy Cognitive Map (FCM); TBM Performance; Supervisory System

1. INTRODUCTION

In recent years, the construction of subway systems and underground utilities has dramatically increased due to the increasing population and lack of surface space. The tunnel boring machine (TBM) technique has found widespread application in tunnel construction, and is used for excavating tunnels in nearly all types of rock masses and geological conditions. However, occurrences of shield cutter-head failure, and even catastrophic failures of drive motors are frequently observed in real projects. In general, there is concern that a decrease in performance of the TBM due to a component failure leads to unacceptable equipment damage, and can significantly adversely impact the overall productivity of tunnel projects.

Often, a failure in the TBM occurs due to the mismatch between the geological variables (e.g., soil density) and the operational variables (e.g., gross thrust of the TBM).

Therefore, timely and appropriate responses to the changing underground conditions are very important in TBM operations. However, precisely predicting the geological conditions for every location in tunnel construction is impossible. Hence, real time sensing and monitoring of the geological conditions is required in tunnel construction as the TBM advances.

Complex systems, like TBM operations, are usually characterized by high dimensions, and involve many different kinds of variables or subsystems that are strongly interconnected and mutually dependent. A large number of complex processes are not fully understood, but their operation is “tuned” successfully through experience, rather than through the application of pure mathematic principles (Stylios & Groumpos, 2000, 2004). In this situation, effectively and efficiently capturing and utilizing experts’ knowledge can provide a solution to develop and improve complex system models (Tianfield, 2001; Wu & Cai, 2000). Soft computing techniques have been proposed to create a platform to capture and integrate experts’ knowledge in an effective and efficient manner. Fuzzy cognitive map (FCM) approach is one of the soft computing modeling techniques. The implementation of FCM for modeling supervisory control systems is a promising area, since FCMs can integrate all kinds of relevant knowledge from domain experts, and simulate thinking processes of human experts by using a more abstract representation of the cause-effect mechanism, general control knowledge and adaptation heuristics, and enhance the performance of the whole complex system (Stylios & Groumpos, 2004).

The supervisory FCM is an augmented model of the complex system, which represents the relationships among the actual systems and their models. The supervisory FCM monitors and organizes all the subsystems in order to accomplish a task, to help the operator make decisions, to plan strategically and to detect and analyze failures (Stylios & Groumpos, 2000). A supervisory FCM for complex systems provides decision support and is efficient in examining what will happen if a state of the system changes (Stylios & Groumpos, 2004).

This paper investigates the potential of using supervisory FCM as a tool to model and simulate the TBM system, as well as assist operators in ensuring satisfactory performance during the TBM operation. A supervisory FCM model for controlling TBM performance is developed. The developed model provides further understanding on how the system behavior evolves quantitatively in the case of changes in variables, and assists engineers (the operators of the system) in adjusting relevant variables to ensure the satisfactory performance of the TBM operation. Also, the dynamic nature of a complex system over time is considered in the developed supervisory FCM models, in order to accurately illustrate the dynamic features of variables in the real world. A tunnel case in China is used to demonstrate the effectiveness of the proposed approach, as well as its application potential.

2. METHODOLOGY

FCM is a modeling methodology for complex systems, which originated from the combination of fuzzy logic and neural networks. The graphical illustration is a signed fuzzy graph with feedback, consisting of nodes and weighted interconnections. Figure 1 illustrates a simple FCM model. Nodes of the graph stand for concepts that are used to describe mainly behavioral characteristics of the system. Nodes are connected by signed and weighted arcs representing the causal relationships that exist among concepts. This simple illustration provides thoughts and suggestions in reconstructing FCM, in regards to adding or deleting interconnections or concepts. In conclusion, an FCM is a fuzzy-graph structure, which allows systematic causal propagation, in particular, forward and backward chaining. Based on the unique reasoning technique of FCM, a systematic decision approach to supervise the TBM performance in tunnel construction is developed in this research, which consists of the following four steps.

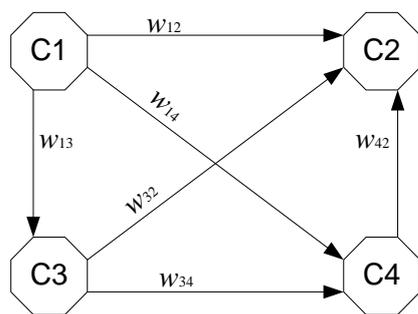


Figure 1: A Simple FCM Model

2.1. Representation of Concepts

FCMs are dynamic systems that have the topology of a directed fuzzy graph (see Figure 1) consisting of nodes and edges and permitting cycles and feedback. Nodes of the graph stand for the concepts that are used to describe the behavior of the system and they are connected by signed and weighted arcs representing the causal relationships that exist among concepts. It must

be mentioned that all the values in the graph are fuzzy, indicating the concept nodes C_i have a fuzzy nature. In the FCM framework, each concept, defined by C_i with an interval of $[-1, +1]$, represents a characteristic of the system (e.g., events, actions, goals, values or trends of the system). Relationships and interdependencies among concepts are represented by the arcs (i.e., interconnections) among concepts in FCM models, as shown in Figure 1.

Concepts in FCMs (C_i, C_j, \dots, C_n) represent the state of a system. At time t , the state of an FCM is defined by the vector A_t of the values of the concepts ($A_t = [C_1, \dots, C_n] \in [-1, +1]$), or in other words, a point in the fuzzy n -dimensional state space. An edge $e_{ij} (\in [-1, +1])$ defines causal flows $C_i \rightarrow C_j$ between the concepts. Based upon the graphical representation, it becomes clear which concept influences other concepts, and the degree of the influence. Therefore, the FCM approach permits thoughts and suggestions to be collected or aggregated in the construction of the graph by adding or deleting an interconnection or a concept.

2.2. Fuzzy Weight Determination

Given an FCM model with N concepts, a weight from the concept C_i to the concept C_j ($1 \leq i, j \leq N, i \neq j$), defined by W_{ij} , is the relative influence from concept C_i that determines the value of concept C_j . These are fuzzy weights that describe the degree of the causal interconnection among concept nodes. To be specific, there are three types of causal interconnections between concepts that represent the degree of influence from one concept to the other:

- $W_{ij} > 0$: means the weight of the arc from C_i to C_j would be positive. An increase in the value of C_i leads to the increase of the value of C_j , and a decrease in the value of C_i leads to the decrease of the value of C_j .
- $W_{ij} < 0$: means the weight of the arc from C_i to C_j would be negative. An increase in the value of C_i leads to the decrease of the value of C_j , and vice versa.
- $W_{ij} = 0$: means that there is no relationship between C_i and C_j .

The causal interrelationships among concepts are declared using a linguistic variable which takes values in the universe $U = [-1, +1]$. It is known that the number of linguistic intervals can largely affect the reliability of the estimation. Specifically, short intervals (or a great number of variables) indicate that the statistic is precisely known, while wide intervals (or a small number of variables) indicate great uncertainty. In order to reach the high precision of the influence estimation among concepts, it is better to conduct a group of short intervals, rather than wide intervals. According to Dawes' (2008) experiment, the use of 5-9 point scales is likely to produce slightly higher mean scores relative to the highest possible attainable score, and besides, a larger number is usually impractical. Thus, in this research, a 9-point linguistic scale is used to provide a basis for an expert to describe in detail the influence of

one concept on another based on his/her knowledge, and then discern between different degrees of influence. Figure 2 illustrates the corresponding membership functions for these 9 linguistic terms: negatively very strong (μ_{nvs}), negatively strong (μ_{ns}), negatively moderate (μ_{nm}), negatively weak (μ_{nw}), zero (μ_z), positively weak (μ_{pw}), positively moderate (μ_{pm}), positively strong (μ_{ps}), positively very strong (μ_{pvs}). In order to exploit the knowledge and experience of experts on the description and modeling of a complex system, it is supposed that there are M experts who evaluate every interconnection and describe it with a fuzzy rule inferring a linguistic weight. The credibility weight for every expert is introduced to increase the objectivity of the FCM developing method, accepting that initially, all the experts are equally credible and have the same credibility weight. Generally, at least $M/3$ of the experts have to fully agree with their suggestions, thus, an average weight of the interconnection is calculated using Equation 1. Otherwise, the credibility weight of an expert should be reduced by $r\%$ every time there is a wrong suggestion for an interconnection. Next, the suggested linguistic weights for an interconnection may be aggregated in accordance with the well-known fuzzy logic method of “SUM,” which can be used to produce an aggregated linguistic weight. Finally, the defuzzification method of center of gravity (COG) is applied, and a numerical weight for the interconnection is calculated (Jang, Sun, & Mizutani, 1997; Nie & Linkens, 1995):

$$W_{ij} = \frac{\sum_{k=1}^M (b_k \times W_{ij}^k)}{M} \quad (1)$$

where, W_{ij} stands for the aggregated fuzzy weight from C_i to C_j ; M stands for the total number of the participating experts; b_k stands for the credibility weight of the k th expert; and W_{ij}^k stands for the fuzzy weight from C_i to C_j based on the judgement of the k th expert.

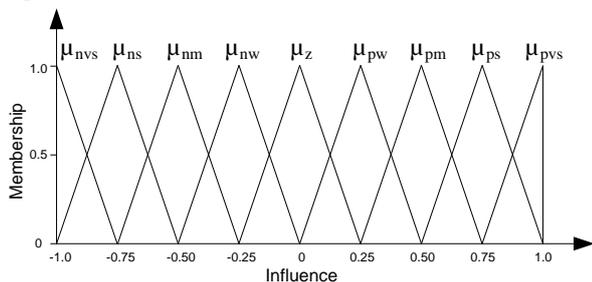


Figure 2: Membership Functions for the 9-point Linguistic Scale Regarding Influence between Concepts.

2.3. Dynamics of FCM

Behind the graphical representation of FCM, there is the mathematical model. The causal interconnections of an FCM model can be expressed by an $n \times n$ matrix W , which contains all of the n^2 rules or pathways in the causal web between n concepts in the FCM. The state of an FCM model can be expressed by a $1 \times n$ vector A .

Then, the dynamics of the FCM model are dictated by these matrices, and can be updated through iteration with other concepts and with its own value. Equation 2 drives this interaction where the strength of the causal relationships is represented by the weight of the summation (Stylios & Georgoulas, 2011; Stylios & Groumpos, 2004). After some iterations, the values of the concepts will evolve until stabilized at a fixed point or a cycle limit.

$$A_i(t+1) = f \left(A_i(t) + \sum_{j=1, j \neq i}^n (W_{ji} \times A_j(t)) \right) \quad (2)$$

In Equation 2, t stands for the interaction index; $A_i(t+1)$ stands for the value of concept C_i at time $t+1$; $A_i(t)$ and $A_j(t)$ stands for the value of concept C_i and C_j at time t , respectively; W_{ji} stands for the interconnection weight from concept C_j to concept C_i ; and $f(\bullet)$ stands for a threshold function.

Generally, there are two kinds of threshold functions used in the FCM framework. One is the unipolar sigmoid function, where $\lambda > 0$ determines the steepness of the continuous function $f(\bullet)$, and squashes the content of the function in the interval $[0,1]$, as shown in Equation 3. Another threshold function can be used to transform the content of the function into the interval $[-1, +1]$, as shown in Equation 4. The selection of the threshold function depends on the method that is used to describe the concepts. In this research, we attempt to assess the impacts of root causes of the reliability of the target event in terms of “positive influence” and “negative influence.” As a result, the threshold function as shown in Equation 4 is selected for this research, and presented later in a case study.

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \quad (3)$$

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (4)$$

2.4. FCM-based Analysis

An FCM represents human knowledge about the dynamic behavior of a complex system. In other words, an FCM is a model about a system’s dynamic behavior in term of concepts and interrelationships among concepts. Once constructed, FCMs can be used to simulate the behavior of a system and perform what-if analysis, due to its effectiveness in representing a complex system’s behaviors, and the ease of use (Taber, 1991).

There is a cause behind every failure or problem, and localization and elimination of its causes should be of the utmost significance, in order to avoid failure occurrence. Predictive analysis aims to forecast the future outcome or effect of an event in light of some evidence being available. It can be used to address the

problem or non-conformance by localizing the true cause and implementing corrective action to prevent recurrence of the problem (Rooney & Heuvel, 2004). Several variables may be involved in the occurrence of a failure; however, what the contributions of those variables are is unknown. In the FCM framework, as shown in Figure 1, at least one concept should represent the target event (decision concept), and some concepts should represent the potential root causes. Causal analysis is able to identify the relationship between the causes and effects of a target event, and further perform predictive analysis. More details regarding the predictive analysis of the TBM performance can be seen later in the tunnel case study.

3. FCM MODELING OF TBM PERFORMANCE

3.1. Concepts Involved in TBM Operation

A TBM is a machine used to excavate tunnels with a circular cross section through a variety of soil and rock strata. TBMs have the advantage of limiting the disturbance to the surrounding ground and producing a smooth tunnel wall, which significantly reduces the cost of lining the tunnel, and makes them suitable to use in heavily urbanized areas. In soft ground, there are two main types of TBMs, earth pressure balance (EPB) and slurry TBM. Compared with the slurry TBM, the EPB TBM is relatively inexpensive and does not need a large amount of space, and thus, is commonly used for a broad range of applications, especially in the construction of urban rail transport systems (Copur et al., 2014).

Generally, TBM performance is measured in terms of *Advance Rate* (C_T), which is the actual excavating and supporting distance divided by the total time, including downtime for TBM maintenance, machine breakdown and tunnel failure (Alber, 1996). The advance rate depends on various types of factors ranging from machine design to geological features. Therefore it may not be possible to capture this complex relationship in an explicit mathematical expression (Zhao, Gong, Zhang, & Zhao, 2007). In accordance with an extensive review of literature, the most important parameters that would be used in TBM performance studies can be divided into two major categories, as follows:

- **Geological variables:** The construction of a tunnel project is very sensitive to geological conditions due to complex tunnel-soil interactions. Some particular geological features may increase the hazardous nature of the TBM advance process. *Soil Cohesion* (C_1), *Friction Angle* (C_2), *Compression Modulus* (C_3) and *Soil Density* (C_4) are four variables frequently used to illustrate the geological features.
- **Operational variables:** In the process of the TBM-driven tunneling excavation, some pressure and speed sensors are installed on the top and middle of the cutter head in advance, and engineers pay close attention to the measurement of those operational variables, in order to maintain the face stability of

the excavation and minimize risk (Ding, Wang, Luo, Yu, & Wu, 2013). Rational and appropriate *Gross Thrust* (C_5), *Cutter Torque* (C_6), *Earth Pressure* (C_7) and *Grouting Speed* (C_8) are considered very important technical variables in tunnel construction practice, which can also be further adjusted to guarantee satisfactory performance in the excavation process.

In total, nine concepts C_i ($i=1, 2, \dots, 8, T$) are created in order to model the complex system of TBM performance. Table 1 illustrates the definitions of the above nine concepts. Due to the imprecise and uncertain nature of the variables in the real world, measurement of the initial values of those concepts remains a problem. Although many models for the prediction of TBM performance have been presented, as mentioned above, a commonly accepted model that can be used in every case regardless of geological conditions has not been reached. One significant reason is that the scope of favorable or unfavorable geological/operational variables varies greatly among different tunnel cases in different areas, leading to challenges in determining the value of a specific concept representing a corresponding geological or operational variable due to uncertainty. In order to solve this problem, a 5-point linguistic scale provides a basis for an expert to describe in detail the favorability of a specific concept in response to the TBM performance based on his/her knowledge. Figure 3 illustrates the corresponding membership functions for these 5 terms: negatively very unfavorable (δ_{nvu}), negatively unfavorable (δ_{nu}), favorable (δ_f), positively unfavorable (δ_{pu}) and positively very unfavorable (δ_{pvu}). Given several experts are involved in the measurement of concepts in a specific project, the credibility weight for every expert is also introduced to increase the objectivity during the integration of expert knowledge from varying backgrounds, as well as to enhance the confidence in using the FCM. In a similar way as shown above in the section of fuzzy weight determination, both the well-known fuzzy logic method of SUM, and the defuzzification method of COG, are then applied to obtain a crisp value of a specific concept. For instance, if the concept C_1 has a value of 0, it means C_1 lies in a very favorable range in response to the TBM performance. Contrary to that, if the concept C_1 has a value of -1 (or 1), it means C_1 lies in a negatively very unfavorable (or negatively unfavorable) range in response to the TBM performance.

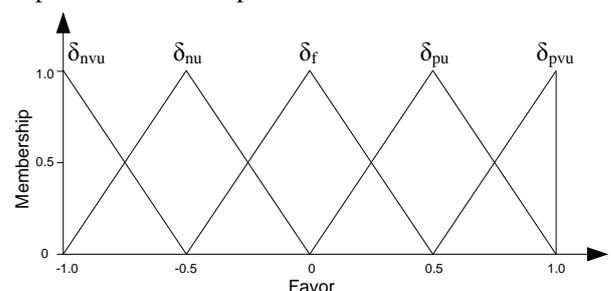


Figure 3: Membership Functions for the 5-point Linguistic Scale Regarding Concept Measurement

3.2. Model Construction

In order to demonstrate the effectiveness and practicality of FCM for supervising the TBM performance, a tunnel project, the Metro Line 7 (ML7) in the Wuhan metro system in China, is used as a case study. A questionnaire was prepared and distributed among ten experts who participated in the construction of ML7, including seven domain experts with at least five years of working experience and three professors in this field. The experts were asked to evaluate the influence/weight between the concept C_i and C_j ($i, j=1, 2, \dots, 8, T; i \neq j$) using the 9-point linguistic terms, as shown in Figure 2. In accordance with the methodology, as mentioned in Section II, the linguistic inputs from the experts were then transformed into crisp values using the fuzzy logic algorithm SUM and the defuzzification method of COG.

Table 1: Definitions of Nine Concepts Involved in TBM Operation

Items	Concepts	Descriptions
C_1	<i>Soil Cohesion</i>	Component of shear strength of a soil
C_2	<i>Friction Angle</i>	Force that resists the relative motion of solid surfaces, fluid layers, and material elements sliding against each other
C_3	<i>Compressive Modulus</i>	Capacity of a material or structure to withstand loads tending to reduce size
C_4	<i>Soil Density</i>	Mass of many particles of the soil divided by the total volume they occupy
C_5	<i>Gross Thrust</i>	Total thrust during the TBM operation
C_6	<i>Cuter Torque</i>	Torque of the cutter disk during the TBM operation
C_7	<i>Earth Pressure</i>	Pressure at the face of the TBM to remain a balanced performance
C_8	<i>Grouting Speed</i>	Speed of the concrete grouting at the shield tail
C_T	<i>Advance Rate</i>	Actual excavating and supporting distance divided by the total time in TBM operation

In order to support the computations, a *Java*-based FCM analysis software tool, *FCM Analyst v1.0*, developed by Margaritis et al. (2002), is used. This software tool has capabilities for creating an FCM, such as creating concept nodes and interconnections between concept nodes, manipulating the concepts/weights matrices, and running simulations to show the behavior of the FCM model. Taking the determination of the influence weight from concept C_1 to C_T as an example, there are five experts who have a judgement of negatively weak (μ_{nw}), two experts have a judgement of negatively moderate (μ_{nm}), and only one expert has a judgement of zero (μ_z). Figure 4 shows an illustration

of how each individual expert's input is entered, and multiple inputs are automatically aggregated and defuzzified in the software tool. The weight of W_{1T} is then determined to be -0.275. With knowledge from all the participating experts integrated in an effective manner, Figure 5 depicts a graphical representation of the FCM model that is used to describe, model and supervise the TBM performance in tunnel construction.

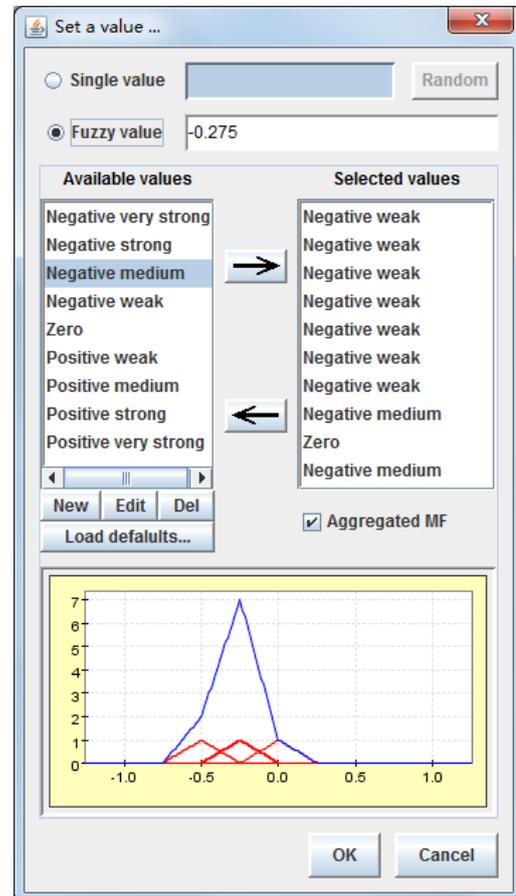


Figure 4: Illustration of Fuzzy Weight Calculation in *FCM Analyst v1.0*

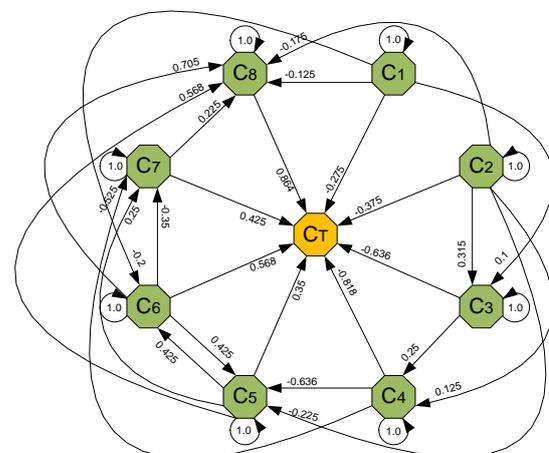


Figure 5: Graphical Representation of the FCM Model for the TBM Performance

4. SAMPLE EXPERIMENTS

Once an FCM is built, many what-if experiments can be conducted for exploring the cause-effect interaction of the TBM performance. In this paper, we present a sample experiment for predictive analysis of the TBM performance in tunnel construction. Predictive analysis aims to forecast the future outcome when evidence of a variable/concept is available. In the FCM framework, the propagation of evidence of a variable allows an update of the outcome of the other variables in the network in light of the newly found evidence. As mentioned above, the values of concepts correspond to real measurements that have been transformed in the interval $[-1, +1]$ using the 5-point linguistic scale (see Figure 3). The concept has a value of 0 (-1 or +1), indicating it lies in a very favorable (negatively very unfavorable or positively very unfavorable) range in response to the TBM performance. Hence, when evidence is set on a specific concept in the FCM model, as shown in Figure 5, the evidence will propagate to the other concepts in the network, and cause the outcome of the affected concept nodes to change. Figure 6 illustrates impacts of variations of different concepts/causes on the TBM performance in the predictive analysis. In other words, this analysis shows the prediction of consequences if a concept (C_1 - C_8) has a non-favorable condition while other concepts remain in the favorable range.

Taking into account the knowledge from domain experts, the developed FCM (see Figure 5) can be used

as a tool to supervise TBM performance. It can be used to examine what will happen if a scenario is running, and what the consequences will be for the whole process if a state of the system changes. In general, the geological concepts C_1 , C_2 , C_3 and C_4 all perform markedly negative corrections with the concept C_T , while the operational concepts C_5 , C_6 , C_7 and C_8 all perform markedly positive corrections with the concept C_T . With respect to the measurement of this kind of correction among different concepts, the concept C_4 displays a strongest negative correction with the concept C_T ; while the concept C_8 displays the strongest positive correction with the concept C_T . That is to say, the concepts C_4 and C_8 are both very sensitive to ensuring satisfactory performance of the TBM operation, and thus, should be given significant attention in tunneling excavation. This is consistent with observed construction practice, indicating the established FCM model is valid to some extent. As a matter of fact, to maintain the face stability of the excavation and minimize risk, engineers need to adjust the operational parameters continuously during the TBM operation so as to fit the surrounding geological conditions in a dynamic manner. The results achieved from the predictive analysis can provide more understanding on how the TBM performance evolves quantitatively in case either the geological or operational variables suffer a change, which can offer insights on the guidance for the reasonable adjustment of operational parameters in tunnel construction.

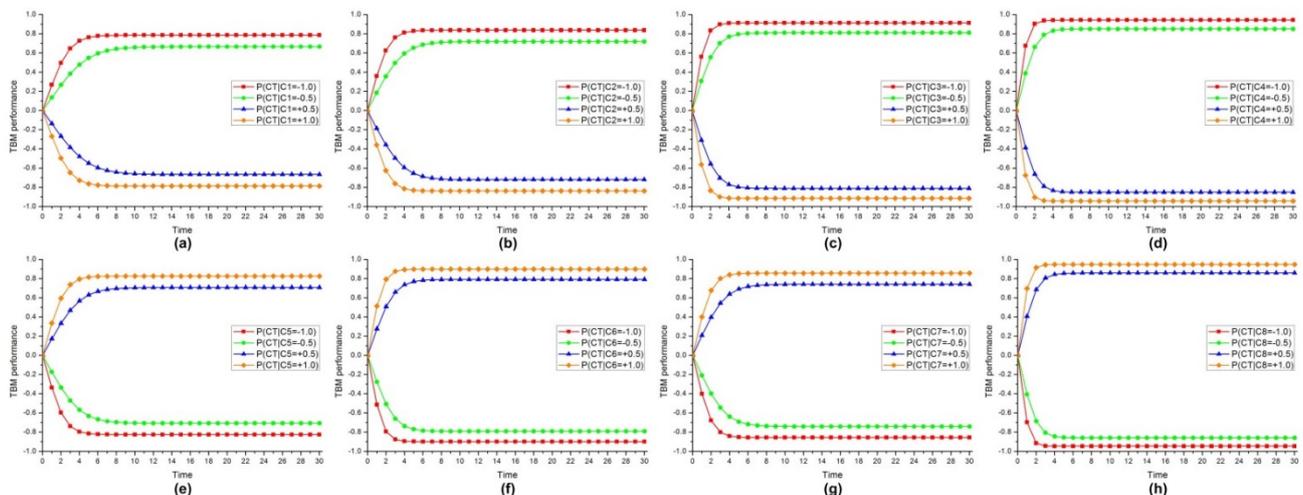


Figure 6: Impacts of variations of different concepts/causes on the TBM performance in the predictive analysis: (a) the concept C_1 ; (b) the concept C_2 ; (c) the concept C_3 ; (d) the concept C_4 ; (e) the concept C_5 ; (f) the concept C_6 ; (g) the concept C_7 ; and (h) the concept C_8

5. DISCUSSION

The FCM model that has been described in this paper takes into consideration expert feedback on the interrelationships between concepts that impact the performance of a TBM. An improved understanding of the relationships between these concepts can extend this theoretical model to practical applications in tunneling construction. One such potential application could be

using the FCM as an intelligent automated supervisor that assesses and reacts to changing conditions. FCMs integrated as an Intelligent Supervisor System have been previously proposed as a method to model a supervisory control system for dynamic environments (Stylios & Groumpos, 2000). If we were to consider a range of optimal TBM advancement rate as an objective, it is possible to design a supervisory FCM based on this model. This would be very useful when the tunneling area under consideration consists of several different

kinds of soil types that are unevenly distributed. Traditionally, engineers would have to conduct rigorous surveys and determine operational parameters for every instance of soil variation and then provide these values integrated with a schedule to the TBM driver. However, if the TBM were equipped with sensors that detect these geological conditions, the FCM could automate this process with greater considerations for the interactions between all the concepts, and provide operational parameters for an optimized TBM advancement rate.

6. CONCLUSIONS AND FUTURE WORK

The operation of TBM systems is very complicated, since various kinds of variables are involved. Balancing the interaction among all the variables as to ensure satisfactory performance of the TBM operation becomes a challenging task. In this research, an FCM-based model is developed to supervise the TBM performance in tunnel construction, with construction experience and knowledge from domain experts taken into account. The developed approach is capable of performing predictive analysis given that an observation of change is available, providing insights into how the system should be adjusted over time. The obtained simulation results for the TBM performance illustrate the feasibility of the proposed approach, as well as its application potential in the modeling and controlling of complex systems in other areas.

The developed FCM-based approach also has some limitations. FCM provides an immediate understanding of causal knowledge in an intuitive way because of its simplicity. It is very helpful in the decision-making processes; however, constructing an FCM causal model by means of knowledge acquisition from experts is time-consuming. In addition, domain experts are generally considered scarce resources, especially in the construction industry, and finding experts in certain domains is challenging. Our subsequent research work would focus on the data mining technique in the knowledge engineering process, as well as the future application of FCM in the construction industry.

REFERENCES

Alber, M. (1996). *Prediction of penetration and utilization for hard rock TBMs*. Paper presented at the Proceedings of the International Conference of Eurock.

Copur, H., Aydin, H., Bilgin, N., Balci, C., Tumac, D., & Dayanc, C. (2014). Predicting performance of EPB TBMs by using a stochastic model implemented into a deterministic model. *Tunnelling and Underground Space Technology*, 42(1), 1-14. doi: <http://dx.doi.org/10.1016/j.tust.2014.01.006>

Dawes, J. (2008). Do Data Characteristics Change According to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research*, 51(1).

Ding, L., Wang, F., Luo, H., Yu, M., & Wu, X. (2013). Feedforward Analysis for Shield-ground System. *Journal of Computing in Civil Engineering*, 27(3), 231-242. doi: 10.1061/(asce)cp.1943-5487.0000219

Jang, J.-S. R., Sun, C.-T., & Mizutani, E. (1997). Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. *Automatic Control, IEEE Transactions on*, 42(10), 1482-1484.

Margaritis, M., Stylios, C., & Groumpos, P. (2002). *Fuzzy Cognitive Map Software*. Paper presented at the SoftCOM 2002: international conference on software, telecommunications and computer networks.

Nie, J., & Linkens, D. (1995). *Fuzzy-neural control: principles, algorithms and applications*: Prentice Hall International (UK) Ltd.

Rooney, J. J., & Heuvel, L. N. V. (2004). Root cause analysis for beginners. *Quality progress*, 37(7), 45-56.

Stylios, C. D., & Georgoulas, G. (2011). *Modeling complex logistics systems using soft computing methodology of fuzzy cognitive maps*. Paper presented at the Automation Science and Engineering (CASE), 2011 IEEE Conference.

Stylios, C. D., & Groumpos, P. P. (2000). Fuzzy cognitive maps in modeling supervisory control systems. *Journal of Intelligent and Fuzzy Systems*, 8(1), 83-98.

Stylios, C. D., & Groumpos, P. P. (2004). Modeling complex systems using fuzzy cognitive maps. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 34(1), 155-162.

Taber, R. (1991). Knowledge processing with fuzzy cognitive maps. *Expert Systems with Applications*, 2(1), 83-87.

Tianfield, H. (2001). Formalized analysis of structural characteristics of large complex systems. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 31(6), 559-572.

Wu, X., & Cai, K. (2000). Knowledge object modeling. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 30(2), 96-107.

Zhao, Z., Gong, Q., Zhang, Y., & Zhao, J. (2007). Prediction model of tunnel boring machine performance by ensemble neural networks. *Geomechanics and Geoengineering: An International Journal*, 2(2), 123-128.

INVESTIGATION OF THERMAL EFFECTS ON ELASTIC CONSTANTS TO SIMULATE CORRECTLY THE THERMAL STRESS IN NANO TRANSISTORS

Abderrazzak El Boukili

Al Akhawayn University in Ifrane, Morocco

a.elboukili@aui.ma

ABSTRACT

Mechanical stress is the main performance booster for the new generation of nano CMOS transistors. Then, accurate models to calculate this stress are requested.

In this paper, we are proposing new and accurate mathematical model to calculate numerically and correctly the thermal induced stress in the new generation of Intel nano PMOSFETs. In this model, we are taking into account, for the first time, the effects of the fabrication temperature on the elastic constants of Silicon and Silicon Germanium materials to calculate the thermal stress. Numerical results from different simulations with different temperatures for thermal induced stress in 14 nm Intel nano PMOS of the year 2014 will be given, discussed, and compared with literature.

Keywords: temperature effects, elastic constants, thermal induced stress , 3D modeling and simulation, Intel nano CMOS transistors

1. INTRODUCTION

For the last decade, mechanical stress is extensively used by many manufacturers as IBM, Intel, AMD, Toshiba, Samsung, and Infineon to increase dramatically the performance of the new generation of nano CMOS transistors (Ghani 2010, Chui 2007, JianLi 2014). It is now admitted by the semiconductor community that the application of mechanical stress to the channel enhances the mobility of both nano NMOS and PMOS transistors by more than 50% (JianLi 2014, Ghani 2010, Chui 2007, Bera 2006, Krivokapic 2003). From physics point of view, this enhancement of the mobility is due fundamentally to the splitting of the conduction and valence energy bands by the mechanical stress (Ghani 2003, JianLi 2014). And, this splitting of the energy bands results in the reduction of band gap energy. When the band gap is reduced, the mobility will be increased. The performance of CMOS transistors have been continuously improved by the traditional geometric scaling of its critical dimensions. The top traditional scaling enablers are: gate oxide thickness scaling, junction scaling, and common collector voltage scaling. The year 2000 marked the

end of this CMOS traditional scaling era that lasts about 40 years. In the year 2000, the CMOS research team at Intel company have faced serious challenges when they have used the traditional geometric scaling to fabricate 90nm CMOS technology node (Ghani 2010). The top challenges for this 90nm CMOS were: the parasitic resistance, the gate leakage, and the mobility degradation due to channel ionized impurity scattering (Ghani 2010). In the beginning of 2000, Intel was the first company to introduce great innovations to overcome the traditional scaling barriers. The first innovation is the application of uniaxial process induced stress to enhance dramatically the channel mobility of 90nm CMOS node and below. The second innovation introduced in 2008 for 45nm CMOS node (45nm SRAM and 45nm CPU) is the use of mechanical stress and the use of high dielectric constant materials for gate insulation to help address gate leakage, and the metal gate to replace poly-silicon. In its innovations, Intel has used a Silicon Nitride capping layer to enhance the mobility of nano NMOS transistors by more than 30% (Ghani 2010). For nano PMOS transistors, Intel has used compressed Silicon Germanium (Si(1-x)Ge(x)) pockets in source and drain to introduce intentionally a compressive uniaxial extrinsic stress σ in the channel. Here x represents the germanium mole fraction. The compressive stress σ enhances the mobility of holes by more than 50% (JianLi 2014, Chui 2007, Ghani 2003). According to the predictions in (Ghani 2010), the application of this compressive stress will continue enhancing the mobility of the future nano PMOS of 16nm node, in 2016, and beyond.

The extrinsic stress σ is due to an initial (or intrinsic) stress σ_0 that builds in Si(1-x)Ge(x) pockets after high temperature epitaxial deposition of these Si(1-x)Ge(x) pockets on top of Silicon substrate. Physically, this intentional initial stress σ_0 is due mainly to material mismatch and thermal mismatch. In this work, we are focusing on how to calculate accurately the intrinsic stress σ_0 and the resulting extrinsic stress σ that are due to thermal mismatch. The thermal mismatch stress is due to the mismatch between the thermal expansion

coefficients of Silicon substrate, $\alpha_{Si}(T)$, and Si(1-x)Ge(x) pockets, $\alpha_{SiGe}(T)$. Here, T represents the deposition temperature.

In this paper, three things are new: (1) The development of new and accurate models to include the effects of the fabrication temperature on elastic constants C11 and C12 of Silicon, Germanium, and Silicon Germanium materials. (2) We use, for the first time, these temperature dependent elastic constants to calculate the thermal induced initial stress σ_0 in Si(1-x)Ge(x) thin films. (3) We use this thermal induced initial stress σ_0 to calculate numerically the resulting extrinsic stress σ in the whole structure of 14nm PMOS transistor based on Intel technology.

In our previous work (El Boukili 2015), we have proposed and used the following model to calculate the thermal induced initial stress σ_0 in Si(1-x)Ge(x) thin films after deposition:

$$\sigma_0 = D \cdot \varepsilon_0(T) \quad (1)$$

where $\varepsilon_0(T)$ represents the initial strain tensor due to temperature. And, D represents the stiffness tensor of the elastic constants C11, C12, and C44. The definition of D will be given in Section 3. In our previous model (El Boukili 2015), D was independent of temperature. It was calculated at room temperature T0. In this paper, we are proposing a new model that includes the effects of temperature on elastic tensor D. Then, the stiffness tensor D in Equation (1) will be replaced by D(T). And, the model in Equation (1), will be replaced by the new model given in Equation (2). The model in Equation (2) does improve and enhance considerably the model in Equation (1) since it does include the effects of temperature on the elastic constants. And, this is what makes this paper relevant for the semiconductor community. As far as we know, no one has used the model in Equation (2) in literature. The use of this new model allows correct calculation of the thermal induced stress in the channel of nano PMOS transistors. And, this thermal stress is used by manufacturers as Intel to enhance dramatically the electrical and optical performances of the new generation of nano PMOS and NMOS transistors (Ghani 2010). And, this will have a great impact on economy and social life. Nanotransistors are currently the backbone for today's economy. Intel estimates about 10 million nanotransistors are shipped every year. The knowledge of the physics of transistors at the nonometer scale is, however, only theoretical. So, the understanding of the physical behavior of the nanotransistors can only happen via theoretical models and computer simulations which is the main objective of this paper.

This paper is organized as follows. Section 2 will outline the main sources of intrinsic stress in Si(1-x)Ge(x) films after high temperature deposition. Section 3 will present the new models for temperature

dependent elastic constants of Silicon substrate and Si(1-x)Ge(x) thin films. It will also present the mathematical models we are using to calculate the initial stress σ_0 due to thermal mismatch, and the resulting extrinsic stress σ . Section 4 will present the validation of the proposed theoretical models for the temperature dependent elastic constants with respect to the experimental data. It will also present the 3D simulation results of the thermal induced extrinsic stress σ , obtained from using these elastic constants. This section will also analyze qualitatively and quantitatively these simulation results for σ and provide comparisons with the results found in the literature. Section 5 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 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 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 $\text{Si}(1-x)\text{Ge}(x)$ films and Silicon substrate shrink and the thermal stresses in the material increase. The initial thermal stress σ_0 in the $\text{Si}(1-x)\text{Ge}(x)$ films depend mainly on the fabrication temperature T , the elastic constants of $\text{Si}(1-x)\text{Ge}(x)$, Germanium ratio x , the substrate orientation, the doping, and the deposition technique used. For the current nano PMOS transistors, $\text{Si}(1-x)\text{Ge}(x)$ 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 $\text{Si}(1-x)\text{Ge}(x)$ films.

3. MODELING OF TEMPERATURE EFFECTS ON ELASTIC CONSTANTS OF SILICON AND GERMANIUM

Since the discovery of the transistor in 1947 many research has been done to understand the electrical properties of Silicon (Si) material and other semiconductor materials like Germanium (Ge) or Silicon Germanium ($\text{Si}(1-x)\text{Ge}(x)$). However, not enough research has been done to understand the mechanical properties of Si, Ge, and $\text{Si}(1-x)\text{Ge}(x)$. Between 1947 and 1999 the performances of CMOS transistors were related mainly to the electrical properties of the materials making up the CMOS devices. Only in the beginning of 2000 that Intel discovered that the performances of a nano CMOS transistor can be enhanced dramatically using the mechanical properties of Si, Ge, $\text{Si}(1-x)\text{Ge}(x)$ or other materials making up the nano NMOS or PMOS transistors (Ghani 2010). The mechanical properties of a material affect significantly its electrical properties. The mechanical properties of Silicon, Germanium, $\text{Si}(1-x)\text{Ge}(x)$ or any other symmetric material are represented mainly by their elastic constants C11, C12, and C44.

In this paper, we are proposing the following new model which includes the effects of temperature on the elastic constants C11, C12 and C44 to calculate correctly the initial thermal stress $\sigma_{\text{SiGe}}^0(T)$ in $\text{Si}(1-x)\text{Ge}(x)$ thin films:

$$\sigma_{\text{SiGe}}^0(T) = D(T) \cdot \epsilon_{\text{SiGe}}^0(T) \quad (2)$$

where $\epsilon_{\text{SiGe}}^0(T)$ represents the initial thermal strain tensor which is defined by:

$$\epsilon_{\text{SiGe}}^0(T) = (\epsilon_0^{xx}, \epsilon_0^{yy}, \epsilon_0^{zz}, \epsilon_0^{xy}, \epsilon_0^{yz}, \epsilon_0^{xz}) \quad (3)$$

where $\epsilon_0^{xx}, \epsilon_0^{yy}, \epsilon_0^{zz}$ are the initial normal strain components, and $\epsilon_0^{xy}, \epsilon_0^{yz}, \epsilon_0^{xz}$ are the initial shear strain components that are assumed to be zero. We also assume that in $\text{Si}(1-x)\text{Ge}(x)$:

$$\epsilon_0^{xx} = \epsilon_0^{yy} = \epsilon_0^{zz} = \epsilon_{\text{SiGe},0}(T) \quad (4)$$

In our previous work (El Boukili 2015), we have proposed the following model to calculate $\epsilon_{\text{SiGe},0}(T)$:

$$\epsilon_{\text{SiGe},0}(T) = \int_{T_0}^T \alpha_{\text{SiGe}}(t) dt + \Delta\alpha(T)\Delta T \quad (5)$$

$$\Delta\alpha(T) = \alpha_{\text{SiGe}}(T) - \alpha_{\text{Si}}(T), \quad (6)$$

$$\Delta T = T - T_0. \quad (7)$$

And the stiffness tensor of elastic constants $D(T)$ is given by the matrix :

$$\begin{bmatrix} c11(T) & c12(T) & c12(T) & 0 & 0 & 0 \\ c12(T) & c11(T) & c12(T) & 0 & 0 & 0 \\ c12(T) & c12(T) & c11(T) & 0 & 0 & 0 \\ 0 & 0 & 0 & c44(T) & 0 & 0 \\ 0 & 0 & 0 & 0 & c44(T) & 0 \\ 0 & 0 & 0 & 0 & 0 & c44(T) \end{bmatrix}$$

As far as we know, no work has been done to include the effects of temperature on the elastic constants to calculate the initial thermal stress σ_0 and the resulting extrinsic stress σ in a nano PMOS transistor. In our previous work (El Boukili 2015), we have used the model given by Equation (1) where the elastic constants C11, C12, and C44 were considered independent of the fabrication temperature T . We have used the values at room temperature T_0 . We believe that this new model will give more accurate values for the initial thermal stress σ_0 that we need to calculate numerically the resulting extrinsic stress σ in the whole structure of an Intel nano PMOS transistor.

We are proposing and using the following models, based on Taylor's expansion, for the temperature dependence of the elastic constants of Silicon and Germanium materials:

$$C_{ij}^{\text{Si}}(T) = C_{ij}^{\text{Si},0} [1 + a_{ij}^{\text{Si}}(T - T_0) + b_{ij}^{\text{Si}}(T - T_0)^2] \quad (8)$$

$$C_{ij}^{\text{Ge}}(T) = C_{ij}^{\text{Ge},0} [1 + a_{ij}^{\text{Ge}}(T - T_0) + b_{ij}^{\text{Ge}}(T - T_0)^2]. \quad (9)$$

Where $C_{ij}^{\text{Si},0}$ and $C_{ij}^{\text{Ge},0}$ represent the elastic constants of Silicon and Germanium at room temperature T_0 . And, $a_{ij}^{\text{Si}}, b_{ij}^{\text{Si}}, a_{ij}^{\text{Ge}}, b_{ij}^{\text{Ge}}$ represent the first-order and the second order temperature coefficients of the elastic constants of Silicon and Germanium respectively. Here, the index i is equal to 1 and j is equal to 1 or 2. C44 is not needed to calculate the initial stress σ_0 .

The values of the Silicon constants $C_{ij}^{Si,0}$, a_{ij}^{Si} , b_{ij}^{Si} are extracted from the measurements values given in (Elding 2015, Antti 2014, Matthew 2010, Bourgeois 1997). The values given in (Antti 2014) are also depending on doping concentrations. In this work, we are taking average values with respect to doping. The results we get for the elastic constants of Silicon ($C_{ij}^{Si}(T)$) are given in Table 1. We believe we are the first to propose and use the model in Equation (9) for Germanium. We should note that not that much work has been done in the literature for Germanium even experimentally.

We are using the least squares method to fit the model in Equation (9) with the measurements from (Madelug 2001) to calculate the values of the constants $C_{ij}^{Ge,0}$, a_{ij}^{Ge} , b_{ij}^{Ge} . The results for $C_{ij}^{Ge}(T)$ are given in Table 1. Finally, to calculate the temperature dependent elastic constants of Si(1-x)Ge(x), we are proposing and using the following model:

$$C_{ij}^{Si(1-x)Ge(x)}(T) = (1-x)C_{ij}^{Si}(T) + xC_{ij}^{Ge}(T). \quad (10)$$

The values we get for $C_{ij}^{Si(1-x)Ge(x)}(T)$ using this model are given in Table 1.

The component form of the initial thermal stress tensor $\sigma_{SiGe}^0(T)$, is defined by:

$$\sigma_{SiGe}^0 = (\sigma_0^{xx}, \sigma_0^{yy}, \sigma_0^{zz}, \sigma_0^{xy}, \sigma_0^{yz}, \sigma_0^{xz}),$$

where $\sigma_0^{xx}, \sigma_0^{yy}, \sigma_0^{zz}$ are the initial normal stress components and $\sigma_0^{xy}, \sigma_0^{yz}, \sigma_0^{xz}$ are the initial shear stress components that are assumed to be zero.

In our simulation program, the thermal initial stress tensor $\sigma_{SiGe}^0(T)$ is used as a source term to calculate, in the whole 3D nano PMOS transistor, the extrinsic stress tensor σ defined by: $\sigma = (\sigma^{xx}, \sigma^{yy}, \sigma^{zz}, \sigma^{xy}, \sigma^{yz}, \sigma^{xz})$. We note that σ^{xx} , σ^{yy} , and σ^{zz} represent the extrinsic stress along the channel, vertical to the channel, and across the channel, respectively. This stress σ in the channel is used to enhance the mobility of holes in 3D nano PMOS based on Intel technology (Ghani 2010). We assume that Silicon and Silicon Germanium are elastic materials. And, to calculate the stress tensor σ we use the elastic stress model based on Newton's second law of motion (motion of atoms), and the following extended Hooke's law relating stress to strain:

$$\sigma = D\varepsilon - \sigma_0 + \sigma_{00}. \quad (11)$$

Here σ_{00} is taken to be zero for simplicity. And, σ_0 is the thermal initial stress which is equal to $\sigma_{SiGe}^0(T)$

on Si(1-x)Ge(x) films and 0 elsewhere. A detailed description and how to use finite volume method to solve numerically the elastic model (11) in 3D is given in (El Boukili 2010).

4. RESULTS AND VALIDATION

4.1 Validation of the proposed models for elastic constants using experimental data

The following Table 1 summarizes the calculated values of the elastic constants, in GPa, of Si, Ge, and Si(1-x)Ge(x) using the models we proposed in Equations (8),(9), and (10) where x=35% and the temperature is ranging from 25°C to 882°C. The values for the elastic constants of Silicon are in good agreement with measurements from (Antti 2014). The values for the elastic constants of Germanium are in good agreement with the measurements from (Madelug 2001). And those for Si(0.65)Ge(0.35) are in good agreement with the measurements from (Siqing 1992).

Table 1: C11(T) and C12(T) of Si, Ge, and Si(0.65)Ge(0.35) in GPa.

T in °C	Ge: C11;C12	Si(0.65)Ge(0.35): C11;C12	Si: C11;C12
25	124; 41.3	115.08;55.70	164.14;63.46
100	123.36; 40.97	149.36;55.29	163.36;63.00
300	121.18;40.09	146.72;54.10	160.47;61.65
500	118.27;39.17	143.20;52.86	156.62;60.23
700	114.64;38.73	138.80;51.57	151.81;58.77
800	112.55;37.73	136.27;50.91	149.04;58.01
882	110.71;37.32	134.03;50.36	146.60;57.38

We can see from the Table 1 that the values of the elastic constants C11(T) and C12(T) are decreasing with increasing temperature. This will have a significant effect on the mechanical and electrical properties of these materials and also on the performances of the nano CMOS devices using these materials. The values of the elastic constants of Si(0.65)Ge(0.35) in Table 1 are between those of Germanium and Silicon which is in a good agreement with the theory and experiments (Siqing 1992). As far as we know, the models we have proposed in Equations (9) and (10) for Germanium and Si(1-x)Ge(x) are new. No theoretical model has been found in the literature for Germanium and Si(1-x)Ge(x). We believe that not many work has been done for Germanium and Si(1-x)Ge(x) because Silicon is the most employed material in semiconductor industry. Integrated circuits, solar cells, and Micro electro mechanical systems (MEMS) extensively use Silicon and not Germanium or Si(1-x)Ge(x).

4.2 Validation of the proposed models for elastic constants using simulation results of stress

The elastic constants we have calculated in Table 1 are used to calculate the thermal initial stress σ_0 using the Equations (2) and (5). And, we have used this initial stress σ_0 to calculate numerically the resulting extrinsic stress σ , using the Equation (10), in the whole structure of the nano PMOS shown in Figure 1. This PMOS belongs to the 14 nm technology node developed by Intel in 2014. The gate length of this nano PMOS is 23 nm. Its source and drain are made up of Si(0.65)Ge(0.35). The Figure 2 shows the stress distribution along channel (σ^{xx}) when the elastic constants are independent of temperature. The Figure 3 shows the stress distribution along channel when the elastic constants depend on temperature T and T=500°C. Figure 4 shows the stress distribution along channel for T=882°C. From Figure 2, the magnitude of the average value of the stress along channel is 4955e+5 Pa and from Figure 3 the magnitude of the average value of the stress along channel is 4330e+5 Pa. Then, the absolute value of the difference is 625e+5 Pa. And, this is a huge difference. We could then say that the values of stress are overestimated when the effects of temperature are not included in the elastic constants. We can then conclude that to get accurate and correct values of the stress in the channel of a nano PMOS or NMOS transistor, we should take into account the effects of the fabrication temperature on the elastic constants. As far as we know, no work has been done in the past to investigate this crucial issue. From Figures 3 and 4 we could note that the magnitude of the stress along the channel is increasing with increasing temperature.

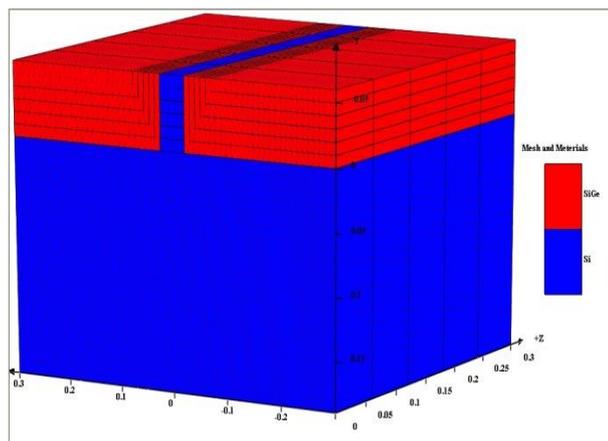


Figure 1: Materials And Mesh Of The Simulated Structure

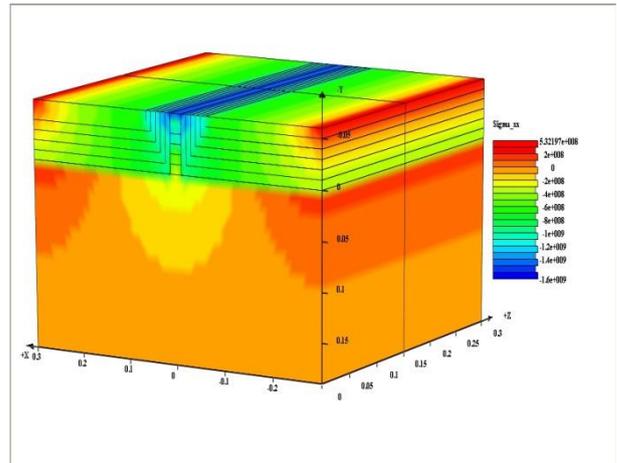


Figure 2: 3D Distribution Of xx Stress Component Without Temperature Effects On C11 And C12

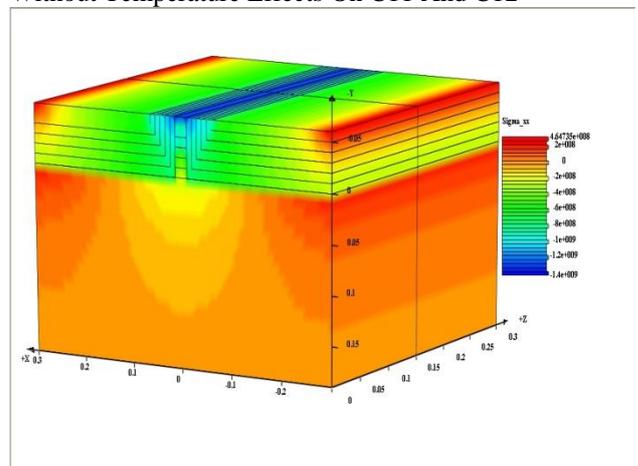


Figure 3: 3D Distribution of xx Stress Component With Temperature Effects On C11 And C12 When T=500°C

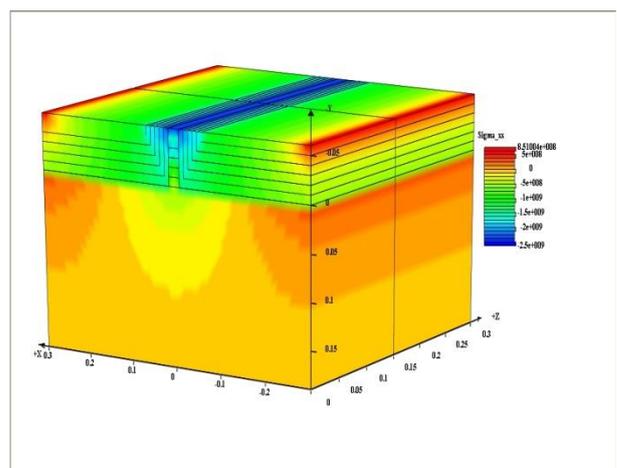


Figure 4: 3D Distribution of xx Stress Component With Temperature Effects On C11 And C12 When T=882°C

5. CONCLUSIONS

In this paper, we have developed new models to calculate accurately the thermal induced initial stress σ_0 and the resulting extrinsic stress σ in the whole structure of the 14nm nano PMOS transistor based on Intel technology. In these models, we have taken into account, for the first time, the effects of the fabrication temperature on the elastic constants C11, and C12 of the Silicon, Germanium, and Silicon Germanium. The values of the temperature dependent C11 and C12 obtained with these models are in good agreement with the measurements from (Antti 2014, Madelug 2001, Siqing 1992). The 3D results showing the effects of the fabrication temperature on the extrinsic stress in the whole 14nm nano PMOS transistor are given in the Figures 3 and 4. The obtained values for the extrinsic stress along channel (σ^{xx}) are qualitatively in good agreement with those found in literature (Victor 2013, 2012, 2004) and (El Boukili 2010, 2015). But, quantitatively these values are much smaller than the values given in our previous work (El Boukili 2010, 2015) where the temperature effects were not included in the elastic constants C11 and C12. This is a great finding for nano PMOS and NMOS manufacturers and design engineers. In the future work, we will investigate the effects of doping and temperature on the elastic constants C11 and C12 and explore their effects on the channel stress.

REFERENCES

- Antti, J. et al., 2014. Determination of doping and temperature dependent elastic constants of degenerately doped silicon from MEMS resonators. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol.61, No.7, 1063-1074.
- 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.
- Bourgeois, C. et al., 1997. Design of resonators for the determination of the temperature coefficients of elastic constants of mono crystalline silicon. *Proceeding of IEEE int. Frequency Control Symp.*, pp. 791-799. May 28-30, Orlando, FL, USA.
- Chui, K.J., 2007. n-MOSFET with silicon-carbon source/drain for enhancement of carrier transport. *IEEE Trans. Electron Devices*, Vol. 54, 249-256.
- Matthew, A. et al., 2010. What's Young modulus of silicon? *Journal of micro electro mechanical systems*, Vol.19, No. 2, 229-238.
- Madelug, O., Rosser, U., Schulz, M., 2001. *Group IV elements, IV-IV and III-V compounds. Part a-lattice properties*. Volume 41, Berlin Heidelberg: Springer.
- 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, pp. 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.
- Elding, J., et al., 2015. Temperature dependence of the elastic constants of doped silicon. Accepted in *Journal of micro electro mechanical systems*.
- 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.
- JianLi, M. et al., 2014. Hole mobility enhancement in uniaxial stressed Ge dependence on stress and transport direction. *Science China, Phys. Mech. & Astr.*, Vol.57, 1860-1865.
- Krivokapic, Z. et al., 2003. Locally strained ultra-thin channel 25nm Narrow FDSOI Devices with Metal Gate and Mesa Isolation. *Proceedings of IEDM, IEEE International*, pp. 445-448. Dec. 8-10, Washington, DC, USA.
- Siqing, W. et al., 1992. Elastic constants of Si/Ge super lattice and bulk Si and Ge. *Physical Review B*, Volume 46, 12411-12420.
- Victor, M., 2004. *Analyzing strained-silicon options for stress-engineering transistors*. *Solid State Technology Magazine*. July, Vol. 47, issue 7, p49.
- Victor, M., 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.
- Victor, M., et al., 2012. 14nm FinFET stress engineering with epitaxial SiGe source/drain. *Proceeding of ISTDM*, pp. 1-2. June 4-6, Berkeley, CA, USA.

MODELING, SIMULATION AND OPTIMIZATION OF LOGISTICS MANAGEMENT OF A CANS PACKAGING LINE

Victoria G. Achkar^(a), Luciana S. Picech^(b), Carlos A. Méndez^(c)

^{(a),(b)} Facultad de Ingeniería Química, UNL, Santiago del Estero 2829, 3000 Santa Fe, Argentina

^(c) INTEC (UNL – CONICET), Güemes 3450, 3000 Santa Fe, Argentina

^(a)victoriaachkar@gmail.com, ^(b)lucianapicech@gmail.com
^(c)cmendez@intec.unl.edu.ar

ABSTRACT

This paper presents the development and application of an advanced discrete event simulation model for logistics management of raw materials and final products of a cans packaging line used in an important brewery company. Advanced techniques for discrete event simulation are utilized to cover a wide range of methods and applications to emulate, advice and predict the behavior of complex real-world systems. The selected tool is SIMIO simulation software, which facilitates, through a modern 3D interface, the verification and validation of the model. Based on the proposed model, a what-if analysis is performed to determine the thresholds values of the current system and determine critical variables. The model analyzes the behavior of a future scenario which considers a significant increase in the packaging line production rate.

Keywords: Logistics, Simulation Model, Packaging Line.

1. INTRODUCTION

Modeling and Simulation (M&S) tools have become an issue of increasing importance to aid complex decision making arising in process design and operations. Recently, many real-world processes evolving over time have started to be studied by using detailed discrete event simulation models (Banks, 2004). In fact, a survey of the current state of the art clearly reveals that a tremendous amount of research works have been published (over the last 50 years) in this area also related to M&S in food industry (Bruzzone and Longo, 2013; Callahan et al., 2005). M&S-based approaches in the food industry can be found in Hussein et al. (2009), Reiner and Trcka (2004), Huda and Chung (2002) and Lipnizki et al. (2002). M&S-based approach does not provide exact or optimal solutions to problems but it allows users analyzing the behavior of complex systems, performing what-if analysis and choosing correctly among different possible scenarios (Banks, 2004; Karacal, 1998). The main advantage of M&S when

studying and analyzing manufacturing and logistics systems is the possibility to take into consideration multiple aspects without introducing restrictive assumptions (Bruzzone, 2004). This is one of the principal differences between M&S and analytical approaches.

Motivated by the decision of increasing the capacity of a cans packaging line, the company under study decided evaluating the future impact of the change by using a simulation model. Major constraints to be considered are limited storage space and material handling capacity.

This work aims to represent the actual process operation in order to analyze critical variables associated to storage and material handling devices to move raw materials and final products. The goal is to evaluate alternative modes of increasing production rate without generating problems into internal logistics. Thus, we developed a simulation model to accurately reflect different scenarios to be studied. The software chosen for this work, SIMIO, is a modern tool for building and running dynamic models of systems (Simio LLC, 2010). When running the model, it is possible to see a 3D animation of the system behavior. This allows observing the operation of the proposed system before building or performing any changes. Since the simulator uses an object-oriented approach, a simpler and more organized model development of complex systems is achieved. Another work has recently used this modern tool to solve similar situations in decision-making processes (Basán et al., 2013).

The paper continues with the following structure: (ii) methodology, (iii) internal logistics system and simulation model, (iv) verification and validation, (v) sceneries definition and experimentation (vi) discussions.

2. METHODOLOGY

To make a model that is able to reproduce the behavior of the real system, it is necessary to collect information from multiple sources using different techniques. This took place in the company under study through: (i) in situ

observation, (ii) interviews with staff of the company and (iii) collection of historical data.

Once obtained the desired information, it was filtered and analyzed to become input data. The input analyzer tool was used to fit arrivals data to proper statistical distributions. Among the most relevant inputs may be mentioned: packaging speed, tasks made by each forklift, operation time available for every element, warehouses and cellars capacity, arrival schedules of different truck types (empty cans, final product dispatches, other raw materials, etc.). Subsequently, this information was used in the verification and validation stages of the model.

Once the operation of the system was fully understood, a conceptual model was generated. This was used to obtain a sufficient abstraction level of the problem and to define the assumptions. Each assumption was reviewed and agreed through successive interviews with specialized personal of every sector of the company.

Finally, the model in the SIMIO software (Pedgen, 2015) was developed. 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.

The simulation model was used to: (i) represent the current operation of the system, (ii) make an analysis of critical variables of the system, evaluating its limits under current operating conditions, and (iii) propose a scenario in which, with a desired production speed, it is possible to identify necessary changes and performance variables impact.

3. INTERNAL LOGISTICS SYSTEM AND SIMULATION MODEL

The process of storing and moving raw materials and finished products includes different sectors, which simultaneously perform different tasks and depend on each other for proper operation. In this paper the sectors that are taken into account are directly related to the process of cans packaging. These sectors are: Cans packaging line, Warehouse and cellar line, Raw material storage, Main loading and unloading area, and Secondary loading and unloading area.

In order to develop the simulation model, it is first necessary to understand the operation of the actual process to clearly identify its main elements. In the system under study the following elements are identified:

- Pallets of final product (Cans and "Other")
- Pallets of different raw materials
- Objects that represent intermediate products or scrap
- Forklifts of different sectors
- Loading and unloading docks

- Warehouses and cellars
- Temporary stock areas
- Paths to connect the different sectors
- Packaging line

The model developed in SIMIO uses a factory scale plan, which can be seen in Figure 1. To start the development of the model, elements from the standard library of SIMIO are used, such as Source, Server and Sink elements. As we investigate into detail, custom elements with internal logic processes are created.

To simulate the internal logistics system, it is necessary to represent the main components of the system as mentioned above. Figure 2 shows the different views of the model developed in SIMIO.



Figure 1: 2D SIMIO Model - Layout. (1) Packaging Line, (2) Line Temporary Stock Area, (3) Warehouse and cellar belonging to line, (4) "Others" Warehouse, (5) Main Loading and Unloading Area, (6) Raw Materials Warehouse, (7) Secondary Temporary Stock Area, (8) Secondary Loading and Unloading Area



Figure 2: SIMIO 3D Model – 3D views of various sectors of the system

Elements modeled in SIMIO are listed below. The first three items represents entities that moves through the system and can be seen in Figure 3.



Figure 3: SIMIO 3D Model – 3D views of pallets, forklifts and trucks

3.1. Trucks

Trucks are differentiated by function (load or unload) as well as their transporting contents. Depending on the vehicle type, it is considered an arrival following a Schedule or a statistical distribution.

In Table 1 different type of trucks and their characteristics are listed.

Table 1: Truck features

Truck	Pallets destination (sectors in Fig. 1)	Truck Arrivals
Empty Cans	Pallets unloaded in (5), stored in (3)	Weekly Schedule
Final Product	Pallets picked-up in (3), loaded in (5)	Monthly Schedule
Others Unload	Pallets unloaded in (5), stored in (4)	Exp(19.1) min
Others Load	Pallets picked-up in (4), loaded in (5)	Exp(20.6) min
Raw Materials (except empty cans)	Pallets unloaded in (8), temporary stored in (7) finally stored (6)	Monthly Schedule

3.2. Forklifts

Eleven forklifts were identified: one for the packaging line, another to the raw materials warehouse, and the remaining belonging to logistic areas. Depending on the forklift type, different speeds, loading and unloading times and carrying capacities are assigned. Unproductive times are also assigned, either by maintenance, breaks and days off. They are represented in the model with Vehicle module. Each type of forklift and their characteristics are listed in Table 2. Then, Table 3 shows every forklift visiting points (places within it moves to perform tasks) and its work schedule.

Table 2: Forklift characteristics

Forklift	Quantity	Speed	Capacity
Line	1	5 km/hr	1 pallet
Primary Logistics	8	10 km/hr	2 pallets
Secondary Logistics	1	10 km/hr	2 pallets
Warehouse	1	10 km/hr	1 pallet

Table 3: Forklift visiting points and work schedules

Forklift	Visiting points (see Fig.1)	Work Schedules (24hr days)
Line	(1) – (2)	7 days a week, 4.7 daily not operative hours
Primary Logistics	(3) – (4) – (5)	From Sunday 22h to Saturday 18hr, 1.5hr of daily work relief
Secondary Logistics	(7) – (8)	From Sunday 22h to Saturday 18hr, 1.5hr of daily work relief
Warehouse	(2) – (6) – (7)	From Monday to Friday, 1.5hr of daily work relief

3.3. Pallets of different materials

Raw material pallets, final products (cans and other materials) are represented through SIMIO entities. They are also identified entities to be dismissed due to rejection rates or secondary entities that are used to represent occupation in certain processes.

3.4. Cans Packaging Line sector

The cans packaging line works 24 hours a day. However, due to unexpected interruptions, only 80% of this time is considered productive. It has a packaging speed of 10.11 pallets/hr with its consequent consumption rate of raw materials.

For immediate supply, it has a close temporary stock area. It also has a cellar and a warehouse. While cans final products pallets are stored in the first one, empty cans pallets are stored in the second one. There is an exclusively forklift designated to this sector.

3.4.1. Packaging line

The line has a production speed in pallets/hour and every day it goes through an unproductive period of time represented by multiple causes. We also model a variable that monitors if there is an unexpected line interruption.

Consumption of raw materials in the line is modeled according to the requirements of materials by final product (BOM matrix). Every raw material in the model has a defined replenish point which, when crossed, triggers a request to bring material to the line. In addition, every pallet leaving the line requests a transfer to the final product warehouse. At most, two pallets can be accumulated in the output of the line.

Concurrently, there are secondary tasks destined to remove raw material pallets and packaging scrap, and then to pack them for its subsequent devolution.

The line forklift is in charge of feeding the line with pallets of the temporary stock area, removing pallets from the output of the line and performing secondary tasks described above.

This line is represented in the model with Workstation, Source, Server and Sink modules, as it can be seen in Figure 3.

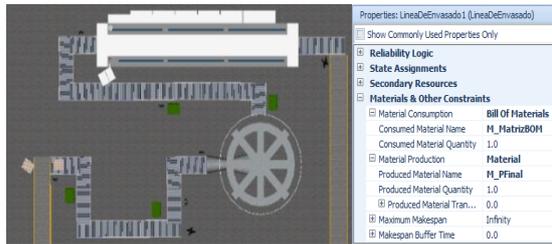


Figure 3: SIMIO 2D Model – Cans Packaging Line and some of its properties

3.4.2. Line temporary stock area

The warehouse forklift transports once a day an enough quantity of every raw material in order to satisfy a certain period of time (depending on the day of the week) and without exceeding its limited capacity. When requested, the line forklift retires pallets of each raw material to feed the line. Figure 4 shows the temporary stock area represented by a set of Servers. They store each raw material with capacity restrictions and processes that wait material requests from the line.

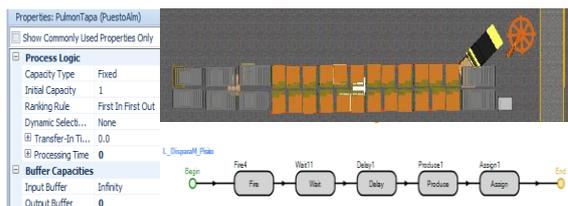


Figure 4: SIMIO 2D Model - Temporary stock area belonging to line, its properties and an internal process

3.4.3. Line warehouse and cellar

This sector is used for storing on the one hand the final product and on the other empty cans. Both sectors are modeled similarly: they receive and deliver pallets to/from different sectors. Each has an assigned capacity which may be diminished by an operational company policy.

The line forklift provisions the final product cellar from one pallet at a time, and this cellar releases 26 pallets each time it is requested. Logistics forklifts remove and carry them to a given dock and perform the truckload.

Logistic forklifts provision the empty cans warehouse whenever there comes a truck and simultaneously a line forklift picks up those pallets to replenish the line.

The cellar and the warehouse are represented in the model with Servers that have internal processes to retain the entities until they are requested and route them to where they should go. Figure 5 shows them.

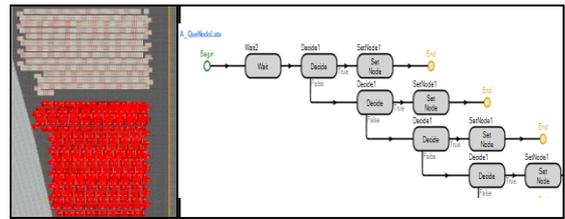


Figure 5: SIMIO 2D Model – Warehouse, cellar and an internal process

3.5. “Others” Warehouse

To approximate the use of logistic forklifts to reality, the model includes the loading and unloading of "Other" products and it would represent anything not related to the production of cans. For this, the elements Source and Server are used. The model creates these products with a certain time interval between them, so that it reaches to charge all trucks which come and load without model saturation. A Sink module is used to discard all products downloaded from the trucks that do not contain raw material for the cans packaging line. The modules have internal processes that indicate, for example, where should be directed forklifts for charging each dock. Figure 7 shows the "Other" warehouse with some internal processes assigned to it.



Figure 6: SIMIO 2D Model – Warehouse “Others” and its properties

3.6. Raw materials warehouse

This warehouse is shared with all packaging lines of the company. It has a defined capacity and it stores all raw materials involved in the cans line, except for empty cans pallets. In the model we consider only storage activities that are related to the line under study.

The exclusive warehouse forklift enters raw materials into the warehouse and then it distributes them to the line once a day.

The raw materials warehouse is modeled in a similar way to the empty cans one mentioned above. It has a certain capacity destined to raw materials used in the cans line, and enough quantity of them in stock in order to supply the line. In the model this is represented, as shown in Figure 7, with a server to every raw material and internal logic processes which retain them until the line request them.

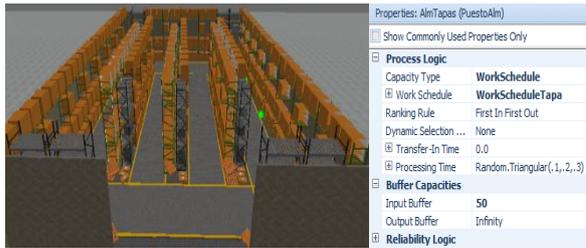


Figure 7: 3D SIMIO Model - Raw Materials Warehouse and one of its Servers properties

3.7. Loading and unloading areas

The company has two of these areas: the main one used to load all final products and unload empty cans pallets, and the secondary one used to unload raw materials pallets.

3.7.1. Main loading and unloading area

This sector has four docks used for loading and unloading pallets from trucks arriving at the company. For proper operation each dock has a maximum of two forklifts assigned to perform such tasks. Therefore, this sector has eight logistic forklifts. In the model all loading and unloading activities of the company are represented, but those corresponding to the line under study are differentiated from others.

3.7.2. Secondary loading and unloading area

Its operation is similar to the area mentioned above, but in this case it is a single dock that only makes downloads of raw materials that, depending on the type of material, require a logistic forklift or not. While this forklift makes an unload, it locates pallets in a stock area next to the dock, and once a day, the exclusive warehouse forklift empties this area carrying pallets into de raw materials warehouse.

For both areas in the model, docks receive different types of trucks and, depending on that, perform a loading or an unloading task. If there is an unload task, a determined quantity of pallets of the material contained in the truck is created and logistic forklifts transport them until the truck is emptied. If a loading task is required, once the truck arrives to the dock, the product to load is requested from its location and the liberation of those entities is ordered, so that forklifts can pick and transport them to the truck. Through variables that monitor the loading or unloading process it is controlled the remaining quantity to load or unload. Once the truck finishes its task, there is a delay representing truck conditioning and then it leaves and liberates the dock. In the model representing the current operation of the company, distributed operations along a month are represented, corresponding to 108 unloads of empty cans pallets, 254 load of can final products pallets, 1889 unloads and 2032 loads of "Other" pallets. In SIMIO, we

represent docks with a simple Server and with many internal logic processes responsible for creating entities, monitoring with variables different restrictions to be accomplished, and destroying the truck once the task is finished. Figure 8 shows docks of the principal loading and unloading area with one of its internal processes.

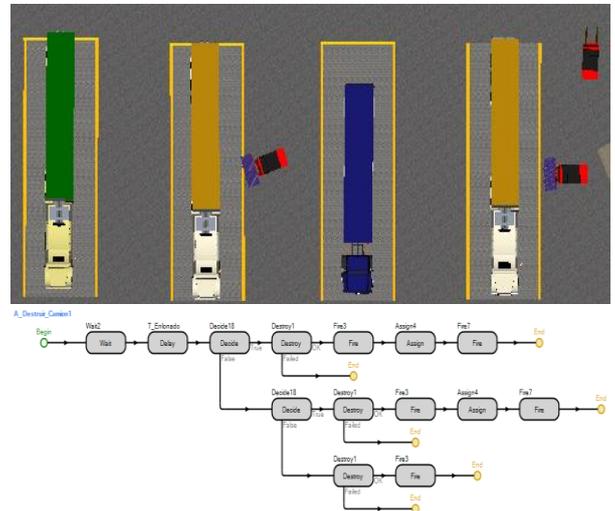


Figure 8: SIMIO 2D Model – Docks of main loading and unloading area, and internal dock process

3.8. Entries

The entries to loading and unloading areas are modeled with Source and Server modules. Source elements are responsible for creating different type of trucks following its corresponding arrival schedules. Figure 9 shows the representation of the entrance to the main loading and unloading area and an internal process.

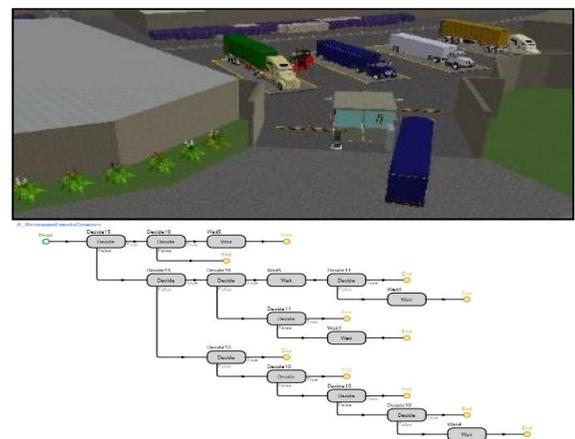


Figure 9: SIMIO 3D Model – Entry and internal logic process

There are defined restrictions that prevent the entry of trucks in the following situations: (i) if it is a day that docks do not operate, (ii) if there is not an available dock, (iii) if there is a certain amount of

trucks of the same type within the loading and unloading area simultaneously (the amount depends on the type) or (iv) if there is not availability of final product. We also assign to the entries a delay representing documentaries issues.

4. VERIFICATION AND VALIDATION

Verification is concerned with determining if the conceptual model with its specifications and assumptions were correctly traduced in computerized representation (Law and Kelton, 2007). To carry on, we perform a detailed analysis of different output variables, within the most relevant: (i) quantity of final product packed, (ii) utilization of every forklift, and (iii) usage of docks. We compare these values with calculations based on assumptions of the conceptual model. In this way, concordance is verified and the model is adjusted in case of finding discrepancies.

Then, validation is carried out, which is concerned with determining how closely the simulation model represents the actual system (Law and Kelton, 2007), to the particular objectives under study. For this process we make an iterative comparison between the values obtained from the computer model and the information provided by the company about the time period represented (critical production month) and through subsequent interviews.

The main values compared are (i) daily occupation of store and warehouse line, (ii) monthly production, (iii) monthly downtime, (iv) number of trucks arriving per day of each type and (v) utilization rate of forklifts. The necessary settings and changes in the model are made to ensure that these values are similar to the records of the company and, in addition, have been accepted by the staff of the company involved with the development of the model.

5. DEFINITION OF SCENARIOS AND EXPERIMENTATION

Output analysis is the examination of data generated by a simulation. Its purpose is either to predict the performance of a system or to compare the performance of two or more alternative system designs (Banks 2004). In this step we generate all data needed and we make an analysis of the impact that decision variables have on response variables through the design and simulation of different scenarios.

For the case under study, we define different scenarios increasing the line speed until achieving the desired one (25.28 pallets/hr). As a consequence of changing the design, we decide to add an extra forklift to the line and to increase the quantity of arrivals of trucks associated to the cans line.

We define the following response variables: (i) monthly production, (ii) the use of the line forklift, (iii) the use of logistic forklifts dedicated to primary load and unload area, (iv) the use of the docks of that area, and (v) increasing of trucks quantity arriving to the main area.

For the definition of different scenarios analyzed increasing packaging speed, we make some modifications to the current operation policies of the company. These are mainly related to the number of trucks that daily arrive to load cans final product and unload empty cans pallets. These new policies redefine: (i) the number of daily arrivals of each type of truck throughout the month and (ii) the number of arrivals calculated in such a way that line warehouse and cellar have a stable occupation in the long term.

The results of the simulation are shown in Table 3.

Table 3: Summary of results

Number of line forklifts	% increase in packaging speed	Performance system variables					
		Monthly production (pallets)	Line forklift usage	Logistic forklifts usage	Docks usage	% increase in empty cans trucks	% increase in final products trucks
1	-	6037	76%	49,5%	58%	0%	0%
	+10%	6635	83,3%	49,7%	58,5%	0%	+0,4%
	+20%	7239	91%	50,1%	59,1%	0%	+11%
	+31%	7949	99,9%	51,0%	60,1%	+7,4%	+21,7%
2	+50%	9048	57,1%	52,1%	61,41%	+25%	+28,2%
	+90%	10253	72,8%	54,3%	63,98%	+53,7%	+74,4%
	+130%	10253	89,2%	56,7%	66,63%	+88%	+110,6%
	+150%	15082	96%	60%	69,97%	+104%	+132%

As we can see, the line speed variation impacts heavily on the use of line forklift. That happens because speed directly affects the frequency of feeding the line and removing final product from it.

The use of logistic forklifts and docks varies in a lower magnitude. This is because the number of trucks related to can's transportation does not significantly increase the total number of trucks.

In the first scenarios there is not a significant increase in the number of trucks because in current situation the required amount is exceeded (we can see in the model that occupation of warehouse increase and of cellar decrease). Moreover, this augmentation is only related to two specific types of trucks (those related to cans production) representing a smaller percentage of the total arrivals in the model. Consequently, it does not significantly affect forklifts and docks usage.

We note that a single line forklift can increase production up to 31% without saturating the system. Subsequent increases are only feasible if we add another line forklift. With two forklifts assigned to the line, the model confirms the possibility to operate in the future at the desired speed by the company (25.28 pallets / h).

6. DISCUSSION

In this work we presented the development of an advanced simulation model to assist the decision making process in materials and final products logistic management related to a cans packaging line of an important brewing company.

We chose the simulation software SIMIO, resulting a flexible, robust and modern discrete event simulation based tool. The proposed model had allowed the company to predict the effect of a future expansion of the production capacity, mainly related to the available storing capacity and handling resources.

The developed work allowed to anticipate and to quantify the impact of future planned changes in different critical variables of current system, finding its limit values. The results obtained were very useful to enable the company to make decisions in advance without altering the current system and avoiding unnecessary costs.

ACKNOWLEDGMENTS

This paper was partially founded by CONICET under Grant PIP-2221 and from UNL under Grant PI-81-481.

APPENDIX

A. GLOSSARY

Path: 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.

Internal logic process: 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.

Server: represents a processing activity in the model. Between its properties they must be set: processing time, resources needed failures, internal process and events associated.

Sink: represents a final point in the model where entities go to be eliminated.

Vehicle: 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.

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.

REFERENCES

- Banks J., 1998. Handbook of simulation. U.S.A.: John Wiley & Sons.
- Banks J., Carson J.S., Nelson B.L., Nicol D.M., 2004. Discrete-Event System Simulation. U.S.A.: Prentice-Hall.
- Basán N., Ramos L., Cóccola M., and Méndez C.A., 2013. Modeling simulation and optimization of the main packaging line of a brewing company. Proceedings of the 25th European Modeling & Simulation Symposium, 551- 560. September 2013, Greece.
- Bruzzone AG., 2004. Preface to modeling and simulation methodologies for logistics and manufacturing optimization. Simulation 2004; 80:119–20. ISSN: 0037–5497. DOI:10.1177/0037549704045812.
- Callahan W.B., Hubbard K.M., Bacoski N.M., 2006. The use of simulation modeling and factorial analysis as a method for process flow improvement. Adv Manufacturing Technol 2006; 29:202–208.
- Hecker F., Mitzscherling M., Becker T., 2007. Simulation and Modeling Analysis. U.S.A.: Mc Graw Hill, Inc.
- Huda AM, Chung CA., 2002. Simulation modeling and analysis issues for high-speed combined continuous and discrete food industry manufacturing processes. Comput Ind Eng; 43:473–483.
- Hussein WB, Hecker F, Mitzscherling M, Becker T., 2009. Computer modelling and simulation of bakeries' production planning. Int J Food Eng; 5: Article number 8.
- Karacal SC., 1998. A novel approach to simulation modeling. Computers Ind Eng; 34:573–587.
- Law A., Kelton W. D., 2007. Simulation and Modeling Analysis. U.S.A.: Mc Graw Hill, Inc.
- Lipnizki F, Olsson J, Trägårdh G., 2002. Scale-up of pervaporation for the recovery of natural aroma compounds in the food industry. Part 1: simulation and performance. J Food Eng 2002; 54:183–95.
- Pedgen C. D., year. An Introduction to Simio for Beginners, Source. Available from: <http://www.simio.com/resources/white->

papers/Introduction-to-Simio/ [Acces: 27 april 2015].

Reiner G, Trcka M., 2004. Customized supply chain design: problems and alternatives for a production company in the food industry. A simulation based analysis. *Int J Production Econ*;89:217–229.

Simio LLC., 2010. Introduction to SIMIO. U.S.A.: Simio LLC.

AUTHORS BIOGRAPHY

Victoria G. Achkar is an advanced Industrial Engineering student conducting research in optimization and simulation tools for production planning and scheduling of automated production systems.

Luciana S. Picech 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/>.

A PROPOSAL OF USING DEVS MODEL FOR PROCESS MINING

Yan Wang^(a), Grégory Zacharewicz^(b), David Chen^(c), Mamadou Kaba Traoré^(d)

^{(a),(b),(c)} IMS, University of Bordeaux, 33405 Talence Cedex, France

^(d) LIMOS, Université Blaise Pascal, 63173 Aubiere Cedex, France

^(a)yan.wang@etu.u-bordeaux.fr, ^(b)gregory.zacharewicz@u-bordeaux.fr,
^(c)david.chen@ims-bordeaux.fr, ^(d)traore@isima.fr

ABSTRACT

Process mining is a relative young research area which consists of process modeling and data mining. Process discovery as a part of the process mining focuses on converting event logs into process models. Petri Nets formalism is identified as the most convenient resulting process model. However, it is not entirely satisfying and needs to be improved with the purpose of covering the temporal aspects of the studied system. Compared with it, DEVS has the advantage of explicit and concurrent time and separating model from simulation. The objective of this paper is to specify DEVS model as the resulting process model. Based on the existing Two-Phased Approaches in process mining, a region-based approach with the suitable mapping is designed to convert the transition system directly to DEVS. A study case is presented to implement this approach. This paper is a position paper and it needs to be completed. In addition, a dynamic semantic should be designed to solve some typical representational limitations and a simulation engine should be selected.

Keywords: process mining, process discovery, DEVS, dynamic semantic, time

1. INTRODUCTION

Nowadays, the development of the information systems cannot be separated by the operational process. Moreover, more and more events with abundant process information are recorded. Despite the omnipresence of event data, most of the organizations analyze and build models based on expert assumption rather than quantified collected fact. Thanks to process mining [Van der Aalst, 2011], it provides the methods to discover monitor and improve actual processes by using event data to extract process-related information. Process mining combines process design by model-driven approaches and data mining. It includes three types: process discovery, process conformance and process enhancement. This paper anticipates and lays a cornerstone to make a breakthrough on the process discovery area.

The purpose of discovery technique is to extract an event log with using recorded historical information and produce a model for simulation. In the process mining,

many models can be selected as the resulting model for business process for example workflow nets and BPMN, whereas Petri Nets [Peterson, 1977] is frequently identified as the direct resulting process models because Petri nets is formal, simple and graphical while still allowing for the modeling of concurrency, choices, and iteration. When considering about the representational bias such as concurrency, loops and even OR-splits/joins, not every process model can be totally satisfied. Petri Nets is able to discover the concurrency in spite of OR-splits. As there are many potentially concurrent activities, Petri Nets may either reach some limits in describing time or distinguish information in the process. The Discrete Event System Specification (DEVS) [Zeigler et al., 2000] provides a hierarchical and modular formalism to describe a state and event based system. It contains the basic components like explicit time, inputs, outputs, states and functions. The characteristic of separating model and simulation makes DEVS able to demonstrate the same function in Petri Nets. The semantic execution of DEVS can satisfy various potential requirements, such as distinguishing different information processing, and solve representational limitations. So we are trying to design a new approach to discover DEVS model as the resulting process model for business process simulation. The paper is organized as follow. Section 2 provides the background of DEVS formalism in comparison with the Petri net and an introduction about transition systems. Section 3 gives the introduction about process discovery, XES standard and DEVS related research. Two-phased approach is also presented. Section 4 proposes to construct a complex and meaningful modeling and simulation structure and presents the methodology design with the mapping. Section 5 makes the simulation based on a study case. Section 6 shows the limitation and indicates the future works. The paper ends with a conclusion.

2. BACKGROUND

2.1. DEVS Formalism

The DEVS formalism for modeling and simulation is based on discrete events, and provides a framework with mathematical concepts based on the set theory and

the system theory concepts to describe the structure and the behavior of the system [Zeigler et al., 2000]. A real system modeled by DEVS is described as a number of connected behavioral (atomic) and structural (coupled) components. A DEVS atomic model is formally defined as:

$$M = \langle X, Y, S, \delta_{int}, \delta_{ext}, \lambda, ta \rangle.$$

Where X is the set of input values; S is the set of states; Y is the set of output values; δ_{int} is the internal transition function; δ_{ext} is the external transition function; λ is the output function; ta is the duration function. From the semantics of DEVS formalism [Zeigler et al., 2000], we know that each atomic model has the duration specified by $ta(s)$. When the elapsed time $e = ta(s)$, the state duration expires and the atomic model will send the output $\lambda(s)$ and performs an internal transition to a new state specified by $\delta_{int}(s)$. However, state transition can also happen due to arrival of an external event which will place the model into a new state specified by $\delta_{ext}(s, e, x)$; where s is the current state, e is the elapsed time, and x is the input value. The time advance function $ta(s)$ can take any real value from 0 to ∞ . A state with $ta(s)$ value of zero is called transient state, and on the other hand, if $ta(s)$ is equal to ∞ the state is said to be passive, in which the system will remain in this state until receiving an external event.

The graphical notation is also used for simulation in this paper [Song et al., 1994]. Each node represents an activity or a state, dotted arc denotes an internal transition and solid arc denotes an external transition. The output events in both internal transition and external transition have different labels. An output event $p!m$ means that a message m is output at the port p . Similarly, an input event $p?m$ means that a message m is input at the input port p .

A coupled DEVS model consists of atomic and other coupled models connected together. They are formally defined as:

$$N = \langle X, Y, D, EIC, EOC, IC \rangle.$$

Both X and Y respectively define the sets of input and output events. D is a set of indices for the components. The external input coupling (EIC) specifies the connections between external and component inputs, while the external output coupling (EOC) describes the connections between component and external outputs. The connections between the components themselves are defined by the internal coupling (IC). The coupling and transformation between separating atomic models make it possible to construct a more complicated hierarchical model.

2.2. Petri Nets

Petri Nets is a modeling formalism originally developed by C.A Petri [Peterson, 1977]. A Petri net is a bipartite graph consisting of place and transition. Arc is used to connect between place and transition. The network

structure is static and token flows through the network. Despite the four parameters, the state of a Petri net is determined by the distribution of tokens and is referred as marking. Enabling and firing are the main operating rules. Comparing with DEVS, Petri Nets can be embedded into DEVS because a DEVS model can represent any discrete event behavior. Also, DEVS model has the timing characteristics connected with the reality whereas Petri Nets cannot present. Zeigler [Zacharewicz et al., 2008] discusses how DEVS modeling is more accurate than Petri nets modeling due to the following facts.

- DEVS gives a more general framework for modeling and simulation of complex systems.
- DEVS integrates naturally the notion of time contrary to Petri nets which require an extension of the formalism.
- DEVS offers a formal (and separated from model) definition of the simulator.

2.3. Transition Systems

Transition systems are considered as the original transformation process model in this paper. Robert [Robert, 1976] gives a formal definition about a transition system. It is a pair (S, \rightarrow) where S is a set of states and \rightarrow is a binary relation on S , called the set of transitions. A named transition system is a triple (S, \rightarrow, Σ) where (S, \rightarrow) is a transition system and each transition system is assigned one or more names in the set Σ . A visualized transition system [Van der Aalst, 2011] consists of states and transitions. The states are represented by black circles. There are one initial state and one final state. Each state has a unique label. This label is merely an identifier and has no meaning. Transitions are represented by arcs. Each transition connects two states and is labeled with the name of an activity. Multiple arcs can bear the same label.

Given a transition system one can reason about its behavior. The transition starts in one of the initial states. Any path in the graph starting in such a state corresponds to a possible execution sequence. A path terminates successfully if it ends in one of the final states. A path deadlocks if it reaches a non-final state without any outgoing transition. The transition system may live-lock if some transitions are still enabled but it is impossible to reach one of the final states.

3. THE STATE OF THE ART

3.1. Process discovery

Process discovery as a part of process mining is the main area we want to develop. General process discovery problem [Van der Aalst, 2011] is defined like this. Let L be an event log specified by the XES standard. A process discovery algorithm is a function that maps L onto a process model such that the model is "representative" for the behavior seen in the event log. The challenge is to find such an algorithm.

Until recently, the de facto standard for storing and exchanging events logs was MXML (Mining eXtensible

Markup Language). MXML emerged in 2003 and was later adopted by the process mining tool ProM. XES is the successor of MXML. Based on many practical experiences with MXML, the XES format has been made less restrictive and truly extendible. An XES document (i.e., XML file) contains one log consisting of any number of traces. Each trace describes a sequential list of events corresponding to a particular case. The log, its traces, and its events may have any number of attributes. To provide semantics for such attributes, the log refers to so-called extensions.

There are many different approaches to do the actual discovery. Two-Phase Approaches is one of the identified approaches. It first constructs a “low-level model” and then converts into a “high-level model” and other more advanced control-flow patterns. There are four steps in this approach: extract event logs; create a transition system based on one abstraction; convert the transition system into a Petri net; convert the Petri net into other notations (e.g., BPMN). In this new approach, we reuse the first two steps as the input and design a new approach of the transformation between transition system and DEVS model.

3.2. DEVS related research

DEVS can be easily transformed from Petri Nets or BPMN although we don't need this transformation. Jacques and Wainer [Jacques et al., 2002] propose an approach of mapping of the Petri Net modeling formalism into the DEVS modeling formalism using an unmodified DEVS simulator. They also show that DEVS simulation results can simply be filtered through a parsing tool to give them a stronger PN flavor. Bazoun et al. [Bazoun et al., 2014] define a transformation approach of BPMN models into DEVS simulation models based on the meta-model approach and describe the enrichment of obtained DEVS models through performance indicators. This approach includes an exhaustive mapping, the transformation architecture and an implementation in SLMToolBox M&S tool. Zacharewicz et al. [Zacharewicz et al., 2008] describe a language for workflow processes and a new transformation algorithm from Workflow XML specification to G-DEVS model. This language supports algorithms to transform the Workflow model.

Some other researches also focus on discovering the interoperability of DEVS. Wainer et al. [Wainer, 2009] standardize the simulation middleware to interface different simulation environments and allow synchronization for the same simulation run across a distributed network regardless of their model representation. They provide several approaches for example DEVS/SOA distributed simulation platform, DDSP, the shared abstract model, RISE and DEVS namespaces which standardize DEVS simulation middleware in different ways.

DEVS has already been widely used in many areas. Until now, there are many professional descriptions about its application. Song et al. [Song et al., 1994] introduces concepts of inverse DEVS and defines

controllability of discrete event systems expressed in the DEVS formalism. A graphical notation is presented to visualize DEVS models. The concurrency is also discussed to analyze the dynamics of a system consisting of several subsystems. Zeigler and Sarjoughian [Zeigler et al., 2005] present the systems entity structure/model base (SES/MB) framework for simulation process. First it sets up a DEVS model base as the organized libraries. There are FIFO, generator, transducer and processor model inside the model base. The knowledge of the desired system is represented by the SES. Then through retrieving component and coupling them together, we can get the synthesized model. At last, this model is evaluated via simulation. This framework can be identified as the basis to construct the semantic of the proposition.

4. DESIGN AND IMPLEMENTATION

4.1. Approach design

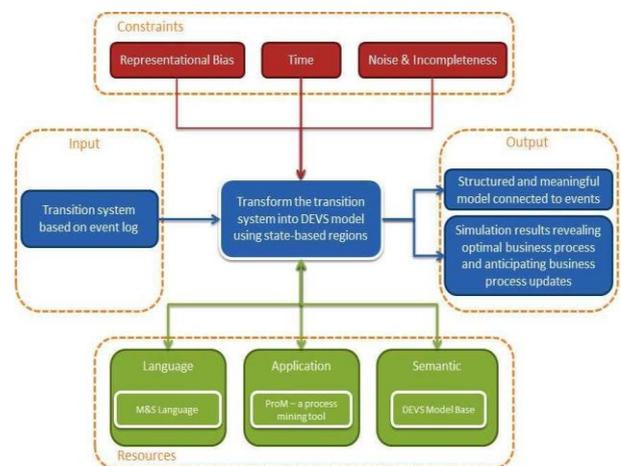


Figure 1: Proposed framework

Figure 1 shows the general view of the proposed framework. We reuse the discovery technique and take the transition system instead of Petri Nets or BPMN as input. In the resource part, we need to provide the DEVS model base to support the new approach. As the reuse of the discovery technique is operated in the process mining tool ProM, we continue to use it in the new approach. A lot of plug-ins implementing various techniques can be used in the ProM and it is a challenge to design a new plug-in for the DEVS model. The representational bias, noise and incompleteness appear when we try to transform the event logs into other formalism because the result isn't always optimal and desired. Furthermore, it will limit the search space of simulation. According to the representational bias, parallel DEVS models with multiple inputs and outputs and time segments can solve the problem of concurrency, a switch network can solve the problem of OR-splits, generator coupled model can solve the problem of loop and duplicate actions. All these models will be designed in the DEVS model base. As the process discovery is tightly connected with the reality,

more and more limitations will describe the enrichment of DEVS model base semantic. The noise defines event logs contain rare and infrequent behavior not representative for the typical behavior of the process. The incompleteness is defined as that the event logs contain too few events to be able to discover some of the underlying control-flow structures. Besides, time is another very important constraint which will lead to desired state scheduling. A time-related approach will be designed to make selections on transitions according to the semantics of DEVS formalism. At last, we wish to implement a structured and meaningful model used for anticipating optimal business process.

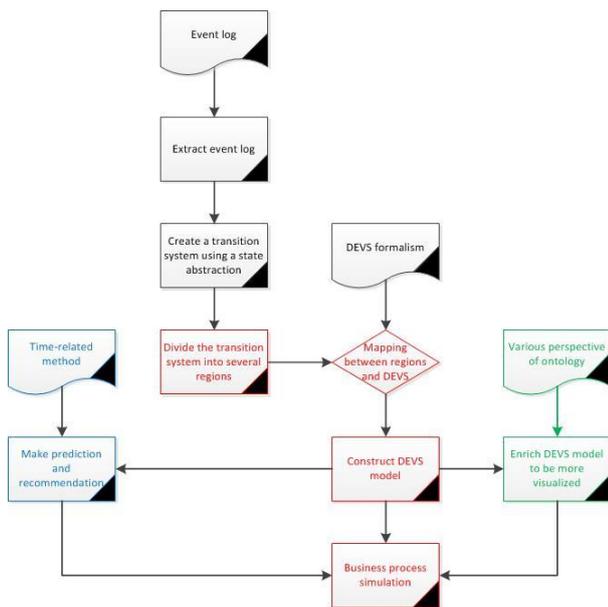


Figure 2: Process of mapping transition system to DEVS model

Figure 2 shows the main process of Modeling & Simulation by integrating DEVS model with process mining. The black part has already been done in the process mining and the red part is the main work of this paper. First we need to extract the useful information from the list of event logs. Using the state abstraction, we can automatically construct a transition system based on the event logs. The main part of this approach is to implement the transformation between transition system and DEVS model. To start to implement the region-based approach, we divide the transition system in several regions. In this case all activities can be classified into entering the region, leaving the region and non-crossing. An activity cannot be entering the region in one part of the transition system and exiting the region in some other part. The mapping in the next section makes contribution on implementing this transformation and a new DEVS model is discovered. The discovered DEVS model will be used for business process simulation. The green part is to extend the ontology of the DEVS model. It is necessary to construct the DEVS model base. As DEVS model has the characteristic of separating model simulation, each region can be mapped with one component in the model

base. It has atomic model and coupled model with special functions which we have talked in the last paragraph. These models will be used as components to support the transformation. Moreover, we need to extend various perspectives (organizational perspective, time perspective, data perspective, etc.) based on ontology in order to be more meaningful and visualized like a map. The blue part is to design a time-related approach in order to predict the desired process flow and provide a more intuitive visualization. As the time is not considered in the transition systems, it is necessary to design a method to obtain the time directly from event logs.

4.2. Mapping design

From Two-Phased Approaches, we know that it converts the transition systems to the Petri Nets. Table 1 shows the mapping between transition systems and Petri Nets. Transition systems are divided by several regions and each region is converted into Place in Petri Nets. Transitions in the transition systems are used to create transition in Petri Nets with the name of activity. Also, transition and place are connected by arc which constructs the whole structure of Petri Nets. If the activities are entering into the region, we set the arc from transition to place. If the activities are leaving the region, we set the arrow from place to transition. Only the first place has the marking which is converted from the initial state.

Table 1: Mapping between transition systems and Petri Nets

Transition Systems	Petri Nets
Region	Place
Transition	Transition
Activity	
State	Marking

Based on table 1, we design the mapping between transition systems and DEVS as illustrated in table 2. Region is transformed to atomic model in DEVS. Each atomic model has a unique label as identify. The initial state is also transformed into the first atomic model. We extract the activities from transition system and transform into the associated input and output event. As the transition in transition systems contains the name and the value of transition, we take the name as the port of each atomic model and the value of transition as the message. To avoid the name repeat, we combine the name of source and target as the name of the transition. If the message is related to “start”, we identify it as internal transition with output event. If the message is related to “finish”, we identify it as external transition with input event. The transition is visualized by arc and connects the source and target atomic model. In the future work, as the transition system don’t have the information about time, we extract the time directly from event logs and set time duration in atomic model by using time related approach. Each execution of the transition is decided by the time and such visualizations

are important to get insight into the desired process flow.

Table 2: Mapping between transition systems and DEVS

Transition system	DEVS
Region	Atomic model
Transition	Internal transition
	External transition
Activity	Input event
	Output event
State	State

4.3. Time concurrency

The transition system in figure 3 was obtained from log $\{[a, b, c, d]^3, [a, c, b, d]^2, [a, e, d]\}$. This demonstrates that region-based approach can be used to convert transition systems to DEVS by using mapping. Consider for example Region $R1 = \{[a], [a, b]\}$. All a labeled transitions in the transition system enter R1 (there is only one), all c labeled transition exit R1 (there are two such transitions), all e labeled transition exit R1 (there are only one), and all other transitions in the transition system do not cross R1. Hence, R1 is a region corresponding to atomic model M2 with input event and output event c and e. In the transition systems, there are six minimal regions and b and c are concurrent. However, in the DEVS model every atomic model is discrete and independent model. Every atomic model has several time durations with associated transition. It shows a strong concurrency because each transition is controlled by the time. For example b and c is OR-split relationship in transition systems and it means we don't know which one happen first. But in the DEVS model, M1 has time duration $t1$ and M2 has time duration $t2$. If $t1$ is bigger than $t2$, c24?complete will execute first. If $t1$ is smaller than $t2$, b13?complete will execute first. This demonstrates that each transition in DEVS is independent corresponding to the time duration with a unique label. The discrete modeling simulation can solve many complicated process discovery.

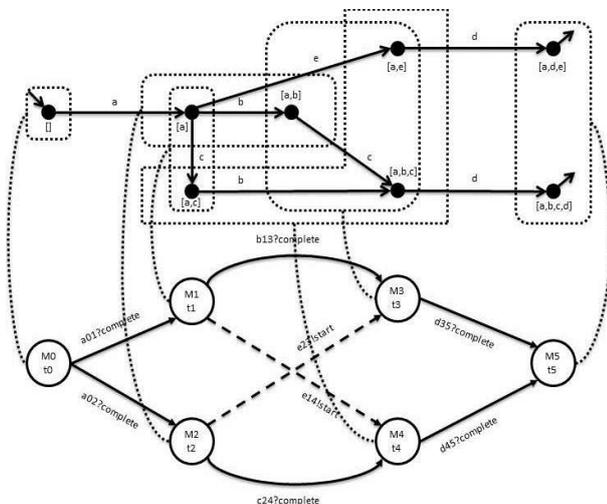


Figure 3: Transition system is converted into DEVS using mapping

5. STUDY CASE

A list of event logs is used as input for process discovery as shown in figure 4. An event can have any number of attributes. However, an extension gives semantics to particular attributes. There are four extensions in these event logs. The concept extension defines the name attribute for traces and events. The lifecycle extension defines the transition attribute for events. Time extension defines a timestamp attribute and both the data and time are recorded. The organizational extension defines a resource attribute. The resource attribute refers to the resource that triggered or executed the event. This example log in figure 4 specifies two lists of global attributes (one is hidden). Traces have one global attribute: attribute concept:name is mandatory for all traces. Events have two global attributes: attributes lifecycle:transition and concept:name are mandatory for all events. It also defines three classifiers in these event logs. Each classifier is specified by a list of attributes. There are four traces in these event logs but only one is visible. Case 1 has four events and each event has four attributes. When mining transition systems starts, the event name and transition is classified as activities and transformed into transition. Time is not considered and resource is all undefined. The information is classified and extracted by the plug-in openXES based on XES standard.

```
<?xml version="1.0" encoding="UTF-8"?>
<log xmlns="http://www.xes-standard.org/" openxes.version="1.0RC7" xes.features="nested-attributes" xes.version="1.0">
  <extension uri="http://www.xes-standard.org/lifecycle.xesext" prefix="lifecycle" name="Lifecycle"/>
  <extension uri="http://www.xes-standard.org/org.xesext" prefix="org" name="Organizational"/>
  <extension uri="http://www.xes-standard.org/time.xesext" prefix="time" name="Time"/>
  <extension uri="http://www.xes-standard.org/concept.xesext" prefix="concept" name="Concept"/>
  <extension uri="http://www.xes-standard.org/semantic.xesext" prefix="semantic" name="Semantic"/>
  <global scope="trace">
    <string value="_INVALID_" key="concept:name"/>
  </global>
  <global scope="event">
    <classifier name="MXML Legacy Classifier" keys="concept:name lifecycle:transition"/>
    <classifier name="Event Name" keys="concept:name"/>
    <classifier name="Resource" keys="org:resource"/>
    <string value="Rapid Synthesizer" key="source"/>
    <string value="exercise3.mxml" key="concept:name"/>
    <string value="standard" key="lifecycle:model"/>
  </global>
  <trace>
    <trace>
      <string value="Case1.0" key="concept:name"/>
    </trace>
    <event>
      <string value="UNDEFINED" key="org:resource"/>
      <date value="2008-12-09T08:20:01.527+01:00" key="timestamp"/>
      <string value="A" key="concept:name"/>
      <string value="complete" key="lifecycle:transition"/>
    </event>
    <event>
      <string value="UNDEFINED" key="org:resource"/>
      <date value="2008-12-09T08:22:01.527+01:00" key="timestamp"/>
      <string value="C" key="concept:name"/>
      <string value="complete" key="lifecycle:transition"/>
    </event>
    <event>
      <string value="UNDEFINED" key="org:resource"/>
      <date value="2008-12-09T08:23:01.527+01:00" key="timestamp"/>
      <string value="G" key="concept:name"/>
      <string value="complete" key="lifecycle:transition"/>
    </event>
    <event>
      <string value="UNDEFINED" key="org:resource"/>
      <date value="2008-12-09T08:23:01.527+01:00" key="timestamp"/>
      <string value="E" key="concept:name"/>
      <string value="complete" key="lifecycle:transition"/>
    </event>
  </trace>
</log>
```

Figure 4: XES document for event logs

After these event logs are extracted, we get the mined transition system as shown in figure 5. It has an initial state represented by dotted line and an acceptance state represented by double line. Each state has an identity label and each transition has the label of activity. The state and the transition are connected to provide several possible execution sequences.

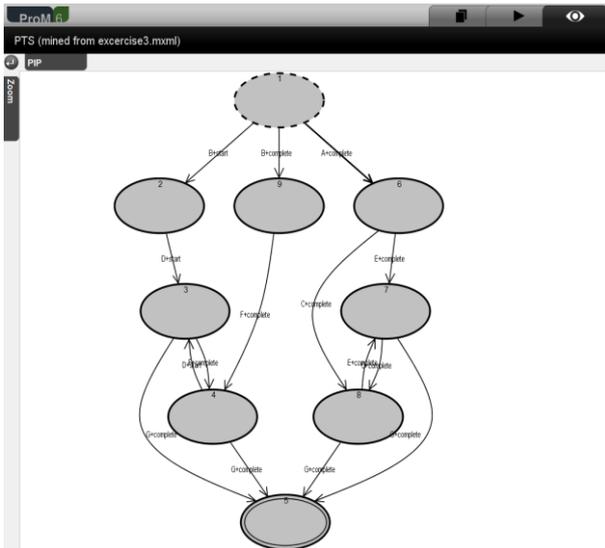


Figure 5: Mined transition system in ProM

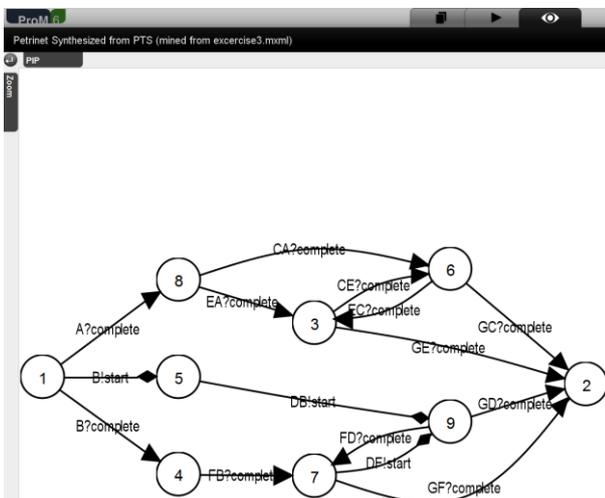


Figure 6: DEVS model is converted from transition system using region-based approach in ProM

After we get the transition system, we start to use the region-based approach to transform transition system into DEVS as shown in figure 6. We transform the components in the transition system to the related DEVS components according to the mapping. First we check if the transition system meets the forward-closure property. Transition system is divided into regions and then transformed to atomic model. Each atomic model has a unique identity number and each transition has a unique name which represents event and message. External transition which contains the message “complete” is visualized as a classical arrow and internal transition which contains the message “start” is visualized as a diamond arrow. The name of the transition is the combination of the source activity and the target activity. Each state (we call here state, the state variable phase) with a leaving internal transition is given a time life function not infinite to be defined. Each state with no internal transition has its time life function set to infinite. The initial state is converted into the atomic model with label of 1. In this study case the

time in the atomic model is not extracted from the event logs. The business process starts with the first atomic model and continues the transitions automatically until it reaches the desired state.

6. PERSPECTIVE

Region-based approach has reached the first step to integrate DEVS model in the process mining. However, it is still a big challenge to design DEVS for business process simulation. How to satisfy the customer’s requirement for simulation is the main objective. This section gives the limitation of the recent work and makes the perspective for the future work.

The DEVS model in the study case is not executed. As the time is not considered, the transition from one state to another state is not controllable. It is necessary to design a method to extract time from event log and integrate with atomic models. If every atomic model has the time duration, the execution of each transition will depend on the semantics of DEVS formalism. The time perspective is concerned with the timing and frequency of events. The presence of timestamps enables the discovery of bottlenecks, the monitoring of resource utilization, and the prediction of remaining processing times of running cases.

Until now, DEVS model in the process mining cannot be used directly for simulation. The graphical view of process model needs to be enriched with temporal information at least very basic one that are time life function values and event planning. After having enriched the model, there are two methods to run it. The first one is to implement the interoperability with other DEVS simulation engine platform like ADEVS, CD++ and so on. We need to export the DEVS model from the ProM and transform into other pattern which another platform can accept. So the discovered DEVS can be simulated in an existing platform. The second one is to implement a DEVS simulation engine inside ProM. Thanks to conformance checking, it defines that the behavior of a process model and the behavior recorded in an event log are compared to find commonalities and discrepancies. So DEVS not only can be used for business simulation, but also it can reduce the deviations from the reality. Deviations from reality include model deviating from reality and case deviating from model. Conformance checking supports deviation by providing algorithms for example token replay. Moreover, in the DEVS model, the atomic model is not explicitly presented. Different atomic models with different functions need to be visualized. A dynamic and abundant DEVS model base semantic will be constructed with the purpose of reducing the representational bias. The DEVS model base is used as a component library in order to construct complicated DEVS models. The enrichment of DEVS model will also use ontology to extend various perspectives. Discovered models may focus on different perspectives (control-flow, data flow, time, resources, costs, etc.) and show these at different levels of granularity and precision.

7. CONCLUSION

In this paper, we presented the current process discovery technique which extracts event logs to discover models. The novelty is coming from DEVS and it is specified as the resulting model for business process simulation. In comparison with Petri Nets, DEVS uses explicit time and separates model from simulation. We integrated DEVS model with process discovery to build an advanced business process model. To achieve this goal, we selected the transition systems as input after extracting event logs. A region-based DEVS transformation approach with a suitable mapping has been designed. Meanwhile, we highlight the advantage of time concurrency in DEVS model. An example is presented to show this approach by using ProM simulation platform. According to the perspective, DEVS models generated need to be improved with the time information. Then a simulation engine has to be selected and finally, we assume that the use of ontology areas will help to improve the model description. So users will be able to build almost automatically the business process model, make simulations, identify various components and supervise the operating process.

REFERENCES

- Bazoun H., Bouanan Y., Zacharewicz G., Ducq Y., Boye H., 2014. Business Process Simulation: Transformation of BPMN 2.0 to DEVS Models. Proceedings of the Symposium on Theory of Modeling & Simulation, Spring Simulation Multi-conference, No. 20.
- Bouanan Y., El Alaoui, M.B., Zacharewicz, G., and Vallespir B., 2014. Using DEVS and Cell-DEVS for modelling of information impact on individuals in social network. In *Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World*, 409-416. Springer.
- Jacques C.J.D., Wainer G.A., 2002. Using the CD++ DEVS Toolkit to Develop Petri Nets. Summer computer simulation.
- Peterson J.L., 1977. Petri Nets. *ACM computing surveys*, 9(3), 223-252.
- Robert M.K., 1976. Formal verification of parallel programs. *Communications of the ACM*, 19(7), 371-384.
- Song H.S., Kim T.G., 1994. The DEVS framework for discrete event systems control. Proceedings of the Fifth Annual Conference on Distributed Interactive Simulation Environment, 228-234. IEEE.
- Van der Aalst W.M.P., 2011. *Process mining: discovery, conformance and enhancement of business processes*. Springer Science & Business Media.
- Wainer G.A., 2009. *Discrete-event modeling and simulation: a practitioner's approach*. New York: CRC Press.
- Zacharewicz G., Frydman C., Giambiasi N., 2008. G-DEVS/HLA Environment for Distributed Simulations of Workflows. Society for computer simulation international, San Diego, CA, USA, 84 (5), 197-213.
- Zeigler B.P., Praehofer H., and Kim T.G., 2000. N-dimensional Cell-DEVS models. *Discrete Event Dynamic Systems*, 12(2), 135-157.
- Zeigler B.P., Praehofer H., and Kim T.G., 2000. *Theory of Modeling and Simulation*. Academic Press: New York, USA.
- Zeigler B.P., Sarjoughian H.S., 2005. *Introduction to DEVS Modeling and Simulation with JAVA: Developing Component-Based Simulation Models*. Technical Document, University of Arizona, 129-147.

AN ACCREDITED ENGINEERING CURRICULUM FOR MODELING AND SIMULATION

Frederic D. McKenzie

Modeling, Simulation and Visualization Engineering Department
Old Dominion University
Norfolk, VA U.S.A.

rdmckenz@odu.edu

ABSTRACT

The Accreditation Board for Engineering and Technology (ABET) is an internationally known organization for recognizing engineering programs that work towards student attainment in learning outcomes that conform to engineering disciplines and have a process for measuring attainment and continuous improvement. There are eight general criteria to be satisfied that include student outcomes, educational objectives, curriculum, facilities and faculty. Additionally, there may be program-specific requirements depending on the particular discipline. The modeling and simulation (M&S) engineering (M&SE) program at Old Dominion University satisfies the program criteria for general engineering in the ABET Engineering Accreditation Commission and declares detailed discipline-specific student outcomes to ensure unambiguous definition of a pioneering program in order to set a standard that can be utilized as a template for future up-and-coming M&SE programs. This paper presents the curriculum and program-specific outcomes that have been incorporated into our continuous improvement process for student attainment of our M&S engineering program outcomes.

Keywords: modeling and simulation, simulation education, engineering accreditation, student outcomes

1. INTRODUCTION

Modeling and Simulation (M&S) is now emerging as a recognized discipline much like the emergence of computer science in the early 1960s with one of the first departments at the University of North Carolina at Chapel Hill founded by Dr. Frederick P. Brooks Jr. Similarly, it would take decades before the emergence of Computer Engineering. Now-a-days, Computer Science and Computer Engineering departments are ubiquitous around the world. We believe M&S is on the same path and, therefore, have started the Department of Modeling, Simulation and Visualization Engineering (MSVE) in 2010 to administer the B.S. degree in M&S Engineering (M&SE). This effort was built upon a successful 12 year interdisciplinary program at the graduate (M.S. and Ph.D.) levels.

2. HOW DO WE DEFINE M&S

M&S is a discipline focused on advancing and using the theories and practices of selecting appropriate modeling techniques, creating associated models, executing models dynamically over time, utilizing visualizations for V&V, and evaluating a range of possible solutions through analytical techniques. Then, applying this knowledge in many domains.

The first step is in creating a model of an existing or perceived system utilizing the appropriate model for the system of interest. Just this step alone can depend upon several theoretical approaches and depends heavily on the objectives of the model but in the end the model will be an abstract representation of the target system. The next step is then simulating the system using the developed model. Computer simulation is an applied methodology in which the behavior of complex systems is described using a model (mathematical or symbolic and often stochastic) and executed on a digital computer. Components of the model are updated in an appropriate manner (in serial, parallel or distributively) and with regards to time and causality. Visualization goes hand-in-hand with simulation and so we would want to visualize the simulation/model to promote understanding and verification & validation (V&V). Scientific visualization of data, model performance and results allows verification and face validation of the model. Finally, possibly the most important step is statistical validation and analysis of the solution space generated by the simulation to be able to choose the best solution and take the best decision. Since simulation allows easy changes to input considerations which results in differing solutions, one must know how to compare solutions and determine which decision is best.

We believe this broad approach to engineering instills the methods and tools supportive of innovation and stresses problem-solving skills applicable to all engineering and science disciplines. After all, engineering is about problem-solving and M&S engineering is about finding and understanding a range

of solutions to provide the best solution over a possibly unforeseen range of options.

3. M&SE CURRICULUM

The M&SE Program at ODU has a curriculum that provides students with strong coverage of the basic sciences, mathematics, and general education as identified in ABET EAC curriculum criteria requirements. Additionally, the curriculum prepares students for professional careers and further study in the M&SE discipline with a core set of engineering topics courses that focus on the development of a variety of models and simulations in software and with the utilization of simulation tools. The curriculum also culminates in a major design experience based on knowledge and skills acquired in the different levels of the curriculum. In the Appendix Tables A-1 and A-2 we show the full curriculum. Below we describe only the M&SE core courses [odu.edu/msve].

MSIM 201. Introduction to Modeling and Simulation Engineering.

This is the first course for Modeling and Simulation Engineering (M&SE) students. M&SE discipline is surveyed at an overview level of detail. Topics include basic definitions, M&S paradigms and methodologies, applications, design processes, and human factors. Information literacy and research methods are addressed. Papers and oral presentations are required and allow the student to investigate different aspects of the discipline. The course provides a general conceptual framework for further M&SE studies.

MSIM 205. Discrete Event Simulation.

An introduction to the fundamentals of modeling and simulating discrete-state, event-driven systems. Topics include basic simulation concepts and terms, queuing theory models for discrete event systems, structure of discrete event simulations, problem formulation and specification, input data representation, output data analysis, verification and validation, and the design of simulation experiments.

MSIM 281. Discrete Event Simulation Laboratory.

A laboratory course designed to provide a hands-on introduction to the development and application of discrete event simulation. Topics include an introduction to one or more discrete event simulation tools, common modeling constructs, data gathering and input data modeling, design of simulation experiments, output data analysis, and verification and validation. The design and implementation of a

series of increasingly complex simulations of various discrete event systems are conducted.

MSIM 320. Continuous Simulation.

An introduction to the fundamentals of modeling and simulating continuous-state, time-driven systems. Topics include differential equation representation of systems, formulation of state variable equations, numerical integration, and techniques for numerical solution of differential equations including the Taylor algorithm and the methods of Runge-Kutta. Application domains considered include physical and biological systems.

MSIM 331. Simulation Software Design.

Introduction to data structures, algorithms, programming methodologies, and software architectures in support of computer simulation. Topics include lists, queues, sets, trees, searching, sorting, reusable code, and order of complexity. Simulation structures developed include event lists, time management, and queuing models. Software models are implemented and tested.

MSIM 382. Continuous Simulation Laboratory.

A laboratory course designed to provide a hands-on introduction to the development and application of continuous simulation. Topics include an introduction to one or more continuous simulation tools, modeling of various physics-based systems, and numerical solution of differential equations. The design and implementation of a series of increasingly complex simulations of various continuous systems are conducted. Written communication skills are stressed; weekly writing assignments are required.

MSIM 383. Simulation Software Design Laboratory.

A laboratory course designed to provide a hands-on introduction to the development of simulation software. Topics include data structures, algorithms, and simulation executives. The students will conclude with the development of a basic simulation executive capable of managing discrete event simulations. Written communication skills are stressed; writing is required for each laboratory assignment.

MSIM 410/510. Model Engineering.

The goal of this course is to develop understanding of the various modeling paradigms appropriate for capturing system behavior and conducting digital computer simulation of many types of systems. The

techniques and concepts discussed typically include UML, concept graphs, Bayesian nets, Markov models, Petri nets, system dynamics, Bond graphs, etc. Students will report on a particular technique and team to implement a chosen system model.

MSIM 441/541. Computer Graphics and Visualization.

The course provides a practical treatment of computer graphics and visualization with emphasis on modeling and simulation applications. It covers computer graphics fundamentals, visualization principles, and software architecture for visualization in modeling and simulation.

MSIM 451/551. Analysis for Modeling and Simulation.

An introduction to analysis techniques appropriate to the conduct of modeling and simulation studies. Topics include input modeling, random number generation, output analysis, variance reduction techniques, and experimental design. In addition, techniques for verification & validation are introduced. Course concepts are applied to real systems and data.

MSIM 487W. Capstone Design I.

Part one of the senior capstone design experience for modeling and simulation engineering majors. Lectures focus on providing professional orientation and exploration of the M&S design process. Written communication, oral communication and information literacy skills are stressed. Individual and group design projects focus on the conduct of a complete M&S project. Industry-sponsored projects are an option. Individual and team reports and oral presentations are required.

MSIM 488. Capstone Design II.

Part two of the senior capstone design experience for modeling and simulation engineering majors. Lectures focus on providing professional orientation and exploration of the M&S design process. Written communication, oral communication and information literacy skills are stressed. Individual and group design projects focus on the conduct of a complete M&S project. Industry-sponsored projects are an option. Individual and team reports and oral presentations are required.

4. PROGRAM SPECIFIC OUTCOMES

Currently, the M&SE Program is evaluated under the ABET General Criteria only. Therefore, there are no specific ABET EAC Program Criteria for the M&SE Program. However, the MSVE Faculty took great care in developing a curriculum that would satisfy both ABET and possible future discipline-specific criteria for modeling and simulation engineers. Along with this curriculum was developed several discipline-specific student outcomes. These outcomes were assessed as to their level of attainment by our students along with the eleven general engineering student outcomes which can be found on the ABET website (www.abet.org). The eleven general engineering outcomes will not be discussed here but rather the discipline-specific outcomes for our M&SE Program are discussed below.

4.1. M&S Capabilities

Students graduating with a B.S. in M&SE are expected to be able to model a variety of systems from different domains using appropriate techniques and tools and communicate those designs to technical peers and non-technical managers or customers. Where available tools are lacking, students should be able to develop needed capabilities in software. Utilizing the experimental process, the M&S student should be able to generate and analyze simulations results after verifying and validating their models and simulations in order to reach an appropriate conclusion. These processes should be supported judiciously where possible with visualization. With these desired capabilities in mind, program-specific outcomes were created above and beyond the eleven ABET general outcomes.

4.2. M&SE Program-Specific Outcomes

Initially, there were 9 different program specific outcomes and those 9 outcomes were assessed as to the program's effectiveness in achieving those outcomes. Homework and test problems that were directly attributable to learning a particular outcome were assessed by the course instructor utilizing a rubric as to whether the work was in the (1) Unacceptable, (2) Marginal, (3) Good, or (4) Excellent category. Then, the results for all students in the class (and likely other classes) were averaged to determine a Student Outcome Category Average (SOCA) score. Our target for all program-specific outcomes was to have a SOCA quantified score of 3.2 out of 4.0 with an average of 75% of students in the Good or Excellent range over all courses assessed for a particular SO, which are reasonably high standards. Our assessment showed that only one program-specific outcome failed to meet either of the two targets namely, "An ability to select and apply appropriate simulation techniques and tools." The statistic for this outcome was a SOCA score of 3.13 with 70.8% at the Good or Excellent range. To address this slight underperformance, the MSVE Assessment Committee recommended changes to MSIM 320 that

included an increase in the number of physical systems models and examples.

Another recommendation that came out of our assessment process was to reduce the high number of program-specific outcomes since many were able to be incorporated under the general ABET criteria outcomes. After careful consideration of desired capabilities and capabilities already contained within the ABET general engineering student outcomes [McKenzie 2015], the MSVE faculty combined and reduced the number of program-specific outcomes and developed the three M&SE program specific outcomes identified below.

1. An ability to model systems across different domains
2. An ability to design and develop appropriate simulation solutions
3. An ability to apply visualization techniques to support the simulation process

The attainment of these newly established outcomes will be measured against a rubric created for each outcome as before using student learning measures such as homework, reports, test questions, etc.

5. CONCLUSIONS

Modeling and simulation is now becoming a recognized discipline with a broad engineering appeal that can be harnessed to find creative solutions to a wide variety of problems. Underlying the techniques and tools is a foundation in mathematical and statistical theories. Such theories bridge the gap between science and engineering and enable the exploration of solution spaces for problems that may be too complex, too costly, or too dangerous to explore by utilizing the real systems themselves. We have presented an educational curriculum to provide students with a well-rounded foundation to apply such methodologies to virtually any domain. We believe this curriculum has shown its viability through its successful ABET accreditation as an engineering program.

ACKNOWLEDGMENTS

MSVE faculty and staff were instrumental in developing the successful department, curriculum, and overall program.

REFERENCES

McKenzie, F. "Preparing for the EAC-ABET Visit for a Novel Undergraduate Program in Modeling and Simulation Engineering." Proceedings of the 13th Latin American and Caribbean Consortium of Engineering Institutions (LACCEI) Conference, Santo Domingo, Dominican Republic, July 29-31, 2015.

ODU.EDU/MSVE,
<http://catalog.odu.edu/courses/msim/>, accessed May 19, 2015.

AUTHORS BIOGRAPHY

Frederic (Rick) D. McKenzie, Ph.D. is Professor and Chair of the Modeling, Simulation and Visualization Engineering (MSVE) Department and joint member of the Electrical and Computer Engineering Department at Old Dominion University where he currently serves as Principal Investigator (PI) and Co-PI on projects involving software architectures for simulation, behavior representation in simulations, and medical modeling and simulation. Dr. McKenzie received his Ph.D. in computer engineering from the University of Central Florida in 1994. Prior to joining ODU, he held a senior scientist position at Science Applications International Corporation (SAIC), serving as Principal Investigator for several distributed simulation projects. At SAIC he was a Team Lead on a large distributed simulation system. Before joining SAIC, Dr. McKenzie worked as a student researcher on research projects involving both NASA Kennedy Space Center and NASA Marshall Space Flight Center. He has several years of research and development experience in the software and artificial intelligence fields, including object-oriented design and knowledge-based systems. Both his M.S. and Ph.D. work have been in artificial intelligence - focusing on knowledge representation and model-based diagnostic reasoning.

APPENDIX

**Table A-1. M&SE Curriculum Sheet with Pre-Co-requisites
(1/2)**

Freshman Year – First Semester					
Course	Grade	Cr.	Sem.	Comments	Pre-/Co-Requirement
MATH 211 – Calculus I		4			P-Placement Exam
ENGL 110C – English Composition I*		3			P-Placement Exam (Not satisfied by the associate's degree)
CHEM 121N – Foundations of Chemistry I Lecture		3			P-MATH 102M ^c
CHEM 122N – Foundations of Chemistry I Lab		1			C-CHEM 121N
ENGN 110 – Engineering and Technology I		2			C-MATH 162M
General Education (Oral Communication): COMM 101R		3			
Freshman Year – Second Semester					
MATH 212 – Calculus II		4			P-MATH 211 ^c
CHEM 123N – Foundations of Chemistry II Lecture		3			P-CHEM 121N ^c
CS 150 – Programming I		4			P-MATH 102M
PHYS 231N – University Physics I		4			C-MATH 211
MSIM 111 – Information Literacy and Research for Modeling and Simulation Engineers		2			P- ENGN 110
Sophomore Year – First Semester					
Course	Grade	Cr.	Sem.	Comments	Pre-/Co-Requirement
STAT 330 – Probability and Statistics		3			P-MATH 211 ^c
PHYS 232N – University Physics II		4			P-PHYS 231N
CS 250 – Programming II		4			P- MATH 162M, CS 150 ^c , C-CS 252
CS 252 – Introduction to Unix		1			P – CS 150 ^c , C-CS 250
MSIM 201 – Introduction to M&S		3			P/C-MATH 163 & CS 150
Sophomore Year – Second Semester					
MATH 307 – Differential Equations		3			P-MATH 212 ^c
ENGL 231C – Technical Writing*		3			P-ENGL 110C ^c (Not satisfied by the associate's degree)
General Education "A" Course (Human Creativity)		3			
General Education "L" Course (Literature)		3			
MSIM 205 – Discrete Event Simulation		3			P-MSIM 201, P/C-STAT 330
MSIM 281 – Discrete Event Simulation Lab		1			C-MSIM 205

^c Requires a grade of "C" or better in pre-requisite course in order to register

**Table A-2. M&SE Curriculum Sheet with Pre-Co-requisites
(2/2)**

Junior Year – First Semester						
Course	Grade	Cr.	Sem.	Comments	Pre-/Co-Requisite	
CS 330 – Object-Oriented Program & Design		3			P-MATH 163, CS 250 ^c & CS 252	
CS 381 – Discrete Structures		3			P-MATH 163 & CS 150 ^c	
MSIM 320 – Continuous Simulation		3			P – MSIM 201; P/C – MATH 307 (280) & PHYS 227N/232N	
MSIM 382 – Continuous Simulation Lab		1			C-MSIM 320	
General Education “S” Course (Human Behavior)		3				
Elective – Approved Program Elective		3				
Junior Year – Second Semester						
MSIM 331 – Simulation Software Design		3			P-CS 330 & CS 381 & MSIM 205; C-MSIM 383	
MSIM 383 – Simulation Software Design Lab		1			C – MSIM 331	
MSIM 451 – Analysis for M&S		3			P – MSIM 205 & STAT 330	
MSIM 410 – Model Engineering		3			P – MSIM 205, P/C – MSIM 320	
General Education “H” Course (Interpreting the Past)		3				
Upper Division General Education – Option D Course 1		3				
Senior Year – First Semester						
Course	Grade	Cr.	Sem.	Comments	Pre-/Co-Requisite	
MSIM 441 – Computer Graphics & Visualization		3			P – CS 250	
MSIM 487W – Capstone Design I*		4			P – MSIM 331, 410, 451; “C” or better in ENGL 231C	
Upper Division General Education – Option D Course 2		3				
ENMA 401 – Project Management		3			C-Junior Standing	
MSIM 4xx – Approved MSIM Elective 1		3				
Senior Year – Second Semester						
ENMA 480 – Engineering Ethics (Satisfies Philosophy & Ethics Gen. Ed. Req.)		3			C-Junior Standing	
MSIM 488 – Capstone Design II		3			P-MSIM 487W and MSIM 441	
Elective – Approved Program Elective		3				
MSIM 4yy – Approved MSIM Elective 2		3			P/C – MSIM 487W	
General Education “T” Course (Impact of Technology)		3			Not necessarily met by the Associate’s Degree	

^c Requires a grade of “C” or better in pre-requisite course in order to register

*Grade of “C” or better (or “TP”/“DP”/“MP”/“XP”/“DN”) required in ENGL 110C, ENGL 231C and a “C” or better in MSIM 487W in order to satisfy University Writing Requirements for graduation.

DEVS-BASED FRAMEWORK FOR MESSAGE DISSEMINATION IN MULTI-LAYER NETWORKS

Youssef Bouanan, Mathilde Forestier, Judicael Ribault, Gregory Zacharewicz, Bruno Vallespir

Univ. Bordeaux, IMS, UMR 5218, F-33400 Talence, France
CNRS, IMS, UMR 5218, F-33400 Talence, France

youssef.bouanan@ims-bordeaux.fr, mathilde.forestier@gmail.com, jribault@gmail.com, gregory.zacharewicz@ims-bordeaux.fr, bruno.vallespir@ims-bordeaux.fr

ABSTRACT

Many research efforts focus on the definition of models and techniques to simulate and predict the reaction of individuals face to an information. In this context, great benefits could derive from the exploitation of individuals' personality and cultural values in the diffusion model process. In this paper we describe a new architecture for agent-based model using the DEVS (Discrete Event System Specification) framework and show how this architecture is flexible and can serve to simulate the dissemination process. In more details, we define a set of models of individuals characterized by a set of state variables to represent the behavior of an individual and the mesh between the individuals within a multi-layer social network. Then, we introduce the platform architecture, specifically designed to simulate message propagation in a multi-layer network. In the end, a military scenario of message diffusion during a stabilization phase is used to test our DEVS models on the platform and the relevancy of the simulation results.

Keywords: Social Simulation, DEVS Formalism, Social Network Diffusion Process

1. INTRODUCTION

Nowadays, many researchers are interested in developing new and more efficient systems for social simulation. The issues explored include psychology, organizational behavior, sociology, political science, economics, anthropology, geography, engineering, archaeology and linguistics (takahashie et al, 2007; Bouanan et al, 2014). In the field of military defense, modeling the human operators in a system is complex because it has to represent the behaviors in critical situations. So, military simulation systems have to support detailed analysis of individual and organizational performance. The SICOMORES project seeks to create a given ethno-cultural community in an artificial society framework. It aims to formalize and simulate the propagation of the psychological effects in a multi-layer social network.

Social networks play an important role in studying the propagation of information, innovation, ideas, and influence among its members. When information appears - for example, the use of cell phones among

students, the adoption of a new information system within the enterprise, or the rise of a political movement in an unstable society: it can either die out quickly or makes significant inroads into a population. Network diffusion process allows us to understand the dynamics and the propagation of information in social network. These networks can consist of individuals, group of person or organizations. The interactions could be done by physical contact, collaboration, religious meeting or some forms of verbal or written communication depending on the situations. This network diffusion process is used to raise numerous research issues such as contamination minimization problem or influence maximization problem (IMP).

Agent-based modeling approach is used to model complex systems composed of interacting and autonomous 'agents'. Agents have behaviors, often described by simple rules and interactions with other agents, which in turn influence their behaviors. By modeling agents individually, the full effects of the diversity that exists among agents in their attributes and behaviors can be observed as it gives rise to the behavior of the system as a whole.

This paper begins by representing the agent-based model approach, The DEVS formalism and the dissemination process in social networks. Then, it provides model of individuals with DEVS characterized by a set of attributes and it presents our architecture to simulate the diffusion process in a multi-layer network using DEVS. At last, the final part concerns our experiments and the conclusion.

2. BACKGROUND AND RELATED WORK

2.1. Agent-Based Social Simulation

Agent-based social simulation (ABSS) consists of social simulations that are based on agent-based modeling, and implemented using artificial agent technologies (Davidsson 2002). ABSS is a scientific discipline concerned with simulation of social phenomenon, using computer-based multi-agent models. In these simulations, people or group of people are represented by agents. Different agent-based platforms have been used for implementing agent-based social simulation (Tobias and Hofmann 2004). NetLogo

is the highest-level platform, providing a simple yet powerful programming language, built-in graphical interfaces, and comprehensive documentation. MASON, Repast, and Swarm are "framework and library" platforms, providing a conceptual framework for organizing and designing Agent-Based Models (ABMs) and corresponding software libraries.

Most of the simulation frameworks are oriented towards applications in the field of artificial intelligence and are not sufficiently dedicated for social scientific applications. In our study we develop a new library based on DEVS formalism in which agents can be modeled as 'free' and complex objects that represent such as institutions as social scientific models do.

2.2. DEVS Formalism

The DEVS formalism for modeling and simulation (Zeigler et al, 2000) is based on discrete events modeling. It tends to represent more mathematical oriented notations consisting of sets and functions in algebraic representation. The sets in DEVS formalism specify the potential states, inputs, and outputs of the model; the functions describe the transitions either from outside input or from the time expiration of the state. The DEVS formalism provides a framework with mathematical concepts based on the sets theory and systems theory to describe the structure and the behavior of a system. With DEVS, there is an explicit separation between a model and its simulator. Once a model is defined, it is used to build a simulator (i.e. a device able to execute the model instructions). Two kinds of models are proposed: the atomic models, which describe behavior, and the coupled models, which describe a hierarchy.

The DEVS formalism supports an open approach to formalism extension, allowing the researcher to explore new extended or specialized formalism (Zeigler and Vahie 1993). These extensions facilitate the development of models for various applications in many different domains such as biology, engineering, and sociology. For example, Barros proposed the dynamic structure DEVS (DSDEVS) formalism which allows changes in model structure during execution (Barros 1996). Chow and Zeigler proposed the parallel DEVS (P-DEVS) for parallel execution benefits (Show and Zeigler 1996). From a network modelling perspective, Uhrmacher proposed Multi-Level-DEVS (ml-DEVS) which supports an explicit description of macro and micro level (Uhrmacher et al. 2007) and Wainer proposed the cell-DEVS formalism which is a combination of cellular automata and DEVS that allows the implementation of cellular models with timing delays (Wainer et al. 2002).

There are a large number of DEVS based simulators. It is difficult to compare the performance or the implementation of these tools. The DEVS standardization group lists on his web site the most used DEVS tools known by the DEVS community (DEVS 2015). Among these platforms, the proposed DEVS models in this study is implemented in the VLE (Virtual

Laboratory Environment) because it supports multi-modeling simulation and analysis by using recent developments in the theory of modeling and simulation proposed by Zeigler. It is also possible to perform statistical analysis of results thanks to a plug-in that allows communication between VLE and R (Quesnel et al, 2007).

2.3. Diffusion process in social network

Social network consists of interconnected individuals linked by informal patterned flows of information and communication that are described as social ties (Tenkasi and Chesmore 2003). Some modern examples of social networks include online social networks, where vertices are user accounts and edges represent a relationship between accounts (e.g., friendship, coworkers), and communications networks, where vertices represent e-mail addresses or phone numbers, and edges represent e-mails sent or phone calls between vertices. Social networks play an important role in studying the propagation of information, innovation, ideas, and influence among its members. An idea will appear - for example, the use of cell phones among students, the adoption of a new information system within the enterprise - and it can either die out quickly or make significant inroads into the population.

The network diffusion processes have a long history in social sciences (Rogers 1962). With the advent of sufficient storage and computational power, this network diffusion process became an emerging research area in computer science (Domingos 2005). Propagation models are designed to reproduce phenomena observed in social networks with applications in viral marketing, spread of disease and diffusion of ideas and innovations. Most models proposed recently are extensions from the independent cascade (IC) (Goldenberg et al, 2001) and the linear threshold models (LT) (Granovetter 1978). The two models characterize two different aspects of social interaction. The IC model focuses on individual (and independent) interaction and influence among friends in a social network. The IC models can also be identified with the so-called Susceptible/Infective/Recovered (SIR) model for the spread of disease in a network (Bailey 1975). The LT model focuses on the threshold behavior in influence propagation, which we can frequently relate to; when enough of our friends bought a new phone, played a new computer game, or used new online social networks, we may be converted to follow the same action.

2.4. Human behaviour

Human behaviour modelling as individuals in groups and in societies is the subject of several fields of research: social science, economics, epidemiology and military service because it has such an important role in many aspects of daily life. Scientific literature abounds in heterogeneous and highly specialized, theoretically founded concepts of human cognition, emotion and other behaviour aspects. A few related works have

provided DEVS models of human behaviour that we use with slight modifications: (Seck et al. 2005) present a DEVS based framework for the modelling and simulation of human behaviour with the influence of stress and fatigue; (Faucher et al. 2012) proposed a first approach using G-DEVS formalism for Civil-Military Cooperation actions (CIMIC) and Psychological actions (PSYOPS), which are actions of influence that take precedence over combat.

3. CONTRIBUTION

In this section, we propose an agent-based model for the information diffusion in a Multi-layer network.

3.1. Problem statement

Several research studies attempted to model and simulate the influence diffusion in a social network. However, the focus is on one layer network (one link or relationship between persons) e.g., the message disseminates only within friends. In addition, they did not consider the individual and cultural factors on diffusion process; the message disseminates in the same way between two friends or two coworkers. Actually, the way how people communicate is highly dependent of the person who people are talking to. In the frame of SICOMORES project, we aim to provide solutions to artificially generate multi-layer social network of realistic population and simulate the effects of information on population with a propagation algorithm of the effects across networks. The challenge is to improve the realism of socially intelligent agents and take into account the impact of individual's personality and cultural values in the diffusion model process. The SICOMORES system can be useful in the domain of military simulation for training and education.

3.2. DEVS-based Social Agent Model Framework

There are several methods to study the information propagation in social network, such as complex network analysis (Ma et al, 2013), cellular automata

(Goldenberg 2001) and agent based modeling (Smith 2007). Based on our needs, we use agent-based models and simulations for this study. Computational models and simulations, especially agent based ones, have been widely used to study a variety of social, organizational and natural phenomena. Agent-based models are capable of simulating macro-level structures resulting from micro-level interactions of heterogeneous agents within complex systems.

We use a low-level language (DEVS) to define a set of models of individuals characterized by a set of state variables to construct their behaviors and the mesh between the individuals within an MSN. Then, we develop an architecture and a methodology that promotes modularity and reusability of these models as a package. We implement these models in a virtual laboratory based on DEVS formalism (VLE). This simulator has been tested and validated with several scenarios (Quesnel et al. 2009).

In the context of this study, we need modularity and reusability. We have to perform the simulations by varying the inputs but also the algorithms of the agents. For example, diffusion algorithms may change depending on the targeted population, the type of message or network on which the message travels. In order not to have to fully develop a model for each combination of algorithms we want to set in the agent, we decided to opt for a modular approach. This approach needs some methodology to be carrying through. Section 3.2.1 presents the architecture of simulation and methodology used. Section 3.2.2 describes the proposed solution in terms of modular and reusable model design.

3.2.1. Architecture and methodology

Figure 1 presents the simulation architecture we have developed.

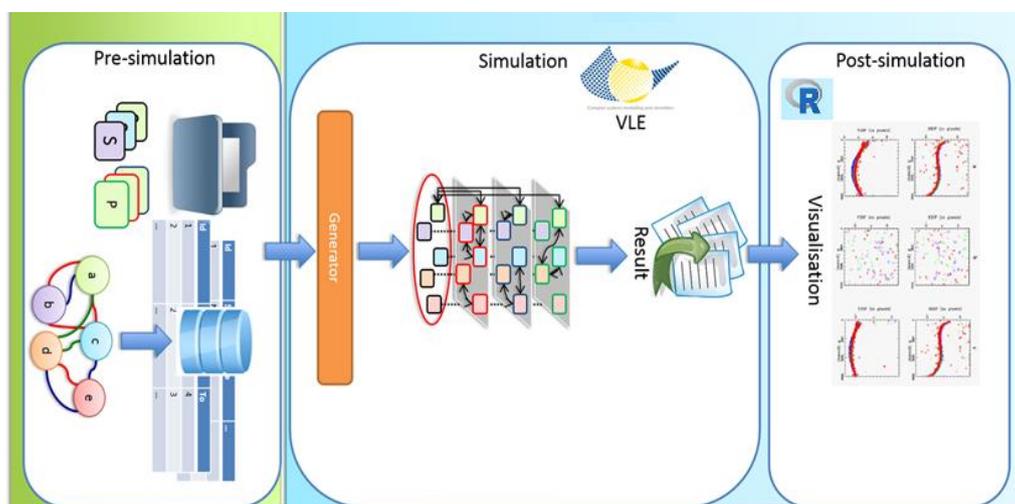


Figure 1: Agent simulation architecture

Pre-simulation: Before making a simulation, we build a new experiment and develop the simulation models we need if they do not already exist in the repository. An experiment includes one or several groups. A group includes many individuals. A graph represents all relations between individuals. We opted for the storage of information in a database so that each experiment is easily accessible and re-playable. The repository contains all the models available to execute the simulations, i.e., all the servers and proxies that will be used by the generator to produce the simulation model.

Simulation: The simulation starts with the experiment to simulate, as well as servers and proxies to use. In this

way, for the same experiment, we can test different behavior algorithms. The generator connects to the database to retrieve all the information from the experience to execute. The generator instantiates agents and related links. The DEVS simulation is executed using VLE and produces a result set in the form of a file.

Post-simulation: The R script post-processes the simulation result. The result file is used to visualize the course of the simulation and to develop an analysis and conclusion. The analysis can lead to a new simulation, i.e., makes a new pre-simulation, simulation and post-simulation.

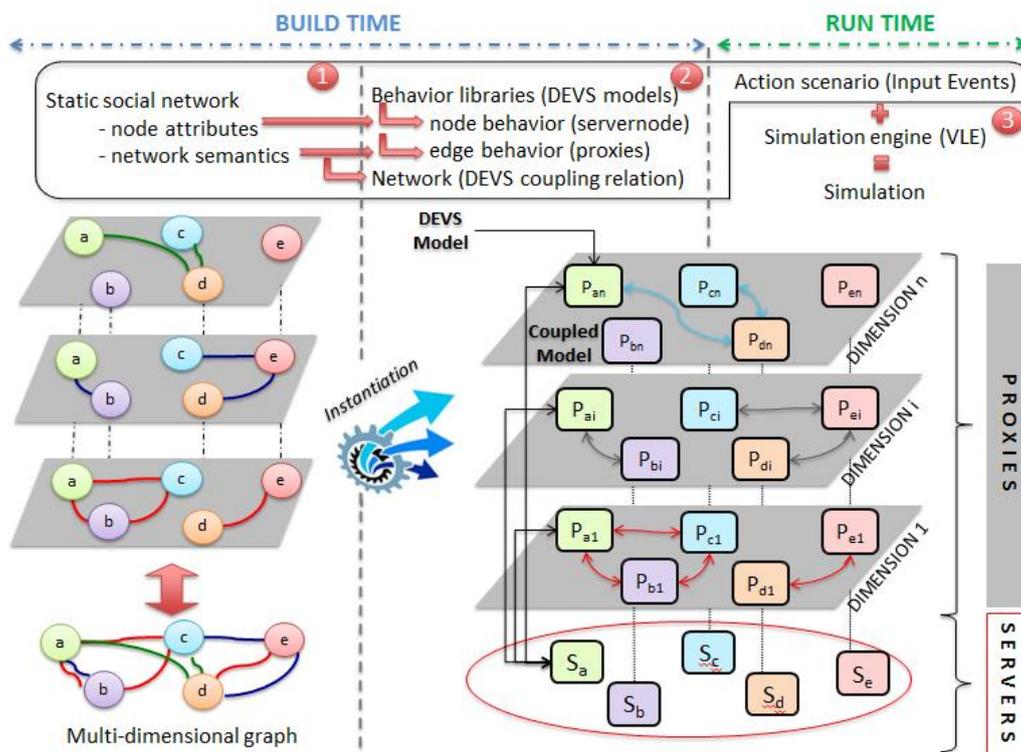


Figure 2: Conceptual overview

3.2.2. DEVS model

Figure 2 presents the solution we have developed to take advantage of a low-level framework allowing the development of algorithms that take into account the specificity of each individual. This solution also considers our needs for modularity and reusability.

Our simulation is based on a static multi-layer graph (1). Specifically, we generate a Multi-layer Social Network (MSN). Our MSN models individuals with attributes (such as gender, age and opinion) and several relations linking them (such as family relation or friendship relation). The idea behind the use of an MSN is to assume that people do not communicate in the same way according to people with whom they are talking, e.g., families or friends. In the second phase of the build run (2), we set the architecture based on a proxy/server framework to model the several layers of relations based on DEVS specifications. Each node of

the MSN is modeled by a server-node and has as much proxies that there are layers.

The node a in the multi-layer graph becomes the server-node S_a in phase (2) and detains P_{a0} to P_{an} proxies. The sever-node contains the individual behavior, i.e., how an individual receives an information and how this information can change his state. The proxies associated to a sever-node contain the rules of message dissemination for each layer. Finally, the phase (3) corresponds to the runtime: using an action scenario of simulation, and a simulation engine we simulate the propagation of information in the multi-layer social network.

Each agent is created at run-time by a generator that has access to the experiment data. This generator initializes each individual, i.e., each server and proxy, and binds the server with all of its proxies. Then, each proxy is

bounded to several proxies depending on the graph between individuals.

3.3. Formalization of human behaviour

In the agent-based model, individuals or group of individuals are represented as agents. Each agent is described by a set of attributes:

- Static attributes: gender, social status, religion, age class, ethnicity, leadership and language.
- Dynamic attributes (variables): opinion, interest, un/satisfied-needs.

Static attributes are intrinsic or unchanged parameters, i.e., time has no effect on them. Dynamic attributes evolve with time or events. For example, individuals can be reached or not by the information depending on its opinion and the social network configuration.

We use DEVS specifications to describe the human behavior. As we presented in section 1.1, DEVS is a well-defined formalism which has numerous advantages over other formalism in the modeling of complex dynamic system. The following notations describe the server model. We do not describe detailed operations of this model; we just explain the role and interface of the server model (Server-Node).

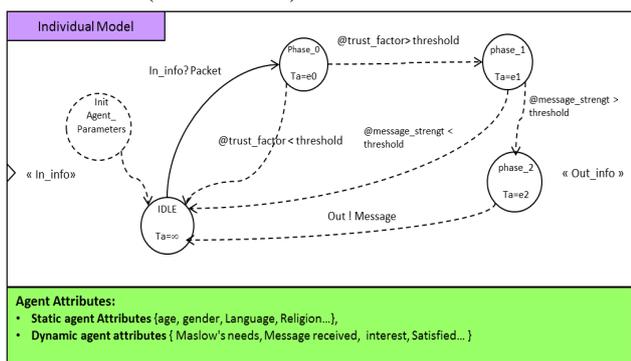


Figure 3: DEVS specification of individual model

Figure 3 describes the message influence on the individual behavior and potentially its dissemination using the DEVS specifications. The first phase is used to configure and initialize the agent's attributes. Then, when the agent is in the $\delta IDLE\delta$ phase and if it receives an external event from another agent on port $\delta In_info\delta$ ($In_info? Packet$), it goes into phase $\delta phase_0\delta$. Here, the agent calculates the trust factor (i,j) between him and the sender depending on the sender's religion, language and age class. If the trust factor is higher than a predefined threshold, the receiver goes into phase $\delta phase_1\delta$ else it returns to $\delta IDLE\delta$. Then, if the message strength is still strong enough, the receiver goes into phase $\delta phase_2\delta$ else it returns to $\delta IDLE\delta$. This message creates an impact on the individual, and eventually its behavior depending on the agent's opinion and the relationship between him and the sender. After a period, *processing-time*, the receiver transmits the message on its ego-network. After the contact between receiver and sender, the receiver's

variables (opinion, interest, satisfaction) change according to the message content, the sender and cultural factors.

3.4. Diffusion process in multi-layer social network

The multi-layer social network is defined as a graph $G=(V,E,L)$ where V is a set of nodes (DEVS model), L is a set of labels (also called layers), E is a set of labeled edges (DEVS coupling relation), i.e., the set of triples (u,v,d) where $u,v \in V$ are nodes and $d \in L$ is a label.

Each node represents an individual with characteristics who can share information on one of its networks L . Each individual who receives a message and does not have enough interest in it becomes "passive" and the message do no spread longer.

In this regard, the state of the nodes can belong to one of these three categories: (type A) informed node: the node has received a message and spread it to its neighbors; (type B) uninformed node: the node has not received any message, (type C) passive node: the node has received a message but does not spread it.

Four conditions can cause the end of the simulation: (1) The individual who receives the information is a type B; (2) The strength of the message to be propagated falls below a given threshold; (3) All nodes are a type C; (4) It has been a long time (higher than a given threshold) since the scenario occurred.

4. EXPERIMENT

4.1. Overview

We illustrate the concepts presented in the paper by modeling and simulating the spread of information in a multi-layer social network. The context of the experiments is motivated by confidential training scenarios provided by the French military. The goal of this scenario is to predict the reactions of people facing events. In more details, the military missions aim to reduce local support for the insurgents by convincing people that it is not in their interest to support or join an insurgency. Using the data described in the scenario, we generate a population of 100 individuals connected on three layers: family, friends and neighbors. These three layers represent the primary groups defined by (Litwak and Szelenyi 1969). The following are types of characteristics that may describe an individual: ethnicity, religion, gender, social status, etc.

Related to the concepts of opinion change described by (Friedkin 1999) and at the contact of a receiver with an emitter, the receiver calculates the trust factor (i,j). This mechanism for opinion change has been validated using an approach based on three issues (Friedkin 1999). The (i,j) represents the trust between agents based on their attributes. The trust factor has value $(i,j)=1$ if the two agents i and j are similar (they have same religion, same ethnicity and same class of age) and $(i,j)=0$, otherwise. In the first case, agents receive the message and their variables (opinion, interest and satisfaction) change over time depending on some rules;

for simplicity, the receiver agent adopts the emitter's opinion. Then, it is able to propagate the message to its neighboring agents. In the other case ($(i,j)=0$), the receiver will enter the informed state and propagate the message, if the emitter is a lead agent. In the all other cases, the message is ignored and the receiver will not propagate anything.

4.2. Settings

To test our architecture, we drive three experiments with the same population but with different algorithms of message propagation (Table 1). In the first experiment, we do not consider the type of the message or the type of the relationship between the emitter and the receiver to implement rules in the proxies (one graph with different connections). In this case, the uninformed agents receive the message from their neighbors without calculating the trust factor. The trust factor between all the agents is equal to one.

In the second experiment, we take into account the trust factor between agents. In this experiment, the receiver calculates the trust factor between him and the emitter before accepting the packet. The diffusion process depends on the strength of the message and on the trust factor.

Tableau 1: Description of simulated experiments

Experiments	Description
Experiment 1	Trust factor between all agents is equal to one
Experiment 2	The diffusion process depends on the trust factor but not on the type of connections
Experiment 3	Each proxy has some specific rules to filter the information

In the third experiment, we use our multi-layer architecture by implementing specific rules to the type of message and to the relationship between different agents. We block the diffusion process in the friends and neighbors layer. We assume that at night and in the absence of communication channels the message will be spread only by face to face between family members.

4.3. Result and discussion

The results show that both the network structure and the node attributes (individual and cultural factors) have an impact on the proportion of informed agents and their opinion.

Figure 4 presents the initialization phase of the simulation using the generator as explained in Figure 2 in Section 3.2.2. Each node in the network represents an agent with different attributes. These entities may be uninformed individuals or information sources. Source nodes initiate the original message. The info-sources are chosen randomly. Each node is colored according to its status: green for informed nodes, red for passive nodes and white for uninformed nodes. At t_0 , all nodes are uninformed except for the three info-sources: nodes V3, V24 and V70. As we are working with an MSN, we

have three kinds of link: family relationships (red links), friendship relationships (green links) and neighborhood relationships (yellow links).

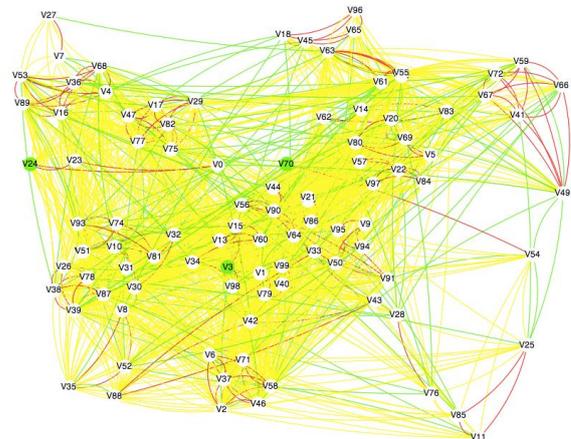


Figure 4: Initializaton of the simulation

Figure 5 shows the simulation results of the first experiment. The green nodes represent the agent who received the message and is interested by it. Red nodes represent the agent who received the message and not interested by it. White nodes represent the agent who did not receive the message. In this case, the diffusion process depends only on the strength of the message. The process is ended when the strength of the message becomes less than zero. The notion of interest used here does not refer to the receiver's emotion causes by the message. At this stage, we use a function that determinates if the agent is interested in the message or not by comparing closeness of the emitter's and receiver's characteristics (e.g., religion, ethnicity, age, language, social status).

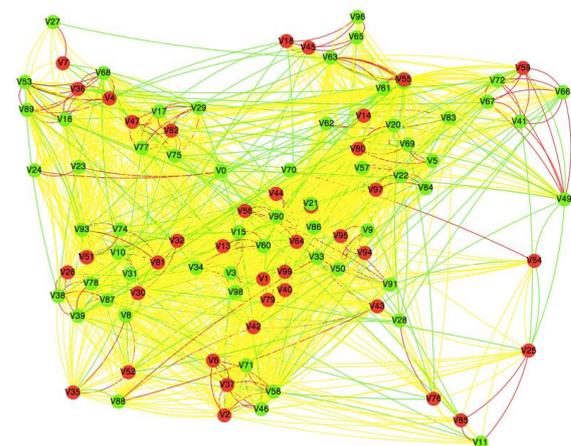


Figure 5: Result of the first experiment

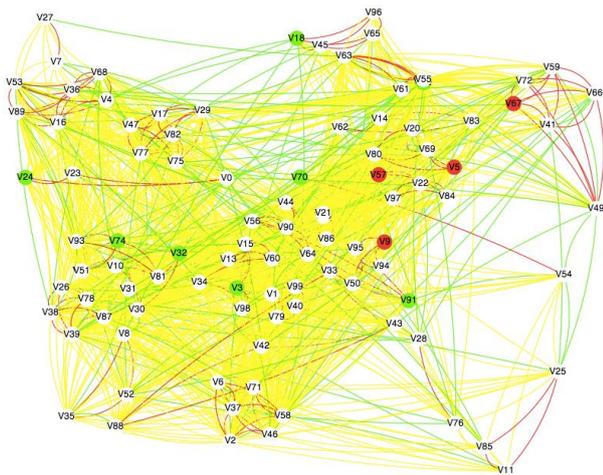


Figure 6: Result of the experiment 2

In the Figure 7 we observe that only the nodes connected by the family relationship are the informed nodes because we add some rules in the other layers to block and filter the diffusion process. We can explain this result by sending an urgent message at night and in the restriction of the use of communication channels (phone and internet).

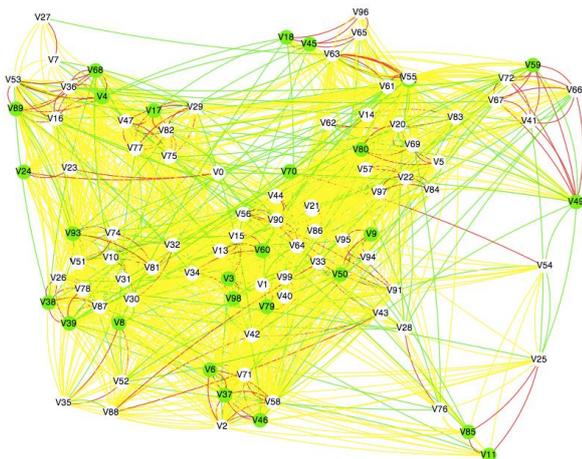


Figure 7: Result of the experiment 3

The different experiments show the flexibility of our Server/proxy architecture based on DEVS. We easily manage the diffusion process according to numerous parameters. We can simulate more realistic phenomena based on cultural and human factors. In addition, we can exploit our architecture to study the information diffusion in a dynamic network.

Our architecture separates the individual behavior modeling from the networks structure modeling. It allows tuning easily the rules of diffusion or the structure of the network as soon as we have to change the context of the study or to test different human's behavior models. Finally, the human behavior algorithms implemented in the models presented in this paper are quite simple but however innovatively used by mixing social sciences studies and multi-layer social networks. The behavior algorithms used in the models were validated in social science studies but in a

different context (Friedkin 1999). The network topology is coming from (Mitchell 1969; Lazega 1998). The VLE simulation platform was also validated (Quesnel et al. 2009; Ramat et al. 2003). Nevertheless, because we merged human behavior and networks structure in a DEVS M&S environment for the first time, the complete framework hasn't been validated yet. We have started to confront the first scenarios models and simulation results to military experts. We have received a positive feedback but we agree that this feedback lies on simple and limited number of models; so it does not permit to validate the genericity of these models of networks and individual behavior. We find only two ways to validate our models. The first, already started, will consist in facing military expert with several sets of input scenarios and results obtained by simulation. The second issue is to set historical situations in models, run them, observe and compare output to historical facts. We are at the moment working on both aspects.

CONCLUSION AND PERSPECTIVES

We presented in this paper a new simulation framework feature for MSN based on the DEVS formalism. This low-level framework allows a full control of the agents' models; it promotes also modularity and reusability. This new architecture is fully dedicated to simulate, as transparently as possible, the propagation of information within a population network. In details, the mechanisms implemented in the models are, at the moment, simple but they are easily upgradeable by modelers. In addition, we assumed that relationships between people are too complex to be modeled by a unique network. Furthermore, information diffusion is dependent of the category of relationship between the people who communicate. So, the DEVS-based agent framework with the Server/Proxy architecture has been proposed to separate individual behavior in Server and perception of the environment in Proxy. It is flexible and sensitive to changes in the environment. In the last section, we managed two experiments showing (1) the importance of managing each relationship separately in the information diffusion process and (2) the simplicity to modify the diffusion rules for each relationship using our Sever/Proxy architecture. We are aware that, even if the different technics employed were previously validated in their respective domain, the models and simulation results proposed in the paper have not been validated. Nevertheless, we already confronted the first results to expert in the domain and will continue in that way. Beyond this, the perspective of our work seems promising to facilitate the setup and run of network structured DEVS models. This framework can be used to solve the problem of influence maximization (IM); find a set of initially activated nodes (info-sources) with the maximum number of activated nodes after the time step t by launching several simulations with different nodes. Then, we will compare the final results to select the nodes that maximum influence or spread the information. As perspective, we are integrating

progressively more variables to describe more accurately human behavior including social and psychological feelings and health. The inner idea is to obtain, using simulation, a better understanding of how information can affect behavior of people and population. We are also looking to adapt our model to other domains such as marketing, logistics and manufacturing.

ACKNOWLEDGMENTS

This work is supported by the SICOMORES Project N° 132936073 funded by French DGA (Direction générale de l'armement). It involves the following partners: IMS university of Bordeaux, LSIS university of Marseille and MASA Group.

REFERENCES

- Bailey, N. T. (1975). *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.
- Barros, F. J. (1996). The dynamic structure discrete event system specification formalism. *Transactions of the Society for Computer Simulation International*, 13(1), 35-46.
- Bouanan, Y., El Alaoui, M. B., Zacharewicz, G., & Vallespir, B. (2014). Using DEVS and CELL-DEVS for Modelling of Information Impact on Individuals in Social Network. In *Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World* (pp. 409-416). Springer Berlin Heidelberg.
- Chow, A. C. H., & Zeigler, B. P. (1994, December). Parallel DEVS: a parallel, hierarchical, modular, modeling formalism. In *Proceedings of the 26th conference on Winter simulation* (pp. 716-722). Society for Computer Simulation International.
- Davidsson, P. (2002). Agent based social simulation: A computer science view. *Journal of artificial societies and social simulation*, 5(1).
- DEVS community. (2015). DEVS tools. Available from: <http://www.cell-devs.sce.carleton.ca/devsgroup/?q=node/8> [accessed 1 July 2015].
- Domingos, P. (2005). Mining social networks for viral marketing. *IEEE Intelligent Systems*, 20(1), 80-82.
- Faucher, C., Zacharewicz, G., Hamri, A., & Frydman, C. (2012, March). PSYOPS and CIMIC operations: from concepts to G-DEVS models. In *Proceedings of the 2012 Symposium on Theory of Modeling and Simulation-DEVS Integrative M&S Symposium* (p. 42). Society for Computer Simulation International.
- Friedkin, N. E., & Johnsen, E. C. (1999). Social influence networks and opinion change. *Advances in group processes*, 16(1), 1-29.
- Goldenberg, J., Libai, B., & Muller, E. (2001). Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3), 211-223.
- Granovetter, M. (1978). Threshold models of collective behavior. *American journal of sociology*, 1420-1443.
- Lazega, E. (1998). *Réseaux sociaux et structures relationnelles*. Presses universitaires de France.
- Litwak, E., & Szelenyi, I. (1969). Primary group structures and their functions: Kin, neighbors, and friends. *American Sociological Review*, 465-481.
- Ma, J. J., Zeng, D., & Huff, R. A. (2013). Complex Network Analysis. *Journal of International Technology and Information Management*, 22(4), 6.
- Mitchell, J. C. (Ed.). (1969). *Social networks in urban situations: analyses of personal relationships in Central African towns*. Manchester University Press.
- Quesnel, G., Duboz, R., & Ramat, É. (2009). The Virtual Laboratory Environment: An operational framework for multi-modelling, simulation and analysis of complex dynamical systems. *Simulation Modelling Practice and Theory*, 17(4), 641-653.
- Quesnel, G., Duboz, R., Ramat, É., & Traoré, M. K. (2007, July). VLE: a multimodeling and simulation environment. In *Proceedings of the 2007 summer computer simulation conference* (pp. 367-374). Society for Computer Simulation International.
- Ramat, E., & Preux, P. (2003). Virtual laboratory environment (VLE): a software environment oriented agent and object for modeling and simulation of complex systems. *Simulation Modelling Practice and Theory*, 11(1), 45-55.
- Rogers, E. M. (1962). *Diffusion of innovativeness*. NY: The Free Press of Glencoe.
- Seck, M., Frydman, C., Giambiasi, N., Oren, T. I., & Yilmaz, L. (2005, July). Use of a dynamic personality filter in discrete event simulation of human behavior under stress and fatigue. In *1st International Conference on Augmented Cognition* (pp. 22-27).
- Smith, T., Coyle, J. R., Lightfoot, E., & Scott, A. (2007). Reconsidering models of influence: the relationship between consumer social networks and word-of-mouth effectiveness. *Journal of Advertising Research*, 47(4), 387.
- Takahashi, S., Sallach, D. L., & Rouchier, J. (Eds.). (2007). *Advancing social simulation: the first world congress* (p. 354). Berlin: Springer.
- Tenkasi, R. V., & Chesmore, M. C. (2003). Social networks and planned organizational change the impact of strong network ties on effective change implementation and use. *The Journal of Applied Behavioral Science*, 39(3), 281-300.
- Tobias, R., & Hofmann, C. (2004). Evaluation of free Java-libraries for social-scientific agent based simulation. *Journal of Artificial Societies and Social Simulation*, 7(1).

- Uhrmacher, A. M., Ewald, R., John, M., Maus, C., Jeschke, M., & Biermann, S. (2007, December). Combining micro and macro-modeling in devs for computational biology. In Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come (pp. 871-880). IEEE Press.
- Wainer, G. A., & Giambiasi, N. (2002). N-dimensional Cell-DEVS models. *Discrete Event Dynamic Systems*, 12(2), 135-157.
- Zeigler, B. P., & Vahie, S. (1993, December). DEVS formalism and methodology: unity of conception/diversity of application. In Proceedings of the 25th conference on Winter simulation (pp. 573-579). ACM.
- Zeigler, B. P., Praehofer, H., & Kim, T. G. (2000). Theory of modeling and simulation: integrating discrete event and continuous complex dynamic systems. Academic press.

in journals and conferences. His email address is bruno.vallespir@ims-bordeaux.fr.

AUTHOR BIOGRAPHY

YOUSSEF BOUANAN is a PhD Student in Electrical and Computer Engineering with the Department of Systems and Computer Engineering at University of Bordeaux. He received his Engineer degree from School of Engineering Science, Morocco and a M.Sc. in Enterprise System Engineering from University of Bordeaux, France. His email address is youssef.bouanan@ims-bordeaux.fr.

MATHILDE FORESTIER is Postdoctoral fellow at the IMS laboratory, University of Bordeaux, France. She obtained her PhD "Enriched social network extraction for social roles analysis on online forum debates" in 2012 from the University of Lyon, France. Her email address is mathilde.forestier@ims-bordeaux.fr.

JUDICHAEL RIBAUT is Postdoctoral fellow at the University of Bordeaux (IUT MP) Lab. IMS. He received his Ph.D. at INRIA Sophia Antipolis, France. He has published several papers in Conferences and is frequent Reviewer in Conferences (SpringSim, TMS/DEVS, etc.). He is involved in European projects. His email address is judicael.ribault@ims-bordeaux.fr.

GREGORY ZACHAREWICZ is Associate Professor HDR at University of Bordeaux (IUT MP) with both competences in enterprise engineering and computer sciences. He is recently focused on Enterprise Modelling and Semantic Interoperability. He has published more than 60 papers in international journals and conferences. His email address is gregory.zacharewicz@ims-bordeaux.fr.

BRUNO VALLESPER is full professor at University of Bordeaux, IMS laboratory. He is member of several international working groups (IFIP, IFAC), he participated to 5 European projects, has directed more than 20 PhD students and written more than 120 papers

A SPECIALIZED LIBRARY FOR SIMULATION OF COAL MINING IN FLAT-LYING COAL SEAM

Victor Okolnishnikov^(a), Sergey Rudometov^(b), Sergey Zhuravlev^(c)

^{(a),(b),(c)} Design Technological Institute of Digital Techniques of Siberian Branch of the Russian Academy of Sciences

^(a)okoln@mail.ru, ^(b)rsw@academ.org, ^(c)s-zhur@yandex.ru

ABSTRACT

The paper describes a new specialized library of simulation models of mining machinery for coal mining in flat-lying coal seam. This library is a part of the new visual interactive discrete simulation system of technological processes MTSS (Manufacturing and Transportation Simulation System). MTSS is effective in solving the task for the rapid creation of correct simulation model by mining engineers. 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, and pumping subsystem. The new library contains models of highwall mining, longwall mining system, coal seam, mining machines, self-moving coal wagon, and storage area. Models of highwall technology and longwall technology for coal mining in flat-lying coal seam were made with the help of the new library. These models are used for developing of process control systems for underground coal mines in Kuznetsk Coal Basin (Russia, Western Siberia).

Keywords: visual interactive simulation, coal mining, longwall mining, highwall mining

1. INTRODUCTION

It is specified in different investigations, that situation in different areas of underground mining that requires automatic control (e.g. coal mining) is that resources become 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 they require? What is the cost of machines installed in mine? Will big universal machines be more appropriate rather than a lot of small, specialized mechanisms? What will the cost of a final product be? These tasks can be solved by methods of computer-based simulation of coal mining process.

The problem of the particular mining installation is that it 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. This 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; Greberg and Sundqvist 2011; Cai et al. 2012; Fioroni et al. 2014). There are also a large number 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 own simulation system. Technologies of coal mining are presented in section 2. Brief description of the simulation system is presented in section 3. Sections 4, 5 contain description of a specialized library and examples of the models.

These models are used for developing of process control systems for underground coal mines in Kuznetsk Coal Basin (Russia, Western Siberia).

2. TECHNOLOGIES OF COAL MINING

There are several well-known technologies of coal mining.

2.1. Longwall mining

Longwall mining system is a highly-automated, very powerful and productive way to mine a product. It is most-widely applied around the world. Its main advantage is that it leaves almost no product inside mines. But it is limited with the depth of the mine (measured from the surface). Also 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.

Longwall system consists at least of armored face conveyor (AFC), a shearer, and roof support sections. The AFC is connected to outbound belt conveyor. The shearer cuts the product from coal seam face, in a series of passes along the AFC. The AFC delivers product 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. This can lead to serious environmental problems. The other problem is that longwall requires significant amount of work to be done before its massive equipment will be installed in production. This work also must be planned.

2.2. Highwall mining

In case when it is impossible to mine a product with longwall or other systems, a relatively new technology can be applied, named Highwall. This is a shearer tool, mounted on the top of the chain of special sections. These sections can be updated to each other, making a long (up to 300 meters) support chain for a shearer. Each section can transport the product developed by shearer to the end of a sections chain.

The shearer and the chain of sections cut the product (coal) from the very thin and curved seams. This is the main advantage of the Highwall technology.

2.3. Coal mining of flat-laying coal seam

This approach uses a number of front-cutting mining machines that cut a coal on a special scheme, and a number of self-moving wagons that move a coal from mining machines to the storage area. It solves the same problem as highwall mining, but requires no specialized equipment. Also this approach can be used in deep mining.

3. THE SIMULATION SYSTEM

A visual interactive Manufacturing and Transportation Simulation System (MTSS) is developed at Design Technological Institute of Digital Techniques of Siberian Branch of the Russian Academy of Sciences (DTIDT). It is a process-oriented discrete simulation system intended to the development and execution of models of technological processes (Okolnishnikov et al. 2010a, 2010b; Okolnishnikov 2011).

MTSS is a set of program interfaces for creating elementary models and for forming complex models from them. The elementary model is a ready-to-use submodel 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 graphic images.
- Input and output parameters.
- Functionality algorithm describing dependence between parameters.
- States which the elementary model can reach during the simulation process.
- Control commands defining switching process between elementary models 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.

This simulation system is effective in solving the task for the rapid creation of correct simulation model by mining engineers. Usually engineers have not enough qualification to create a simulation models in details, but they know how to connect correctly elementary models to create the required topology. MTSS uses the 2D as the graphical editor and 2D, 3D for the 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 and the upper level of complex control of production. Therefore one of the distinguishing features of MTSS is a separation of the logic of simulation model into two parts: low-level logic and an upper level logic.

Such separation allows us not only to correspond to the usual structure of the process control systems but 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. This separation into upper and lower logics allows also organizing a switch between various implementations of the decomposition. It allows 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 allows simulating various accidents. This allows minimizing time and costs on site for commissioning.

4. SIMULATION OF COAL MINING IN FLAT-LYING COAL SEAM

A specialized library of simulation models of mining machines for coal mining was developed. This library is a part of MTSS and its prime goal is to simulate interactively and visualize various aspects of coal mining in flat-lying coal seam. The library consists of new elementary models of:

- Highwall mining system.
- Longwall mining system.
- Coal seam.
- Mining machines.
- Self-moving coal wagon.
- Storage area.

The library contains also a simulation model of a flat-lying coal seam. This model is a source of the product in a simulation model, while storage area is a consumer of a product. The product itself is coal.

These new components can communicate with existing libraries of MTSS (Okolnishnikov et al. 2013a, 2013b) which simulate mines subsystems like:

- Belt conveyor subsystem.
- Power supply subsystem.

- Ventilation subsystem.
- Pumping subsystem.

A simulation model of coal mining subsystem in a flat-lying coal seam was developed using the library of mining machines. Figure 1 contents a sample layout in the simulation model of the flat-lying coal seam (2D and 3D view combined in different views).

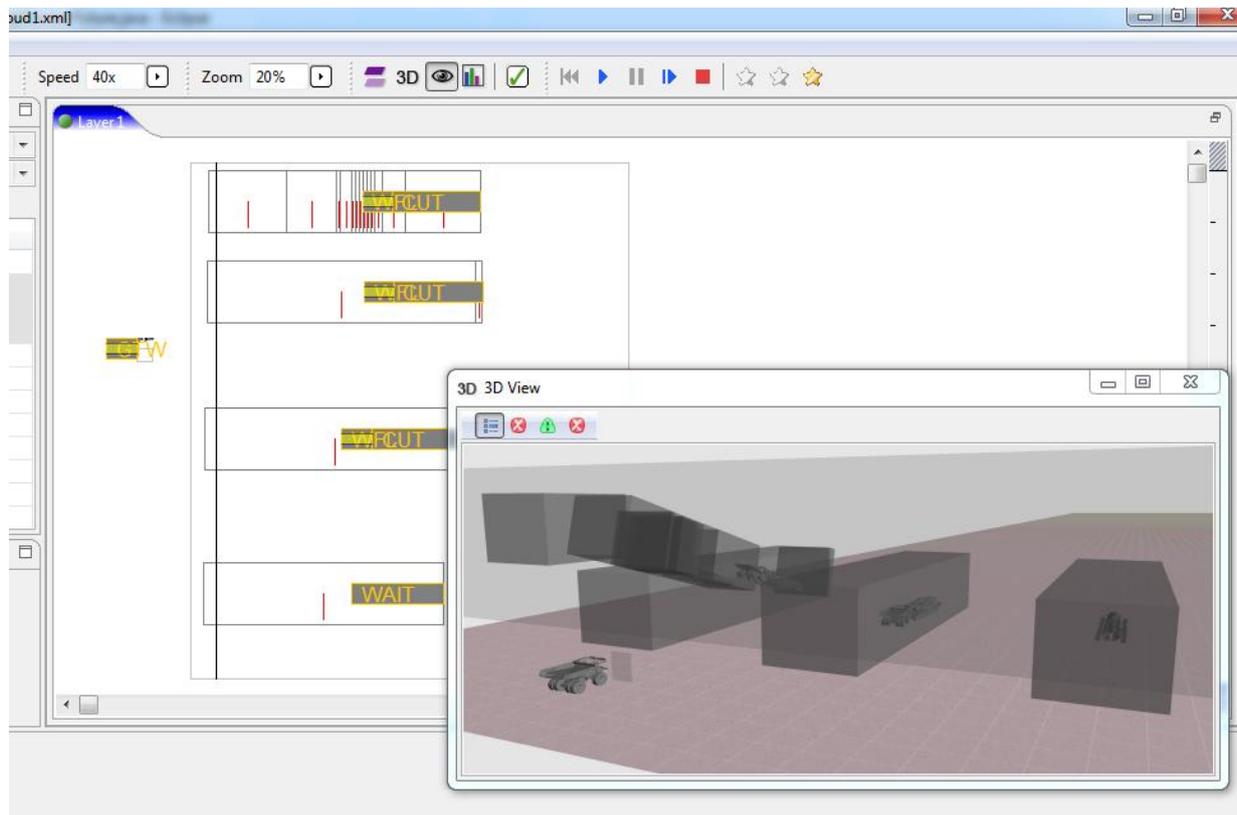


Figure 1: Simulation model of coal mining in flat-lying coal seam (4 mining machines and 4 self-moving coal wagons)

The simulation model built from the content of the library will simulate movement of all mobile objects of the model. Both 2D (top view) and 3D visualization are available. Statistical data is also collected.

Figure 2 shows a sample layout in another simulation model of the flat-lying coal seam. Main window consists of 6 areas:

1. 2D top view. It contains: main mine, side mines. The origin coordinates is at the left top corner of this view.
2. 3D view. On Fig. 2 there is a view from point 3).
3. Point of view for the 3D.
4. Parameters of a simulation model.
5. Specialized view for fast navigation in simulation model.
6. Settings for the time start, time end and current model time.

The model was created to visualize new technology of coal-mining in flat-lying coal seam (Fedorin et al. 2013). This simulation model also allows receiving statistical results for usage of this technology.

As patent describes, mining is done by frontal winning machine, paired with self-moving coal wagons. Patent describes the directions of cut of a flat-lying coal seam for the frontal winning machine. Simulation model of this process allows visualizing the process described.

5. SIMULATION OF LONGWALL MINING

For detailed simulation of longwall mining system we finished with next decomposition: armored face conveyor, 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.

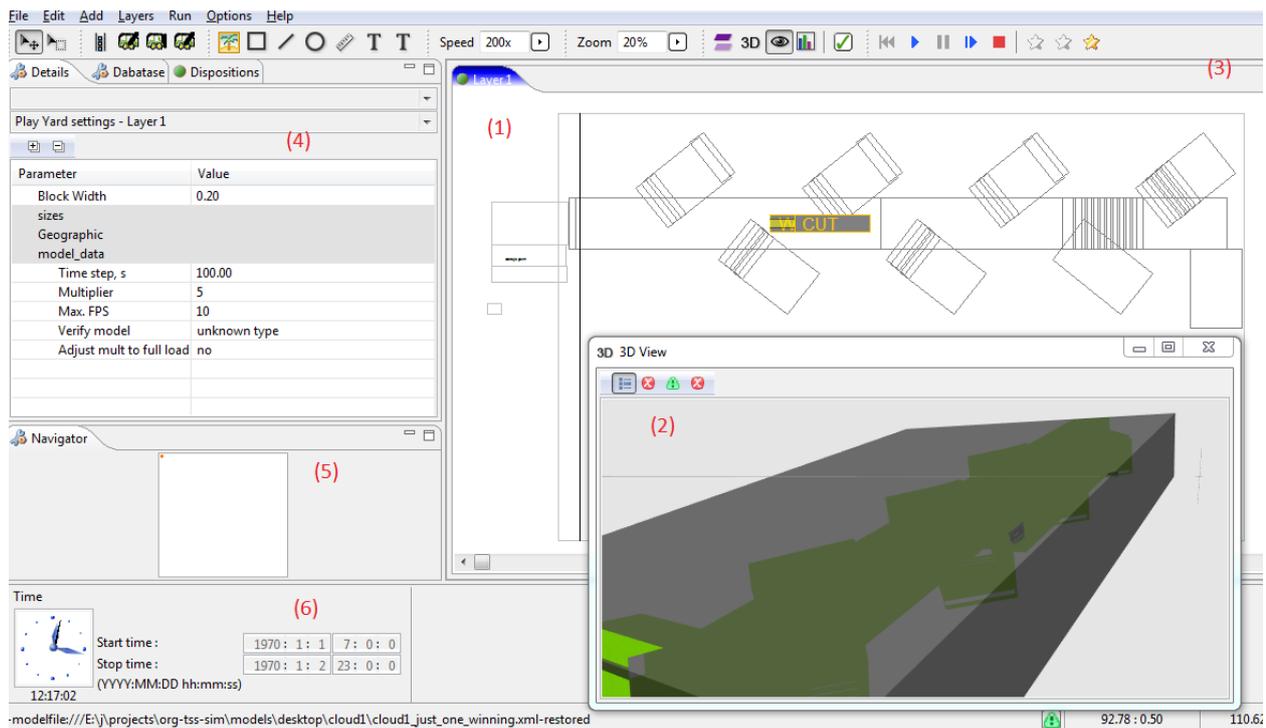


Figure 2: Main window of MTSS system while running of simulation model of coal mining in flat-lying coal seam

The algorithm for longwall mining system simulation is done closer to the control algorithms of the real longwall mining system. The task will define the amount of time for the shearer to move to the state while it will allow one roof support section to be free to move, and an amount of a product to be mined. Then it will advance the simulation to the time defined, and then simulate the roof support sections movement. The amount of a product will be moved to the AFC that will deliver it to the belt conveyor. Task will repeat these steps until the “done” or “postpone” conditions will be achieved.

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 product or gas level is not safe.

During any of these steps, the simulation model of flat-lying coal seam will simulate roof fall (behind the roof support) and gas level increasing. Also ventilation 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 product for belt conveyor system, for example. All that is really needed in such cases are:

- To define that longwall or flat-lying coal mining is a source of a product for a big mining system like conveyor.
- To know the performance of the longwall coal mining during some time period (working day, 8-hours time interval, 1-hour time interval).

Note that the emergency stops (like gas or coal dust) are already included in this statistical data.

But if the goals of simulation are:

- To predict how longwall or flat-lying coal seam 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 impact to the overall performance of the mines.
- To define scenarios of broken or temporarily inaccessible parts interactively.

Then achieving these goals will require detailed decomposition of common longwall (or highwall) mining system and a detailed visualization of all its parts.

In our work both detailed and statistical approaches are presented.

6. 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.

ACKNOWLEDGMENTS

This work is supported by the Russian Foundation for Basic Research (Project 13-07-98023 r_siberia_a).

REFERENCES

- Cai D., Baafi E. and 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.
- Fedorin V., Shachmatov V., Anferov B. and Kuznetsova L., 2013. The method of open-underground mining of thick flat coal seam. Patent for invention of Russian Federation № 2490456.
- Fioroni M., Santos L., Franzese L., Santana I., Telles G., Seixas J., Penna B. and Alkmim G., 2014. Logistic Evaluation of an Underground mine Using Simulation. Proceedings of the 2014 Winter Simulation Conference, pp. 1855-1865. December 07-10, Savannah, GA, USA.
- Greberg J. and Sundqvist F., 2011. Simulation as a tool for mine planning. Proceedings of Second International Future Mining Conference, pp 273-278. November 22-23, Sydney, New South Wales.
- Kizil M.S., McAllister A. and 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.
- Okolnishnikov V., Rudometov S. and Zhuravlev S., 2010a. 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., Rudometov S. and Zhuravlev S., 2010b. Monitoring System Development Using Simulation. Proceedings of the 2010 IEEE Region 8 International Conference, vol. II, pp. 736-739. July 11-15, Irkutsk, Russia.
- Okolnishnikov V., 2011. Development of Process Control Systems with the Use of Emulation Models. International Journal of Mathematics and Computers in Simulation, issue 6, vol. 5, pp. 553 – 560.
- Okolnishnikov V., Rudometov S. and Zhuravlev S., 2013. 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., 2013. Simulation Environment for Development of Automated Process Control System in Coal Mining. International Journal of Systems Applications, Engineering & Development, issue 5, vol. 7, pp. 255-262.

Sturgul J.R., 2001. Modelling and Simulation in Mining - Its Time Has Finally Arrived. Simulation, 76(5), pp. 286-288.

MODELLING AND SIMULATION EXAMPLES BASED ON COMPUTER ALGEBRA SYSTEMS

Stefanie Winkler^(a), Andreas Körner^(b), Andreas Bauer^(c), Martin Bicher^(d), Felix Breitenecker^(e)

^{(a),(b),(c),(e)} Technische Universität Wien, Wiedner Hauptstraße 8-10, 1040 Vienna
^(d) dwh GmbH, Neustiftgasse 57-59, 1070 Vienna

^(a)stefanie.winkler@tuwien.ac.at, ^(b)akoerner@asc.tuwien.ac.at, ^(c)andreas.e101.bauer@tuwien.ac.at,
^(d)martin.bicher@tuwien.ac.at, ^(e)felix.breitenecker@tuwien.ac.at

ABSTRACT

Nowadays education has to satisfy different requirements regarding the variety of demands of industry but also pedagogical. The students have to learn as well as apply different skills. Another important part is deep understanding of the environment. It is not enough to know how to apply a formula they should also be able to connect that formula to the mathematical background to enable connecting different topics. They should be able to apply the learnt methods in real-life scenarios. This all should be done in a minimum of time during the course. The lectures are forced to accomplish the impossible. Therefore universities are encouraged to use online environment to support their courses to enable individual learning. Using platforms and other e-tools the learning process can be translocated to leisure time. This also enables students to learn in their own speed and to their own conditions. In this paper three different e-learning environments will be presented and their usage and development at the TU Wien will be presented.

Keywords: b-learning, mathematical modelling, mathematics, education

1. INTRODUCTION

The research group for mathematical modelling and simulation (MMS) at the TU Wien started to use online tools in 2004. The first step was the usage of a moodle based platform. The moodle platform was initialized by the university and therefore also some modification or add-ons according to university's requirement were made. This webpage is mainly used to organize all the materials of the lectures. On the one hand the lecture materials but also additional documents and links are provided using moodle. On the other hand the administration of students can be done using this system. There are different tools which are useful to organize exercise tasks, dues and appointments. The person in charge of the system for TU Wien also programmed some plugins if needed. For example the administration of the exercises required a certain plugin. In the exercises organized by the MMS the students

have one week to prepare their homework. Task information is provided in form of pdf documents uploaded onto the moodle platform. At the end of the week the students have to tell the lecturer or tutor which examples he or she has prepared. This is realized using a special plugin where the students can tick off which examples they have prepared.

This plugin is mainly used in mathematical courses. Due to the fact that the research area of MMS covers as well different applications of modelling and simulation in various fields the group also organizes different lectures dealing with modelling and simulation. For these courses moodle cannot fulfil all requirements to ensure an adequate e-learning support. Therefore a combination of moodle and an e-learning system, called MMT, is used. The MMT server, which was developed by a spinoff company and the group itself in 2006, stores various modelling and simulation examples. It can be compared to a CMS for modelling examples. In the following section it will be described in detail.

In 2008 the vice rector for academic affairs initialized a refresher course for mathematics which should be supported by an e-learning tool. Due to the experience in this field this course idea was executed by MMS. The used e-learning system is called Maple T.A. which will be explained later. The refresher course is now held for nearly every first semester student at TU Wien which would not be possible without the usage of an additional online learning platform.

2. MMT

In the modelling and simulation courses the MMT, which stands for Mathematics, Modelling and Tools, is used. It was developed in 2006. Before that the research group used a server which was hosted by MATLAB itself. After MATLAB shut the server down the research group invented a new system by themselves.

2.1. Structure

MMT is partly a content management system. The content is organized in a tree structure. Additionally it is possible to execute MATLAB files on the system. There are basic mathematical examples realized as well

which are still used in mathematical courses for students of geodesy in the first and second semester.

The main advantage of this system is the parametrization of the examples. If the students face an example, it does not matter if it is a basic mathematical one or a simulation example, the parameter used for calculations of this example can be changed and the example can be executed again. In this way the students see the influence of the variables using different values. This online system is used in several modelling and simulation courses. The realized examples for these courses deal with different model approaches. In the beginning this simulation examples with variable parameters where only implemented in MATLAB. In the last year some improvements were made. There are as well examples implemented using other software for example open source Octave, Simulink, Anylogic or System Dynamics are now available. Nearly for the last 10 years this platform was used to support the learning process and understanding. In the lectures the examples are used to present the different modelling aspects. Using different application examples enables students to gain a connection between real-life problems and the mathematical definition of the according model.

2.2. Examples

In the following example the structure of the system as a CMS is shown.

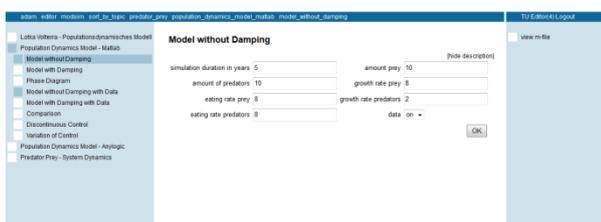


Figure 1: The structure of the MMT server is shown.

In Figure 1 the principle structure of the MMT server is shown. The left side displays the tree structure of the content. The students can see the examples connected to their course sorted by topic. In the middle of the picture the parameter section is shown. There the students can fill in the parameter values they want to explore. After submitting the values the simulation starts by clicking the OK button. Depending on the complexity of the example the calculations can take some seconds. In this example that is not the case. Then the result plot appears and the student can study the data.

In the right column there is a link to the source code which was used for the program. With this code the students have the possibility to execute the example after some modifications on their own MATLAB or Octave installation. Regarding the other program languages the model can be rebuilt using the representation provided by the MMT system.

Figure 2 shows the result of an implementation of the Lotka-Volterra equation. This differential equation system of first-order describes the dynamical interaction

of prey and predator. The following two equations characterize the change of the populations.

$$\frac{dx}{dt} = \alpha x - \beta xy$$

$$\frac{dy}{dt} = \delta xy - \gamma y$$

This predator prey model takes into account two populations, predators, for example foxes, and preys, for example rabbits. The existence of predators is dependent on the number of the prey, because their survival is dependent on how much they can eat (prey). The survival of the preys in return is also dependent on the amount and eating habits of the predators.

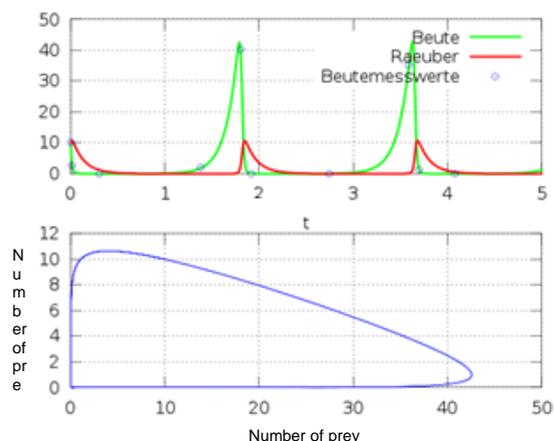


Figure 2: The results of the prey-predator model are shown.

In the Figure 2 in the upper graphic the red line gives the amount of predators and the green line shows the number of prey.

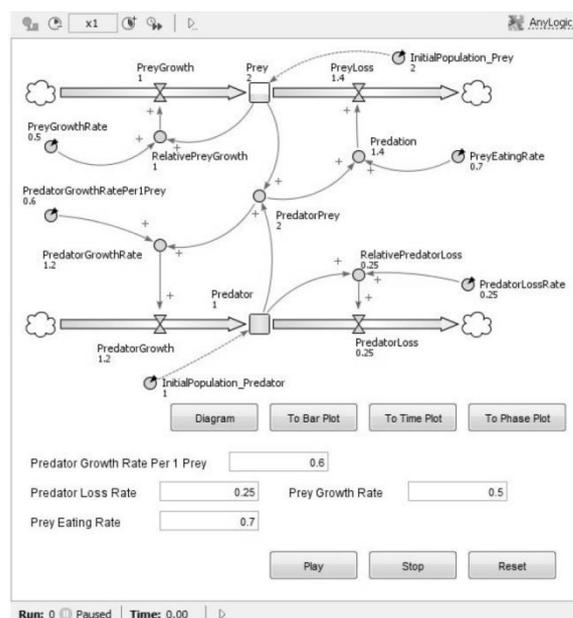


Figure 3: Implementation of the prey-predator model in AnyLogic.

The dependency mentioned before can be noticed easily. When the predators are less the preys can reproduce. Due to the increasing number of preys the predators can hunt again.

In Figure 3 the same example is developed in AnyLogic where Java-applets are used. AnyLogic is a simulator which can export the created models as Java-applets. Therefore it is possible to produce models using this software for the MMT.

Currently these examples are not only used to present them during the lecture to explain certain modelling aspects but also for practising the influence of variables in a model.

In the courses about modelling and simulation the students have to pass a test. As mentioned before this test was performed on the moodle platform. There are some randomized multiple choice questions realized in moodle itself. Other questions are dealing with the determination of certain parameter values in order to generate certain behaviour of the model. These parameter values have to be entered in moodle. The grading of the test is also realized in moodle.

In order to generate a great range of examples and also different applications more examples were gathered through cooperation with a research group in Ljubljana, Slovenia. They also use an e-learning tool to improve the learning process of the students. Due to their profession they were also able to provide some real-life experiments. The students have the possibility, using a step-by-step environment to prevent experimental errors, to set parameters for a real experiment and watch the results live. This principle we could not include directly up to now but the experiments gave us some new example ideas. For further cooperation an enlarging of the example pools of both groups is planned. An inclusion of their experiments in form of a webcam connection is also in mind.

3. MAPLE T.A.

In 2008 the vice rector for academic affairs asked for a so-called refresher course for mathematics. The necessity of such course is still controversial. Due to the fact that most of the academic institutions in Germany have already initiated such courses it seems to be necessary not only in Austria.

This kind of course should remind the students of their perhaps forgotten knowledge from school especially in mathematics. The refreshment should then help the students to surmount difficulties which can appear in the first year of study. In order to offer this course for nearly every newcomer at the TU Wien the course combines ordinary lectures with hands-on practice and e-learning tools.

3.1. Moodle Connection

Maple T.A. stands for Maple Teaching and Assessment. It is a computer algebra system which is based on Maple. Therefore it is a very powerful mathematical tool. In the provided interface the management of examples as well as user administration come in handy. The examples are organized in different classes which are related to the different courses in university. This system can not only be used for exercising but is also, as mentioned in the name, used to perform tests. All the test results are stored in the system and can also be accessed by the student. Since some years there is also the possibility to connect Maple T.A. to a moodle platform. As mentioned before TU Wien uses a moodle based website to organize course materials. Regarding modelling and simulation courses the moodle platform is also used to realize small tests.

In order to perform a test in the mathematical courses the students have to provide three different logins to connect to the different platforms. Using this point of intersection the Maple T.A. system uses the moodle platform to identify the students and to make sure that only registered students are using Maple T.A. Therefore the students need only access to the university platform to exercise examples on Maple T.A. The layout of moodle can be used to arrange a certain course structure and include the related examples from Maple T.A. directly to this structure. This ensures a more instructed learning process which should support students' progress.

3.2. In General

Maple T.A. offers different prepared question types as multiple choice, numerical graded, maple graded but also drop down menu questions. All these question types can be created separately but can also be combined in a single question. This range of question types enables a variability of examples focusing on different learning phases.

In order to create an example an algorithm is necessary. The algorithm which the example is based on can be realized in two different ways. On the one hand there are some special commands which are implemented for the use in Maple T.A. For example the randomization of integer values in a certain range is possible with a special command which is only working inside Maple T.A. There are other commands like if and switch which are different or unique in Maple T.A. On the other hand it is also possible to use commands and procedures programmed in Maple itself. In the algorithm Maple can be accessed in every step and for every variable. But in order to assure nice performance, part of the algorithm should be done in Maple T.A. itself. It is possible to integrate a whole Maple procedure into the algorithm as well if it is necessary. After programming the algorithm the next step is to define the right grading for the student's response. There are different predefined grading routines. For example a multiple choice question is automatically graded using partial grading in terms of multiple

possible answers. Other forms of grading are, for example, numeric, equations and maple-graded. The numeric routine enables the tutor to define a certain range of accuracy for the student's solution. Using the equation type allows the student to enter an algebraic equivalent answer as an equation. In the maple-graded section the grading routine can be an algorithm written in Maple. This last choice offers many different possibilities regarding the type of student responses.

3.3. Library

Most examples are graded using the maple-graded response. In the answer window of the question students enter their results which can be either a value or a whole function. If the example deal with Linear Algebra it is also possible that the result is a vector or matrix. These results can be graded using every available Maple command. The grading should depend on the purpose of the examples therefore the didactical aim of the example should be focused.

In order to guarantee a student friendly grading routine MMS created two different libraries. One library deals with grading of different types of answers. The main reason was an example dealing with partial fraction expansion. The calculation for students is complicated and prolonged but one small mistake ruins the whole result. The first step was to grade this example partially. Every coefficient should be considered and not only the whole answer. Furthermore also the grading routine of vector and matrix follows this principle. If the example asks for the inverse matrix the grading algorithm considers every single entry and not only the matrix as a whole. Additionally, grading routines for sets and integration results are included.

In Figure 4 the grading routine for vectors is applied. This example is a special one because the routine determines how many entries of the resulting vector of the matrix vector multiplication are zero. In this example only the multiplication of the first row is zero. Therefore the student receives 25 % of the points.

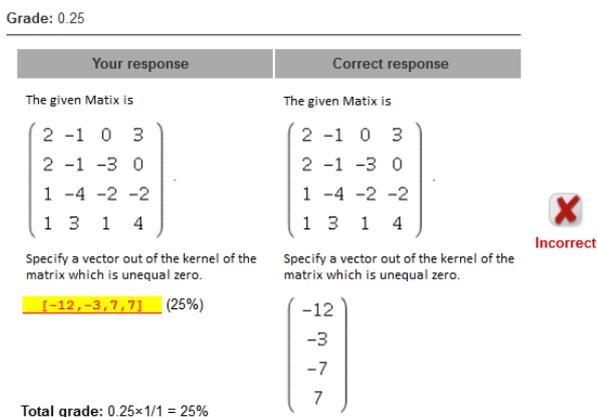


Figure 4: The grading routine for vectors is used in this example.

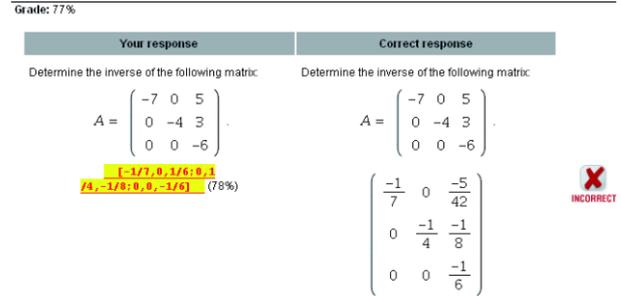


Figure 5: Using the matrix grading routine every entry is graded separately.

In Figure 5 shows the example mentioned before. The student has to determine the inverse of a certain matrix. Due to the fact that the last entry of the first row and the second entry in the second row are wrong the student gains 78% of the points.

Another advantage of using this online tool lies in the automatic generation of a various number of examples by creating only one. In the algorithm the used variables can be picked randomly out of a certain range. If the students choose one of these examples twice they get new actual numbers every time. Due to this fact, memorizing of example solutions is prevented. Sure they might remember how to calculate the solution but that is exactly one of the aims of this system. Students can practice as long as they want. Sometimes the numbers might repeat but in a sequence of several attempts. It is not only possible to vary the variables but also to randomly generate function with every attempt. Using the second developed library the function can be picked out of a defined set. This does not explain why the need of a separate library was necessary. But as seen before the examples dealing with Linear Algebra are in some aspects more difficult than others. Therefore this library also includes some creation routines which support the development of questions about matrix properties and eigenvalues. Some of the properties are size, rank and zerocount which defines how many zero entries the matrix contains. Another command defines the definition of the matrix which is generated. Randomization of integer vectors is also included in the library. This library eases the work of the developer but also guarantees a comparability of the different attempts and assures a greater variety of the examples because the different cases are not separated but randomized in one example.

The development of the libraries and the following increased randomization has two main consequences. The students have literally the possibility to practice examples all day long. The main goal of the examples remains but the calculations changes every time. For teachers and lecturers this fact changes the arrangement of tests and exams. The variable variation reduces the chances to cheat in examination. The neighbouring student gets examples with other numbers so he or she

cannot copy the results from him or her. The students have to find a way to calculate it.

Since usage of the online platform in the research group students attending mathematical courses had to pass two of three tests realized in Maple T.A. These tests were distributed throughout the semester. In the last two years, the experiment executing exams in Maple T.A. has started. Due to the fact that the environment requirements for online exams for around 300 students are not given at the university this form of exam is voluntary. The students can decide if they want to perform the exam in Maple T.A. or as an ordinary written exam. After this exam the second part is an oral examination of the professor in order to determine the theoretical knowledge of the student.

In the following two examples of the different mathematical levels are presented.

3.4. Examples

The first example is used in the refresher course as well as in the mathematical course in the first semester of electrical engineering. The question shows either a cosine or sine and the student have to fill in the right parameters. In Figure 6 two different question types available in Maple T.A. are used. This example is every year a challenge for the students. Additionally this example is very important for electrical engineers to learn how to read a graphic right.

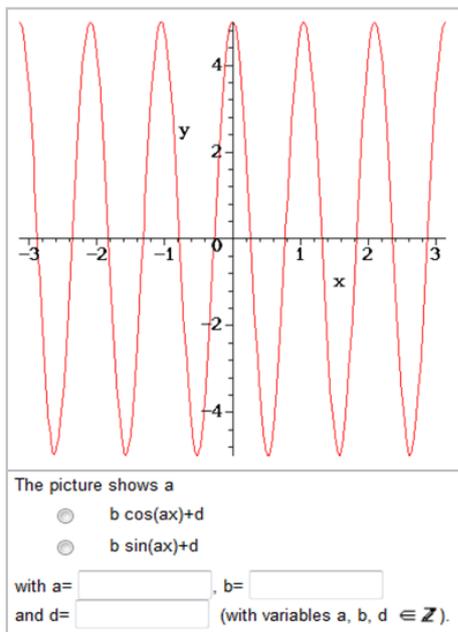


Figure 6: The example is used in the refresher course and basic mathematical course.

Question:

Let's observe the matrix

$$A = \begin{pmatrix} 2 & 3 \\ 4 & 1 \end{pmatrix}.$$

- Compute the characteristic polynomial:

$$p(\lambda) = \text{[input field]}$$

- Calculate the eigenvalues and insert them as set.

$$\{ \text{[input field]} \}$$

- Compute the eigenvector v to the eigenvalue $\lambda = -2$

$$v = \text{[input field]}$$

- Choose the right answer. The matrix A is:

- positiv definit
- negativ definit
- indefinit

Figure 7: The example deals with the matrix properties and is used in the higher mathematical courses.

The question of Figure 7 shows a Linear Algebra example. Firstly, the student has to enter the characteristic polynomial to determine the eigenvalues for the second step. In the third step the eigenvector should be calculated. Due to the used library the last step is easy to develop and asks for the definition of the matrix.

All in all there are many different examples dealing with various topics which are part of the mathematical courses.

4. EXAMPLE TRANSFERATION

The testing tool of moodle is quite useable. In order to be more flexible regarding the creation and development of tests there are some unsatisfying gaps. In the description of the two used systems some differences can be outlined. On the one hand the MMT system offers many different possible programming languages and provides an interface where students can experiment with the modelling examples. A small disadvantage is the location of the examples. It is not possible to link them directly to the moodle platform. The students have to enter a certain login to get to the right course examples. The testing environment is realized in moodle and the students have to experiment meanwhile on the MMT System.

The testing and assessment environment of Maple T.A. is quite mature. There are many properties which can be adapted to the needs of students and lecturers. The example tool was up to now only dealing with mathematical tasks. The usability of these questions can be shown due to the fact that at the last exam more than 60% of the students chose the online exam over the

written exam. There are different reasons for this development. On the one hand the students can practice all examples as preparation for the exam. On the other hand the online exams provide the students' results immediately after finishing the exam. This enables the students to make the oral exam shortly after the written one.

Due to the new developments of Maplesoft, Maple T.A. is updated every year. The last update includes some nice features which led to a new step regarding the two e-learning environments. In the newest version it is possible to integrate MathApps into the Maple T.A. question. The main advantage of this feature is the possibility for students to explore and experiment with graphical outputs. Additionally this change leads to a completely different structure of the questions. The MathApp part can be used to convey the mathematical theory or background of the question.

This possibility enables the research group to merge both systems into one. The simulation examples are transformed into MathApp questions. The disadvantage of an external grading of the MMT system using moodle disappears.

Figure 8 shows one of the first examples transformed from MMT to Maple T.A. As seen before the students had to fill in their parameter values and generate the new image using an OK button. In the MathApp implementation the parameter variation is realized using sliders. The position of these sliders can be changed by the student. Immediately the system generates the plot again to show the student the modified output.

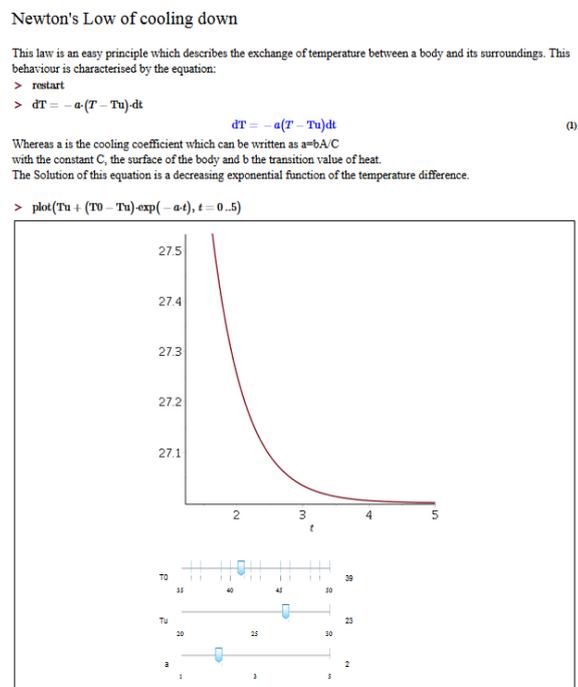


Figure 8: A modelling example realised as MathApp in Maple T.A. is shown here.

In the modelling and simulation courses a test is performed. The fact that the grading of these MathApps can now be done in Maple T.A. directly eliminates the usage of moodle for testing. But moodle is still used to organize the lecture dates and materials and links directly without further login to the according examples in Maple T.A.

5. CONCLUSION

The new features of Maple T.A. support the merging of the three e-learning systems. Unfortunately the transition of the MMT example to the Maple T.A. system will take more time than expected due to the fact that the examples have to be implemented again. This time span can be bridged using both systems until the transition is done. In the end all the existing examples will be gathered inside one system with an easy administration and assessment.

REFERENCES

- Pöll C., Bicher M., Hafner I., Winkler S., Körner A., 2013. Making Modelling Teachable MMT. Proceedings of ERK – Internation Electronical and omputer Science Conference, pp. 164-168. Sept. 16-18, Portoroz (Slovenia)
- Atanasijević-Kunc M., Logar V., Karba R., Papić M., Kos A., 2011. Remote multivariable control design using a competition game. IEEE Transactions on Education, Vol. 54, No. 1, 97–103.
- Logar V., Karba R., Papić M., Atanasijević-Kunc M., 2011. Artificial and real laboratory environment in an e-learning competition. Mathematics and computers in simulation, Vol. 82, No. 3, 517–524.
- Kristöffel P., Körner A., Breitenecker F., Urbonaite V., 2010. An E-Learning Study: Using Maple T.A. to improve the Mathamatic Skills of Students. Proceedings of 7th Congress on Modelling and Simulation, pp. 1204-1208. Sept. 06-10, Prague (Czech Republic)

AUTHORS BIOGRAPHY

Stefanie Winkler passed her bachelor study in technical mathematics in 2012 and her master study as well in technical mathematics in 2014. Both theses covered applications of mathematical modelling and simulation. The master thesis in particular covers a comparison of different approaches to simulate groundwater pollution. Her fields of activity include numerical aspects of simulation, comparison and evaluation, simulation of partial differential equations as well as applied mathematics in engineering. Since 2014 she is project assistant in the research group of Mathematical Modelling and Simulation at the Institute for Analysis and Scientific Computing at Vienna University of Technology. She is currently working on her PhD in mathematics.

THE IMPACT OF PROCUREMENT AND INVENTORY MANAGEMENT IN OPERATIVE PERFORMANCE IN A SUPPLY CHAIN

Jorge Luis García-Alcaraz^(a), Aidé Aracely Maldonado-Macías^(b) Emilio Jiménez Macías^(c) Julio Blanco Fernández^(d)

^(a) Autonomous University of Ciudad Juarez, Ciudad Juarez, Mexico

^(b) Autonomous University of Ciudad Juarez, Ciudad Juárez, Mexico

^(c) University of La Rioja, Logroño, Spain

^(d) University of La Rioja, Logroño, Spain

^(a) jorge.garcia@uacj.mx, ^(b) amaldona@uacj.mx, ^(c) emilio.jimenez@unirioja.es, ^(d) julio.blanco@unirioja.es

ABSTRACT

Supply chain is a major opportunity area to reduce costs and obtain operative performance, and that's why companies are looking to have a proper administration of the same. In this paper appears a structural equation model that integrates a total of 14 observed variables integrated in three latent variables, which are the raw materials procurement process, inventory management and operative performance of supply chain, generating three hypotheses related their relationships. A dataset was used containing 306 cases from maquiladoras in México. The model is run using partial least squares method. The results indicate that there is a direct and positive from raw material procurement on inventory management administration of 0.74 standard deviations, but in turn the first impact on 0.50 to operative performance.

Keywords: procurement, supply chain, causal model, PLS.

1. INTRODUCTION

Some authors indicate that at present times the competition between companies is not really in their production systems, but rather in their supply chains (CS), and this is due to the globalization processes that are currently (Hadjimarcou, Brouthers, McNicol, & Michie, 2013). Therefore, companies make a great effort to get a competent and efficient supply chain management (SCM), so there is a need to know the metrics on relations that exist between variables in CS.

A supply chain is a group of activities and entities in a production system ranging from the supply of raw materials to distribution as finished product. However, in a CS exists not only a flow of raw materials as discussed above but also there are flows of information and financial resources (Caridi, Moretto, Perego, & Tumino, 2014), as can be seen in Figure 1.

It is important to note that for some members in the supply chain a component or piece may be the end product, but for another, maybe is a raw material and that is why it is given chain name, which must be analyzed from a scientific point of view, albeit a complex problem (Feng, 2012) given the interactions

between components or members of it, where everyone has an important role; however, this complexity has not been discouraged for many researchers, as there are currently a lot studies that focus on the study of some of its components (Bertolini, Bottani, Rizzi, & Bevilacqua, 2007; Hartmann & De Grahl, 2011; Tang, Goetschalckx, & McGinnis, 2013).



Figure 1. Traditional supply chain

It is also important to note that supply chains from one product to another are completely different, but there is general agreement on several points, such as the activities to be carried out (Lu & Swaminathan, 2015). In that sense, a supply chain starts with a procurement process, which is performed by a supplier in the manufacturer warehouses, who by some kind of modeling and administrative rules, is responsible for giving the best use and maintain his inventory. From here, from the raw material warehouses, it goes to the production process to begin their transformation by a set of activities that will make it a finished product that will be shipped to a warehouse or store again. From here, through a distribution system, the company will send the finished product to end customers.

In the northern region of Mexico, specifically in Ciudad Juarez, there are a total of 324 maquiladoras, which are companies with foreign capital mainly, and import 100% of their raw materials from other countries, those companies assemble their products in México and they distribute their final product in countries like United States of America and European Union. This process makes the supply and inventory management are very peculiar and worthy of study (Sargent & Matthews, 2009).

Management of materials throughout the entire production process described above and the efficiency with it will make the company a competitive entity and hence the importance of good management thereof

(Lin, Chow, Madu, Kuei, & Pei, 2005). But surely, that all these activities are interrelated, and that the success of one of them has an impact on efficiency performance ratios for the company. So the aim of this article is to analyze the impact of the activities carried out during the raw material procurement process and their inventory management by the companies in the operating performance index for the supply chain.

1.1 Hypothesis

As mentioned earlier in this article three latent variables which are reported: *Procurement*, *Inventory management* and *Operative performance*. For know the impact from one latent variable on another, the following working hypothesis that are illustrated in Figure 2 are proposed.

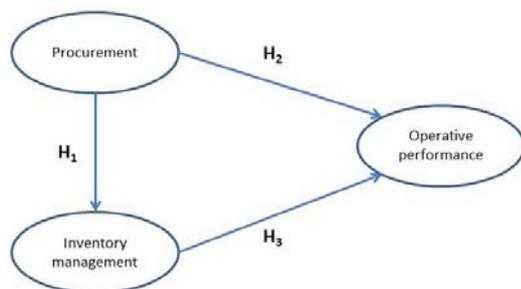


Figure 2: Proposed hypotheses

As mentioned above, supply chain starts with the supply of raw material from a supplier (Miocevic & Crnjak-Karanovic, 2012), that has been selected following strict quality standards and always looking to integrate him into the production process, so in large companies the supplier for a product are a few just because confidence levels generated between producer - provider are very high, enabling the establishment of philosophies as just in time purchasing. That relationships and how manufacturers are making the purchasing process will surely have an impact on inventory management by the company throughout the entire production system (David & Eben-Chaime, 2003).

Here it is important to note that the company manages inventory not only in their warehouses, but it should be done throughout whole production process (Dong, Carter, & Dresner, 2001; Selvarajah & Zhang, 2014). Very often, the companies for achieve their goals by managing their inventories; have established some regional distribution centers to help them do so quick shipments to their workplaces. However, to achieve all this, it requires the integration of many information and communication technologies to follow up the status of materials in each production stages (Jiang & Liu, 2015). Thus there are two types of policies here, those that relate to the process of sourcing from a supplier and that relate to the inventory management by manufacturer, which without doubt are related, therefore, it is proposed the following hypothesis.

H₁. There is a direct and positive relationship between the process of raw material *Procurement* by supplier and the *Inventory management* by manufacturer.

However, cooperation and good relations that have between manufacturer and supplier and their confidence levels, as well as the integration of both production systems, may also result in a number of benefits, which are reflected in operating performance index (Dey & Giri, 2014; Tseng, Wu, Lin, & Chiu, 2014). Thus, timely delivery by the supplier can represent very low cycle times, but also those deliveries to customers on time and complete production orders (Bertolini et al., 2007). But if production systems are properly integrated by information and communications technologies, then the SC has a virtual visibility to both, and that will facilitate decision making along the same (Caridi et al., 2014; Musa, Gunasekaran, & Yusuf, 2014). Thus, it appears that the process of raw materials procurement and the way how it is made, it will impact the operating efficiency ratios in manufacturer SC surely, so the following hypothesis is proposed.

H₂. There is a direct and positive relationship between the *Procurement* system of a company and the *Operative performance* in a supply chain.

In last hypothesis is assumed that there is a relationship between *Procurement* policies and *Operative performance* indices; however, there must be a relationship between *Inventory management* policies that the company has with the *Operative performance* (Brox & Fader, 1997), because if an *Inventory management* throughout the company is applicable in every departments, making use of just-in-time philosophy (Alcaraz, Maldonado, Iniesta, Robles, & Hernández, 2014) and with adapted information and communications technology, then it is possible to obtain timely deliveries to the end customer, have visibility in supply chain throughout the entire company and shorter cycle time (Karimi & Davoudpour, 2015; Selvarajah & Zhang, 2014). Therefore, the following hypothesis is proposed:

H₃. There is a direct and positive relationship between *Inventory management* policies and *Operative performance* indices in supply chain.

2. METHODOLOGY

The methodology that is used in this research involves the design of a data collection instrument (questionnaire) and identification of benefits that are obtained after a successful JIT implementation process. Then the survey has been applied to active managers in manufacturing industries to collect information, do some statistical analysis and get a conclusion based on findings, so the work is executed on different stages described below.

2.1 Survey development

This stage is focused on the design of a survey based on findings reported by (Soin, 2004) and a literature review is conducted. Three latent variables are analyzed in this research, but each one is integrated by another observable variables or item. In Table 1 appears their distribution, *Procurement* with 4 items, inventory management with 6 items and finally, Operative performance with 4 items.

Table 1: Latent variables and items

Procurement
Cooperation and relationship with suppliers (Ahi & Searcy, 2015; Bhagwat & Sharma, 2007; Cho, Lee, Ahn, & Hwang, 2012; Merschmann & Thonemann, 2011; Soin, 2004)
Reduction in the number of suppliers (Blome & Schoenherr, 2011; Jayaram, Dixit, & Motwani, 2014; Soin, 2004)
Delivery of raw material using JIT philosophy (JL. García-Alcaraz, Rivera, Blanco, Jiménez, & Martínez, 2014; Inman, Sale, Green Jr, & Whitten, 2011; Soin, 2004)
Raw material buying process are looking the best price (Mansoornejad, Pistikopoulos, & Stuart, 2013; Soin, 2004)
Inventory management
There is coordination and inventory management in whole Company (Soin, 2004) (Cirtita & Glaser-Segura, 2012; Falk & Hogström, 2000)
Deliveries inside the company are made using a JIT philosophy (Inman et al., 2011; Singh & Garg, 2011; Soin, 2004)
There is a management inventory control in all production process (Avelar-Sosa, García-Alcaraz, Vergara-Villegas, Maldonado-Macias, & Alor-Hernández, 2014; Caridi et al., 2014; Ramanathan, 2014)
The Company is focused in use low inventory level (Y.-M. Lee, Mu, Shen, & Dessouky, 2014; Pan & Liao, 1989)
There are regional distribution centers (Cheng & Tsai, 2009; Hu, Lu, Li, Zhang, & Zhang, 2014; Ou & Chou, 2009)
There are automatized warehouse systems (Berné, García-González, García-Uceda, & Múgica, 2015; Bertolini et al., 2007; Mensah, Merkurjev, & Longo, 2015)
Operative performance
Deliveries to customers are on time and complete (Cater & Cater, 2009; Fawcett, Calantone, & Smith, 1997)
The lead time between vendor and manufacturer are low (Bertolini et al., 2007; Li, Xu, & Ye, 2011)
There is a visible supply chain (Caridi et al., 2014; H. Lee, Kim, & Kim, 2014; Williams, Roh, Tokar, & Swink, 2013)
There is a high product customization level (Brun & Zorzini, 2009; Ngniatedema, Fono, & Mbondo, 2015; Yao & Liu, 2009)

The questionnaire answered on a Likert-based-scale on subjective assessments, where the lower value (1) indicated that the task never is done, and the highest value (5) represents that the task or operative index is always obtained. But also in the judge's validation, the first questionnaire contains blank spaces where the respondents could incorporate some other specific task that are not included in the initial questionnaire.

2.2 Data Collection

For data collection, the sample is stratified and focused on maquiladora industries that have a maturity supply chain. 324 companies are contacted via email.

For the survey application, three strategies are applied. The first one consists in face to face interviews with managers who work in supply chain departments or relate to material flow in industries established in Chihuahua, Mexico.

The second strategy consists of e-mails sent to some company managers to survey and answer within two weeks. After that time, a reminder is send and after three unsuccessful attempts, the case is abandoned. The third strategy consists in sending to every manager a link to answer the survey in a specialized web page for surveys application.

2.3 Capturing Information and Questionnaire Validation

At this stage the information is captured and analyzed using SPSS 21® software. Internal consistency or reliability of the questionnaire for each latent variable is performed using the Cronbach coefficient and composite reliability index (Cronbach, 1951; Liu, Ke, Wei, & Hua, 2013), considering a minimum cutoff values of 0.7 (Fornell & Larcker, 1981; Nunnaly, 1978; Nunnaly & Bernstein, 1994; Rexhausen, Pibernik, & Kaiser, 2012). Additionally, some tests are also performed at this stage to improve the quality of the questionnaire and the reliability in analyzed dimensions, since analyzing the elimination of some items, often the reliability increases (Nunnaly & Bernstein, 1994) and the procedure used were by (Blome, Schoenherr, & Eckstein, 2014; Lin et al., 2005; Ramanathan & Gunasekaran, 2014; Zailani, Jeyaraman, Vengadasan, & Premkumar, 2012) in supply chain surveys.

Also, this stage included a data screening process in order to detect missing values, which are then replaced using the median, because data is obtained by using an ordinal scale (Likert-based scale), although it is always kept in mind that there should be a maximum of 10% missing values for every item (Hair, Anderson, & Tatham, 1987; Hair, Black, Babin, & Anderson, 2009). Also, the values in the database are analyzed for outliers or extreme values and for this, a standardization process is executed for every item considering a standardized value as an outlier if its absolute value is bigger than 4 (Giaquinta, 2009; Hair et al., 2009; Kaiser, 2010; Rosenthal & Rosnow, 1991; Wold, Trygg, Berglund, & Antti, 2001). A similar procedure was used in supply chain research by (JL García-Alcaraz, Maldonado-Macias, Iniesta, Robles, & Hernández, 2014).

The discriminant validity is measured by the average variance extracted (AVE), and this measure is used by (Avelar-Sosa et al., 2014; JL García-Alcaraz, Maldonado, Alvarado, & Rivera, 2014) in a supply chain survey, and the minimum cutoff acceptable value for AVE is 0.5; while, for convergent validity assessment, the AVE's and correlations among latent variables are used (Fornell & Larcker, 1981; Kock,

2013). This procedure was used for discriminant validity by in supply chain research (Avelar-Sosa et al., 2014; JL García-Alcaraz, AA Maldonado-Macias, et al., 2014; JL García-Alcaraz, AidéA Maldonado, et al., 2014).

However, in regression analysis is important to measure the collinearity among latent variables, because if there is a high collinearity, then the inverse matrix estimation is very difficult, and then the full collinearity VIFs (variance inflation factor) value is used and the maximum cutoff value is 3.3 (Cenfetelli & Bassellier, 2009; Petter, Straub, & Rai, 2007). But some authors suggest more relaxed values, i.e. bigger than 10 (Hair et al., 1987; Hair et al., 2009; Kline, 1998; Petter et al., 2007).

Also, considering that the survey is answered on an ordinal scale using only assessments and not measurements, then the Q-squared coefficient is used since it is a nonparametric measure traditionally calculated via blindfolding. Q-squared coefficient is also used for the assessment of the predictive validity (or relevance) associated to each latent variable in the model. Acceptable predictive validity in connection with an endogenous latent variable is suggested by a Q-squared coefficient greater than zero (Kock, 2013) and preferably, must be similar to R-Squared values.

2.4 Descriptive Analysis

This stage focuses on a univariate analysis for identifying the central tendency and deviation measures in items collected. As a central tendency measure, the median or percentile 50th is obtained; where high values indicate that the task is always done; lower values indicate that those tasks are not done or the operative index is not obtained. Also, as deviation measure, the interquartile range (IR) is obtained (difference between percentile 75th and percentile 25th). High values in IR indicate that the task listed does not present agreement among respondents, while lower values represent little dispersion in those items (Tastle & Wierman, 2007) and therefore, a greater consensus among respondents. Similar interpretations were applied in supply chain research by (JL García-Alcaraz, AA Maldonado-Macias, et al., 2014; JL García-Alcaraz et al., 2014).

2.5 Structural Equation Model

In order to prove the hypotheses stated in Figure 1, the model is evaluated using the Structural Equation Modelling (SEM) technique, due to its widely and recent use in causal relations validations and specifically in the supply chain. For example, the impact of JIT in supply chain performance (Green Jr, Inman, Birou, & Whitten, 2014), the flexibility, uncertainty and firm performance in supply chain (Merschmann & Thonemann, 2011) and the effect of green supply chain management on green performance and firm competitiveness (Yang, Albert, & Carlo, 2013).

The SEM model is executed in WarpPLs 3.0® software because its main algorithms are based on Partial Least Squared (PLS), widely recommended for low sample size (Kock, 2013). The model here presented is specifically executed using the WarpPLs3 PLS algorithm, with a bootstrapping resampling method for a better coefficients values convergence and diminish the effect of possible outliers.

Three model fit indices are analyzed: average path coefficient (APC), the average R-squared (ARS) and average variance inflation factor (AVIF) that are proposed by (Kock, 2013) and used by (Ketkar & Vaidya, 2012) in the supply chain environment. For the APC and ARS, the p-values are analyzed in determining the model efficiency, establishing a maximum cutoff p-value of 0.05, which mean that the inferences are made with 95% of confidence level, testing the null hypotheses that APC and ARS are equal to 0, versus the alternative hypotheses that APC and ARS are different to zero; while for AVIF, values low of 5 are desirable.

Three different effects are measured in the structural equation model: (1) direct effect (that appears in Figure 1 as arrows from a latent variable to another), (2) indirect effect (given for paths with two or more segments), and (3) total effects (the sum of direct and indirect effects), and with the aim to determine their significance, the P values are analyzed, considering the null hypothesis: $\beta_i = 0$, versus the alternative: hypothesis $\beta_i \neq 0$.

3. RESULTS

3.1 Survey validation

Before starting with the information analysis obtained through the questionnaire, the validation statistical validation process was done in order to see if it was reliable. In Table 2, the indexes obtained for each of the latent variables are illustrated.

Table 2: Latent variables coefficients

Index	Procurement	Inventory Management	Operative Performance
R-squared		0.54	0.521
Adj. R-Squared		0.539	0.518
Composite Reliability	0.857	0.903	0.86
Cronbach's alfa	0.778	0.87	0.783
Avg. var. extracted	0.601	0.607	0.606
Full collin VIF	2.654	2.306	1.999
Q-squared		0.541	0.517

Note that Cronbach's alpha index is greater than 0.7 in every latent variables analyzed, which also occurs with the composite reliability index, so it is concluded that the measuring instrument has internal validity. Regarding convergent validity, it is observed that the

average variance extracted (AVE) is greater than 0.5 in every variables, so also it is concluded that the questionnaire is adequate. Likewise, it is observed that the square R-squared and R - square adjusted have values above 0.2, so it is concluded that the questionnaire also had predictive validity.

Regarding the linearity in dataset, it appears that none variable is higher than 3.3, the maximum allowed value and this led to conclude that there is no problem with collinearity. Given the above, it is concluded that the used questionnaire for obtain information from companies is adequate, so it can be used for further analysis.

3.2 Descriptive analysis

Table 3 briefly illustrates the descriptive analysis for variables analyzed, which contain the median as measure of central tendency and interquartile range (IR) for deviation.

Table 3: Descriptive analysis

Procurement	Median	IR
Cooperation and relationship with suppliers	4.39	1.31
Reduction in the number of suppliers	3.81	1.54
Delivery of raw material using JIT philosophy	4.23	1.48
Raw material buying process are looking the best price	4.27	1.38
Inventory management		
There is coordination and inventory management in whole company	4.28	1.40
Deliveries inside the company are made using a JIT philosophy	4.3	1.38
There is a management inventory control in all production process	4.13	1.40
The Company is focused in use low inventory level	4.17	1.49
There are regional distribution centers	3.96	1.46
There are automatized warehouse systems	3.81	1.72
Operative performance		
Deliveries to customers are on time and complete	4.39	1.21
The lead time between vendor and manufacturer are low	4.14	1.43
There is a visible supply chain	4.12	1.37
There is a high product customization level	4.18	1.35

In relation to the *Procurement* process, it appears that the *Cooperation and relationship with suppliers* has the highest median value, followed closely by *Raw material buying process are looking the best price*. Note that in this dimension interquartile ranges have all values above the unit, indicating a high dispersion in assessments from responders.

In relation to *Inventory management*, the items that have the highest value in the medium concern inside the company *Deliveries are made using a JIT philosophy* and *There is coordination and inventory management in whole company*, which indicates that just in time (JIT) programs, coordination and materials management throughout the production system are of vital importance for the company. This latent variable is also noted that all values and interquartile ranges in items are greater than one, and even in the latter, which refers *There are automatized warehouse systems* has a value very close to two, which indicates that there were variability by respondents.

Finally, the third dimension refers to Operative performance, is the only one with medium highest than four in all items that integrate it and the most important item concerns is *Deliveries to customers are on time and complete*, that also has the smallest interquartile range.

3.3 Structural equation model

In Figure 2, it was proposed a structural equation model containing three different hypotheses. In Figure 3 appears the solved model using the algorithm explained in the methodology section.

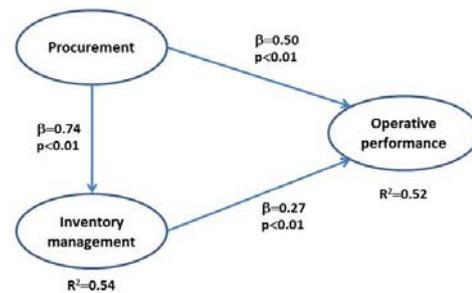


Figure 3: Structural equation model evaluated

In Figure 3 is observed a value for beta and a value for p for each relationship between the latent variables analyzed. Importantly, the beta values are standardized, thus representing an intensity change measured in standard deviations. P values represent the hypothesis test performed to determine if the beta value is zero or not. Also remember that all hypothesis tests were performed with a confidence level of 95%, indicating that the level of significance was only 5%. Accordingly, since all p-values are less than 0.05, we conclude that all relationships are statistically significant.

Similarly in each dependent latent variable appears a value for the R-squared, which is a measure of the dependence between a variable and another. However, it is important to split the value of R-square *Operative performance*, as it is influenced by two variables, the raw materials *Procurement* and *Inventory management*. In this case, the total R-square is 0.52, but 0.35 is from *Inventory management* and 0.17 comes from

Procurement, leading to the conclusion that first variable given its level of influence is much more important.

4. CONCLUSIONS

As mentioned in paragraphs above, based on the values related to relationships between latent variables, we can conclude in relation to the initial hypotheses, but also inferences applicable to the industry can be done, which are described below.

4.1 Conclusions regarding hypothesis

H₁. There is enough statistical evidence to declare that the raw materials *Procurement* has a direct and positive impact on *Inventory management*, because when the first latent variable increases its standard deviation by one unit, the second one goes up in 0.74 units.

H₂. There is enough statistical evidence to declare that the raw materials *Procurement* has a direct and positive impact on *Operative performance* indexes of supply chain, because when the first latent variable increases its standard deviation in one unit, the second one goes up by 0.50 units. However, there is also an indirect effect of two segments that is given through the *Inventory management* process, which is 0.1998 units. Thus, the total effect is 0.6998, indicating a strong relationship between these variables.

H₃. There is enough statistical evidence to declare that *Inventory management* has a direct and positive impact on *Operative performance* in supply chain, because when the first latent variable increases its standard deviation in one unit, the second one goes up 0.27 units.

4.2 Industrial inferences

Based on conclusion regarding the hypotheses and the statistical results obtained, the following industrial inferences can be done:

1. Depending on the relationships the company has with its suppliers, the ability to supply they have and the level of integration with the manufacturer, the latter must generate a policy for handling materials inventories received.
2. According to univariate statistical analysis, it was found that the most important item in dimension associated with *Procurement* is the relationship between purchaser and manufacturer, so both should pay attention to it.
3. While the raw material deliveries are made by applying a just in time philosophy, it is important that the manufacturer perform inventory management throughout the

production process and not only in storage at the beginning of production process.

4. The company will always make a great effort to deliver production orders on time and complete, and that was the item that received the median with a higher value, and apparently is an index highly valued by managers.
5. The manufacturer must pay special attention to the raw materials *Procurement* process, since the total effect (sum of direct and indirect effects) is 0.6998, a very high value.

REFERENCES

- Ahi, P., & Searcy, C. (2015). An analysis of metrics used to measure performance in green and sustainable supply chains. *Journal of Cleaner Production*, 86(0), 360-377. doi: <http://dx.doi.org/10.1016/j.jclepro.2014.08.005>
- Alcaraz, J. L. G., Maldonado, A. A., Iniesta, A. A., Robles, G. C., & Hernández, G. A. (2014). A systematic review/survey for JIT implementation: Mexican maquiladoras as case study. *Computers in Industry*, 65(4), 761-773. doi: <http://dx.doi.org/10.1016/j.compind.2014.02.013>
- Avelar-Sosa, L., García-Alcaraz, J., Vergara-Villegas, O., Maldonado-Macias, A., & Alor-Hernández, G. (2014). Impact of traditional and international logistic policies in supply chain performance. *The International Journal of Advanced Manufacturing Technology*, 1-13. doi: 10.1007/s00170-014-6308-3
- Berné, C., García-González, M., García-Uceda, M. E., & Múgica, J. M. (2015). The effect of ICT on relationship enhancement and performance in tourism channels. *Tourism Management*, 48(0), 188-198. doi: <http://dx.doi.org/10.1016/j.tourman.2014.04.012>
- Bertolini, M., Bottani, E., Rizzi, A., & Bevilacqua, M. (2007). Lead time reduction through ICT application in the footwear industry: A case study. *International Journal of Production Economics*, 110(1-2), 198-212. doi: <http://dx.doi.org/10.1016/j.ijpe.2007.02.016>
- Bhagwat, R., & Sharma, M. K. (2007). Performance measurement of supply chain management: A balanced scorecard approach. *Computers & Industrial Engineering*, 53(1), 43-62. doi: <http://dx.doi.org/10.1016/j.cie.2007.04.001>
- Blome, C., & Schoenherr, T. (2011). Supply chain risk management in financial crises—A multiple case-study approach. *International Journal of Production Economics*, 134(1), 43-57. doi: <http://dx.doi.org/10.1016/j.ijpe.2011.01.002>
- Blome, C., Schoenherr, T., & Eckstein, D. (2014). The impact of knowledge transfer and complexity

- on supply chain flexibility: A knowledge-based view. *International Journal of Production Economics*, 147, Part B(0), 307-316. doi: <http://dx.doi.org/10.1016/j.ijpe.2013.02.028>
- Brox, J. A., & Fader, C. (1997). Assessing the impact of JIT using economic theory. *Journal of Operations Management*, 15(4), 371-388. doi: [http://dx.doi.org/10.1016/S0272-6963\(97\)00005-3](http://dx.doi.org/10.1016/S0272-6963(97)00005-3)
- Brun, A., & Zorzini, M. (2009). Evaluation of product customization strategies through modularization and postponement. *International Journal of Production Economics*, 120(1), 205-220. doi: <http://dx.doi.org/10.1016/j.ijpe.2008.07.020>
- Caridi, M., Moretto, A., Perego, A., & Tumino, A. (2014). The benefits of supply chain visibility: A value assessment model. *International Journal of Production Economics*, 151(0), 1-19. doi: <http://dx.doi.org/10.1016/j.ijpe.2013.12.025>
- Cater, B., & Cater, T. (2009). Relationship-value-based antecedents of customer satisfaction and loyalty in manufacturing. *Journal of Business & Industrial Marketing*, 24(7-8), 585-596. doi: 10.1108/08858620910999457
- Cenfetelli, R., & Bassellier, G. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly*, 33(4), 19.
- Cirtita, H., & Glaser Segura, D. A. (2012). Measuring downstream supply chain performance. *Journal of Manufacturing Technology Management*, 23(3), 299-314. doi: 10.1108/17410381211217380
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi: 10.1007/BF02310555
- Cheng, Y.-H., & Tsai, Y.-L. (2009). Factors influencing shippers to use multiple country consolidation services in international distribution centers. *International Journal of Production Economics*, 122(1), 78-88. doi: <http://dx.doi.org/10.1016/j.ijpe.2008.12.022>
- Cho, D. W., Lee, Y. H., Ahn, S. H., & Hwang, M. K. (2012). A framework for measuring the performance of service supply chain management. *Computers & Industrial Engineering*, 62(3), 801-818. doi: <http://dx.doi.org/10.1016/j.cie.2011.11.014>
- David, I., & Eben-Chaime, M. (2003). How far should JIT vendor-buyer relationships go? *International Journal of Production Economics*, 81-82(0), 361-368. doi: [http://dx.doi.org/10.1016/S0925-5273\(02\)00274-8](http://dx.doi.org/10.1016/S0925-5273(02)00274-8)
- Dey, O., & Giri, B. C. (2014). Optimal vendor investment for reducing defect rate in a vendor-buyer integrated system with imperfect production process. *International Journal of Production Economics*, 155(0), 222-228. doi: <http://dx.doi.org/10.1016/j.ijpe.2014.02.004>
- Dong, Y., Carter, C. R., & Dresner, M. E. (2001). JIT purchasing and performance: an exploratory analysis of buyer and supplier perspectives. *Journal of Operations Management*, 19(4), 471-483. doi: [http://dx.doi.org/10.1016/S0272-6963\(00\)00066-8](http://dx.doi.org/10.1016/S0272-6963(00)00066-8)
- Falk, H., & Hogström, L. (2000). *Key success factors for a functioning supply chain in e-commerce B2B*. (MsD.), Göteborg University. School of Business, .
- Fawcett, S. E., Calantone, R., & Smith, S. R. (1997). Delivery capability and firm performance in international operations. *International Journal of Production Economics*, 51(3), 191-204. doi: [http://dx.doi.org/10.1016/S0925-5273\(97\)00051-0](http://dx.doi.org/10.1016/S0925-5273(97)00051-0)
- Feng, Y. (2012). System Dynamics Modeling for Supply Chain Information Sharing. *Physics Procedia*, 25(0), 1463-1469. doi: <http://dx.doi.org/10.1016/j.phpro.2012.03.263>
- Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 18(1), 39-50.
- García-Alcaraz, J., Maldonado-Macias, A., Iniesta, A. A., Robles, G. C., & Hernández, G. A. (2014). A systematic review/survey for JIT implementation: Mexican maquiladoras as case study. *Computers in Industry*, 65(4), 761-773. doi: <http://dx.doi.org/10.1016/j.compind.2014.02.013>
- García-Alcaraz, J., Maldonado, A., Alvarado, A., & Rivera, D. (2014). Human critical success factors for kaizen and its impacts in industrial performance. *The International Journal of Advanced Manufacturing Technology*, 70(9-12), 2187-2198. doi: 10.1007/s00170-013-5445-4
- García-Alcaraz, J., Rivera, L., Blanco, J., Jiménez, E., & Martínez, E. (2014). Structural equations modelling for relational analysis of JIT performance in maquiladora sector. *International Journal of Production Research*, 52(17), 4931-4949. doi: 10.1080/00207543.2014.885143
- Giaquinta, M. (2009). *Mathematical analysis: An introduction to functions of several variables*. New York, NY: Springer.
- Green Jr, K. W., Inman, R. A., Birou, L. M., & Whitten, D. (2014). Total JIT (T-JIT) and its impact on supply chain competency and organizational performance. *International Journal of Production Economics*, 147, Part A(0), 125-135. doi: <http://dx.doi.org/10.1016/j.ijpe.2013.08.026>
- Hadjimarcou, J., Brouthers, L. E., McNicol, J. P., & Michie, D. E. (2013). Maquiladoras in the 21st

- century: Six strategies for success. *Business Horizons*, 56(2), 207-217. doi: <http://dx.doi.org/10.1016/j.bushor.2012.11.005>
- Hair, J., Anderson, R., & Tatham, R. (1987). *Multivariate data analysis*. New York, NY.: Macmillan.
- Hair, J., Black, W., Babin, B., & Anderson, R. (2009). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.
- Hartmann, E., & De Grahl, A. (2011). THE FLEXIBILITY OF LOGISTICS SERVICE PROVIDERS AND ITS IMPACT ON CUSTOMER LOYALTY: AN EMPIRICAL STUDY. *Journal of Supply Chain Management*, 47(3), 63-85. doi: 10.1111/j.1745-493X.2011.03228.x
- Hu, X.-H., Lu, C.-Z., Li, M., Zhang, C.-H., & Zhang, H. (2014). Mathematical modeling for selecting center locations for medical and health supplies reserve in Hainan Province. *Asian Pacific Journal of Tropical Medicine*, 7(2), 160-163. doi: [http://dx.doi.org/10.1016/S1995-7645\(14\)60014-6](http://dx.doi.org/10.1016/S1995-7645(14)60014-6)
- Inman, R. A., Sale, R. S., Green Jr, K. W., & Whitten, D. (2011). Agile manufacturing: Relation to JIT, operational performance and firm performance. *Journal of Operations Management*, 29(4), 343-355. doi: <http://dx.doi.org/10.1016/j.jom.2010.06.001>
- Jayaram, J., Dixit, M., & Motwani, J. (2014). Supply chain management capability of small and medium sized family businesses in India: A multiple case study approach. *International Journal of Production Economics*, 147, Part B(0), 472-485. doi: <http://dx.doi.org/10.1016/j.ijpe.2013.08.016>
- Jiang, X., & Liu, Y. (2015). Global value chain, trade and carbon: Case of information and communication technology manufacturing sector. *Energy for Sustainable Development*, 25(0), 1-7. doi: <http://dx.doi.org/10.1016/j.esd.2014.12.001>
- Kaiser, H. (2010). *Mathematical programming for agricultural, environmental, and resource economics*. Hoboken, NJ: Wiley.
- Karimi, N., & Davoudpour, H. (2015). A branch and bound method for solving multi-factory supply chain scheduling with batch delivery. *Expert Systems with Applications*, 42(1), 238-245. doi: <http://dx.doi.org/10.1016/j.eswa.2014.07.025>
- Ketkar, M., & Vaidya, O. S. (2012). Study of Emerging Issues in Supply Risk Management in India. *Procedia - Social and Behavioral Sciences*, 37(0), 57-66. doi: <http://dx.doi.org/10.1016/j.sbspro.2012.03.275>
- Kline, R. (1998). *Principles and practice of structural equation modeling*. New York, NY, USA: The Guilford Press.
- Kock, N. (2013). Using WarpPLS in e-collaboration studies: What if I have only one group and one condition. *International Journal of e-Collaboration*, 9(3), 12.
- Lee, H., Kim, M. S., & Kim, K. K. (2014). Interorganizational information systems visibility and supply chain performance. *International Journal of Information Management*, 34(2), 285-295. doi: <http://dx.doi.org/10.1016/j.ijinfomgt.2013.10.03>
- Lee, Y.-M., Mu, S., Shen, Z., & Dessouky, M. (2014). Issuing for perishable inventory management with a minimum inventory volume constraint. *Computers & Industrial Engineering*, 76(0), 280-291. doi: <http://dx.doi.org/10.1016/j.cie.2014.08.007>
- Li, Y., Xu, X., & Ye, F. (2011). Supply chain coordination model with controllable lead time and service level constraint. *Computers & Industrial Engineering*, 61(3), 858-864. doi: <http://dx.doi.org/10.1016/j.cie.2011.05.019>
- Lin, C., Chow, W., Madu, C., Kuei, C., & Pei, Y. (2005). A structural equation model of supply chain quality management and organizational performance. *International Journal of Production Economics*, 96(3), 355-365.
- Liu, H., Ke, W., Wei, K. K., & Hua, Z. (2013). The impact of IT capabilities on firm performance: The mediating roles of absorptive capacity and supply chain agility. *Decision Support Systems*, 54(3), 1452-1462. doi: <http://dx.doi.org/10.1016/j.dss.2012.12.016>
- Lu, L. X., & Swaminathan, J. M. (2015). Supply Chain Management. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 709-713). Oxford: Elsevier.
- Mansoornejad, B., Pistikopoulos, E. N., & Stuart, P. (2013). Metrics for evaluating the forest biorefinery supply chain performance. *Computers & Chemical Engineering*, 54(0), 125-139. doi: <http://dx.doi.org/10.1016/j.compchemeng.2013.03.031>
- Mensah, P., Merkurjev, Y., & Longo, F. (2015). Using ICT in Developing a Resilient Supply Chain Strategy. *Procedia Computer Science*, 43(0), 101-108. doi: <http://dx.doi.org/10.1016/j.procs.2014.12.014>
- Merschmann, U., & Thonemann, U. W. (2011). Supply chain flexibility, uncertainty and firm performance: An empirical analysis of German manufacturing firms. *International Journal of Production Economics*, 130(1), 43-53. doi: <http://dx.doi.org/10.1016/j.ijpe.2010.10.013>
- Miocevic, D., & Crnjak-Karanovic, B. (2012). The mediating role of key supplier relationship management practices on supply chain orientation—The organizational buying effectiveness link. *Industrial Marketing Management*, 41(1), 115-124. doi:

- <http://dx.doi.org/10.1016/j.indmarman.2011.11.015>
- Musa, A., Gunasekaran, A., & Yusuf, Y. (2014). Supply chain product visibility: Methods, systems and impacts. *Expert Systems with Applications*, 41(1), 176-194. doi: <http://dx.doi.org/10.1016/j.eswa.2013.07.020>
- Ngniatedema, T., Fono, L. A., & Mbondo, G. D. (2015). A delayed product customization cost model with supplier delivery performance. *European Journal of Operational Research*, 243(1), 109-119. doi: <http://dx.doi.org/10.1016/j.ejor.2014.11.017>
- Nunnally, J. (1978). *Psychometric theory*. New York, NY: Mc Graw Hill.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York, NY: Mc Graw Hill.
- Ou, C.-W., & Chou, S.-Y. (2009). International distribution center selection from a foreign market perspective using a weighted fuzzy factor rating system. *Expert Systems with Applications*, 36(2, Part 1), 1773-1782. doi: <http://dx.doi.org/10.1016/j.eswa.2007.12.007>
- Pan, A. C., & Liao, C. J. (1989). An Inventory Model under Just-in Time Purchasing Agreements. *Production and Inventory Management Journal*, 30(1), 49-52.
- Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 31(4), 623-656.
- Ramanathan, U. (2014). Performance of supply chain collaboration – A simulation study. *Expert Systems with Applications*, 41(1), 210-220. doi: <http://dx.doi.org/10.1016/j.eswa.2013.07.022>
- Ramanathan, U., & Gunasekaran, A. (2014). Supply chain collaboration: Impact of success in long-term partnerships. *International Journal of Production Economics*, 147, Part B(0), 252-259. doi: <http://dx.doi.org/10.1016/j.ijpe.2012.06.002>
- Rexhausen, D., Pibernik, R., & Kaiser, G. (2012). Customer-facing supply chain practices—The impact of demand and distribution management on supply chain success. *Journal of Operations Management*, 30(4), 269-281. doi: <http://dx.doi.org/10.1016/j.jom.2012.02.001>
- Rosenthal, R., & Rosnow, R. (1991). *Essentials of behavioral research: Methods and data analysis*. Boston, MA: Mc Graw Hill.
- Sargent, J., & Matthews, L. (2009). China versus Mexico in the Global EPZ Industry: Maquiladoras, FDI Quality, and Plant Mortality. *World Development*, 37(6), 1069-1082. doi: <http://dx.doi.org/10.1016/j.worlddev.2008.10.02>
- Selvarajah, E., & Zhang, R. (2014). Supply chain scheduling at the manufacturer to minimize inventory holding and delivery costs. *International Journal of Production Economics*, 147, Part A(0), 117-124. doi: <http://dx.doi.org/10.1016/j.ijpe.2013.08.015>
- Singh, S., & Garg, D. (2011). JIT System: Concepts, Benefits and Motivation in Indian Industries. *International Journal of Management and Business Studies*, 1(1), 26-30.
- Soin, S. S. (2004). *Critical success factors in supply chain management at high technology companies*. (Ph.D.), University of Shouthern Queensland, Australia, 2004.
- Tang, Z. E., Goetschalckx, M., & McGinnis, L. (2013). Modeling-based Design of Strategic Supply Chain Networks for Aircraft Manufacturing. *Procedia Computer Science*, 16(0), 611-620. doi: <http://dx.doi.org/10.1016/j.procs.2013.01.064>
- Tastle, W. J., & Wierman, M. J. (2007). Consensus and dissent: A measure of ordinal dispersion. *International Journal of Approximate Reasoning*, 45(3), 531-545. doi: <http://dx.doi.org/10.1016/j.ijar.2006.06.024>
- Tseng, C.-T., Wu, M.-F., Lin, H.-D., & Chiu, Y.-S. P. (2014). Solving a vendor-buyer integrated problem with rework and a specific multi-delivery policy by a two-phase algebraic approach. *Economic Modelling*, 36(0), 30-36. doi: <http://dx.doi.org/10.1016/j.econmod.2013.09.013>
- Williams, B. D., Roh, J., Tokar, T., & Swink, M. (2013). Leveraging supply chain visibility for responsiveness: The moderating role of internal integration. *Journal of Operations Management*, 31(7-8), 543-554. doi: <http://dx.doi.org/10.1016/j.jom.2013.09.003>
- Wold, S., Trygg, J., Berglund, A., & Antti, H. (2001). Some recent developments in PLS modeling. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 131-150. doi: [http://dx.doi.org/10.1016/S0169-7439\(01\)00156-3](http://dx.doi.org/10.1016/S0169-7439(01)00156-3)
- Yang, S., Albert, R., & Carlo, T. A. (2013). Transience and constancy of interactions in a plant-frugivore network. *Ecosphere*, 4(12), art147. doi: 10.1890/ES13-00222.1
- Yao, J., & Liu, L. (2009). Optimization analysis of supply chain scheduling in mass customization. *International Journal of Production Economics*, 117(1), 197-211. doi: <http://dx.doi.org/10.1016/j.ijpe.2008.10.008>
- Zailani, S., Jeyaraman, K., Vengadasan, G., & Premkumar, R. (2012). Sustainable supply chain management (SSCM) in Malaysia: A survey. *International Journal of Production Economics*, 140(1), 330-340. doi: <http://dx.doi.org/10.1016/j.ijpe.2012.02.008>

CEREBELLUM FUNCTION FOR MSAAS

Erdal Cayirci^(a), Hakan Karapinar^(b), Lutfu Ozcakir^(c)

^(a) Electrical Engineering and Computer Science Department, University of Stavanger, Norway

^{(b),(c)} Simulation, Training and Test Systems Department, HAVELSAN, Turkey

^(a) erdal.cayirci@uis.no, ^(b) hakank@havelsan.com.tr, ^(c) ozcakir@havelsan.com.tr

ABSTRACT

Modelling and simulation as a service (MSaaS) offers elasticity, better utilization, higher reusability and many more advantages. However, MSaaS has a number of challenges studied in the literature. One of these challenges, namely time sensitivity of the services in an MSaaS federation, is introduced. Our scheme, called cerebellum function, is designed to address this new challenge. The conditions related to time sensitivity and the algorithm for the cerebellum function are also presented.

Keywords: modelling and simulation as a service, cloud, service, MSaaS, cerebellum function, quality of service, QoS

1. INTRODUCTION

Modelling and Simulation as a Service (MSaaS) is a model for provisioning, modelling and simulation (M&S) services on demand from an MSaaS provider (MP), which keeps the underlying infrastructure, platform and software details hidden from the MSaaS Customers (MC) (Cayirci, 2013a; Johnson, 2013). MP is responsible for licenses, software upgrades, scaling the infrastructure according to evolving requirements and accountable to the MC for providing grade of service (GoS) and quality of service (QoS) specified in the service level agreements (SLA). MSaaS introduces better utilization, ease in technical administration and therefore cost reduction. It also implies a big paradigm shift in computing and a long list of challenges related to both its ecosystem and technical requirements. Academia and industry have already tackled with many of those challenges (Cayirci, 2013a; Cayirci, 2013b; Cayirci, 2013c; Cayirci, 2014; Cayirci, 2015; Jensen, 2009; Subashini, 2012). In this paper, we focus yet another challenge for MSaaS: delay sensitivity of some M&S services, specifically military MSaaS; and present the preliminary results from our research on a new scheme that we call as cerebellum function for MSaaS.

A cloud can provide three basic service types (i.e., service models) as shown in Figure 1 (Ambrust, 2010; Badger, 2012; Cayirci, 2013a):

- Infrastructure as a Service (IaaS),
- Platform as a Service (PaaS) and
- Software as a Service (SaaS).

There are also many other service types introduced in literature, such as, Network as a Service (NaaS), Trust as a Service, Authorization as a Service (Laborde 2013). These are derivations of PaaS and SaaS in various combinations and forms. MSaaS can be perceived as one of these derivatives. MSaaS is in essence a special form of SaaS. The inter-relations between MSaaS, SaaS, PaaS and IaaS are depicted in Figure 1. We consider three types of MSaaS:

- Modelling as a service,
- Model as a service and
- Simulation as a service.

An MC may develop models by using modelling as a service, use previously developed models to run simulations in their enterprise (i.e., model as a service) or run simulations by using simulation as a service.

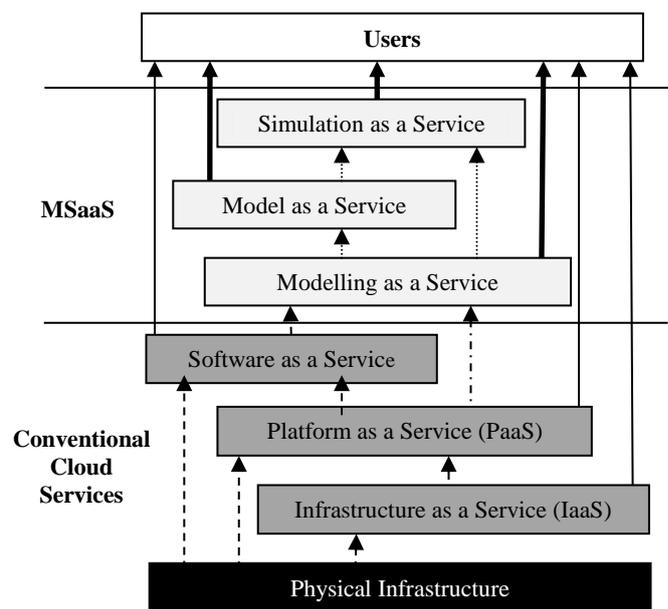


Figure 1: The Inter-relations of Cloud Services Including MSaaS

Please note that a cloud (i.e., a cloud service provider) typically maintains multiple data centers remotely located from each other. A data center is a facility that houses server pools and infrastructure to store, to process and to communicate large volumes of data.

Hence, MSaaS can be designed in one of the following forms:

- Standalone MSaaS applications: Standalone applications, such as business process modelling and supply chain simulation (Rosetti 2012), are already available as simulation as a service in the Internet.
- Federated standalone MSaaS applications: Standalone MSaaS applications can be federated. These applications can be from the same data center or multiple data centers.
- Composed MSaaS: Not standalone applications, but services and data that can be integrated into a composite service are offered as MSaaS.
- Automatically composed MSaaS: As the technology and interoperability among the services mature, composed MSaaS become automatically discoverable and composable with each other.

In this paper, we focus on composed and automatically composed MSaaS. As explained in (Cayirci 2013a; Cayirci 2013b), composing an MSaaS from the services provided by multiple data centers is a challenging task. In the literature, composed cloud services are called as service mash-ups or service federations. We will use the term “MSaaS federation” for a composed MSaaS, and the term “federate” for each service that the federation is composed of. Federation has a different meaning in cloud computing from M&S. In cloud computing, the term “federation” is used not only for federating models but also for infrastructure or platforms, and therefore a federation may also mean a cloud service that integrates various resources in the form of IaaS (e.g., memory, processor time, etc) from multiple data centers (Buyya, 2010; Cayirci, 2013b; Singhal, 2013; Toosi 2011). On the other hand, in MSaaS domain, federations integrate multiple MSaaS either in standalone application or service module form. Please note that we do not imply any architecture, such as high level architecture (HLA), when we use the term federation, but integration of various M&S services for composing an MSaaS. We categorize MSaaS federations into four broad classes as depicted in Table 1 (Cayirci 2013a):

- Type 0: Federation of standalone applications located in the same data center (Toosi 2011)
- Type 1: Composite MSaaS of service modules located in the same data center
- Type 2: Federation of standalone applications from multiple data centers (Cayirci 2013b)
- Type 3: Composite MSaaS of service modules from multiple data centers (Cayirci 2013b).

Table 1: Types of MSaaS Federations

Nature of Federates	Intra datacentre	Inter datacenter
Standalone applications	Type 0	Type 2
Composed services (SOA)	Type 1	Type 3

MSaaS intrinsically introduces a new challenge, namely the physical distance and therefore the propagation delay in between the user device and the cloud. This challenge is exacerbated by the additional computational delay due to service federating. In the Internet, few hundred milliseconds of round trip times (RTT) can be expected. Based on the physical distance and other computational delays, RTT can be significantly higher than few hundred milliseconds, which may become an important issue for interactive MSaaS. In this paper, we elaborate the dynamics of this challenge, and introduce our new cerebellum function for MSaaS scheme, which is developed to tackle with that.

In the following section, we examine various potential services for military MSaaS federations. Their quality of service requirements are also discussed and listed qualitatively. In Section 3, we explain our cerebellum function scheme briefly and introduce a practical algorithm for configuring and locating cerebellum functions. We conclude our paper in Section 4.

2. POTENTIAL SERVICES IN A MILITARY SERVICE ORIENTED MSAAS

In this Section, we will focus on the potential services for a service oriented military MSaaS. Table 2 depicts a preliminary list of services. We expect that this list will have hundreds of services, and therefore, Table 2 is not even close to be exhaustive. However, Table 2 includes key MSaaS services which are already implemented by various organizations including HAVELSAN. These services can also be differentiated based on several factors, such as, level of fidelity, level of resolution, the type of simulation that they are designed for (i.e., live, virtual and constructive). Therefore, Table 2 is only a preliminary version of our work, and based on a focus group study. We are currently developing a more detailed and quantitative version of the same table.

In Table 2, we focus on four QoS parameters: reliability, bandwidth, jitter and delay. All the services in the Table are highly sensitive against reliability because they are based on the transfer of digital data. Bandwidth and jitter requirements/constraints are low for almost all services. On the other hand, almost all of these services have highly constrained delay requirements. Especially, the sensitivity of interactive visualization services (IVS) against delay is very high. It is clear that one of the most challenging M&S services with respect to the quality of service parameters is IVS. Therefore, we use IVS as our example in the later sections of this paper.

IVS is a key service for immersive M&S, which visualizes the virtual environment including the simulated entities (i.e., all objects including people) interactively, such that the angle and point of projection can be changed instantaneously based on user commands. Its quality and responsiveness is paramount for the immersion of the virtual simulation users, such as, the trainees in the aircraft and tank simulators. Moreover, the response time of IVS to the controls must

be the same as the real system. Otherwise, although the virtual simulation provides a perfect immersion, negative training can be given. For example, if a virtual aircraft responds the controls more rapidly or slowly than the real aircraft, the trainee may develop wrongly conditioned reflexes.

Table 2: Potential Cloud Services for Military MSaaS

Service	Reliability	Bandwidth	Jitter	Delay
Weapon Effects	High	Low	Low	High
Exterior Ballistics	High	Low	Low	High
Common Effects	High	Low	Low	High
Synthetic Environ.	High	High	Low	Medium
Synthetic Dynamic Environment	High	Medium	Low	Medium
Weather	High	Low	Low	Medium
Geography/Hydrograph	High	High	Low	Medium
Line of Sight	High	Low	Low	Medium
Order Translation	High	Low	Low	Low
Autonomous Planning	High	Low	Low	Low
Movement	High	Low	Low	Medium
Supply	High	Low	Low	Medium
Mainten.	High	Low	Low	Medium
Detection and Recognition	High	Low	Low	High
Comms	High	Low	Low	Medium
Human/Social Behavior	High	Low	Low	Medium
Troop/Platform/Unit Behavior	High	Low	Low	Medium
Threat Network	High	Low	Low	Medium
Recognized Picture	High	Medium	Medium	Medium
C4 Population and Stimulation	High	Medium	Medium	Medium
Visualiz.	High	High	High	High
Interactive Visualiz.	High	High	Very High	Very High

A cloud has two ends: front-end and back-end as shown in Figure 2. The front-end is where the user interfaces are. That is the only part of a cloud visible to the users, and should not need any special hardware. The user sees the back-end as a cloud without knowing any details about its internal architecture. The back-end includes

various components (i.e., storage space, processors and platforms) loosely coupled to each other through a mechanism that allows elasticity. The delineation between front and back ends introduce longer propagation delays between the user interface and the service comparing to conventional computation schemes where user interface and server are physically co-located. In addition to that, a composed MSaaS (i.e., an MSaaS federation) may have many MSaaS received from multiple data centers. Even an MSaaS federation may have a federate, which itself is a federation with federates from the other data centers. This nested architecture can definitely introduce jitter and delay, which is not manageable for services like IVS.

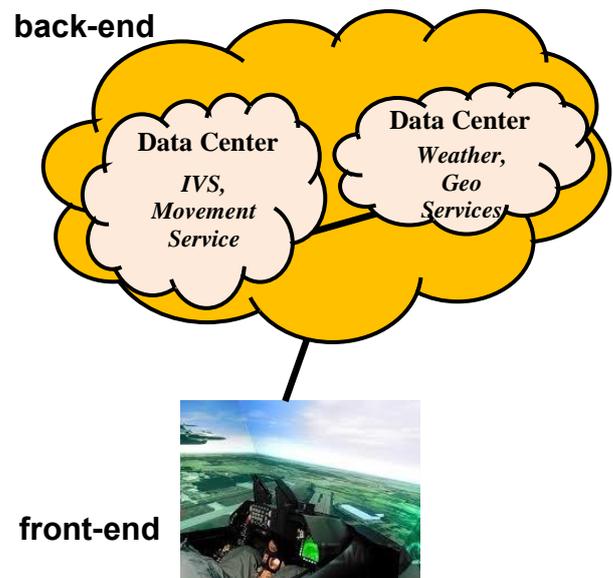


Figure 2: An Example for Back-end and Front-end in MSaaS (the services in the figure are just an example)

3. CEREBELLUM FUNCTION

Cerebellum function includes the part of an MSaaS federation which is time sensitive in responding the user commands (i.e., inputs). Please note that we do not mean shortest delay by time sensitivity, but the delay arranged the same as the delay in response by the real system to the user commands. For example, if the delay in real system d_r is between 90 and 100 msec, the delay in virtual system needs to be within exactly 90-100 msec window. It is expected that r is a random variable, which (i.e., both distribution and parameters) may change for various systems. Our scheme is based on the idea that the maximum delay between the user interface and cerebellum function d_{max} must be shorter than the lower bound of the real life system delay r_{amin} according to a given confidence level α . Hence, we can manage the delay such that negative training is avoided and immersion is maintained. The maximum delay d_{max} includes not only the propagation delay p_{max} introduced

by the physical distance between two ends of a communications link but also all sort of computational delays c_{max} due to processes, such as encryption, decryption, routing, service federating, etc. It is clear that we need to treat also $d_{\alpha max}$ as a random variable, and make our computations based on the upper bound according to the given confidence level α .

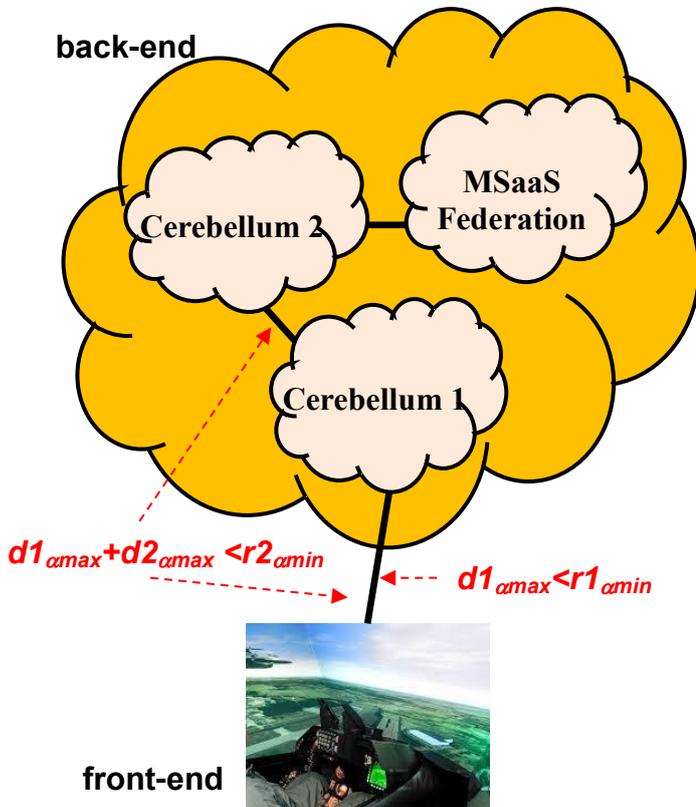


Figure 3: Cerebellum Function

When the services are designed, the designer should think about if a service is time sensitive, and if the time sensitive part of the service can be separated from the rest of the service. If there is a time sensitive part of a service, that part needs to be designed decomposable. Hence we do not need migrating all the service and data closer to the front end but only the time sensitive part of the service. For example, the part of IVS that fetches the terrain data and weather conditions and creating three dimensional virtual environments can be designed separately from the part that makes the projections based on the user commands. The later part, which is time sensitive, becomes the cerebellum function for IVS. Please note again that this is only a simplified example to clarify what we mean by cerebellum function. In some cases, not only the cerebellum function of a service, but all of the service may be treated as a cerebellum function depending on the configuration of an MSaaS federation. If an input of Service s_a uses another Service s_b , which has a part that needs to be treated within the cerebellum function, s_a as a complete service has to be within the cerebellum

function. Moreover a cerebellum function may also have a nested structure, which means that the inputs of a cerebellum function may be coming from another cerebellum function. Therefore, the location of a cerebellum function is selected such that the conditions in Equations 1 and 2 are met.

$$dn_{\alpha max} = \sum_{k=1}^n u(pk_{\alpha max}) + u(ck_{\alpha max}). \quad (1)$$

$$dn_{\alpha max} < u(rn_{\alpha max}). \quad (2)$$

where $n-1$ is the number of cerebellum functions preceding the cerebellum function n in the nested structure.

Algorithm 1 is designed for configuring and locating cerebellum functions. After selecting all the services in an MSaaS Federation, a dependency tree is constructed starting from the service that directly interacts with the user. That is typically the user interface for an MSaaS Federation. Then all the services in the service fan in (the services used by the service) of each service are inserted into the tree as a child node to the service.

Select all the services required for an MSaaS federation
Create dependency tree for the services in MSaaS Federation

Insert the top level service as the root node
 $n=0$

Repeat until none of the services at level n has another service in their service fan in

Insert all the services in the service fan in of a service at level n into the tree

$n=n+1$

Run depth first traverse of the tree and create cerebellum function structure

If Service a has a cerebellum function

If Service a has a parent Service b

If $u(ra_{max}) < u(rb_{max})$

Merge cerebellum functions

Service a = Service a – Cerebellum a

Cerebellum b = Service b + Cerebellum a

$u(rb_{max}) = u(ra_{max})$

Remove Cerebellum a

Locate the cerebellum functions according to their time sensitivity

Algorithm 1: Cerebellum Function Configuration Algorithm

When the dependency tree is complete, it is traversed in the depth first order. If a cerebellum function of a service has more stringent time sensitivity constraint comparing to its parent service, all of the parent service is treated as a cerebellum function together with the cerebellum function of the child service. The time sensitivity of the merged cerebellum function is assigned with the time sensitivity parameter of the child service. After the traverse of the tree and merging of the cerebellum functions are complete, the cerebellum

functions are migrated to appropriate machines. A cerebellum function can be located in any data center that accepts the function and satisfies the constraints in Equations 1 and 2 within the cloud or in the front end machine itself if none of the data centers or servers in the Cloud can provide the required conditions.

The cerebellum function for IVS can also support the design of architectures that fulfil security constraints of military MSaaS. Although the environmental data, and specifications of military equipment, such as maximum speed and altitude that a military aircraft can reach, are unclassified, the turn rates and similar data about the aircraft may be classified. Since the effects like turn rates are time sensitive and therefore will be typically treated by a cerebellum function in IVS, the cerebellum function approach may become useful also for tackling with the security related challenges of MSaaS. Since this is a different topic, we do not further elaborate on the cerebellum function for enhanced security in MSaaS in this paper.

4. CONCLUSION

MSaaS is an emerging approach for M&S following the latest trends in information technologies. It promises many advantages, such as rapid elasticity, ease in technical administration and licensing, better utilization, pays per use, and therefore enables considerable cost reduction. However, it also introduces many challenges including security, privacy, accountability, risk and trust management and service composition. Industry and academia have tackled with these challenges for almost a decade. In this paper, we introduce yet another challenge related to quality of service guarantees, specifically delay and jitter. When a service is a time sensitive service, it needs to be physically close enough to the front end. However, migrating all the service and data closer to the front end may not always be feasible. In such a case, a service may be divided into two parts: a time sensitive part, and the other part which is not time sensitive. We call the time sensitive part that we need to locate closer to the front end, and sometimes at front end, as cerebellum function of the service. A practical algorithm that is designed for selecting and configuring the services with cerebellum function is presented. We are currently implementing a cerebellum function for IVS as a prototype for experimentation.

REFERENCES

Armbrust M., A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia. 2010. A view of Cloud computing. *Communications of the ACM*, vol. 53, no. 4, pp. 50–58.

Badger L., T. Grance, R.Patt-Corner and J.Voas. 2012. Draft Cloud Computing Synopsis and Recommendations. *National Institute of Standards and Technology, Special Publication 800-146*.

Buyya R., R. Ranjan, R.N. Calheiros. 2010. InterCloud: Utility-oriented federation of Cloud computing environments for scaling of application services.

Proceedings of the 10th International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP'10), pp. 13–31.

Cayirci E. 2013a. Modelling and Simulation as a Cloud Service: A Survey. In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl. Washington DC, December.

Cayirci E. 2013b. Configuration Schemes for Modelling and Simulation as a Service Federations. *Simulation Transactions of the Society for Modelling and Simulation International*, Vol. 89, Issue 11, pp. 1388 – 1399, November 2013.

Cayirci E. 2013c. A Joint Trust and Risk Model for MSaaS Mashups. In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl. Washington DC, December.

Cayirci E. , A. Garaga, A.S.Oliveira and Y. Roudier. 2014. Cloud Adopted Risk Assessment Model. *International Workshop on Advances in Cloud Computing Legislation, Accountability, Security and Privacy (CLASP)*.

Cayirci E. and A.S.Oliveira. 2015. Modelling Risk and Trust for Cloud Service Mashups. *IEEE Transactions on Cloud Computing* (submitted).

Jensen M., J.Schwenk, N.Gruscka and L.L.Iacono. 2009. On Technical Security Issues in Cloud Computing. *IEEE International Conference on Cloud Computing*, pp. 109-116.

Johnson H. and Tolk A., “Evaluating the Applicability of Cloud Computing Enterprises in Support of Next Generation of Modelling and Simulation Architectures,” *Spring Simulation Multi-Conference*, April 2013.

Rosetti M. And Chen Y., “Cloud Computing Architecture for Supply Chain Network Simulation,” *Winter Simulation Conference*, December 2012.

Singhal, M., S. Chandrasekhar, G. Tingjian, R., S. Sandhu, R. Krishnan, G-J. Ahn, E. Bertino. 2013. Collaboration in Multicloud Computing Environments: Framework and Security Issues. *IEEE Computer Magazine February 2013*, pp 76-84.

Subashini, S. and Kavitha, V. 2011. A survey on security issues in service delivery models of cloud computing. *Elsevier Journal of Network and Computer Applications*, Vol. 34, Issue 1, pp. 1-11.

Toosi A.N., R.N.Calheiros, R.K.Thulasiram, R.Buyya. 2011. Resource Provisioning Policies to Increase IaaS Provider's Profit in a Federated Cloud Environment. *HPCC 2011*.

AUTHORS BIOGRAPHY

Erdal Cayirci graduated from Army Academy in 1986 and from Royal Military Academy, Sandhurst in 1989. He received his MS degree from Middle East Technical University, and a PhD from Bogazici University both in computer engineering in 1995 and 2000, respectively.

He retired from the Army when he was a colonel in 2005. He was a faculty member and a researcher at Istanbul Technical University, Yeditepe University, Naval Sciences Institute and Georgia Institute of Technology between 2000 and 2005. He is currently Head, CAX Support Branch in NATO's Joint Warfare Center in Stavanger, Norway, and also a professor in the Electrical and Computer Engineering Department of University of Stavanger. His research interests include computer simulations, cloud computing, risk and trust modelling, mobile communications and sensor networks.

He received the **"2002 IEEE Communications Society Best Tutorial Paper" Award** for his paper titled "A Survey on Sensor Networks" published in the IEEE Communications Magazine in August 2002, the **"Fikri Gayret" Award** from Turkish Chief of General Staff in 2003, the **"Innovation of the Year" Award** from Turkish Navy in 2005 and the **"Excellence" Award** in ITEC 2006.

He co-authored two books titled as "Security in Wireless Ad Hoc and Sensor Networks," and "Computer Assisted Exercises and Training: A Reference Guide" both published by John Wiley & Sons in 2009.

Hakan KARAPINAR graduated from Hacettepe University Electrical & Electronics Engineering Department in 1996. He received his MS degree from Bilkent University in 1998 about antenna simulation. He started working at HAVELSAN Company; Simulation, Training and Test Systems department in 1998 and he is still working as Program Director in that division. He started his PhD studies at Hacettepe University Electronics Engineering Department and MBA studies at Cankaya University, Turkey. His research interests include Real Time Simulation, Modelling, Distribution Interactive Simulation, Electronics Warfare, Radar, Antenna, Tactical Environment Simulation and Sensor Simulation.

In HAVELSAN, between 1998 and 2005, he worked as System Engineer, System Team Leader, System Group Manager working actively on simulation of electronic warfare and weapon systems. Between 2005-2011, he worked as a Project Manager and Program Manager for fixed-wing and rotary-wing platform training simulators projects. Till 2011 he is working as Programs Director responsible for all Simulator and Simulation programs at HAVELSAN, managing around 22 projects and programs with more than 600 M \$ value for local and foreign customers.

Lutfu OZCAKIR, graduated from Hacettepe University (Ankara) Electronics Engineering Department in 1996 as honour faculty student. He received his MS degree from of Bilkent University in Electronics Engineering Medical Signal Processing in 1998. In 1998 he has joined HAVELSAN company Simulation and Training Systems Division as responsible engineer of Air Force Wargaming systems.

He started his PhD studies about simulation of anatomical structures at Hacettepe University, Ankara. His research interests include Medical Signal Processing, Tactical Environment Simulation, Real Time Modelling, Distributed Joint Simulation Systems and C2 Simulation.

Between 1998-2005 as a team leader, group manager and project manager he has completed many programs in simulation and training systems of HAVELSAN. From 2005-2011 he worked as Project Manager and Program Director of all projects at HAVELSAN Simulation Systems division. Till 2011 he is working as Executive Vice President of Simulation, Training and Test Systems division and board member of HAVELSAN Technology Radar (HTR).

CHEMICAL RISK EVALUATION: APPLICATION OF THE MOVARISH METHODOLOGY IN AN INDUSTRY OF THE TEXTILE SECTOR

Duraccio V.^(a), Di Bona G.^(b), Falcone D.^(b), Silvestri A.^(b), Forcina A.^(c)

^(a) University Niccolò Cusano, Rome, via Don Carlo Gnocchi 3, 00166 Rome – Italy

^(b) DiCeM University of Cassino and Southern Lazio, Via G. Di Biasio, 43, 03043 Cassino (FR)-Italy

^(c) University Parthenope, Department of Engineering Centro Direzionale - Isola C4 80143 - Naples - Italy

^(a)vincenzo.duraccio@unicusano.it, ^(b)falcone@unicas.it, silvestr@unicas.it, dibona@unicas.it,
^(c)antonio.forcina@uniparthenope.it

ABSTRACT

The present paper analyses the minimum requirements for workers' protection against health and safety chemical risks. The potential risks from exposure are divided into two categories which are often simultaneously present:

Health risks due to chronic exposure to toxicological or harmful substances;

• Safety hazards due to acute exposure: fire, chemical burns, eye injury, poisoning, suffocation.

A case study concerning chemical compounds used in a textile serigraphy department is proposed. A recently installed machine, with high level of automation, was analyzed. An operative spreadsheet tool combining "Workers' Safety Requirements" and "Risk Reduction Actions", has been defined. By evaluating the risk degree of each chemical product, if the assessed risk is acceptable, only safety requirements were considered.

Otherwise, if level and duration of exposure determine a harmful risk for workers, specific prevention and protection actions were suggested, such as appropriate work processes, technical planning, individual protections, Training and information.

Keywords: Chemical Risk, Health Risk, Risk reduction actions

1. INTRODUCTION

The concept of chemical risk and, more generally, of risks of dangerous substances, is considered in Title IX of the Italian Legislative Decree n. 81/2008 (Title IX of Legislative Decree 81/2008), which indicates the minimum requirements for the protection of workers against risks to health and safety arising from the effects of chemical agents present in the workplace or as a result of any work activity involving the presence of chemical agents (European agency for safety and health at work, 2002). The requirements identified in this Title can be applied to all hazardous chemical agents present in the workplace, with the exception of agents related to radiation protection (Bengtsson, 2010). A chemical agent is anything, substance or preparation, which constitutes a danger to workers; also lead and carcinogens, in general, and those substances can cause a hazard in certain conditions, though not dangerous by their very

definition. In general, the production, handling and storage of chemicals involves a number of potential risks by chemical exposure (Kogg *et al.*, 2010). In assessing workers' risks, we have to consider all the activities related to the production process (Di Bona *et al.*, 2014), such as transport, handling or production of waste products that can cause a particular exposure to certain employees. In order to evaluate, in a first approximation, the chemical risk in the company, we can use different models of computation that refer to algorithms (Gnoni *et al.*, 2010). In general, an algorithm (or model) is a procedure that assigns a numeric value to a number of factors or parameters involved in determining the risk weighting for each of them, and the absolute importance of mutual evaluation on the final result.

The model used for the Chemical Risk assessment was elaborated by Emilia Romagna, Lombardy and Tuscany Regions and within Legislative Decree n. 81/08 (Title IX - Chapter I "protection against chemical agents"). The model suggests a process of "facilitation" which enables the risk classification over or under the threshold "IRRELEVANT FOR HEALTH"; it is the first level of investigation and must be followed, if necessary, by additional investigations, also instrumental. The evaluation includes a comparison between the potential source of danger and exposed persons; specifically, we evaluate the probability of occurrence of an event (accident or illness) due to the use of a single chemical substance and compound (Weaver, 1993), identified through the air analysis in the various workplaces; the analysis of the company preventive systems in each process phase; the analysis of protective systems (PPE) provided to workers. According to the considered method, the risk depends on three factors

- Gravity (negative intrinsic quality of the chemical agent)
- Duration (duration of exposure to chemical agent)
- Level Of Exposure (qualitative and quantitative)

2. CASE STUDY

The studied company belongs to an industrial group leader in the wear production and global distribution of

international brands. In particular, the analyzed department for the application of the methodology mentioned is the serigraphy department. Serigraphy or screen printing is an artistic printing technique of images and graphics on any material or surface, through the use of a fabric (printing fabric), by depositing the ink on a support through the free areas of the fabric.

2.1 Chemical Risk Onset in Movarish Methodology

The chemical risk happens when in the workplace both the following two risk factors are simultaneously present:

- presence of hazardous chemical agents (chemical risk factors);
- presence of conditions of exposure (risk factors exposition).

The chemical risk R of exposure to hazardous chemicals is the product of hazard and exposure:

$$R = H \times E \quad (1)$$

The danger H is the index of intrinsic hazard of a substance or preparation, identified through the risk phrases defined by European Directive 67/548/EEC. For each R phrase a score is assigned taking into account the criteria for the classification of substances and dangerous preparations, as indicated in Legislative Decree 28 July 2008, 145. The danger H represents the potential hazard of a substance regardless of the levels which people are exposed to (intrinsic hazard). The exposure E is the level of exposition of individuals in the specific work activity. In this model, the risk R can be separately calculated for inhalation exposure and dermal exposure:

$$R_{\text{INAL}} = H \times E_{\text{INAL}} \quad (2)$$

$$R_{\text{DERM}} = H \times E_{\text{DERM}} \quad (3)$$

In case there are both the above types of exposure, the cumulative risk R (R_{CUM}) is obtained thanks to the following equation:

$$R_{\text{CUM}} = (R_{\text{INAL}}^2 + R_{\text{DERM}}^2)^{1/2} \quad (4)$$

The proposed methodology for identifying the H index (hazard of chemical risk to the health of workers) is based on the knowledge of the toxicological properties of chemical agents used or that are released into the workplace according to the exposure of workers, depending on the quantity, the method of use and the frequency of exposure. In order to evaluate the toxicological properties of chemical agents present in the productive activities, a score between 1 and 10 is assigned to each substance, single or combined, representing the danger H. Subsequently, a risk ranking is possible. In other words, the above index summarize in a single number the health hazards of a chemical agent. Another important aspect for a proper graduation of danger is the classification and labelling of substances and preparations which are dangerous according to Directive 67/548/EEC. The main principle is that the effects in the long-term (e.g. hazard category of toxic for reproduction) are more important than the acute effects. The numerical index that determines the hazard ranking must consider this general principle. Because of the low probability of occurrence, it was decided to give a quite

low score, but not zero, according to the assessment of the intrinsic danger in the case of effects due to ingestion. If a chemical agent exerts its dangerousness exclusively in the case of ingestion, the risk can be eliminated using proper sanitation and behaviour. Then, to consider an inequality between dermal and inhalation, a "weight" was introduced.

The inhalation is the main way of entry into the body of dangerous substances at work. The risk of inhalation exposure to dangerous chemicals occurs when processes or operating procedures cause the dissemination in the air of chemical pollutants. The index of exposure via inhaled E_{INAL} is determined by the product of a Sub-index I (intensity of exposure) and a Sub-index d (distance of the worker from the source of intensity I)

$$E_{\text{INAL}} = I \times d \quad (5)$$

The calculation of the Sub-index I involves the use of the following five factors: Physical and Chemical Properties, Quantity in use, type of use, type of control, time of exposure. In particular, it is possible to use the criteria identified in S.C. Maidment, "Occupational Hygiene Considerations in the Development of a Structured Approach to Select Chemical Control Strategies."

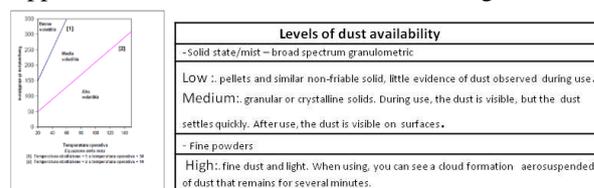


Figure 1: Level of dust availability

Type of use

It identifies four levels, always in ascending order, regarding the possibility of leakage in the air, the type of use of the substance, which identifies the source of exposure.

- Use in closed system: the substance is used and/or stored in reactors or sealed containers and transferred from one container to another through pipes watertight.
- Use in inclusion in the matrix: the substance is incorporated in materials or products to which it is prevented or limited dispersal in the environment.
- Controlled and non-dispersive use: this category includes processes in which only limited selected groups of workers are involved, experts of the specific process and adequate control systems to monitor and control the exposure are available.
- Use with significant dispersion: this category includes processes and activities that may cause an uncontrolled exposure not only to people involved but also other workers and possibly the general population. Jobs with paints and other similar activities can be classified in this category of use.

Type of control

- Complete Containment: it corresponds to a situation in a closed cycle.
- Ventilation - Local exhaust of discharges and emissions (LEV): this system removes the contaminant at its release source, preventing dispersion in areas with human presence, where it could be inhaled.
- Separation: the worker is separated from the source of release of the contaminant by an appropriate security space, or there are adequate time intervals between the presence of the contaminant in the environment and the presence of staff in the same area.
- Dilution: natural or mechanical. This method allows minimising the exposure.
- Direct Manipulation (with individual protection systems): in this case the worker operates in direct contact with the hazardous material, adopting mask, gloves or other similar equipment.

2.2 Time of Exposure

Five ranges define the time of exposure to the substance or preparation:

- less than 15 min
- between 15 minutes and 2 hours
- between 2 and 4 hours
- between 4 and 6 hours
- more than 6 hours

Through the identification of physico-chemical properties of the substance or preparation and the amount in use, inserted into matrix 1, a first indicator D is defined on four levels increasing potential availability to aero-dispersion

- Once you have the D indicator and identified the type of use, as defined in paragraph 3, the following indicator U on three levels of increasing actual availability to aero-dispersion can be obtained through matrix 2;
- Once you have identified the indicator D and the "type of control", as defined in paragraph 4, through matrix 3, it is possible to obtain a subsequent C indicator that takes into account the compensation factors relating to the prevention measures and protection measures adopted in the work environment;
- Finally, using the indicator C, obtaining the time of actual exposure of the worker, the value of the sub-index I can be assigned through matrix 4, distributed on four different levels, which correspond to different "intensity of exposure," regardless of the distance from the source of exposed workers.

The sub-index d takes into account the distance between a source of intensity I and the exposed worker: if these are close to the source (<1 meter) the sub-index I is unchanged ($d = 1$); gradually, during which the employee is away from the source, the sub-index of the intensity of exposure must be reduced proportionately until arriving

at a value of one tenth of I for distances greater than 10 metres. The values of d to be used are indicated in the following table:

Table 1: Values of d

Distance in meters	Value of d
less than 1	1
1 to less than 3	0,75
3 to less than 5	0,50
5 to less than 10	0,25
Greater than or equal to 10	0,1

Chemicals are absorbed by the skin more slowly than the gut or lungs. However, the chemicals (especially organic solvents) can enter the body either directly or via contaminated clothing. The risk of exposure for skin contact can occur during handling of dangerous substances. Values of E_{cute} index are a criterion for the assessment of risk from hazardous chemicals. In agreement with those reported in the previous paragraphs, it is possible to classify the use of preparations on the basis of the value of the risk index R.

Table 2: Classification risk

	RISK VALUES (R)	CLASSIFICATION
INSIGNIFICANT RISK TO HEALTH	$0,1 \leq R < 15$	Insignificant risk to health
	$15 \leq R < 21$	Uncertainty interval: it is necessary, first of classification insignificant risk to the health, review with care the assignment of various scores, review the measures of prevention and protection measures adopted and consult a qualified doctor
HIGHER RISK THAN IRRELEVANT TO HEALTH	$21 \leq R \leq 40$	Risk outweighs the chemical risk irrelevant to health
	$40 < R \leq 80$	Area of high risk
	$R > 80$	Area of serious risk to reconsider the path of the identification of the measures of prevention and protection with a view to their possible implementation. Intensify controls such as health monitoring, measurement of chemical agents and maintenance periods.

The model can also be applied to exposures to hazardous chemical agents arising from work activity. In this case the algorithm for the choice of the score and in the calculation of the exposure E must be used very cautiously. P is also necessary to take into account the model that may not always be specific for all the activities in which they can develop chemical agents. In particular, applying the model to choose the P score is extremely important to know if the amount of the development of the pollutants from the workforce is high or low and which classification can be attributed to chemical agents that are developed. After choosing the size of the emission, to assign a score P, it is necessary to identify chemical agents, assign relative rankings (very toxic, toxic, harmful, irritant inhalation) and to use for the calculation of R, the highest value of P. For the allocation of the value of E_{inal} value, it is necessary to use a modified matrix system:

- in the matrix 1 / bis using the amount in use, daily and overall, of the material from which they can develop hazardous chemical agents.
- in the matrix 2 / bis is used the index value obtained from the matrix 1 / a and the exposure time, according to the criteria previously defined, obtaining the value of the sub-index of intensity I

multiplied by the distance d which indicates the distance of the worker exposed to the source of emission. The risk R for inhalation of hazardous chemical agents developed from work must be considered against a conservative estimate and is calculated by:

$$R = P \times E_{\text{inal}} \quad (6)$$

Examples

Table 3: Matrix 1/bis

Amount in use	Type of Control			
	Complete containment	Local exhaust	Segregation/separation	General ventilation
<10Kg	low	Low	Low	Medium
10 to 100Kg	low	Medium	Medium	High
> 100Kg	low	Medium	High	High

Table 4: Matrix 2/bis

	Time of exposure				
	<15 min	Between 15 min and 2 h	Between 2 and 4 h	Between 4 and 6 h	>6 h
C1	low	low	Medium/low	Medium/low	Medium/high
C2	low	Medium/low	Medium/high	Medium/high	High
C3	Medium/low	Medium/high	High	High	High

2.3 Risk Evaluation For Safety

The evaluation of exposure to chemical risk is based on two methods of analysis, one that takes into account the health and the other which takes into account the safety. In assessing the safety risk due to the use of hazardous chemical agents a qualitative assessment is proposed.

2.4 Application Note for Case Study

In order to assess the risk of individual chemical products used in screen printing, we use an Excel spreadsheet where, through the use of safety data sheets, and can be determined as a function of exposure time, if the product constitutes a significant risk to the health of workers (Bureau of labor statistics, 2004; Bureau of labor statistics, 2005). Consider one of the most used products in screen printing: Printperfekt 226. This product is used in two distinct departments of screen printing: in dep. B, the department for the preparation and storage of paints, and in dep. C, the department in which printing occurs. Suppose that Printperfekt 226 is to be used in department B, specifying the employees in the department B, the date of the Safety Data Sheets, the total amount present in the department and the quantities used daily. Subsequently, through the material safety data sheet occurs if the product is made from hazardous substances. The considered product is made from a single dangerous substance namely ethane-1,2-diol. With regard to this substance, it is necessary to specify the concentration, the associated symbolism (if it is a corrosive substance, irritant, flammable, harmful, toxic, etc.) – Xn in this case (the substance is harmful) – the CAS number, which is an identifying number that unambiguously identifies a

chemical substance, the code EC, or EINECS code that is a registration code, present in European inventory, of commercial chemicals and, finally, the risk phrases R. Through the risk phrase R22, which in our case is that the product is harmful if swallowed, it is possible to determine the hazard index P. It is important to take the risk phrase that has the highest score to evaluate the risk in a worse situation. After determining the hazard index P, the risk of inhalation exposure and the risk of dermal exposure are calculated. In order to determine the risk of inhalation exposure, the intensity of inhalation exposure is calculated and to do this it is necessary to:

- Determine the physical state of the product that we are using, which can be:
 - Solid state/mists;
 - Liquid of low volatility;
 - Liquid medium and high volatility;
 - Gaseous state.

If the product is a liquid, to determine the level of volatility, it is necessary to use a graph that shows the performance of the boiling temperature as a function of operating temperature. After defining the physical state of the product that we are using in the amount in use (expressed in kg / day) is calculated by multiplying the daily amount for the concentration;

- Determine the type of use that is supposed to be always controlled non-dispersive;
- Determine that the type of control is always local exhaust ventilation
- Specify the time of daily exposure of workers to the product, in this case 8 hours / day;
- Enter the table in which we have the amount in use and the physical-chemical properties so that, depending on the amount in use and the physical-chemical properties, fixed value Indicator of Availability (D) is estimated (in our case this is 3);
- Enter the table indicating the type of use and the value of Availability Indicator (D) so that, by entering the previously determined value of D and knowing that the type of use is always controlled use, the value Indicator of Use (U) is determined (in this case 3);
- Enter the table indicating the type of control and the value of (U) so that, by entering the value of (U) previously determined and knowing that the type of control is always localised aspiration, Compensation Indicator (C) can be determined (in this case 2);
- Entering in the table indicates that the exposure time and the value of Indicator Compensation (C) so that entering the value of (C) previously determined and knowing the exposure time, the

value of the Sub-Index Intensity (I) is estimated (in our case this is 10);

- Fix the index of distance (d) which happens to be always equal to 1 m;
- After fixing all these parameters, it is possible to calculate the index of inhalation exposure as E_{INAL}

$$E_{INAL} = I \times d \quad (7)$$

At this point it is possible to calculate the risk of inhalation exposure:

$$R_{INAL} = P \times E_{INAL} \quad (8)$$

After determining the risk of inhalation exposure, this may determine the risk of dermal exposure progressing in a similar manner and then evaluating the intensity of dermal exposure:

- Determine the type of use that is supposed to be always controlled non-dispersive;
- Determine the level of skin contact, which happens to be always accidental contact;

Entering the table that shows the type of use, which we know is always controlled non-dispersive use, and the level of skin contact, which we know is always accidental contact, we have $E_{CUTE} = 3$. Once you know the risk of inhalation exposure and the risk of dermal exposure, the cumulative risk R can be calculated:

$$R_{CUM} = (R_{INAL}^2 + R_{CUTE}^2)^{1/2} \quad (9)$$

In this case $R_{CUM} = 18,27$ and is shown in the table where the criterion for assessing the risk from hazardous chemicals shows that $15 \leq R < 21$. Then, we are in the range of uncertainty. Therefore, before classifying the risk as irrelevant to health, it is necessary to review with care the assignment of various scores, the measures of prevention and protection measures adopted and consult a doctor. The analysis carried out for Department B is repeated in a similar manner to Department C (pressroom). The only difference is in the time of exposure, which in B is eight hours, but 1.5 hours in C. Being less than the exposure time, the cumulative hazard is minor. In fact, in this case, $R_{CUM} = 7.42$ and, therefore, the risk is irrelevant to the health and we must simply comply with the requirements of the safety data sheet. In the case of non-negligible risk, the employer must undertake specific measures of prevention and protection, such as:

- The design of appropriate work processes and engineering technicians (Falcone *et al.*, 2014).

2.5 Conclusion

In order to reduce the risk from exposure to chemicals in the workplace, important measures of prevention and protection can be taken, which help protect the health and safety of workers. As for preventive measures to be implemented to reduce the risk resulting from the

presence of chemical agents, it is necessary to distinguish between:

- Preventive measures to be implemented to reduce the risk at source (elimination or substitution of dangerous products).
- Preventive measures relating to the workplace (local to local exhaust ventilation and separate in depression).
- Preventive measures to be implemented to reduce the organisational risk of work systems (reduction of exposure duration and intensity).
- Preventive measures to be implemented for the management of emergencies (first aid equipment and kits washing).
- Preventive measures to be implemented to reduce the residual risk (information, training, procedures).
- Preventive measures to be implemented to reduce the residual risk (use of PPE).
- Preventive measures to be implemented to reduce the residual risk (health monitoring).

Table 5: Scheme Excel sheet

N°	Descrizione sostanza	Aspetti	Scheda tecnica (SIC)	Quantità locale presente	Quantità prodotta (SIC)	Componente principale (SIC)	Conc. (%)	Sintossicità	CAS	EC	Class. di rischio (SIC)	Indice I*
1	PRINT PERFECT 228	3000	Produttore: 4 del 09/12/2005	800 Kg	10,000	Stearil-3-oleato	3,50%	30x	107-21-1	204-172-2	R22	1,75
2	BIANCO STAMPANTE	3000	Produttore: 20/03/2010	80 Kg	8,000	N.P.	---	---	---	---	---	---
3	PRINT PERFECT LAC 107 (SIC)	3000	Produttore: 2 del 14/12/2005	30 Kg	1,000	Nitro (giallo)	10,00%	30x	84742-48-0	205-150-0	R22, R50	3,50
4	PRINT PERFECT C-1	3000	Produttore: 2 del 15/03/2005	30 Kg	1,000	N.P.	---	---	---	---	---	---
5	PRINT PERFECT C-2 (SIC)	3000	Produttore: 1 del 15/03/2012	30 Kg	1,000	Nitro (giallo)	10,00%	30x	84742-48-0	205-150-0	R22, R50	3,50
6	SEALANT FOR S Kg	3000	Produttore	10 Kg	1,000	Metilpropileno acetato acetato	7,00%	30x	---	---	R50	2,50
7	TRIO METILCO 228 (SIC)	3000	Produttore: 1 del 20/12/2012	30 Kg	1,000	Acetato di 2-butanone	3,00%	30x	112-07-2	203-632-0	R22, R11	4,25
8	BIANCO STAMPANTE (SIC)	3000	Produttore: 4 del 11/03/2010	20 Kg	1,000	Acido citrico	3,50%	30x	59-49-25-1	201-025-1	R36	3,50
						Tetraossido di etilico diclorato	1,00%	T	1333-98-4	215-424-4	R22, R51	1,00

MISURE PREVENTIVE E PROTETTIVE												
Proprietà chimica	Quantità (Kg)	Tossicità	Tossicità (SIC)	Tossicità (SIC)	D	L	C	Indice di esposizione (SIC)	Indice di esposizione (SIC)	Rischio (SIC)	Indice di esposizione (SIC)	Rischio (SIC)
Liquido irritante	0,2000	Non controllato	Ventilazione aspirazione	1,5 h/gg	3	2	3	1,00	3,000	3,250	L30	3,250
---	---	---	---	---	---	---	---	---	---	---	---	---
Solida	0,1000	Non controllato	Ventilazione aspirazione	1 h/gg	1	1	1	1,00	1,000	3,000	L30	10,000
Solida	0,1000	Non controllato	Ventilazione aspirazione	2 h/gg	1	1	1	1,00	1,000	1,750	L30	5,250
---	---	---	---	---	---	---	---	---	---	---	---	---
Solida	0,1000	Non controllato	Ventilazione aspirazione	1 h/gg	1	1	1	1,00	1,000	3,000	L30	10,000
Liquido irritante	0,2000	Non controllato	Ventilazione aspirazione	1 h/gg	1	1	1	1,00	1,000	2,500	L30	1,500
Liquido	0,2000	Non controllato	Ventilazione aspirazione	1 h/gg	1	1	1	1,00	1,000	2,750	L30	5,250
Solida	0,2000	Non controllato	Ventilazione aspirazione	1 h/gg	1	1	1	1,00	1,000	4,250	L30	10,500
Solida	0,2000	Non controllato	Ventilazione aspirazione	1 h/gg	1	1	1	1,00	1,000	2,500	L30	1,500
Solida	0,2000	Non controllato	Ventilazione aspirazione	1 h/gg	1	1	1	1,00	1,000	10,000	L30	20,000

2.6 List of References

Title IX of Legislative Decree 81/2008

European agency for safety and health at work. 2002. safety and health good practice on line for the healthcare sector. facts 29.

Bengtsson, G. 2010. Global trends in chemicals management. In *Regulating chemical risks*, edited by J. Eriksson et al. New York, NY, USA: Springer Science + Business Media.

Kogg, B. and A. Thidell. 2010. Chemicals in products - An overview of systems for providing information regarding chemicals in product and of stakeholders' needs for such information. www.chem.unep.ch/unepsaicm/cip/Documents/Kogg_Thidell_CiP%20report_final.pdf. Accessed 20 August 2011

Di Bona, G., Duraccio, V., Silvestri, A., Forcina, A. 2014. Productive line reengineering through simulation techniques (2014). Proceedings of the IASTED International Conference on Modelling, Identification and Control pp. 291-296

Gnoni, M.G., Elia, V., 2013 An environmental sustainability analysis in the printing sector, *International Journal of Sustainable Engineering*, Volume 6, Issue 3, 2013, Pages 188-197

Weaver, V.N. et al. 1993. occupational chemical exposure in an academic medical center. *J. occup. med.* 35: 701-707

Bureau of labor statistics, U.S. department of labor. 2004. tomorrow's jobs. Washington, dc.

Bureau of labor statistics, U.S. department of labor. 2005. injuries, illnesses and fatalities program. <http://www.bls.gov/iif/>

Falcone D., Cerbaso C., Duraccio V., Silvestri A., Forcina A. 2014 Maintenance critical analysis and priority index: A new model for maintenance policy (2014) 26th European Modeling and Simulation Symposium, EMSS2014. Pp 432-437

NONLINEAR SCALING IN SMART ADAPTIVE MODELLING

Esko K. Juuso

Control Engineering, Faculty of Technology, University of Oulu, Finland

esko.juuso@oulu.fi

ABSTRACT

Modelling methodologies provide a good basis for the integration of intelligent systems. Small, specialised systems have a large number of feasible solutions, but developing truly adaptive, and still understandable, systems for highly complex systems require domain expertise and more compact approaches at the basic level. The nonlinear scaling approach extends the application areas of linear methodologies to nonlinear modelling and reduces the need for decomposition with local models. The close connection to the fuzzy set systems provides a good basis for understandable models. Data-based methodologies are suitable for developing smart adaptive applications. Complex problems are solved level by level to keep the domain expertise as an essential part of the solution.

Keywords: nonlinear systems, linguistic equations, smart adaptive systems, statistical analysis

1. INTRODUCTION

Models are understood as relationships between variables and used to predict of properties or behaviours of the system. Variable interactions and nonlinearities are important in extending the operation areas (Juuso 2014). Phenomenological models based on physics, chemistry and mathematics require domain expertise (Figure 1). Linear methodologies extended with principal components (Jolliffe 2002; Gerlach et al. 1979) and semi-physical models (Ljung 1999) provide a feasible solution for many applications. Nonlinearities have been handled commonly with interaction and quadratic terms (Box and Wilson 1951). Artificial neural networks (ANNs) starting from (Rummelhart et al. 1986) continue this by using more complex architectures.

Knowledge-based information can be handled with fuzzy set systems introduced by Zadeh (1965): numerous methodologies have been developed, see (Takagi and Sugeno 1985; Driankov et al. 1993; Dubois et al. 1999), and combined with neural networks (Fullér 2000). Different fuzzy approaches can be efficiently combined (Juuso 2014).

First order ordinary differential equations are solved by numerical integration and special solutions have been developed for identification (Ljung 1999). These approaches, which are also used in ANNs and fuzzy set systems (Babuška and Verbruggen 2003), define structures for hybrid dynamic models (Figure 1).

Local models need to be combined in complex systems (Sontag 1981; Ljung 2008; Jardine et al. 2006).

The linguistic equation (LE) approach originates from fuzzy set systems (Juuso and Leiviskä 1992): rule sets are replaced with equations, and meanings of the variables are handled with scaling functions which have close connections to membership functions (Juuso 1999a). The nonlinear scaling technique is needed in constructing nonlinear models with linear equations (Juuso 2004c). Constraints handling (Juuso 2009) and data-based analysis (Juuso and Lahdelma 2010), improve possibilities to update the scaling functions recursively (Juuso 2011; Juuso and Lahdelma 2011). The LE approach together with knowledge-based systems, neural networks and evolutionary computation form the computational intelligence part (Figure 1).

Three levels of smart adaptive systems (SAS) identified in (Anguita 2001):

1. adaptation to a changing environment;
2. adaptation to a similar setting without explicitly being ported to it;
3. adaptation to a new or unknown application.

Smart use of intelligence by integrating specific intelligent systems is essential in development of complex adaptive applications.

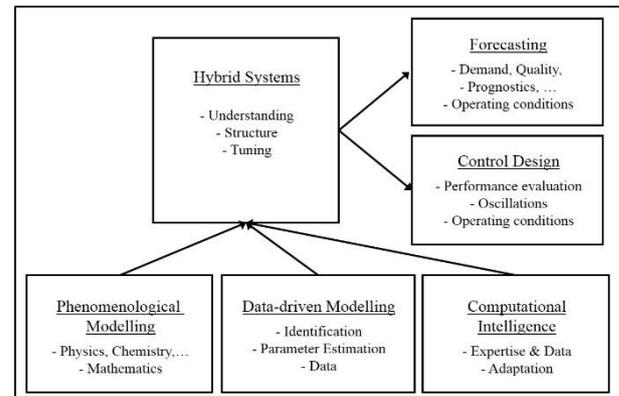


Figure 1: Methodologies and application types of modelling and simulation, modified from (Juuso 2004b)

Technically, an automatic black box modelling could be possible in various Big Data problems by using combinations of these methodologies. The domain expertise is an essential part in integrated solutions to understand and assess the applicability.

This paper classifies modelling methodologies and focuses on the nonlinear scaling and integrates the LE approach in developing modelling applications for complex systems. Various applications are shortly discussed.

2. MODELLING METHODOLOGIES

Steady-state modelling with linear and nonlinear methodologies are the basis of modelling. Dynamic modelling introduces additional model structures. Decomposition is needed to extend the solutions to multiple operating conditions.

2.1. Steady-state modelling

The steady-state simulation models can be relatively detailed nonlinear multiple input, multiple output (MIMO) models $\bar{y}=F(\bar{x})$, where the output vector $\bar{y}=(y_1, y_2, \dots, y_n)$ is calculated by a nonlinear function F from the input vector $x=(x_1, x_2, \dots, x_m)$.

2.1.1. Linear methodologies

Statistical modelling in its basic form uses linear regression for solving coefficients for a linear function. Linear methodologies are suitable for large multivariable systems. Linear methodologies can be extended to more complex areas by handling inputs. Principal components compress the data by reducing the number of dimensions with minor loss of information (Jolliffe 2002). Partial least squares regression (PLS) is an extension of these ideas (Gerlach et al. 1979). Semi-physical models of inputs are important in linear modelling, see (Ljung 1999).

2.1.2. Nonlinear methodologies

Exponential and logarithmic functions are suitable for the modelling of steep changes.

In the response surface method (RSM), the coefficients of linear, interactive and quadratic terms are obtained by the multiple regression for several input variables (Box and Wilson 1951). The models can include several multiple input, single output models.

Artificial neural networks (ANNs) are input output models: the most popular architecture, multilayer perceptron (MLP) has a very close connection to the backpropagation learning (Rummelhart et al. 1986). The response of a neuron is

$$y_i = F_i \left(\sum_{j=1}^m w_{ij} p_j + b_i \right), \quad (1)$$

where w_{ij} is the weight factor of the element p_j in the input vector of the neuron i and b_i a scalar bias. Linear networks correspond to the models with linear terms in RSM models.

Normalisation and principal components are essential in various applications. Various extensions of the principal component analysis (PCA) are referred in (Jolliffe 2002). A function expansion

$$y_i = \sum_{k=1}^{m_f} w_{ik} F_k(\bar{x}) = \sum_{k=1}^{m_f} w_{ik} f(\bar{\beta}_k \bullet (\bar{x} - \bar{\gamma}_k)), \quad (2)$$

with some basis functions $F_k(\bar{x})$, $k=1, \dots, m_f$, provides a flexible way to present several types of black box models (Ljung 2008). The functions are generated from one and the same function characterised by the scale (dilation) parameters $\bar{\beta}_k$ and location (translation) parameters $\bar{\gamma}_k$. The expansion can contain, for example, radial basis functions, one-hidden-layer sigmoidal neural networks, neurofuzzy models, wavenets, least square support vector machines (SVMs), see (Ljung 1999).

2.1.3. Knowledge-based methodologies

Zadeh (1965) presented the fuzzy set theory to form a conceptual framework for linguistically represented knowledge. Origins of the fuzzy logic are in approximate reasoning and the connection of fuzzy rule-based systems and expert systems is evident (Dubois et al. 1999). Membership functions provide a key to expand expert systems. In linguistic fuzzy models, both the antecedent and consequent are fuzzy propositions (Driankov et al. 1993). In Takagi-Sugeno (TS) fuzzy models, each consequent is a crisp function of the antecedent variables x , can be interpreted in terms of local models (Takagi and Sugeno 1985). The extension principle generalises arithmetic operations if the inductive mapping is a monotonously increasing function of the input. Type-2 fuzzy models introduced by Zadeh in 1975 take into account uncertainty about the membership function (Mendel 2007).

In neurofuzzy systems, fuzzy neurons combine the weight factors and the inputs. The activation function is handled with the extension principle from the fuzzy input, which is obtained by the fuzzy arithmetics. Different combinations with fuzzy and crisp weight factors and elements can be used in these models (Fullér 2000). Neurofuzzy systems can be represented with (2).

2.2. Dynamic modelling

First order ordinary differential equations (ODEs),

$$\frac{dx}{dt} = f(t, x), \quad (3)$$

are solved by integration:

$$x = \int_0^T f(t, x) dt + x_0, \quad (4)$$

where T is the time period for integration and x_0 the initial condition. The function $f(t, x)$ can be linear or nonlinear. Additional algebraic equations are needed, e.g. for handling material.

Linear methodologies are used in the time series modelling is to fit the waveform data to a parametric time series model and extract features based on this parametric model. For parametric models, the output y at time t is computed as a linear combination of past inputs u and

past outputs y . The signal values should be chosen according to the appropriate time delays. The time step is not adapted in these models. The number of delayed inputs and outputs is usually referred to as the model order(s). Several types of models can be obtained from the general parametric model

$$A_q(q) y(t) = \frac{B_q(q)}{F_q(q)} u(t - n_k) + \frac{C_q(q)}{D_q(q)} e(t), \quad (5)$$

where $A_q(\cdot)$, $B_q(\cdot)$, $C_q(\cdot)$, $D_q(\cdot)$ and $F_q(\cdot)$ are polynomials of the delay operator q^{-1} . The orders of these polynomials are n_a , n_b , n_c , n_d and n_f , respectively, and the number n_k is the number of delay from input to output. Autoregressive (ARX) and autoregressive moving average (ARMAX) models, both with the exogenous input u , are special cases of (3). If the data are presented as a time series, which has no input channels and only one output channel y , then ARX and ARMAX models become AR and ARMA models, respectively. (Ljung 1999)

Fuzzy and neural models are based on the same structures. The most common structure for the input-output models is the NARX /Nonlinear AutoRegressive with eXogenous input model, in which the input and output values are chosen according to appropriate system orders, as in the ARX model. The regressor vector consists of a finite number of past inputs and outputs (Babuška and Verbruggen 2003). Another possibility is to use recurrent networks, e.g. the Elman networks are two-layer feedforward networks, with the addition of a feedback connection from the output of the hidden layer to its input (Elman 1990).

2.3. Decomposition

Linear models are approximations of a nonlinear system in different neighbourhoods. Composite local models combine local linear models to construct a global model. If the partitioning is based on a measured regime variable, the partitioning can be used in weighting the local models. Linear parameter varying (LPV) models, where the matrices of the state-space model depend on an exogenous variable measured during the operation, are close related to local linear models (Ljung 2008). Piecewise affine (PWA) systems are based on local linear models, more specifically in a polyhedral partition (Sontag 1981). The models can be state-space models or parametric models. The model switches between different modes as the state variable varies over the partition (Ljung 2008).

Gradual transitions between models can be done with weight factors in function expansion (2) or with fuzzy rules in TS fuzzy models. Common model structures are beneficial in applications. In prognostics, calculations can be based on a time-dependent proportional hazard model (PHM)

$$h(t) = h_0(t) \exp\left(\sum_{j=1}^m \eta_j x_j(t)\right), \quad (6)$$

where changing operating conditions are handled with exponential functions: $h_0(t)$ is a baseline hazard function, $x_j(t)$, $j = 1, \dots, m$, are covariates which are functions of time and η_j , $j = 1, \dots, m$ are coefficients.

The baseline hazard function $h_0(t)$ can be in non-parametric or parametric form, e.g. a Weibull hazard function, which is the hazard function of the Weibull distribution. The covariates $x_j(t)$, $j = 1, \dots, m$ are any condition variables such as health indicators and features in condition monitoring. Maximum likelihood estimation is usually used to build a PHM from event data and condition monitoring data. Modelling a PHM is more or less like the process of regress analysis: a set of significant covariates is finally found and only these significant observations for the “dependent” variable $h(t)$, instead of observations, are available as event data. (Jardine et al. 2006)

Cascade modelling divides the problem into sequential parts to further alleviate the problem of parameters: TS fuzzy models use fuzzy reasoning for weighting local linear models; radial basis networks are linear combinations of the outputs of the RBF; learning vector quantization (LVQ) combines a competitive layer with a linear model. Neurofuzzy systems can be constructed as sequential combinations of neural and fuzzy parts. Variable grouping is important in cascade model structures.

The need for decomposition is evident since the modelling methodologies have limitations, like operating areas of linear methodologies and highly complex structures of nonlinear systems. Fuzzy set systems are natural tools in the management of the decomposed models. Several fuzzy modelling approaches are combined in Figure 2: fuzzy arithmetics is suitable both for processing the fuzzy inputs and outputs of the rule-based fuzzy set system; fuzzy inequalities produce new facts; fuzzy relations can be represented as sets of alternative rules, where each rule has a degree of membership (Juuso 2014).

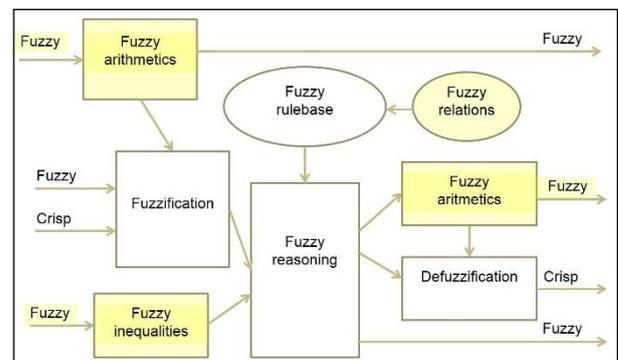


Figure 2: Combined Fuzzy Modelling (Juuso 2014)

3. LINGUISTIC EQUATION MODELS

The key of the linguistic equation (LE) methodology is the nonlinear scaling developed to extract the meanings of variables from measurement signals.

3.1. Nonlinear scaling

Normalisation or scaling of the data is needed since measurements with considerably different magnitudes cause problems in modelling. The nonlinear scaling extends modelling to various statistical distributions and allows recursive tuning.

3.1.1. Generalised norms

Arithmetic mean and standards deviation, which are the key features in statistical analysis, are special cases of generalised norms

$$\|{}^\tau M_j^p\|_p = (\tau M_j^p)^{1/p} = \left[\left(\frac{1}{N} \sum_{i=1}^N (x_j)_i^p \right) \right]^{1/p}, \quad (7)$$

where the order of the moment $p \in \mathfrak{R}$ is non-zero. The signal is measured continuously for the analysis is based on consecutive equally sized samples. Duration of each sample is called sample time, denoted τ . The number of signal values $N = \tau N_s$, where N_s is the number of signal values which are taken in a second. This norm, which has the same dimensions as the signal x_j . The generalised norms were introduced for condition monitoring (Lahdelma and Juuso 2011a, 2011b). The norm values increase monotonously with increasing order if all the signals are not equal.

3.1.2. Scaling functions

Scaling is based on the z-score

$$p_j = \frac{x_j - c_j}{\Delta c_j} = \frac{x_j - \|{}^\tau M_j^1\|_1}{\|{}^\tau M_j^2\|_2}, \quad (8)$$

which is calculated about the arithmetic mean, $c_j = \|{}^\tau M_j^1\|_1$, by the standard deviation $\Delta c_j = \|{}^\tau M_j^2\|_2$.

The arithmetic mean and standard deviation are optimal if the data sample comes from a normal distribution. This approach is sensitive to data entry errors, e.g. outliers. The geometric mean and harmonic mean are useful when the sample is distributed lognormal or heavily skewed. The median and trimmed mean are two measures that are resistant (robust) to outliers.

The z-score based linear solutions are extended to asymmetric nonlinear cases by two second order polynomials. The parameters of the polynomials are defined with five parameters corresponding the operating point c_j and four corner points of feasible range represented by a fuzzy number: core $[(c_l)_j, (c_h)_j]$ and support $[\min(x_j), \max(x_j)]$ (Juuso 2004c).

These points can be defined manually or obtained from data by using generalised norms (7) and moments

$$\gamma_k^p = \frac{\sum_{i=1}^N ((x_j)_i - \|{}^\tau M_j^p\|_p)^k}{N \sigma_j^k}, \quad (9)$$

where the order k is a positive integer. The k th moment is generalised by calculating it about the generalised norm (5), and normalised by the standard deviation σ_j which is calculated about the origin. The operating point c_j is the central tendency taken from the point, where the skewness γ_3^p changes from positive to negative, i.e. $\gamma_3^p = 0$. Then the data set is divided into two parts, a lower part and an upper part, and then the same analysis is done for these two data sets. The estimates of the corner points, $(c_l)_j$ and the operating point $(c_h)_j$, are the points where the skewness goes to zero. The iteration is performed with generalised norms (5) to get the corresponding order of norm. (Juuso and Lahdelma 2010)

The shape of the second order polynomials are defined by the ratios

$$\alpha_j^- = \frac{(c_l)_j - \min(x_j)}{c_j - (c_l)_j} = \frac{(c_l)_j - \min(x_j)}{\Delta c_j^-}, \quad (10)$$

$$\alpha_j^+ = \frac{\max(x_j) - (c_h)_j}{(c_h)_j - c_j} = \frac{\max(x_j) - (c_h)_j}{\Delta c_j^+},$$

which are limited to the range $\left[\frac{1}{3}, 3 \right]$ by resizing the core

or the support. Then the scaling functions are monotonously increasing throughout the feasible range. (Juuso 2009).

3.1.3. Recursive tuning

The computation of the generalised norms can be divided into the computation of equal sized sub-blocks, i.e. the norm for several samples can be obtained as the norm for the norms of individual samples. The same result is obtained using the moments

$$\|{}^{K_S \tau} M_\alpha^p\|_p = ({}^{K_S \tau} M_\alpha^p)^{1/p} = \left\{ \frac{1}{K_S} \sum_{i=1}^{K_S} \left[({}^\tau M_\alpha^p)_i \right]^p \right\}^{1/p}, \quad (11)$$

where K_S is the number of samples $\{x_i^{(\alpha)}\}_{i=1}^N$. Each sample has N signal values. Weights can be introduced by means of density functions. It is useful to calculate the norms from short samples since the number of signal values per second is quite high. The sample time τ is an essential parameter in the calculation of moments and norms. (Juuso and Lahdelma 2010)

For the arithmetic mean, the calculation based on sub-blocks is the normal practice in automation systems. This approach can be extended to all generalised norms.

3.2. Interactions

The linguistic equation (LE) models are linear equations

$$\sum_{j=1}^m A_{ij} X_j + B_i = 0, \quad (12)$$

where X_j is the linguistic level for the variable $j, j=1..m$. Each equation i has its own set of interaction coefficients $A_{ij}, j=1..m$. The bias term B_i was introduced for fault diagnosis systems. (Juuso 2004c) The scaled values can be used in the same way as any other variables since the nonlinear scaling extends the normalisation and the z-score approach. The dimensionless variables are suitable for various nonlinear methodologies discussed in Section 2. However, linear methodologies have been sufficient in various applications.

3.2.1. Steady-state LE models

A multiple input single output model is represented by

$$x_{out} = f_{out} \left(- \frac{\sum_{j=1, j \neq out}^m A_{ij} f_j^{-1}(x_j(t-n_j)) + B_i}{A_{iout}} \right) \quad (13)$$

where the functions f_j and f_{out} are scaling functions of the variables j and out , respectively. An appropriate time delay n_j needs to be taken into account for each variable.

The scaling functions can be understood as new type basis functions to be used in expansion (2).

3.2.2. Dynamic LE models

The basic form of the linguistic equation (LE) model is a static mapping in the same way as fuzzy set systems and neural networks, and therefore dynamic models can include several inputs and outputs originating from a single variable (Juuso 2004c). External dynamic models provide the dynamic behaviour, and LE models are developed for a defined sampling interval in the same way as in various identification approaches defined by (5). Nonlinear scaling reduces the number of input and output signals needed for the modelling of nonlinear systems.

For the default LE model, all the degrees of the polynomials become very low: $n_a = 1, n_b = 1, n_c = 0, n_d = 0$ and $n_f = 0$ in the parametric models (3) resulting

$$Y(t) + a_1 Y(t-1) = b_1 U(t-n_k) + e(t) \quad (14)$$

where Y and U are scaled variables and coefficients a_1 and b_1 coefficients of the polynomials $A_q(\cdot), B_q(\cdot)$.

Alternatively, a new value for the derivative (3) can be calculated with a LE model and then integrated by (4). This approach allows the adaptation of the integration step.

3.3. Uncertainty in LE models

The LE approach originates from the fuzzy set systems which keeps the connections of the methodologies strong. Compact LE models provide a good basis for multimodel systems, where local LE models are combined with fuzzy logic, to handle transitions between models, some special situations and uncertainty with fuzzy set systems. Fuzzy reasoning is an important part of the LE based fault diagnosis and the decision making in the recursive adaptation.

The coefficients of the model (12) and the parameters of scaling functions can be represented as fuzzy numbers, which are used in the calculations with the extension principle and fuzzy arithmetics (Juuso 2014).

3.4. Smart adaptive LE models

Recursive updates of the norm values discussed in Section 3.1.3 provide a real time solution to the adaptation to a changing environment (SAS level 1). Strong changes in statistical distributions can be taken into account by obtaining the orders of the norm, which realises the adaptation to a similar setting without explicitly being ported to it (SAS level 2). Similar settings are understood as unchanged interaction models. The settings are based on the analysis of the interactions (Section 3.2), which is the key in the adaptation to a new or unknown application (SAS level 3).

4. APPLICATIONS

Nonlinear scaling forms the basis for the LE modelling: an important benefit of the linear approach is that the models can be inverted, technically to any direction. The compact basic solution makes extensions to dynamic and case-based systems possible. Complex models for steady-state and dynamic systems can be built with the cascade and interactive structures.

4.1. Steady-state LE models

Steady-state LE models are mainly used in adaptation and feedforward control (Table 1). In most cases, the models include only a single equation (13). The first LE model developed for designing submerged arc furnaces was an exception which used well known relations represented by five equations (Juuso and Leiviskä 1992). A steady-state LE model was developed in an early control application from the process measurements of a lime kiln (Juuso et al. 1997). The working point model presented in (Juuso et al. 1998) is still an essential part of the model-based LE control of a solar power plant (Juuso and Yebra 2013). For continuous cooking, a LE model has been developed for predicting the Kappa number, which is widely used quality variable (Leiviskä et al.

2001). Stress-cycle (S-N) curves, also known as Wöhler curves, are represented by a linguistic equation (Juuso and Ruusunen 2013).

Table 1: Steady-state LE model applications

Case	Application area
Electric furnace	Decision support for process design
Lime kiln	Feedforward control of fuel feed for changing capacity
Solar collector field	Adaptation based on steady-state working point model
Continuous cooking	Forecasting model for quality control
Fatigue	Stress contribution obtained from LE based stress-cycle curve
Water treatment	Feedforward control
Wastewater treatment	Forecasting

4.2. Dynamic LE models

The basic dynamic LE model is represented by (14). The approach was first tested in a gas furnace data provided by (Box & Jenkins 1970). The dynamic models of the solar plant are based on test campaigns, which cannot be planned in detail because of changing weather conditions (Juuso 2003a). The basic dynamic flotation model is the core of the quality indicator in water treatment (Ainali et al. 2002, Joensuu et al. 2005). A dynamic LE model has been used for fatigue prediction in (Juuso and Ruusunen 2013). In all these models, only one equation is needed. The applications are indirect measurements and controller tuning (Table 2).

Table 2: Dynamic LE model applications

Case	Application area
Gas furnace	Modelling
Solar collector field	Controller tuning for oil flow
Fatigue	Forecasting fatigue risk from stress contributions
Water treatment	Water quality indicator Controller tuning for two chemicals

4.3. Decomposition in LE models

The multimodel LE system can include several submodels and complex interactions (Table 3). All basic models are represented by (14).

The model with a fuzzy decision module was first used for a lime kiln (Juuso 1999b) and then for a solar thermal power plant Juuso (2003a). The lime kiln model had six operating areas defined by the production level and the trend of the fuel feed (increasing, decreasing). The model of the collector field includes four operating areas: start-up, low, normal and high operation. For handling special situations in the solar plant, additional fuzzy models have

been developed by using the Fuzzy-ROSA method (Juuso et al. 2000).

Interactive dynamic models were needed in several cases: batch cooking (Juuso 2003b), fluidised bed granulator (Mäki et al. 2004), industrial fed-batch fermenter (Saarela et al. 2003) and wastewater treatment. Linguistic equations, neural networks and fuzzy modelling with several variants have been compared by using the process data obtained from the fed-batch fermenter.

Table 3: Dynamic LE model applications

Case	Application area
Lime kiln	Fuel quality Controller tuning by using multiple models
Solar collector field	Controller tuning for oil flow by using multiple models
Batch cooking	On-line forecasting by using three interactive models: alkali, lignin and dissolved solids
Fluidised bed granulation	Forecasting by using three interactive models: temperature, humidity and granular size
Fed-batch fermentation	On-line forecasting by using submodels of three growth phases, each including three interactive models
Wastewater treatment	Detection of operating conditions and trend analysis by using three submodels: load, treatment and settling
Condition monitoring	Prognostics with recursive tuning

4.4. Distributed parameter LE models

In distributed parameter models, the solar collector field is divided into modules, where the dynamic LE models are applied in a distributed way (Juuso 2004b). The same single equation model is used in all modules. Element locations for partial differential equations (PDEs) are defined by the flow rate. In cloudy conditions, the heating effect can be strongly uneven.

5. CONCLUSIONS

The nonlinear scaling approach extends the application areas of linear methodologies to nonlinear modelling: the meanings of variables and interactions are analysed sequentially. Local nonlinear models reduce the need for decomposition with local models is needed. The close connection to the fuzzy set systems provides a good basis for understandable models. Data-based methodologies are suitable for developing smart adaptive applications. Big Data problems are solved level by level to keep the domain expertise as an essential part of the solution. The basic models are compact and additional properties, including dynamics, uncertainty and decomposition are included if needed.

ACKNOWLEDGMENTS

The combined approach has been developed within the research program "Measurement, Monitoring and Environmental Efficiency Assessment (MMEA)" on the basis of several earlier and parallel projects.

REFERENCES

- Ainali I., Piironen M. and Juuso E., 2002. Intelligent water quality indicator for chemical water treatment unit. Proceedings of SIMS 2002 - the 43rd Scandinavian Conference on Simulation and Modelling, pp. 247–252, September 26-27, Oulu, Finland.
- Anguita D., 2001. Smart adaptive systems - state of the art and future directions for research. Proceedings of Eunite 2001 - European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, pp. 1-4. July 13-14, Tenerife, Spain.
- Babuška R. and Verbruggen H., 2003. Neuro-fuzzy methods for nonlinear system identification, Annual Reviews in Control, 27(1), 73-85.
- Box G.E.P. and Jenkins G.M., 1970. Time Series Analysis, Forecasting and Control, San Francisco: Holden Day.
- Box G.E.P. and Wilson K.B., 1951. On the experimental attainment of optimum conditions. Journal of the Royal Statistical Society, Series B, 13(1), 1–45.
- Driankov D., Hellendoorn H. and Reinfrank M., 1993. An Introduction to Fuzzy Control. Berlin, Germany: Springer.
- Dubois D., Prade H. and Ughetto L., 1999. Fuzzy logic, control engineering and artificial intelligence. In: Verbruggen H.B., Zimmermann H.-J. and Babuska R., eds., Fuzzy Algorithms for Control, International Series in Intelligent Technologies, Boston: Kluwer, pp. 17–57.
- Elman J.L., 1990. Finding structure in time, Cognitive Science, 14(2), 179-211.
- Fullér R. 2000. Introduction to Neuro-Fuzzy Systems. Advances in Soft Computing. Springer.
- Gerlach R.W., Kowalski B.R. and Wold H.O.A., 1979. Partial least squares modelling with latent variables. Analytica Chimica Acta, 112(4), 417–421.
- Jardine A.K.S., Lin D. and Banjevic D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing, 20(7), 1483–1510.
- Joensuu I., Piironen M. and Juuso E., 2005. Dynamic simulator for dosing of water treatment chemicals. Proceedings of European Symposium on Computer Aided Process Engineering-15 (Escape-15), pp. 301–306, Computer-aided chemical engineering, 20A, May 29 - June 1, Barcelona, Spain.
- Jolliffe I.T., 2002. Principal Component Analysis. 2nd edition, New York: Springer.
- Juuso E.K., 1999a. Fuzzy control in process industry: The linguistic equation approach. In: Verbruggen H.B., Zimmermann H.-J. and Babuska R, eds., Fuzzy Algorithms for Control, International Series in Intelligent Technologies, Boston: Kluwer; pp. 243–300. Doi: 10.1007/978-94-011-4405-6 10
- Juuso E.K., 1999b. Intelligent dynamic simulation of a lime kiln with linguistic equations. ESM'99: Modelling and Simulation: A tool for the Next Millennium, 13th European Simulation Multiconference, pp. 395–00, June 1–4, Warsaw, Poland.
- Juuso E.K., 2003a. Intelligent dynamic simulation of a solar collector field. In: Verbraeck A. and Hlupic V., eds., Simulation in Industry, 15th European Simulation Symposium ESS 2003, SCS, Gruner Druck, Erlangen, Germany, pp. 443–449.
- Juuso E.K., 2003b. Intelligent dynamic simulation of batch cooking. 44th Scandinavian Conference on Simulation and Modeling SIMS 2003, pp. 163–169, September 18 -19, Västerås, Sweden.
- Juuso E. K., 2004a. Adaptive intelligent modelling in hybrid process simulation. Proceedings of the fifth EUROSIM Congress on Modelling and Simulation, 6 pp, Sept. 6-10, Paris, France.
- Juuso E.K., 2004b. Dynamic simulation of a solar collector field with intelligent distributed parameter models. Proceedings of SIMS 2004 - the 45th Scandinavian Conference on Simulation and Modelling, pp. 141–153, September 23-24, Copenhagen, Denmark.
- Juuso E.K., 2004c. Integration of intelligent systems in development of smart adaptive systems. International Journal of Approximate Reasoning, 35(3), 307–337. DOI: 10.1016/j.ijar.2003.08.008
- Juuso E.K., 2009. Tuning of large-scale linguistic equation (LE) models with genetic algorithms. In: Kolehmainen M., ed., Revised selected papers of the International Conference on Adaptive and Natural Computing Algorithms - ICANNGA 2009, Kuopio, Finland, Lecture Notes in Computer Science, vol. LNCS 5495, Heidelberg: Springer-Verlag, pp. 161–170. DOI: 10.1007/978-3-642-04921-7_17.
- Juuso E.K., 2011. Recursive tuning of intelligent controllers of solar collector fields in changing operating conditions. Proceedings of the 18th World Congress The International Federation of Automatic Control; 2011 August 28 – September 2; Milano, pp. 12282–12288. doi: 10.3182/20110828-6-IT-1002.03621
- Juuso E.K., 2014. Intelligent Methods in Modelling and Simulation of Complex Systems. Simulation Notes Europe SNE, 24(1), 1-10. Selected SIMS 2013. DOI: 10.11128/sne.24.on.102221
- Juuso E. and Lahdelma S., 2010. Intelligent scaling of features in fault diagnosis. Proceedings of the 7th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Vol. 2, pp. 1358-1372, 22-24 June, Stratford-upon-Avon, UK.

- Juuso E. and Lahdelma S., 2011. Intelligent trend indices and recursive modelling in prognostics. Proceedings of the 8th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Vol. 1, pp. 440-450, June 20-22, Cardiff, UK.
- Juuso E.K. and Leiviskä K., 1992. Adaptive expert systems for metallurgical processes. In: Jämsä-Jounela S.L., Niemi A.J., eds., Expert Systems in Mineral and Metal Processing, Proceedings of the IFAC Workshop, August 26-28, 1991, Espoo, Finland; IFAC Workshop Series, Number 2, Oxford, UK : Pergamon, pp. 119-124.
- Juuso E. and Ruusunen M., 2013. Fatigue prediction with intelligent stress indices based on torque measurements in a rolling mill. Proceedings of CM 2013 and MFPT 2013, Vol. 1, pp. 460-471, 19-20 June, Krakow, Poland.
- Juuso E.K. and Yebra L., 2013. Model-based intelligent control of a solar energy collector field, Proceedings of 8h EUROSIM Congress on Modelling and Simulation, pp. 513-518, 10-13 September, 2013, Cardiff, UK, doi: 10.1109/EUROSIM.2013.92.
- Juuso E.K., Balsa P. and Valenzuela L., 1998. Multilevel linguistic equation controller applied to a 1 MW solar power plant. Proceedings of the ACC'98, vol. 6, pp. 3891-3895, June 24-26, Philadelphia, PA.
- Juuso E.K., Schauten D., Slawinski T. and Kiendl H. 2000. Combination of linguistic equations and the fuzzy-rosa method in dynamic simulation of a solar collector field. Proceedings of TOOLMET 2000 Symposium - Tool Environments and Development Methods for Intelligent Systems, pp. 63-77, April 13-14, Oulu, Finland.
- Lahdelma S. and Juuso E., 2011a. Signal processing and feature extraction by using real order derivatives and generalised norms. Part 1: Methodology. The International Journal of Condition Monitoring, 1(2), 46-53. DOI: 10.1784/204764211798303805.
- Lahdelma S. and Juuso E., 2011b. Signal processing and feature extraction by using real order derivatives and generalised norms. Part 2: Applications. The International Journal of Condition Monitoring, 1(2), 54-66. DOI: 10.1784/204764211798303814.
- Leiviskä K., Juuso E. and Isokangas A., 2001. Intelligent modelling of continuous pulp cooking. In: Leiviskä K., ed., Industrial Applications of Soft Computing, Heidelberg: Springer, pp. 147-158.
- Ljung L., 1999. System Identification - Theory for the User, 2nd edition, Upper Saddle River, N.J.: Prentice Hall.
- Ljung L., 2008. Perspectives on system identification. Plenary papers, milestone reports & selected survey papers, 17th IFAC World Congress, pp. 47-59, July 6-11, 2008, Seoul, Korea.
- Mendel J.M., 2007. Advances in type-2 fuzzy sets and systems, Information Sciences, 177(1), 84-110.
- Mäki T., Juuso E. and Leiviskä K., 2004. Fuzzy modelling and dynamic simulation of a fluidised bed granulator. Proceedings of AFNC'04 - the 2nd IFAC Workshop on Advanced Fuzzy/Neural, pp. 133-138, September 16-17, Oulu, Finland.
- Rummelhart D.E., Hinton G.E. and Williams R.J., 1986. Learning internal representations by error propagation. In: Rummelhart D.E. and McClelland J., eds., Parallel Data Processing, Cambridge, MA 4: M.I.T. Press, pp. 318-362.
- Saarela U., Leiviskä K., Juuso E. and Kosola A., 2003. Modelling of a fed-batch enzyme fermentation process. IFAC International Conference on Intelligent Control Systems and Signal Processing, April 8-11, Faro, Portugal.
- Sontag E., 1981. Nonlinear regulation: The piecewise linear approach. IEEE Transactions Automatic Control, 26(2), 346-358.
- Takagi T. and Sugeno M., 1985. Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics, 15(1), 116-132.
- Zadeh L.A., 1965. Fuzzy sets, Information and Control, 8(June), 338-353.

AUTHORS BIOGRAPHY

Esko K. Juuso is the Team Leader of Industrial process and Energy in Research Community of Automation, Faculty of Technology, University of Oulu. He has D.Sc.Tech. in Control and Systems Engineering and MSc in Technical Physics, both from University of Oulu. He has worked several years in metal industry and as a teacher and a project manager in several research projects on intelligent systems applications in various fields of industry. He is the developer of the linguistic equation (LE) approach and the nonlinear scaling methodology, which are currently used in various applications. His research interests are in the modelling and control of industrial processes with a special emphasis on combining intelligent control, fault diagnosis and performance monitoring into smart adaptive systems. He has published more than 300 technical and scientific publications. He has a long experience in European networks of excellence and scientific organizations in simulation, control and condition monitoring. He is currently the President of Juuso is the President of the European Federation of European Simulation Societies (EUROSIM) and a Member of the Management Committee of the International Society of Condition Monitoring (ISCM).

ALTERNATIVES AGGREGATION PETRI NETS APPLIED TO MODULAR MODELS OF DISCRETE EVENT SYSTEMS

Juan Ignacio Latorre Biel^(a), Emilio Jiménez Macías^(b), Jorge Luis García-Alcaraz^(c),
Juan Carlos Sáenz Díez Muro^(d), Mercedes Pérez de la Parte^(e)

^(a) Public University of Navarre, Tudela, Navarre, Spain

^(b,d,e) University of La Rioja, Logroño, Spain

^(c) Autonomous University of Ciudad Juarez, Ciudad Juarez, Mexico

^(a) juanignacio.latorre@unavarra.es, ^(b) emilio.jimenez@unirioja.es, ^(c) jorge.garcia@uacj.mx,
^(d) juan-carlos.saenz-diez@unirioja.es, ^(e) mercedes.perez@unirioja.es

ABSTRACT

The use of modelling formalisms for the design of discrete event systems presents many advantages, such as the possibility of structural analysis of the model or performance evaluation. However, the difficulty of the process to obtain an appropriate model of the system require the use of methodologies to ease the work of the designers. In this paper, two main subjects are discussed. On the one hand, the modular construction of Petri nets, alleviate the design process by the use of blocks that can be assembled to build up a complete Petri net model. On the other hand, the development of decision support systems may require the assessment of the performance and properties of complete models obtained from different combinations of modular blocks. The formalism of the alternatives aggregation Petri net may help in the development of compact and efficient models that may reduce the use of scarce computer resources.

Keywords: modular Petri nets, alternatives aggregation Petri nets, decision support systems, performance evaluation

1. INTRODUCTION

The application of modelling and simulation methodologies as base for the development of decision support systems, requires in many cases making suppositions on the structure of the real system itself. However, it is usual that the structure of the real system is not completely defined, but it should be clarified after making the subsequent decisions.

For this reason, the modular construction of models, allows the designer to use encapsulated blocks to construct the model of the system, easing the process of modelling. Moreover, it is common that the designer does not know, which is the best combination of blocks for the purposes of the system in process of being designed. For this reason, an automatic testing of the different possible combinations of the blocks for build up complete models would alleviate the modelling process.

Furthermore, as this problem is intensive in the use of computer resources, the development of adequate methodologies for obtaining compact and efficient models is a crucial issue in the development of decision support systems based on modelling and simulation.

The development of decision support systems based on modelling and simulation has been discussed by (Bruzzone and Longo, 2010), (Longo et al. 2013). The range of application of these decision support systems is broad, including the food industry (Latorre et al. 2014b), (Latorre et al. 2013b).

The use of the Petri nets as a versatile paradigm for modelling discrete event systems is considered in (Silva et al., 1993), (David and Alla 2005), and (Jensen and Kristensen 2009). In particular (Piera et al. 2004), (Latorre et al. 2013a) describe Petri net models for simulation. Moreover, (Mújica et al. 2010) is oriented to the application of simulation for quantifying a performance evaluation in the context of an optimization process.

A particularly difficult problem consists of designing a discrete event system, whose model should be chosen among a set of alternatives. In this case, it is convenient that the model of the system includes a set of exclusive entities (Latorre et al. 2010).

Moreover, an optimization process may be based on the simulation of a set of selected feasible decisions, chosen from a solution pool. In this case, the choice of the most promising decisions may be performed, for instance, by means of a search methodology guided by a metaheuristic (Latorre et al. 2013c).

In the following section, the topic of the modular Petri nets is discussed. Moreover, in section 3 brief comments on the concept of alternatives aggregation Petri net are provided, while in section 4, a discussion of the application of this formalism to represent a modular Petri net constructed as a sequence from combinations of four Petri subnets is given.

The next section is focussed on the conclusions and the future research work, while the last one is devoted to the bibliography.

2. MODULAR PETRI NETS

The construction process of models of discrete event systems that approximate complex real systems may be considered more an art than a precise and algorithmic procedure.

One of the most common methodology for coping with the modelling process of a complex discrete event system is the so called bottom-up approach. According to this idea a model is developed for every one of the subsystems in which the complete system can be divided. The level of detail required for every model depends its purpose and application.

This methodology derives naturally to the concept of modular construction of the Petri net model of a discrete event system. This idea implies the definition of a set of Petri net modules, ready to be assembled for the construction of complex models of real systems.

As an example, let us consider a set of four Petri subnets called R_A , R_B , R_C , and R_D . These Petri nets should present a compatible interfase to be connected to other subnets. In the subnets considered in this paper, a single input link transition and a single output transition appear. Moreover, the input link transition presents a single output place, belonging to the considered subnet. Similarly, the only output link transition presents a single input place, also belonging to the considered subnet.

See *figure 1* for a simplified representation of three of the mentioned Petri subnets, as well as some of their constituent elements, such as the input and output transitions and their output and input places.

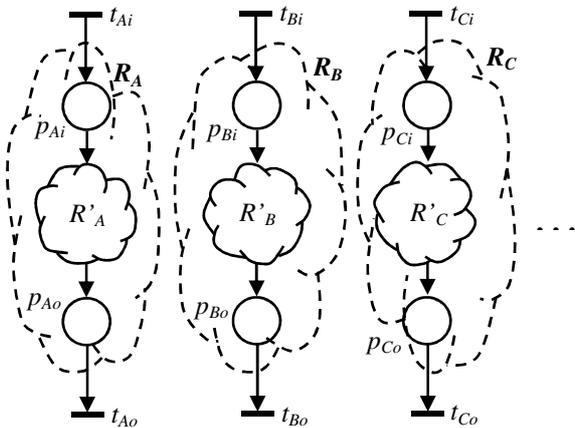


Figure 1: Three Petri subnets defined for the construction of a modular Petri net

In *Figure 1*, the input link transitions of the Petri subnets R_A , R_B , and R_C are, respectively, t_{Ai} , t_{Bi} , and t_{Ci} . On the other hand, the output link transitions of these same Petri subnets, in the same order, are t_{Ao} , t_{Bo} , and t_{Co} .

These subnets might be combined in different ways to build up the complete model of a real system. Prior to a

detailed analysis of every resulting model, it may be difficult to foresee the performance of any of them.

For this reason, a procedure can be defined in order to construct a set of feasible solutions for the complete model of the system by combining the subnets in appropriate ways. A second step in this procedure would be to develop a performance analysis of every complete model and, eventually, to compare the desired performance parameters calculated for every complete model in order to decide the best combination of subnets. *Figure 2* and *figure 3* show examples of combinations of the Petri subnets R_A , R_B , R_C , and R_D leading to complete models for a real system.

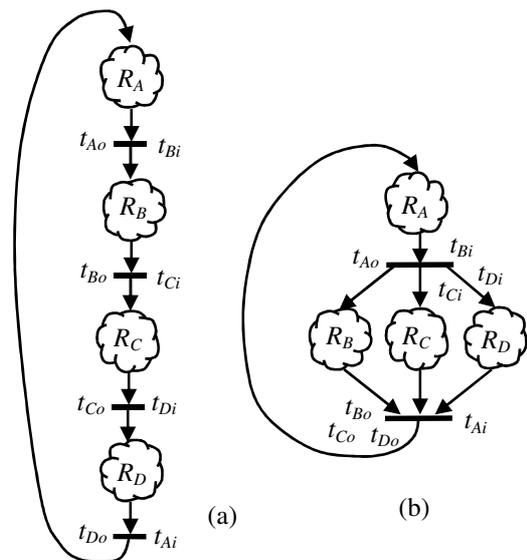


Figure 2: Different feasible combinations of 4 subnets

A Petri net model of a discrete event system can be developed for different purposes, such as performing structural analysis, calculating a certain subset of the set of reachable states, or for performance evaluation.

Specially in this last case, it is crucial for the success of the operation, to use an efficient algorithm, able to cope with the, sometimes, very costly process in terms of computer resources and time. One methodology, broadly used, that can virtually cope with every model, no matter how complex it is, is simulation.

Regarding the previous considerations, an important goal in the process of modeling a Petri net for performance evaluation is to obtain a formal description of the original system, as simple and reduced as it is possible. Hence, the costly process of simulation might be developed in affordable time and computer resources.

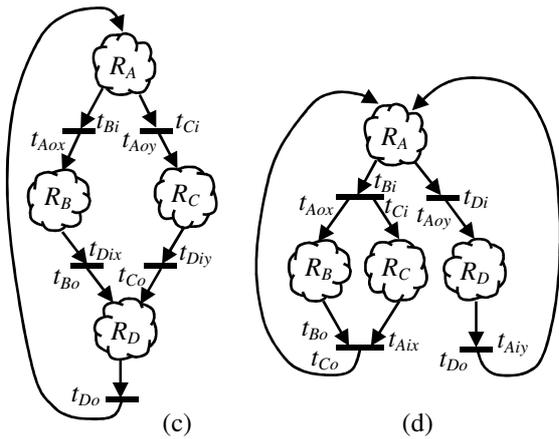


Figure 3: More feasible combinations of 4 subnets

It may be noticed the possibility of having different input or output link transitions for a given Petri subnet in a certain complete model. Also, consider that these multiple input or output link transitions present, respectively, a single output or input place. This possibility is illustrated in both Petri net models depicted in figure 3.

One important application for performance evaluation of Petri net models using simulation, consists in decision-making support with the purpose of designing a real system. The feasible models of the system in process of being designed can be compared by means of the result of a performance evaluation of every candidate model.

In the modular construction of a Petri net model, it may be interesting to test different or even all the feasible combinations of subnets that can be obtained. Every feasible solution is a candidate for being selected as the final model of the system in a design process; hence, every solution is an alternative model for the system. For this reason, a Petri net formalism able to represent alternative Petri nets, such as one containing a set of exclusive entities, should be considered.

3. ALTERNATIVES AGGREGATION PETRI NETS

The existence of alternative models for the development of a given discrete event system, require the use of particular formalisms, able to cope with the particularities of this kind of design problems.

A family of formalisms, based on the Petri net paradigm, specially developed for this purpose are the ones based on the idea of exclusive entities, deriving in formalisms such the set of alternative Petri nets or the alternatives aggregation Petri nets (Latorre et al., 2014a, 2014c, 2012,).

Both formalisms will be extensively used in this paper.

4. SEQUENCE OF FOUR SUBNETS

In this section, it will be considered the modular Petri net model constructed as a strict sequence from the four

Petri subnets called R_A , R_B , R_C , and R_D as they were mentioned in section 2.

All the possible complete Petri net models, built up from different combinations of the four Petri subnets in a sequence will be considered, as feasible solutions for the design process of a real system. One of the feasible solutions is presented in figure 2.(a), called R_{a1} , while most of the rest of them, R_{a2} to R_{a5} are presented in figure 4. The only remaining modular Petri net, not represented in a figure and called R_{a6} , presents the sequence of Petri subnets R_A , R_D , R_C , and R_B and its representation is less interesting than the remaining five feasible modular Petri nets, as it will be shown when constructing the alternatives aggregation Petri net.

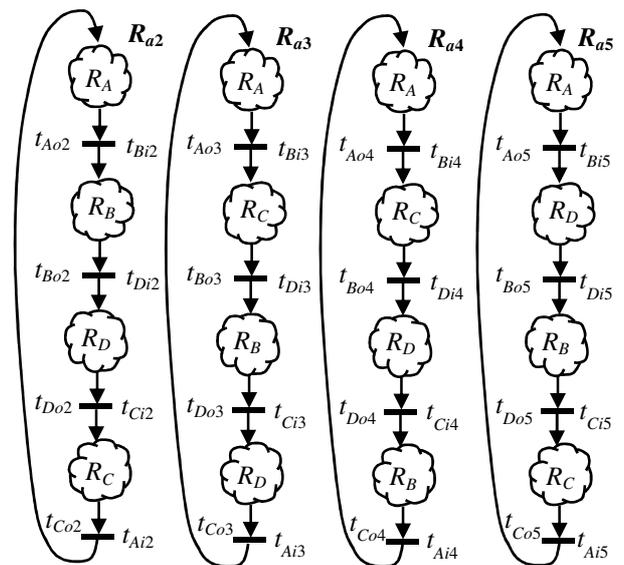


Figure 4: Different combinations of four Petri subnets for constructing modular Petri nets.

The six modular Petri nets may be feasible models of a real system in process of being designed. In this design process, it should be made a decision regarding the modular Petri net that best complies with the objectives of the real system, usually measured or quantified by means of performance parameters.

In fact, the modular Petri nets are alternative Petri nets; hence, the model of the real system in process of being designed can be represented by a formalism containing a set of exclusive entities. Furthermore, using the appropriate formalism may reduce considerably the computational resources required to solve the associated decision-making problem.

Regarding previous results in other case-studies, the formalism of the alternatives aggregation Petri nets is chosen for modelling the real system in process of being designed, that is to say, to represent in a single model the six alternative Petri nets, removing from the model the redundant information.

One of the algorithms for obtaining an alternatives aggregation Petri nets from a set of alternative Petri nets (Latorre et al., 2013c) states that any of the alternative Petri nets may be chosen as the seed for the resulting the alternatives aggregation Petri net model. See *Figure 2(a)*, where R_{a1} has been selected for this purpose. In this seed, every link transition should be associated to a choice variable a_1 as a guard of the transition itself.

The next steps of the algorithm for the construction of an alternatives aggregation Petri net from a set of alternative Petri nets belong to an iterative procedure, where every new alternative Petri net is added to the seed of the alternatives aggregation Petri net by including the new subnets (in this case-study there is not any of them) and all the link transitions associated to a guard function, given by the choice variable a_i that corresponds to the alternative Petri net R_{ai} containing the link transitions.

Following this algorithm, its second step consists of adding to the seed of the alternatives aggregation Petri net the alternative Petri net called R_{a2} (see *figure 4*). Due to the fact that this alternative Petri net does not present any subnet that is not already included in R_{a1} , then the only modification of the seed of the alternatives aggregation Petri net introduced by the alternative Petri net R_{a2} is the addition of the link transitions. In the case of R_{a2} , the link transitions are (see *figure 4*) t_{A02} , t_{B02} , t_{D02} , and t_{C02} , or, what is the same, the transitions called t_{B12} , t_{D12} , t_{C12} , and t_{A12} . This transitions are added to the seed of the alternatives aggregation Petri net associated to the choice variable a_2 .

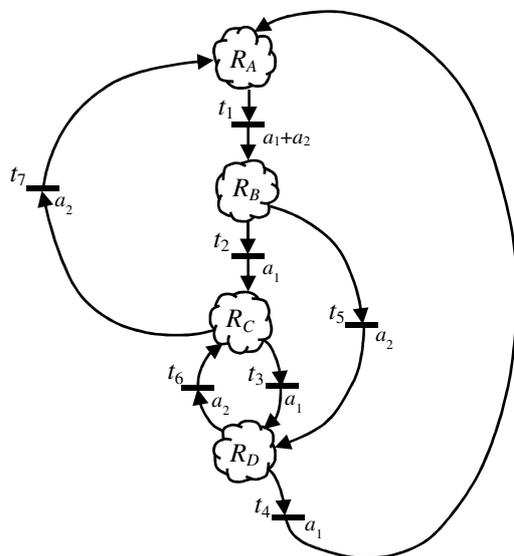


Figure 5: Second step in the construction of the alternatives aggregation Petri net

Once the new link transitions have been included in the seed of the alternatives aggregation Petri net, it is possible to apply a reduction rule, which groups together the quasi-identical transitions, modifying the associated function of choice variables. If it is possible to apply this reduction rule to a given operation, then

the resulting model will be more simple, since it contains a lower number of transitions. In fact, a couple of quasi-identical transitions verify, having the same set of input and output places, as well as the same weight in the input and output arcs. Moreover, the transitions should be associated to different functions of choice variables, otherwise the transitions would be identical instead of quasi-identical ones.

In the example of this second step of the algorithm, a link transition of R_{a2} has been merged with t_1 , just by constructing an associated function of choice variables with the logic operator OR applied to both choice variables a_1 and a_2 (see *figure 5*). The link transitions that have been added in this second step of the algorithm appear in *figure 5* and are named t_5 , t_6 , and t_7 .

In *figure 5*, the second step of the algorithm can be seen, while the complete alternatives aggregation Petri net is shown in *figure 6*.

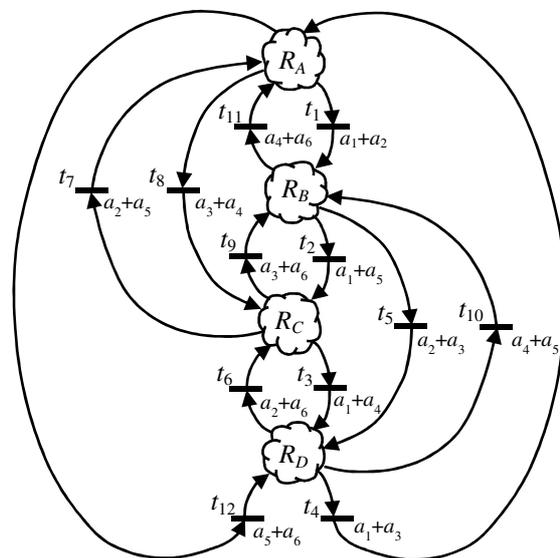


Figure 6: Complete alternatives aggregation Petri net

In order to complete the alternatives aggregation Petri net depicted in *figure 7*, it has been necessary to include five more link transitions, in addition to the four ones introduced by R_{a1} and to the three new transitions delivered by R_{a2} .

In the following paragraphs, it will be discussed the compacity in the matricial representation of both models of a system in process of being designed by means of a set of alternative Petri nets (see *figure 4*). The compacity of the alternatives aggregation Petri net is based in the fact that a large amount of redundant information, present in the set of alternative Petri nets, has been removed from the model: the Petri subnets, which appear in every alternative Petri net.

The incidence matrix of any of the six alternative Petri nets has a dimension that can be calculated as follows:

Let us consider that the dimension of a Petri net is given by the multiplication of the number of rows and the number of columns of the associated incidence matrix.

According to this idea, the dimensions of the four Petri subnets of this case-study are:

$M(R_A) \in M_{A_r \times A_c}$, where A_r is the number of rows of the incidence matrix of R_A and A_c is its number of columns.

Analogously:

$M(R_B) \in M_{B_r \times B_c}$; $M(R_C) \in M_{C_r \times C_c}$; $M(R_D) \in M_{D_r \times D_c}$

On the other hand, any of the alternative Petri nets R_{a1} , R_{a2} , R_{a3} , R_{a4} , R_{a5} , and R_{a6} present the same dimension, which can be calculated as it is described below.

$M(R_{a1}), M(R_{a2}), M(R_{a3}), M(R_{a4}), M(R_{a5}), M(R_{a6}) \in M_{r \times c}$ where

$$r = A_r + B_r + C_r + D_r \quad (1)$$

$$c = A_c + B_c + C_c + D_c + 4 \quad (2)$$

It should be considered that r is the number of rows of the incidence matrix of R_{ai} , with $i = 1, \dots, 6$. On the other hand, c is the number of columns of the incidence matrix of R_{ai} , with $i = 1, \dots, 6$.

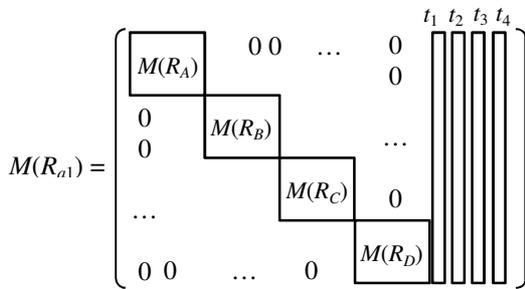


Figure 7: Incidence matrix of an alternative Petri net

The number 4 that appears in the expression (2) is originated by the four link transitions included in every R_{ai} , with $i = 1, \dots, 6$.

Figure 7 shows a representation of the incidence matrix of any of the alternative Petri nets

The calculation of the resulting alternatives aggregation Petri net, R_{AA} , can be developed in a similar way.

$M(R_{AA}) \in M_{r' \times c'}$, where

$$r' = A_r + B_r + C_r + D_r \quad (3)$$

$$c' = A_c + B_c + C_c + D_c + 4 + 8 \quad (4)$$

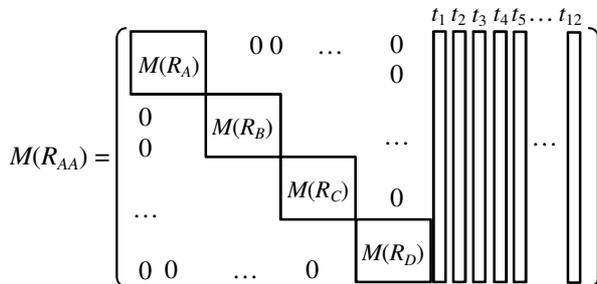


Figure 8: Incidence matrix of the alternatives aggregation Petri net

The comments on the previous expressions (3) and (4) are the same as done before. However, the number 8 in (4) is a consequence of the fact that the aggregation of

alternative Petri nets to the seed of the alternatives aggregation Petri net introduces 8 new link transitions.

Figure 8 shows the incidence matrix of the alternatives aggregation Petri net.

In order to compare the convenience of using one of both models of a modular Petri net for developing a decision support system, the computer resources required to execute an optimisation algorithm using the simulation of one of both models can be compared.

In fact, some important computer requirements depend on the size of the model itself. Hence, the comparison between the set of alternative Petri nets and the alternatives aggregation Petri nets can be performed calculating a size ratio, defined in the following way:

$$size\ ratio = \frac{size\ of\ the\ alternatives\ aggregation\ Petri\ net}{size\ of\ the\ set\ of\ six\ alternative\ Petri\ nets}$$

$$size\ ratio = \frac{r' \times c'}{6 \times r \times c} \quad (5)$$

$$size\ ratio = \frac{(A_r + B_r + C_r + D_r) \times (A_c + B_c + C_c + D_c + 12)}{6 \times (A_r + B_r + C_r + D_r) \times (A_c + B_c + C_c + D_c + 4)}$$

Due to the fact that the number of rows is the same in the alternatives aggregation Petri net and in the alternatives aggregation Petri net, it is possible to cancel this number in the numerator and the denominator of the expression. As a result, it is possible to see that the size ratio does not depend on the number of places of the Petri nets.

$$size\ ratio = \frac{(A_c + B_c + C_c + D_c + 12)}{6 \times (A_c + B_c + C_c + D_c + 4)}$$

As an example, if every subnet presents 5 internal transitions, the size ratio has the value:

$$size\ ratio = \frac{(5 + 5 + 5 + 5 + 12)}{6 \times (5 + 5 + 5 + 5 + 4)} = \frac{1}{4.5}$$

In other words, for a small size of the Petri subnets, 5 internal transitions, the alternatives aggregation Petri net is 4.5 times smaller than the equivalent set of alternative Petri nets.

The calculation of the amount of redundant information removed from the set of alternative Petri nets is:

$$100 \times (1 - size\ ratio) = 100 \times \left(1 - \frac{1}{4.5}\right) \cong 77.8\%$$

Finally, it is possible to calculate the upper bound in both parameters: the size ratio and the amount of removed redundant information in the model of the system:

Let us call x_A , x_B , x_C , and x_D the number of columns of the incidence matrices of the Petri subnets R_A , R_B , R_C , and R_D respectively. Let us call $x = x_A + x_B + x_C + x_D$.

$$\lim_{x \rightarrow \infty} (size\ ratio) = \lim_{x \rightarrow \infty} \left(\frac{(x + 12)}{6 \times (x + 4)} \right) = \frac{1}{6}$$

On the other hand, the upper bound of the percentage of redundant information removed from the set of alternative Petri nets is:

$$100 \times (1 - \text{size ratio}) = 100 \times (1 - \frac{1}{6}) \cong 83.3\%$$

As it can be seen, the alternatives aggregation Petri net outperforms the complete set of alternative Petri nets, while this last formalism is more intuitive and easy to apply for the modelling of discrete event systems.

It is also interesting to point out, that the model based on the alternatives aggregation Petri net should add some additional information to the model itself: the functions of choice variables associated to every link transition.

5. CONCLUSIONS AND FUTURE RESEARCH LINES

In this paper, some considerations on the modular construction of Petri net models have been introduced. Furthermore, the application of these models to decision-support systems based on simulation requires the development of exigent algorithms in terms of computer resources.

In order to overcome or at list palliate this problem, a transformation of a non-efficient model based on a set of alternative Petri nets into an alternatives aggregation Petri net is discussed. Two parameters have been defined and calculated: the size ratio, to quantify the relative size between both models and the percentage of redundant information that has been removed in the alternatives aggregation Petri net but not in the original set of alternative Petri nets.

As conclusions, it can be stated that the modular construction of Petri net models is a promising research line to develop decision support systems to construct models of discrete event systems. On the other hand, there are formalisms, such as the alternatives aggregation Petri nets, able to reduce significantly the size of a model in the case of models composed of sequences of four Petri subnets.

As future research actions, it can be considered to extend the discussion of these methodologies and results to other layouts in the modular Petri nets, as well as considering a larger number of subnets and different constitutions of the subnets.

REFERENCES

- Bruzzzone A.G. and Longo F. 2010. An advanced system for supporting the decision process within large-scale retail stores. *Simulation*; 86: 742–762.
- David R and Alla H. 2005. *Discrete, Continuous and Hybrid Petri Nets*. Berlin: Springer.
- Jensen, K., Kristensen, L.M. 2009. *Colored Petri nets. Modelling and Validation of Concurrent Systems*, Springer.
- Latorre, J.I., Jiménez, E., Pérez, M. 2014. Sequence of decisions on discrete event systems modeled by Petri nets with structural alternative configurations. *Journal of Computational Science*. 5(3): 387-394 (2014).
- Latorre, J.I. and Jiménez, E., Blanco, J., Sáenz, J. C. 2014. Optimal Design of an Olive Oil Mill by Means of the Simulation of a Petri Net Model. *International Journal of Food Engineering*. Published online, May 2014.
- Latorre, J.I., Jiménez, E., de la Parte, M., Blanco, J., Martínez, E. 2014. Control of Discrete Event Systems by Means of Discrete Optimization and Disjunctive Colored PNs: Application to Manufacturing Facilities. *Abstract and Applied Analysis*. Volume 2014, 16 pages.
- Latorre, J.I. and Jiménez, E. 2013. Simulation-based optimization of discrete event systems with alternative structural configurations using distributed computation and the Petri net paradigm. *Simulation*. November 2013 89 (11), pp. 1310-1334
- Latorre, J.I. and Jiménez, E., Blanco, J., Sáenz, J. C. 2013. Decision Support in the Rioja Wine Production Sector. *International Journal of Food Engineering*. Volume 9, Issue 3 (Jun 2013). Page 267.
- Latorre, J.I., Jiménez, E., Pérez, M. 2013. The optimization problem based on alternatives aggregation Petri nets as models for industrial discrete event systems. *Simulation*. March 2013 89 (3), pp. 346-361.
- Latorre, J.I., Jiménez, E. 2012. Colored Petri Nets as a Formalism to Represent Alternative Models for a Discrete Event System. 24th European Modelling and Simulation Symposium (EMSS 12). Vienna, 2012.
- Latorre, J.I., Jiménez, E., Pérez, M. 2010. Colored Petri Nets as a Formalism to Represent Alternative Models for a Discrete Event System. 22nd European Modelling and Simulation Symposium (EMSS 10). Fez, Morocco, 247-252, 2010.
- Longo, F., Nicoletti, L., Chiurco, A., Solis, A. O., Massei, M., Diaz, R. 2013. Investigating the behavior of a shop order manufacturing system by using simulation. *SpringSim (EAIA) 2013*: 7
- Mújica M. A., Piera M.A., and Narciso M. 2010. Revisiting state space exploration of timed coloured Petri net models to optimize manufacturing system's performance. *Simulation Modelling Practice Theory* 2010; 18: 1225–1241.
- Piera, M.Á., Narciso, M., Guasch, A., and Riera, D. 2004. Optimization of logistic and manufacturing system through simulation: A colored Petri net-based methodology. *Simulation*, vol. 80, number 3, pp 121-129, May 2004
- Silva, M. "Introducing Petri nets", In *Practice of Petri Nets in Manufacturing*", Di Cesare, F., (editor), pp. 1-62. Ed. Chapman&Hall. 1993.

INSIDE VIRTUAL: A NEW APP FOR INTERACTIVE AND INTELLIGENT CULTURAL HERITAGE FRUITION

Francesco Longo^(a), Letizia Nicoletti^(b), Gaetano Florio^(c), Marco Vetranò^(d), Luigi Bruno^(e), Lorenzo Caputi^(f)

^{(a), (c), (e)} DIMEG, University of Calabria, Italy

^(f)Department of Physics, University of Calabria, Italy

^{(b)(d)}CAL-TEK SRL, Italy

^(a) f.longo@unical.it, ^(b) l.nicoletti@cal-tek.eu, ^(c) gaetano.florio@unical.it, ^(d) m.vetranò@cal-tek.eu,
^(e) l.bruno@msc-les.org, ^(f) lorenzo.caputi@unical.it

ABSTRACT

Technology is even more embedded into everyday life and its pervasiveness grows ceaselessly. People are becoming more and more addicted to technology that is seen a mean to improve life quality, extend human capabilities and enjoy free time. In this framework, the proposed research work seeks to combine information and communication technologies, web technologies, mobile technologies and serious games to create new models for cultural contents fruition and dissemination. The proposed approach results in a desktop and web application, INSIDE VIRTUAL, that can be installed on mobile devices. The user of INSIDE VIRTUAL, is involved in a learning path under the guise of a virtual museum (archaeological site) visitor that, in proximity of an artifact, receives information and cultural contents from dedicated avatars (driven by intelligent agents). Such avatars act as guides and are able to interact with the user avatar in several ways (even vocally through questions) and deliver custom contents based on the user preferences gathered at the game start and during the game evolution. In this paper, some aspects of the INSIDE VIRTUAL design and development process are covered and preliminary implementation activities are introduced.

Keywords: Cultural Heritage, Serious Games, Modeling.

1. INTRODUCTION

The proposed research is aimed at introducing some ongoing research activities belonging to a joint research project that involves CAL-TEK SRL and the MSC-LES Lab of the University of Calabria (The overall project presentation and architecture can be found in Longo et al. 2014). Being aware of the crucial role that cultural heritage has in education and tourism, the main idea behind this work is to promote new strategies for cultural contents dissemination and fruition. In more detail, the proposed solution combines cultural contents with digital, mobile and immersive technologies with the aim of making contents fruition a unique and engaging experience. One of the main outcomes of the aforementioned research project is INSIDE VIRTUAL, a Serious Game, delivered both in the form of mobile and desktop app, for virtual interactive visits in virtual museums/archaeological sites. INSIDE VIRTUAL is envisaged as an educational tool able to transfer contents in an effective manner and capture the interest of its users.

Indeed, INSIDE VIRTUAL recreates not only a museum/archeological site and the artifacts in it but includes also advanced functionalities to let the player (that plays the role of a virtual visitor represented by an avatar) interact in a dynamic way with the virtual world as well as other avatars acting as guides. As a matter of facts, when the game starts, the player can choose to visit the museum by himself or rather to be supported by a virtual guide. In such a case the player has the opportunity to communicate and interact with the guide as it happens in the real world through questions and answers. To this end, guide avatars are driven by Intelligent Agents that are meant to provide the guide with a proactive behavior and to be responsive to any kind of user inputs (such as questions, preferences, etc.). This way, leveraging on visual aspects (3D reconstructions), on advanced interaction modes (vocal interaction between the player and the guide), and proactive functions (along the visit information on user preferences is gathered in order to drive the player toward artifacts he may be interested in) INSIDE VIRTUAL seeks to engage and involve the user making is attention and interest toward cultural heritage grow. To fulfill such a challenging research, the design and development process has been accurately and critically planned. At the early stage, conceptual modeling has been a crucial phase. Indeed, conceptual modeling has resulted in the identification and representation of all the entities and their mutual relationships setting a reference framework for the whole development process. In other words, conceptual modeling has enabled the identification and formalization of the logics that drive the game evolution based on the inputs provided by its users. It has been carried out with the purpose of:

- creating a common and shared knowledge base for the entire development team;
- documenting specifications, features and functionalities the app has been (will be) provided with;
- facilitating maintenance;
- promoting sustainability over time even if requirements may change.

Taking into account all these aspects, conceptual modeling has been carried out through UML (Unified Modeling Language) whose characteristics in terms of expressiveness and simplicity proved extremely useful for the integration of the research team members and for

setting up effective communication strategies with the museums that have been involved as end-users of the research outcomes namely the Museum of Operation Avalanche located at Eboli, and the Silk Museum of Mendicino, both located in Italy. After conceptual modeling, the issues of providing guide avatars with intelligent behaviors have been explored and the reference frameworks for Intelligent Agents implementation have been set out. After addressing these aspects, some preliminary implementation activities have been introduced. Thus the reminder of the paper is organized as follows: section 2 introduces and discusses the main INSIDE VIRTUAL conceptual models; section 3 is centered on the main reference approaches for Intelligent Agents implementation; lastly, section 4 introduces the development environment and the implementation of some basic features.

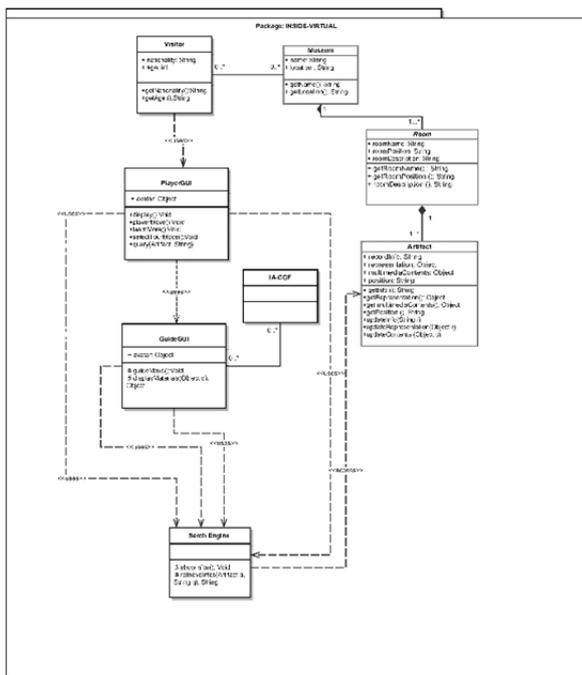


Figure 1 – Class Diagram

2. STATE OF THE ART

To date, digital technologies applied to cultural heritage have been mainly used for the reconstruction of the original appearance of artifacts as well as the storage and archiving of information in media format. Indeed, reviewing the state-of-the-art, it is possible to ascertain that, most of the time, researchers and research communities are mainly concerned with the preservation of cultural heritage through modeling techniques, graphics and virtual reality (Addison, 2000). As shown by Anderson et al. (2010), the use of Modeling, Simulation and Serious Games in the cultural heritage for teaching/learning historical contents has been little considered.

Anyway, talking about the digitization of cultural and historical heritage, it is worth considering virtual museums and information kiosks that are meant to make the contents more appealing. In addition, Sylaiou et al. (2009) point out

that museums have different possibilities, in terms of tools and methodologies, to make contents fruition more engaging. Mostly, virtual reality and/or augmented reality technologies are deployed in museums/archeological sites to enhance visitor's experience.

As for virtual reality, the most common solutions range from simple interactive screens to full-immersive systems that may include projection systems and 3D stereoscopic glasses. Augmented reality applications, instead, offer a greater variety of solutions. One of the very first solutions dates back to 1995 and was proposed by Bederson. This solution was based on the imposition of audio contents based on the user's location. Anyway, since 1995, many other applications were proposed such as Mase et al. (1996), Brogni et al. (1999), White et al. (2004), Liarokapis et al. (2008), Sylaiou et al. (2009), and many others. It is worth mentioning also that remarkable research efforts have been oriented toward more advanced solutions based on the coexistence of real and virtual objects (Mixed Augmented Reality). Consider for instance Hall et al. (2001), Hughes et al. (2004) and many others.

Besides, when dealing with digital technologies applied to cultural heritage, it is possible to notice that Serious Games are gaining increasing attention. According to the taxonomy adopted by Mortara et al. (2013), Serious Games can have various purposes including cultural awareness, historical reconstruction and can vary according to the learning objectives, genre (e.g. adventure, simulation) and application context. Representative examples in this regard have been proposed by Froschauer et al. (2010). Here, the INSIDE VIRTUAL contribution to the state of art relies not only on the innovation potential of Serious Gaming technologies deployed through mobile and desktop devices for cultural contents dissemination and promotion but also, and above all, on the combination of SG with Intelligent Agents able to drive the scenario evolution based on events sequences (actions and reactions) generated by the reciprocal and continuous interaction between real users and Intelligent Agents.

In this sector, the analysis of the state-of-the-art reveals that agents are a powerful metaphor for the modeling and analysis through simulation of complex systems in various application domains (Michel et al., 2009) and for the implementation of complex software systems (Jennings, 2000). According to this paradigm, a system is modeled as a set of entities (agents) capable of operating in a given environment with different degrees of autonomy and intelligence cooperating and/or competing to achieve their goals or those of the company/organization they belong to (Woolridge, 2002).

A relevant example on agent-based environments for cultural experience in museums can be found in Costantini et al. (2008) where the agent technology is used to offer suggestions in line with the profile and interests of visitors by monitoring (via satellite) the movements of visitors within an archaeological park.

At international level, many projects have been proposed and developed in the field of advanced fruition of cultural heritage. I.e. Anderson et al. (2010) explore some existing projects focused on ancient historical sites

visualization and virtual reconstruction such as Nova Roma (Frischer, 2008), Ancient Pompeii (Maim et al., 2007), the Parthenon (Debevec, 2005). Nova Roma is a multi-year project on the 3D reconstruction of the Ancient Rome as it was in 320 AD and it has been integrated with a game engine in order to allow interaction between player and non-player avatars. Ancient Pompeii is a project based on the reconstruction of ancient Pompeii, through procedural modeling with real-time simulated entities. Analyzing in detail the technical specifications of both projects, it can be ascertained that: greater interactivity should be ensured as well as a more consistent web technology and the possibility of having apps for mobile devices.

There are numerous other projects such as the PEACH project (Stock and Zancanaro, 2010) and the MINERVA project (Amigoni e Schiaffonati, 2009). The PEACH project seeks to take advantage of synergies among different technologies (multi-agent architecture, audio-visual technologies, sensors for 3D visualization, etc..) in order to enhance the experience of cultural heritage fruition whereas MINERVA is a system that, in its latest release, supports the curators in the automatic creation of virtual museums. Many other projects, funded by the European Community under the Seventh Framework Programme, are also reported below.

The V.MUST.NET project aims to define a knowledge base, a common language, an ontology dedicated to the creation of virtual museums as well as to identify innovative visualization and interaction tools for virtual museums of the future. The 3D-COFORM project deals with all aspects of the 3D digitizing, the semantics of shapes and material properties of cultural heritage objects, with the aim of moving forward the 3D digitizing frontiers. A similar project (but smaller in size and partnership) is DASI that has the objective to digitalize the inscriptions found in the Arabian Peninsula before the Islamic period. Finally, the DECIPHER project promotes the development of new solutions that combine rules, engines, virtual environments and interfaces in order to support curators in the preparation and fruition of cultural goods which are as much as possible in line with the interests of the visitors.

This analysis of the state art allows ascertaining that INSIDE VIRTUAL contribution is the ability to combine, in a multidisciplinary approach, different methodologies (serious games, intelligent agents, virtual reality) and technologies (web technologies, mobile devices) resulting in a unique tool that can be used for cultural heritage contents and site promotion and dissemination. Indeed, INSIDE VIRTUAL will be deployed as an app for mobile and desktop devices and therefore with a great extent of accessibility even considering that it is meant to be used wherever and whenever the user wants. In addition, contents delivery entails greater interactivity (compared to the existing solutions mentioned above) from the users and offers advanced features in terms of visualization, contents customization, proactivity and responsiveness to users inputs. In addressing such requirements, the research and development team of the INSIDE VIRTUAL app takes

advantage of many experiences gained along the years in the field of simulation (Longo 2012; Bruzzone and Longo, 2013; Bruzzone and Longo, 2014;), virtual environments (Longo et al. 2013) and agent-driven systems (Bruzzone et al. 2012; Bruzzone et al 2014).

3. INSIDE VIRTUAL CONCEPTUAL MODELS

As mentioned before, conceptual modeling has been carried out through UML diagrams. To have a clear operational picture of how the game works, of the functionalities it has to provide, of the logics that rule its evolution over time as well as of the basic components and entities involved, the following diagrams have been drawn up: the class diagram, the use case diagram, the state machine diagram, the sequence diagram and the activity diagrams have been drawn. The class diagram is shown in Fig. 1 and highlights the main entities and interactions that occur within the scope of INSIDE VIRTUAL.

Therefore it is a very useful mean to represent in a rather direct and immediate way the application domain and its key elements such as the visitor, the museum environment, the halls, the exhibits and the Graphic User Interface (GUI). In particular, the GUI makes specific functions available to the users such as the request for additional information, the option to choose between guided and free tours inside the virtual site, the possibility to select how the user and the guide avatars look, etc. Furthermore, other basic components include intelligent agents, computer generated actions and a search engine. Such components allow defining the logic that drives the game evolution as time goes by as well as the guide behavior and its interaction with the user when guided tours are activated. On the other hand, the search engine role is to allow the user to ask for further information when visiting the virtual museum/archeological site both in the guided tour and in the free tour mode.

As well-known, the class diagram provides a static representation and as a consequence does not show the game evolution over time based on user interaction patterns. To overcome this limitation, the use case diagram has been drawn in Fig. 2, where the app functional requirements are depicted.

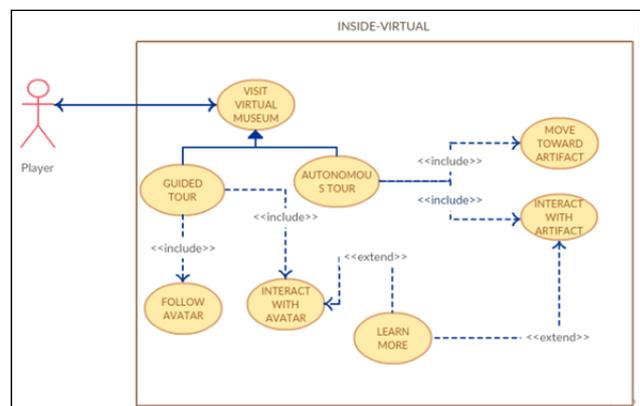


Figure 2 – Use case diagram

As shown in the figure, the virtual tour in a museum can be driven by a guide or by the player. In the first case,

the guide leads the player through a virtual tour in which the user can interact with the guide to acquire information and eventually ask questions on specific aspects and curiosities that arise while visiting a particular artifact. On the other hand, the player has also the possibility to move freely in the museum as an independent visitor and when an artifact is reached he may decide to interact with the object through dialogue and / or information boxes. Moreover, he may seek additional information and / or ask specific questions.

System states and transitions are, represented in the State Machine diagram shown in Fig 3. This chart is useful for understanding the serious game dynamics and evolution patterns based on the actions taken by the user/player. In fact, the picture highlights that the transitions between states are purely determined by the choices and actions of the user (e.g. The selection of initial parameters, of the visit type, the transition from one object to another, the passage from one level to the next, etc.) putting in evidence the "event-driven" nature of INSIDE VIRTUAL.

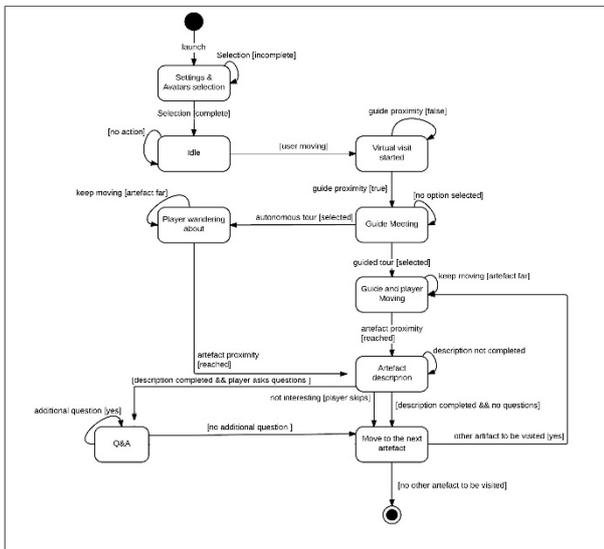


Figure 3 – State machine diagram from the player perspective

A relevant aspect when dealing with the evolution of the game is represented by the intelligent logics that govern the guide's behavior. For this reason, a State Machine Diagram, intended to represent what happens from the guide point of view, has been drawn up (Figure 4). In particular, this diagram is meant to show the states that characterize the guide such as: the movement from one artifact to another, the waiting for the player arrival before information delivery, the resumption of the tour when the user ceases to ask questions, the reaction to user inputs, etc.

Another crucial aspect is the need to identify and model the interactions between objects, actors and entities in the system with a particular focus on information exchanges. To this end, temporal and casual links are displayed within the sequence diagram in Figure 5.

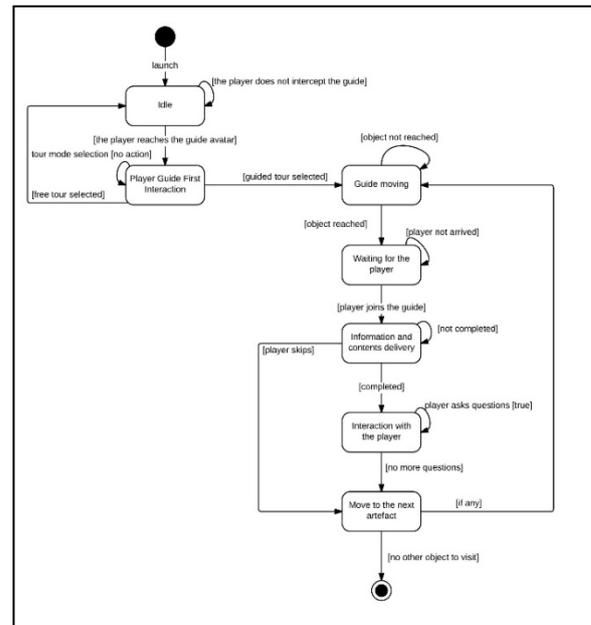


Figure 4: State machine diagram for the guide perspective

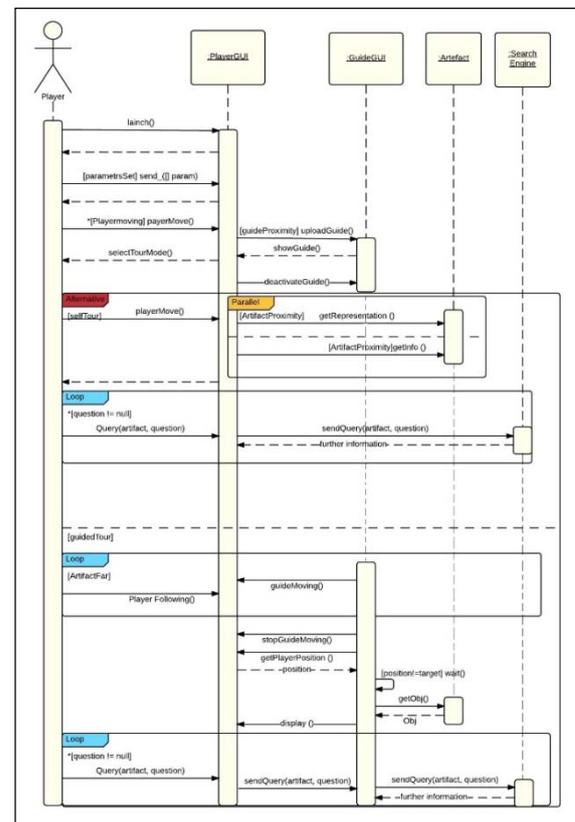


Figure 5 – Sequence diagram

Thus the INSIDE VIRTUAL design has acquired a greater level of detail as the analysis of interactions and their timeline clarifies the nature and the way communications occur establishing which communications are synchronous and which asynchronous. On the other hand, a relevant achievement of the sequence diagram is to have methods and classes mapped in order to give evidence of the pieces of code required to make the game evolve properly over time. In other words, the sequence diagram, has had a

critical role to characterize and express all the INSIDE VIRTUAL scenarios. Finally, the Activity Diagram shown in Fig. 6 describes the INSIDE VIRTUAL workflow completing the overall representation of INSIDE VIRTUAL conceptual models.

4. INTELLIGENT AGENTS AND COMPUTER GENERATED FORCES IN INSIDE VIRTUAL

As shown in Section 2, great attention has been paid on Conceptual Modeling that has ended up with the UML diagrams reported so far. Such diagrams provide a detailed and intuitive description of INSIDE VIRTUAL in terms of components, features, requirements and dynamics but not only. As a matter of facts, this phase has been crucial to establish the requirements as well as the behavioral patterns of the virtual guide (that is driven by Intelligent Agents) along the interaction process with the virtual visitor. In other words, UML diagrams and the underpinning modeling effort have highly contributed to clarify the structural and behavioral models that Intelligent Agents (IA) have to embody.

Recalling the definition of Goi and Torre (2009) an agent is a computational system that:

- interacts with and is responsive to its surroundings;
- is able to take decisions and to act, in an autonomous way, to achieve a goal (which may be predefined or negotiated). Therefore it is proactive.
- is able to communicate (coordinate, cooperate, negotiate) with other agents (and / or with human beings). Therefore it is capable of social interaction.

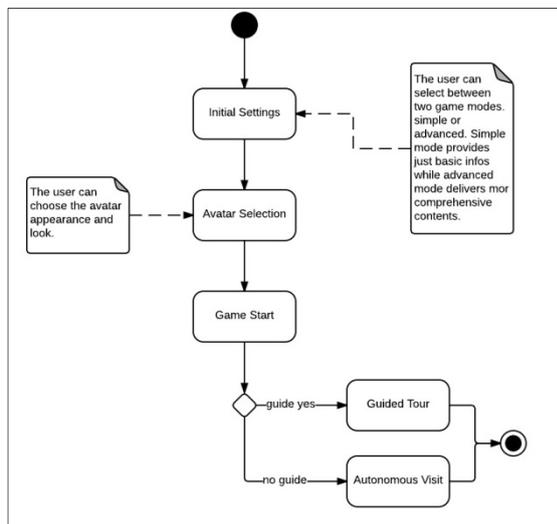


Figure 6 – Activity diagram

This definition fits very well the role and the features of the guide within INSIDE VIRTUAL. In fact, the guide interacts with the museum environment and the objects in it, communicates with the player and acts autonomously choosing how to design the guided tour based on the inputs it gains from the user and the user profile. Needless to say that the guide can attain the user interest to the

extent that it is able to deliver custom contents that fall under the user’s preferences. Thus Intelligent agents driving the guide behavior are developed through a symbolic deliberative approach that provides the agent with both declarative (or domain) and prescriptive knowledge. Here prescriptive knowledge consists of a set of rules that allow identifying a proper sequence of artifacts and information to be delivered. As a consequence, building agents with this approach requires a formal description of the domain specific knowledge (i.e. the museum and the artifacts in it), of the objectives to achieve and of the actions (along with related preconditions and effects) that can contribute to reach such objectives. In greater detail and from a methodological point of view, the reference framework is that of practical agents that are agents governed by cognitive processes in two steps: the goal selection (which depends on the input provided by the user when the game starts for instance the age) and the planning phase where the agent establishes which actions have to be implemented to achieve the goal. For planning purposes the agent uses a dedicated algorithm that allows selecting the most suitable plan from a selection of possible plans. In addition, considering that an intelligent behavior is, by its nature, related to user interactions, such an approach has been integrated in order to make the agent responsive to the user inputs. For example, if the user chooses to skip an object the agent has to revise its action plan passing over similar objects. Conversely, if the user asks a lot of questions about a particular object, it is likely that objects of the same type /period are of interest to the user. Thus the agent has to be able to infer user preferences collecting as many inputs as possible and changing the plan or selecting a new one accordingly. Implementing such capabilities requires a great effort to organize and structure the plans as well as to manage metadata properly so as to ensure an effective and dynamic rescheduling during guided tours.

4. INSIDE VIRTUAL DEVELOPMENT ENVIRONMENT AND DEVELOPMENT ACTIVITIES

As discussed so far, in order to encode the conceptual models previously introduced, an extremely flexible development environment is required. It should be a cross-platform environment able to support the development of interactive Serious Games with advanced functionalities and properties. To this end, after a thoroughly evaluation of both open and owner systems, UNITY3D platform has been selected. UNITY3D has a number of extremely important features such as a visual editor with advanced features for display and live previews, extensibility, ability to create optimized contents ensuring excellent performances and high-quality graphics. Thus UNITY 3D is currently used to recreate a virtual museum that will implement all the INSIDE-VIRTUAL functionalities. The virtual museum will be based on a real museum: to this end, the museum involved in the project is the Museum of Operation Avalanche (MOA), located in Eboli, Italy. This museum was founded in 2012 to remember some historical events of the World War II. In September 1943, during the

Second World War, Salerno Coast (in South Italy) was a war theatre for one of the most relevant landings in the history, called Operation Avalanche by Allied Forces (a similar landing was the Normandy landing, in 1944). "Operation Avalanche" was characterized by 1000 ships (400 Army ships and 600 merchant ships adapted for military operations) and 170.000 soldiers (100.000 English Soldiers and 70.000 Americans). The MOA is a place where visitors can relive some events of World War II. Figures 7 shows some military helmets used by soldiers during the Operation Avalanche landing; figures 8 and 9 respectively show a German Machine Gun and an American Ammunition Box.



Figure 7 – Military Helmets at MOA



Figure 8 – German Machine Gun at MOA



Figure 9 – Ammunition Box at MOA

Realistic 3D models are currently under development and will be imported in the virtual environment to give the sensation to experience a real museum visit. Figures 10 and 11 respectively show the 3D model of a military helmet and an ammunition box; both models will be successively imported within the 3D virtual environment of the INSIDE-VIRTUAL app.



Figure 10 – An example of 3D models (a military helmet) that will be used within the INSIDE-VIRTUAL app



Figure 11 – An example of 3D model (an ammunition box) that will be used within the INSIDE-VIRTUAL app

Besides, the avatars movement management system (for the player and for the guide) was set up along with the animation system, ad hoc search functions and dialog windows etc. Some of these features are shown in Fig. 7. The avatars shown in Figure 12 are those currently used for tests and ongoing research activities; however, from a graphic point of view, these avatars will be later replaced with more realistic ones.



Figure 12 – INSIDE VIRTUAL Preliminary implementation

As far as the animations are concerned, scripts for movement have been developed using C#. Such scripts include interpolation functions that offer many possibilities in terms of directions and movement types for both the user and the guide avatar. Some of the main motion parameters are reported on Figure 13. Furthermore, some default paths have been identified and implemented: these are some of the paths that will be used when the game evolves according to the "guided tour" mode. Lastly, collisions between avatars and objects have been implemented.

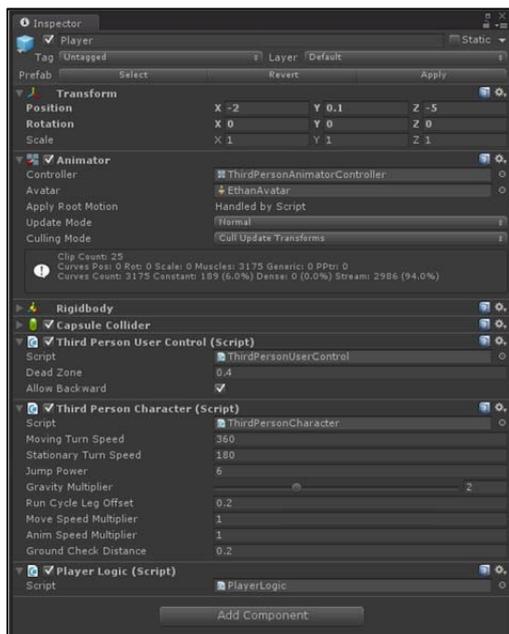


Figure 13 – Motion parameters for avatars

5. CONCLUSIONS

The proposed research introduces the design and implementation process of an advanced Serious Game and app, INSIDE VIRTUAL, for cultural heritage promotion and fruition. The game logic and functionalities have been carefully designed and as a result conceptual models have been drawn up. This phase has been documented through UML diagrams that have allowed expressing in a rather immediate and intuitive way all the game functional requirements and properties. Such diagrams have been a fundamental premise for evaluating and detecting suitable development frameworks. It applies to the intelligent logics that drive the game evolution, to the development environment selection and to the basic building blocks of the INSIDE VIRTUAL app.

In particular, a symbolic deliberative approach is adopted for implementing the intelligent logics that drive the guide behavior along the game while, considering the game functionalities and properties, UNITY3D has been chosen as authoring tool. Here, the virtual environment is currently under development as well as artifact reconstructions. Some functionality such as animations, collisions, search functions, dialog and information windows and the movement management system have been already shaped up.

The INSIDE VIRTUAL app is part of a two years research project financed by the Italian Ministry of Education, University and Research. As shown, INSIDE VIRTUAL app is provided with advanced functionalities that require a considerable effort both in terms of research, modeling and coding. However preliminary results look promising and the whole research team, that involves both industry and academia, is confident that the final project outcomes will be in line with the expectations.

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., Liarakapis, 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.
- Bederson B.B., 1995. Audio Augmented Reality: A Prototype Automated Tour Guide. *Proceedings of ACM Human Computer in Computing Systems conference (CHI'95)*, pp. 210-211.

- Brogni A., Avizzano C., Evangelista C., Bergamasco M., 1999. Technological Approach for Cultural Heritage: Augmented Reality. The IEEE 8th International Workshop, pp. 206-212
- Bruzzone, A.G., Frascio, M., Longo, F., Massei, M., Siri, A., Tremori, A. (2012) MARIA: An agent driven simulation for a web based serious game devoted to renew education processes in health care 1st International Workshop on Innovative Simulation for Health Care, IWISH 2012, Held at the International Multidisciplinary Modeling and Simulation Multiconference, I3M 2012, pp. 188-194.
- Bruzzone, A.G., Longo, F. (2013) An advanced modeling & simulation tool for investigating the behavior of a manufacturing system in the hazelnuts industry sector (2013) International Journal of Food Engineering, 9 (3), pp. 241-257.
- Bruzzone, A., Longo, F. (2014) An application methodology for logistics and transportation scenarios analysis and comparison within the retail supply chain (2014) European Journal of Industrial Engineering, 8 (1), pp. 112-142.
- Bruzzone, A., Massei, M., Longo, F., Poggi, S., Agresta, M., Bartolucci, C., Nicoletti, L. (2014) Human behavior simulation for complex scenarios based on intelligent agents. Simulation Series, 46 (2), pp. 71-80.
- Costantini S., Mostarda L., Tocchio A., Tsintza P., 2008. DALICA: Agent-Based Ambient Intelligence for Cultural-Heritage Scenarios. *IEEE Intelligent Systems*, 23 (2): pp. 34-41.
- Debevec P. 2005. Making “The Parthenon”. *6th international symposium on virtual reality, archaeology, and cultural heritage*.
- Frischer, B., 2008. *The Rome Reborn Project. How Technology is helping us to study history*, OpEd, November 10, 2008. University of Virginia.
- Froschauer, J., Seidel, I., Gartner, M., Berger, H., Merkl D., 2010. Design and evaluation of a serious game for immersive cultural training. In: *Proceedings of the 16th International Conference on Virtual Systems and Multimedia*, 2010, pp. 253–260.
- Hall T., Ciolfi L., Bannon L., Fraser M., Benford S., Bowers J., Greenhalgh C., Hellstrom S., Izadi S. and Schnadelbach H., 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.
- Hughes Ch., Smith E., Stapleton Ch. and D. Hughes, 2004. Augmenting Museum Experiences with Mixed Reality. In: *Proceedings of Knowledge Sharing and Collaborative Engineering*, 2004, St. Thomas, US Virgin Islands.
- Jennings N.R., 2000. On agent-based software engineering. *Artificial Intelligence* 117(2): pp. 277-296, Elsevier.
- Liarokapis F., Sylaiou S., Basu A., Mourkoussis N., White M. and P.F. Lister, 2004. An Interactive Visualisation Interface for Virtual Museums. In: *Proceedings of the 5th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, Brussels, pp. 47-56.
- Longo, F. (2012) Sustainable supply chain design: An application example in local business retail. *Simulation*, 88 (12), pp. 1484-1498.
- Longo, F., Nicoletti, L., Chiurco, A., Solis, A., Spadafora, F. (2013) Drivers and parkers training in car terminals 25th European Modeling and Simulation Symposium, EMSS 2013, pp. 704-712.
- Longo, F., Nicoletti, L., Vena, S., Padovano, A. (2014) Serious games at increased impact on culture and tourism. 26th European Modeling and Simulation Symposium, EMSS 2014, pp. 641-648.
- Mase K., Kadobayashi R., Nakatsu R., 1996. Meta-Museum: A Supportive Augmented-Reality Environment for Knowledge Sharing. In: *Proceedings of the International Conference on Virtual Systems and Multimedia '96* in Gifu, pp. 107-110.
- Michel F., Ferber J., Drogoul A., 2009. Multi-Agent Systems and Simulation: A Survey from the Agent Community's Perspective. In: *Multi-Agent Systems: Simulation and Applications*, Cap.1, CRC Press, 2009.
- Mortara M., Catalano C.E., Bellotti F., Fiucci G., Houry-Panchetti M., Petridis P., 2013. Learning Cultural Heritage by Serious Games. *Journal of Cultural Heritage*.
- Stock O., Zancanaro M., 2010. Personalized Active Cultural Heritage: The PEACH Experience. In: *Handbook of Research on Culturally-Aware Information Technology: Perspectives and Models*.
- Sylaiou S., Liarokapis F., Kotsakis K., Patias P. 2009. Virtual museums, a survey on methods and tools. *Journal of Cultural Heritage*, 10(4): pp. 520–528.
- White M., Mourkoussis N., Darcy J., Petridis P., Liarokapis F., Lister P.F., Walczak K., Wojciechowski R., Cellary W., Chmielewski J., Stawniak M., Wiza W., Patel M., Stevenson J., Manley, J., Giorgini F., Sayd P. and F. Gaspard, 2004. ARCO: An Architecture for Digitization, Management and Presentation of Virtual Exhibitions. In: *IEEE Proceedings 22nd International Conference on Computer Graphics*, Hersonissos, Crete, June 16-19, pp. 622-625.
- Woolridge M., (2002). *Introduction to Multiagent Systems*. John Wiley & Sons, Inc., NY, USA.

SYNTHETIC ENVIRONMENT FOR INTEROPERABLE ADVANCED MARINE SIMULATION

Agostino G. Bruzzone, Marina Massei
DIME University of Genoa
agostino@itim.unige.it, massei@itim.unige.it
URL: www.itim.unige.it

Alberto Tremori
NATO STO CMRE
alberto.tremori@cmre.nato.int
URL: www.cmre.nato.int

Riccardo Di Matteo, Giovanni Luca Maglione
Simulation Team
riccardo.dimatteo@simulationteam.com, gianluca.maglione@simulationteam.com
URL: www.simulationteam.com

ABSTRACT

This paper proposes a new virtual simulation environment designed as element of an interoperable federation of simulator to support the investigation of complex scenarios over the Extended Maritime Framework (EMF) including sea surface, underwater, air, coast, space and cyberspace.

The paper proposes different complex cases as examples to validate this approach related both to deep waters and to coast and littoral protection. These simulation environment involves different kind of traditional assets (e.g. satellites, helicopters, ships, submarines, underwater sensor infrastructures, etc.) interacting dynamically and collaborating with innovative autonomous systems (i.e. AUV, Gliders, USV and UAV). The proposed synthetic environment based on the use of virtual simulation supports validation of new concepts as well as the investigation of engineering and collaborative solutions. So the proposed approach enables the creation of dynamic interoperable immersive frameworks useful for Man-in-the-Loop training, education and tactical decision making by creating a support for the Man-on-the-Loop concept.

Keywords: *Interoperable Simulation, Maritime Simulation, Heterogeneous Networks, Autonomous Systems, Modeling & Simulation, Port Security, Man on the Loop*

1. INTRODUCTION

The research and development of the Autonomous Systems is continuously growing up because of their increasing performances and versatility thanks to recent progress of the technology. Today it is possible replace humans with such vehicles in many actions which makes them faster and more secure and their use is becoming popular in a large quantity of sectors from defense to civil protection, oil and gas, etc.



Figure 1 – EMF and corresponding Autonomous System Examples

In the external environments the autonomous systems deals with all the different domains including air (UAVm Unmanned Aerial Vehicles), sea surface (USV, Unmanned Surface Vehicles), land (UGV, Unmanned Ground Vehicles), underwater (AUV, Autonomous Underwater Vehicles.), etc. In particular the underwater environment emphasizes the importance of autonomy due to the heavy limitations in underwater communications affecting remote control and coordination capabilities over multiple domains (Tether 2009; DARPA 2012; Shkurti et al. 2012); among the others autonomous systems, the AUVs introduce additional challenges being a crucial element into the Extended Maritime Framework (EMF, sea surface, underwater, air, coast, space and cyberspace) (Shkurti et al. 2009) so it is evident the necessity to rely on Modeling and Simulation (M&S) to investigate the different possible solutions in such innovative sector and to test them integrated within real systems; indeed this context is pretty interesting for many strategic sectors such as Defense, Security, Port Activities and Oil & Gas (Bruzzone et al.2010; Bruzzone et al. 2011b; Bruzzone et al. 2014a). Obviously a similar approach could be easily extended to be applied to other sectors, therefore the complexity of the marine environment (especially underwater)

makes this solution pretty unique within this framework.

2. AUTONOMOUS SYSTEMS WITHIN EMF & JOINT SIMULATION ENVIRONMENT

In general, the autonomous systems development requires continuous test so a time-effective approach might result very useful. In this context, simulation fosters a better understanding of the behavior of Unmanned Vehicles and avoid useless experimentations and costs by allowing to identify and solve the problems in advance by virtual world experimentation (Stilwell et al. 2004).

Indeed the engineering processes gain from the simulation a strategic advantage for instance the creation of an interoperable synthetic environment for EMF allows to study the most effective and efficient behaviors for joint operations among multiple autonomous systems interacting with traditional assets; by this approach it becomes possible to optimize them and thus find critical points of the systems and to engineer their design. This paper presents the creation of a virtual environment with the aim to simulate and understand a Joint Naval Scenario over the EMF.

The scalability is a very crucial element considering that this EMF should be able to address from Hardware and Software in the loop issues up to operational procedures and policies.

This study especially focuses on the integration of Autonomous Systems with traditional assets; the proposed simulation deals with collaborative operation involving different types of Autonomous Underwater Vehicles (AUV), Unmanned Surface Vehicles (USV) and UAV (Unmanned Aerial Vehicle) with surface and underwater vessels as well as with sensor networks.

The authors designed an interoperable virtual simulation devoted to integrate different systems within one environment able to simulate and represent the overall situation. The simulators are combining discrete event stochastic simulator with continuous simulators (Bruzzone et al. 2007; Longo F. 2012, Longo et al., 2013); intelligent agents are used to create a distributed control of the autonomous systems (Bruzzone et al. 2011b; Ören et al. 2009; Feddema et al. 2002). The proposed system, designed MAMA (Multiple Advanced Maritime Architecture) is able to guarantee the simulation of the different systems as well as their supervision, but considering the different capabilities of sensors, communications systems and platforms (Bruzzone et al. 2011d); this simulation enables to address complex Measure of Merits including mission effectiveness as dynamic results dependent of the different variables and behaviors affecting the complex heterogeneous network of available assets.

The researchers aim to define also guidelines and standards to develop this flexible virtual simulation solution; in this case it was decided to create an HLA federation as crucial element to address the complexity of the EMF and to integrate them with IA-CGF (Intelligent Agent Computer Generated Forces)

developed by Simulation Team (Kuhl et al.1999; Bruzzone & Massei 2007; Bruzzone 2008; Bruzzone et al. 2011e). In general use of Intelligent Agents enable the possibility to create complex behavior to investigate best policies and criteria for finalizing decision related to effective and reliable Course of Actions (Bruzzone et al. 2011a).

The general architecture is proposed by figure 2; indeed the author decided to use HLA as already successful done in flexible marine simulation environments such as ST_VP (Bruzzone et al. 2011c); indeed HLA was also successfully used in marine ports applications for training purposes (Longo et al. 2013). In this case it was decided to implement multiple RTIs including Pitch, Mäk and Portico for being available for a wide spectrum of integration possibilities; the federation was extensively tested by integrating different federates such as virtual simulator JEANS (Join Environment for Advanced Naval Simulation, using Unity 3D Graphic Engine), constructive simulation MCWS (Marine Cyber Warfare Simulation), physical model MSTPA (Multi-static Tactical Planning Aid acoustic engine), SPARUS-ROS (AUV Real Control System based on Robot Operating System) and open to other models (Bruzzone et al. 2013b). Different kind of FOM was investigated and tested this architecture respect its capabilities including MCWS-MSTPA, RPR (Real-time Platform-level Reference for integrating DIS legacy systems) and preliminary STANAG 4684 (Virtual Ship).

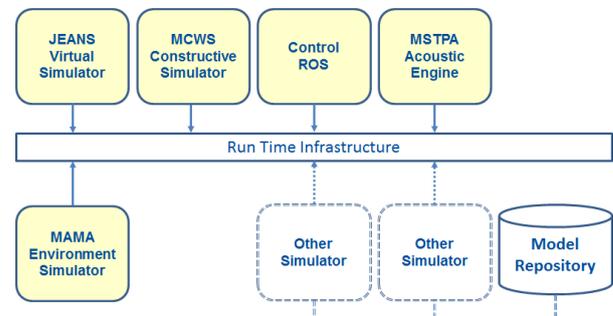


Figure 2 - Architecture Interoperable Simulation for EMF

Indeed this new generation of marine interoperable simulation represents an important strategic advantage being able to introduce new assets within the scenario as well as models characterized by different level of fidelity based on the simulation objectives; this allow to use it as a flexible approach for experimentation and investigating over a wide spectrum of problems related to the operational use of autonomous systems and to find new ways to use them respect different scenarios.

The paper proposes different scenarios: one related to military operations and another one on coastal and littoral protection where the virtual simulation proposes the overall situation and allows to experiment the different solutions into the virtual world considering the complex physics affecting movement, perception, interactions and communications (Bruzzone et al. 2012; Bruzzone 2013a).

By this approach it becomes evident the capability to identify, by experimental analysis within the virtual world, the new solutions in terms of engineering and technological configuration of the different systems and vehicles as well as new operational models and tactics to address the specific mission environment.

The proposed case study is about a maritime scenario with a representation of heterogeneous network frameworks that involve multiple vehicles both naval and aerial including AUVs, USVs, UAVs, gliders, helicopters, ships, submarines, satellites, buoys and sensors (Bruzzone et al. 2013c).

The simulator is developed according to High Level Architecture (HLA) Standard (IEEE 1516 evolved) to ensure interoperability and allow its extensibility for other case studies (Bruzzone et. al.2007c; Joshi et al.2006).

In some way this approach adopt the MS2G (Modeling, interoperable Simulation and Serious Game) paradigm combining M&S and Serious Game (SG) by framing integrating by HLA also a virtual world easily deployable on multiple distributed solutions (Bruzzone et al. 2014a). Involvement of Subject Matter Experts (SME) as users during the development process is fundamental to take advantage of their Know-How along the validation process in such sense the synergy with CMRE resulted very useful.

The creation of a versatile simulation engine is thus effective to achieve the highest knowledge collecting capability to allow expert to face the widest range of possible situations.

In MAMA areas are defined where weather condition are defined including sea current, wind, waves, surface temperature, fog, rain, salinity and thermal layers, etc (Lundquist 2013). For instance the sea current is introduced as additional velocity component for all objects operating within the water, under the assumption that it provokes a body shifting as if bodies are moving respect to a steady drifting inertial coordinate system.

MAMA framework support integration with real systems this allows to federate also hardware and software in the loop as well as man in the loop controls.

3. MODELS OF PLATFORMS & SYSTEMS

To simulate the different assets operating within EMF it is necessary to adopt real-time models able to deal with a large number and variety of platforms. Several alternative models has been tested, therefore, as expected, the main issue in this phase is to properly scale fidelity with computational workload. The guideline is obviously defined by the specific requirements and objectives of the simulation; indeed the authors by MAMA allowed to use different models with different resolution levels and fidelity; for instance the acoustic models for the passive sonar have been developed ad hoc for this project, while active acoustic could be based on simplified meta-models or by federating a specific acoustic engine.

In the deep sea scenario models of AUV and Gliders

have been developed including physics, control, communications and sensors; the different models of each new asset or system is defined within a repository that allow to use it for different scenarios based on the specific needs (Maravall et al. 2013).

Indeed the simulation instantiates all the n-copies of library objects and assigns their control to the IA (Intelligent Agent) related to the specific system or platform. In most case it was decided to use simplified meta-models (e.g. sea keeping, underwater communications) that consider just main elements in order to guarantee the possibility to conduct experimentations also with an high number of objects to be simulated.

For instance it was implemented a model for the gliders that are autonomous underwater vehicles without propeller with long autonomy able to operate and move just by changing buoyancy in order to glide underwater up and down (Bhatta et al. 2005); for these platform a special pitch control was activated allowing them to sail underwater by triggered based on vehicle depth and sea bed configuration; in this way the vehicle glides avoiding impact on the bottom and move towards. Those depths are computed as a function of the advance speed and the trim angle of the vehicle considering a circular trajectory (conservative) and then adjusted through safety factors.

The glider stops pitching (angles positive clockwise) when the limit on the trim angle is reached. Second derivative of trim angle over the time (angular acceleration) is given as a descriptive parameter of the system to be modeled.

As a matter of example is reported the portion of the code responsible for the computation of the safe distance where to pitch ‘rkm3’:

$$R_{cr} = S / As_2$$

$$Hs = k_1 R_{cr} (1 - \cos(|a_2|)) + k_2;$$

R_{cr}	curvature radius,
S	advancing speed of the vehicle along his longitudinal axis
As_2	Angular velocity respect pitch
a_2	Pitch
k_1, k_2	safety factors (e.g. $k_1 = 1.2, k_2 = 1$)
Hs	safe distance to change the pitch for gliding

Indeed the advance speed that is one of the governing parameter for the current control; the model considers that this speed is influenced by the oscillation on lift direction, resulting in reducing velocity when the glider surge direction is close to be horizontal.

Indeed the control allows for heading on targets and waypoints

Course angles in the horizontal plane (angles positive clockwise) are:

a_1	the course as angle between North and platform advance
d_{rd}	direction between north and the vector pointing at the target/waypoint
d_{rd_0}	the difference of the two and indicates the yaw

angle to be covered to head on the target.

If drd_0 module is higher than 180 degrees than his complementary is chosen, meaning the vehicle yaws on the opposite direction.

Once determined drd_0 as the most convenient angle to head on the target, the vehicle experiences an angular acceleration representing the lift force generated by the rudder blade.

To allow the vehicle to keep the course heading on the target, the angle A_{IRB} when to invert the rudder blade (when an opposite angular acceleration start to be experienced on the center of gravity) is computed as

$$A_{\text{IRB}} = k_3 |Sa_1^2 / Aa_1|$$

Sa_1	angular yaw velocity
Aa_1	acceleration on yaw
k_3	safety factor (e.g. $k_3=0.9$)

Target is considered reached when the vehicle gets to a certain distance in the horizontal plane; normally to speed up simulation this distance is proportional to platform speed based on a parameter specific for the platform that affects the precision of the targeting; the gliders and AUV could be assigned to path based on waypoint or on patrolling an area autonomously based on different scheme such as creeping line and random multi-agent search.

A different and integrated control is defined for course and pitch respect the presence of other vessels in order to avoid collisions.

Indeed obstacle avoidance is currently under development to combine the vertical management of this issues with course adjustment when the longitudinal projection of body cross section intercept a physical object and there is not enough depth into the area to glide down.

These controls are integrated within an intelligent agent that drives each of the glides within the scenario and take care of activating SATCOM when surfacing for data transfer.

The IA uses collisions in order to generate HLA interactions based on box colliders able manage the phenomena.

4. MS2G & MAN-ON-THE-LOOP

MS2G is a powerful paradigm able to guarantee the engagement of SMEs in many different areas; in MAMA case this allow to get a comprehensive representation of the EMF within the virtual reality; obviously this support the man-on-the-loop concept (Magrassi 2013).

This means that decision makers could observe in the virtual environment a 3D representation of assets and their capabilities (e.g. discovering spheres, autonomy ranges, communication ranges) and assign high level task to the IA controlling each autonomous system without need to remotely control in details their activity directly or through waypoints (Cooke et al. 2006); this leads towards introduction of orders and collaborative behavior through combined task

assignment to autonomous vehicles operating within an area (Bruzzone et al. 2013b; Ferrandez et al. 2013; Kalra et al. 2007; Shafer et al. 2007a; Vail et al. 2003); for instance it could become possible to assign a USV to recharging and fast data collection over a set of AUV within an area or a set of AUV to create a dynamic underwater dynamic cascade communication network over a zone; indeed in the future collaborative tasks are expected to become very important (Richards et al.2002; Ross et al. 2006; Tanner et al. 2007; Shafer et al. 2007b).

Obviously this approach supports improvements in distributed exercising providing a more clear understanding of the events and situation; in this case the users could be able to figure visually the different COAs (Courses of Actions) in the Extended Maritime Framework, while the interoperable stochastic simulator evaluates consequences of the different alternatives and estimates risks.

The expertise accumulated via MS2G constitutes a base knowledge for future projects.

The communications among the entities are defined through a dynamic evolving set of nodes and links connecting real and virtual assets; each of these links is defined in terms of type (e.g. Radio frequency, Satcom, Acoustic Modem, optic fiber, etc); capability, bandwidth, basic model, background standard traffic model, reliability, confidentiality, availability, integrity, mutual interference; the messages are managed as pockets moving along the heterogeneous networks and even visualized in terms of status and connection over the virtual representation of the EMF; each assets could define a connection point to visualize the communication by some kind of augmented reality; it is important to outline that underwater communications are usually strongly constrained in terms of range and speed going often down to 100 bauds. Based on the communication model adopted the availability of the communication is computed even considering the capability to use other nodes as bridges; in similar way a standard traffic model is used to considering the bandwidth utilization where to add the specific communication packages simulated, allowing to use simplify meta-models and networks; therefore MAMA structure allows to integrate specific communication models in case the need for an high fidelity simulation of detailed large communication architectures; in the past for instance the authors have developed federation integrating Opnet based simulators to deal with these aspects.

5. GRAPHICS

The authors evaluated different solutions for the graphic engine including among the others Vega Prime, VBS2 and 3, Unity 3D; also in this case the use of MAMA architecture keeps open the possibility to use different visualizers and currently the author are working on a distributed application integrating both Unity 3D and Vega Prima for AUV docking simulation using legacy simulators (Bruzzone et al. 2011e; Zini

2012).

However all the cases proposed here are using Unity3D as graphic engine and Suimono as graphic tool for water effect; indeed Suimono tool does not consider influence of external forces on ship hull, but hydrostatic pressure distribution given directly as buoyancy force as an input parameter. Therefore the authors are using their physical models to deal with these aspects and in particular to develop the dynamic description of buoyant or submerged body; indeed in this case is necessary to create an own model taking into account hydrostatics and hydrodynamic forces addressing at hull stability, motions of body in still water, sea keeping, maneuverability, ship propulsion and all what Naval Architecture concerns.

The 3D representation of the scenario include visualization of assets and communications as well as capabilities; these are supported by a free camera management system implemented within the Graphic User Interface (GUI); this entitles the users to fly around and observe the whole EMF scenario and to move freely in the synthetic environment in order to investigate different phenomena by best perspective of the scene.

Moreover, each object has its own inertial camera; so the players are enabled to adopt the visual from that specific object and to observe what is going on the scene from that point of view. This should allow players to have a better involvement in the simulation and to understand the COA by a direct visual perception of what is happening in every moment.

In facts the simulator is supplied with features to improve perception of the scenario and to see beyond the limit of real awareness, providing visual representation for the limitation of aerial and underwater communications, underwater sight, etc.; in addition the whole water can be toggled removing it to see the whole assets over and under the surface, additional markers are used to identify the platforms and understand their operational status.

This approach was extensively used with SME for supporting VV&A (Verification, Validation and Accreditation) of the models and of their interactions; in addition the proposed graphic supports the training and after action review for future applications (Massei & Tremori 2010; Kennedy 2010; Kracke et al. 2006)

6. SCENARIOS

Different scenarios have been simulated to validate MAMA approach. Deep water scenario includes difference cases and aims to simulate interactions between traditional assets in EMF (i.e. satellites, navy ships, submarines, NATO Research Vessels (NRVs) helicopters) with new generation unmanned assets (i.e. AUV, Gliders, UAV, USV) and the mutual advantage the subjects involved in the scenario can have. In other word, the increase in persistence, interoperability and efficacy.

The littoral scenario is mostly devoted to model the behavior of unmanned assets and traditional asset involving AUV, USV and vessels to control the coast

and to patrol a harbor to find possible threats through a persistent surveillance. These cases aims to develop and test new solutions for autonomous vehicle release and docking as well as for patrolling; for instance it was studied an algorithm to lead patrolling toward an optimum, guaranteeing an high probability of success in the safest way reducing human involvement in the scenario for marine C-IED missions (Counter Improvised Explosive Device).

In all scenarios the authors are experimenting the different engineering solutions and integration technologies in order to guarantee interoperability among the different systems; in particular experimentation are ongoing by testing them within HIL and SIL simulations (Hardware and Software in the loop).

Deep Water Scenario - Glider Fleet Management Case

In this case a Destroyer and a oceanographic ship are sailing while a fleet with up to 200 underwater gliders are moving autonomously underwater to collect salinity and temperature data and using passive acoustic sensors; over them a Predator and a Global Hawk are available as sensor platform and communication nodes as well as a Satellite; on the sea bed in this scenario a sensor infrastructure is collecting passive data while sonobuoy are active for multi-static operations.

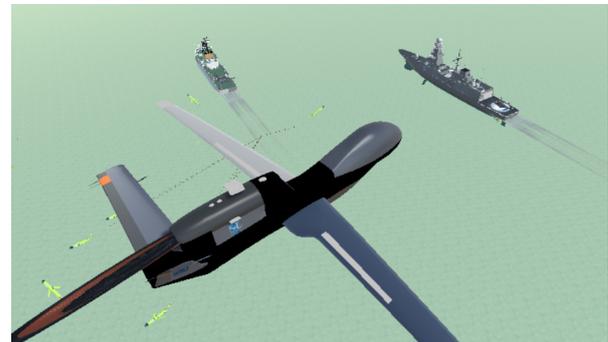


Figure 3 - Deep Water Scenario - Glider Fleet Management Case

Deep Water Scenario – Task Force Defense

This scenario involves a surface task force including 2 frigates, 2 destroyers, a medium size aircraft carrier; the task force and its aerial resources is supported also by a submarine, some AUV and an air patrol provided by UAV; the task force has to face up to over 30 supersonic missile threats launched behind the horizon by an hostile attack submarine that is receiving satellite information on task force; also in this case helicopters, sonobuoy and underwater sensor networks could be activated; the scenario allows to investigate different cooperative engagement policies as well as reliability of the processes (Calfee & Rowe 2004)

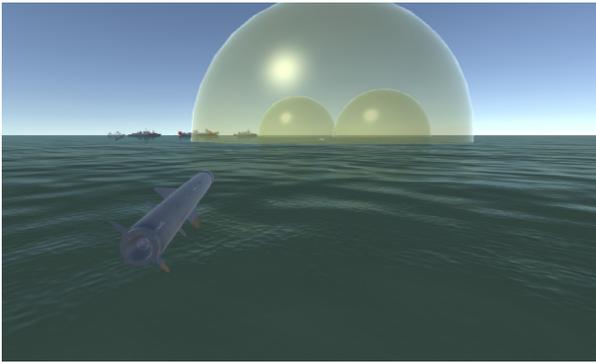


Figure 4 - Deep Water Scenario – Task Force Defense

Littoral Scenario – Patrolling, Deploying and Recovery

This scenario includes an oil platform, a frigate and a set of AUV; the AUV patrol the area by different logic and could dock or board the frigate autonomously avoiding collision and properly board by approaching the vessel' stern gate usually used for its larger RHIB (Rigid Hull Inflatable Boat).



Figure 5 - Littoral Scenario – Patrolling, Deploying and Recovery

Littoral Scenario – Port Protection

In this case a USV is acting as gateway for a UAV patrolling the docks and ships into a port for marine C-IED; a UGV on the ground is used to patrol the piers from land and they are connected through an UAV flying over the area and guaranteeing a bridge with ships sailing over the horizon and for inland targeting (Martins et al. 2011; Michael 2007; Dogan and Zengin 2006; Grocholsky et al. 2006).

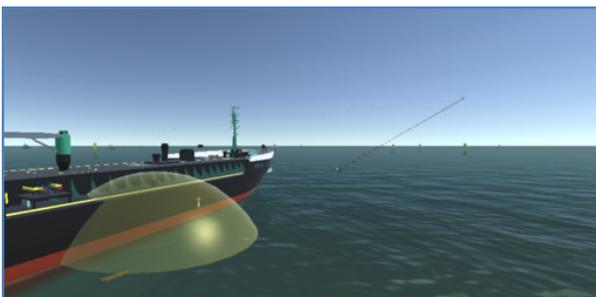


Figure 6 - Littoral Scenario – Port Protection

7. CONCLUSIONS

The proposed environment was completed and experimented confirming the validity of the proposed approach; the scenario developed include pretty advanced scenarios interoperating over a wide spectrum of simulators and systems as confirmation of the flexibility achieved. The key of success was based on availability of previous researches used for creating the new MAMA environment and the approach based on multi resolution meta-models to be integrated based on specific simulation need.

Currently the authors are proceedings in further developing the proposed models as well as in integrating new systems and new aspects.

REFERENCES

- Bhatta P., E. Fiorelli, F. Lekien, N. E. Leonard, D. A. Paley, F. Zhang, R. Bachmayer, R. E. Davis, D. Fratantoni and R. Sepulchre (2005) "Coordination of an underwater glider feet for adaptive sampling, International Workshop on Underwater Robotics", Genova, Italy, pp 61-69
- Bruzzone A.G., Massei M. (2007) "Polyfunctional Intelligent Operational Virtual Reality Agent: PIOVRA Final Report", EDA Technical Report
- Bruzzone A.G. (2008) "Intelligent Agents for Computer Generated Forces", Invited Speech at Gesi User Workshop, Wien, Italy, October 16-17
- Bruzzone A.G., Cantice G., Morabito G., Mursia A., Sebastiani M., Tremori A. (2009) "CGF for NATO NEC C2 Maturity Model (N2C2M2) Evaluation", Proceedings of I/ITSEC2009, Orlando, November 30-December 4
- Bruzzone A.G., Tremori A., Bocca E., (2010) "Security & Safety Training and Assessment in Ports based on Interoperable Simulation", Proceedings of HMS2010, Fes, Morocco, October 13-15
- Bruzzone A.G., Massei M., Tarone F., Madeo F. (2011a) "Integrating Intelligent Agents & AHP in a Complex System Simulation", Proceedings of the international Symposium on the AHP, Sorrento, Italy, June.
- Bruzzone A.G., Tremori A., Merkurjev Y. (2011b) "Asymmetric marine warfare: PANOPEA a piracy simulator for investigating new C2 solutions" Proceedings SCM MEMTS 2011, St.Petersburg June 29-30
- Bruzzone A.G., Fadda P, Fancello G., Massei M., Bocca E., Tremori A., Tarone F., D'Errico G. (2011c) "Logistics node simulator as an enabler for supply chain development: innovative portainer simulator as the assessment tool for human factors in port cranes", Simulation October 2011, vol. 87 no. 10, p. 857-874, ISSN: 857-874, doi: 10.1177/0037549711418688.
- Bruzzone A.G., Massei M. Tremori A., Longo F., Madeo F., Tarone F, (2011d) "Maritime Security: Emerging Technologies for Asymmetric Threats", Proceedings of EMSS, Rome, Italy, September 12 -14
- Bruzzone, A.G., Tremori, A., Massei, M., (2011e),

- "Adding Smart to the Mix," Modeling, Simulation & Training: the International Defence Training Journal, 3, 25-27.
- Bruzzone A., Tremori A., Longo F., (2012) "Interoperable Simulation for Protecting Port as Critical Infrastructures", Proc. of HMS2012, Wien, September 19-21
- Bruzzone A.G. (2013a) "New Challenges for Modelling & Simulation in Maritime Domain", Keynote Speech at SpringSim2013, San Diego, CA, April
- Bruzzone A.G., Berni A., Cignoni A., Massei M., Tremori A., Dallorto M., Ferrando A. (2013b) "Virtual Framework for Testing/Experiencing Potential of Collaborative Autonomous Systems", Proceedings of IITSEC2013, Orlando, FL, November
- Bruzzone A.G., Merani D., Massei M., Tremori A., Bartolucci C., Ferrando A. (2013c) "Modeling Cyber Warfare in Heterogeneous Networks for Protection of Infrastructures and Operations" Proceedings of EMSS2013, Athens, Greece, September
- Bruzzone, A., Massei, M., Longo, F., Tremori, A., Nicoletti, L., Poggi, S., Bartolucci, C., Picco, E., Poggio, G. (2014a) "MS2G: Simulation as a Service for Data Mining and Crowd Sourcing in Vulnerability Reduction", Proceedings of WAMS2014, Istanbul, Turkey, September
- Bruzzone A.G., Frascio M., Longo F., Chiurco F., Zaroni S., Zavarella L., Fadda P., Fancello G., Falcone D., De Felice F., Petrillo A., Carotenuto P. (2014b) "Disaster and Emergency Management Simulation in Industrial Plants", Proceedings of EMSS2014, Bordeaux, France, September
- Calfee, S.H., Rowe, N.C., 2004. "Multi-agent simulation of human behavior in naval air defense," Naval Engineers Journal, 116, no.4, 53-64.
- Cooke N.J., Pringle H.L., Pedersen H.K., Connor O. (2006) "Human Factors of Remotely Operated Vehicles", JAI Press, NY
- DARPA (2012) "Making Connections At 45,000 Feet: Future UAVs May Fuel Up In Flight", DARPA Information, October 05
- Dogan A., U. Zengin, (2006) "Unmanned aerial vehicle dynamic-target pursuit by using a probabilistic threat exposure map", AIAA Journal of Guidance, Control and Dynamics 29 (4), pp. 944-954.
- Feddema, J.T.; Lewis, C.; Schoenwald, D.A., (2002) "Decentralized control of cooperative robotic vehicles: theory and application, "Robotics and Automation, IEEE Transactions on, vol.18, no.5, pp.852,864, Oct
- Ferrandez J.M., De Lope H., De la Paz, F. (2013) "Social and Collaborative Robotics", International Journal Robotics and Autonomous Systems, 61, 659-660
- Grocholsky, B.; Keller, J.; Kumar, V.; Pappas, G., (2006) "Cooperative air and ground surveillance," Robotics & Automation Magazine, IEEE, vol.13, no.3, pp.16,25, Sept.
- Joshi R., Castellote G.P., (2006) "A Comparison and Mapping of Data Distribution Service and High-Level Architecture", Real Time Innovations, 1-5.
- Kalra N., D. Ferguson and A. Stentz (2007), "A generalized framework for solving tightly-coupled multirobot planning problems, Proc. of the IEEE International Conference on Robotics and Automation, April, pp.3359-3364.
- Kennedy, K.P., (2010). "Training: the key to keeping your head in a crisis situation," Naval Engineers Journal, 122, no.3, 73-85.
- Kracke, R., Hauge, J.B., Duin, H., (2006) "Training of strategic decisions in collaborative networks through serious games," Network-Centric Collaboration and Supporting Frameworks Book Series, International Federation for Information Processing, 224, 305-312.
- Kuhl, F., Weatherly, R., Dahmann, J., (1999) "Creating Computer Simulation Systems: An Introduction to the High Level Architecture", Prentice Hall, Upper Saddle River, USA.
- Lundquist, E.H (2013) "Navy Currents", Sea Technology, February, Vol 54, Issue 2, p77
- Magrassi C. (2013) "Education and Training: Delivering Cost Effective Readiness for Tomorrow's Operations", Keynote Speech at ITEC2013, Rome, May 22-24
- Longo F., (2012). Supply chain security: An integrated framework for container terminal facilities. International Journal of Simulation and Process Modeling, Volume 7, Issue 3, September 2012, Pages 159-167.
- 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.
- Longo F., Huerta A., Nicoletti L., (2013). Performance analysis of a Southern Mediterranean seaport via discrete-event simulation. Strojnicki Vestnik/Journal of Mechanical Engineering, Volume 59, Issue 9, pp. 517-525
- Maravall D., J. de Lopea,b, R. Dominguez, (2013) "Coordination of communication in robot teams by reinforcement learning", Robotics and Autonomous Systems 61, pp.661-666
- Martins, R.; de Sousa, J.B.; Afonso, C.C.; Incze, M.L., (2011) "REP10 AUV: Shallow water operations with heterogeneous autonomous vehicles," Oceans, IEEE - Spain, vol., no., pp.1,6, 6-9 June
- Massei, M., Tremori, A., (2010) "Mobile training solutions based on ST_VP: an HLA virtual simulation for training and virtual prototyping within ports." In Proceedings of the International Workshop on Applied Modeling and Simulation, WAMS, Buzios, Brazil, May 5-7
- Michael N., J. Fink and V. Kumar (2007) "Controlling a team of ground robots via an aerial robot", Proc. of the IEEE International Conference on Intelligent Robots and Systems, San Diego, CA, Oct 2007.

- Ören, T.I. and L. Yilmaz (2009) "On the Synergy of Simulation and Agents: An Innovation Paradigm Perspective", Special Issue on Agent-Directed Simulation. The International Journal of Intelligent Control and Systems (IJICS), Vol. 14, Nb. 1, March, 4-19
- Richards A., J. Bellingham, M. Tillerson, and J. P. (2002) "How: Co-ordination and control of multiple UAVs", Proc. of the AIAA Guidance, Navigation, and Control Conference, Monterey, CA, Aug., AIAA-2002-4588.
- Ross, S., D. Jacques, M. Pachter, and J. Raquet, (2006) "A Close Formation Flight Test for Automated Air Refueling," Proceedings of ION GNSS-2006, Fort Worth, TX, Sep
- Shafer, A.J.; Benjamin, M.R.; Leonard, J.J.; Curcio, J., (2008) "Autonomous cooperation of heterogeneous platforms for sea-based search tasks," Oceans, vol., no., pp.1,10, 15-18 Sept.
- Stilwell D. J., A. S. Gadre, C. A. Sylvester and C. J. Cannell (2004) "Design elements of a small low-cost autonomous underwater vehicle for field experiments in multi-vehicle coordination", Proc. of the IEEE/OES Autonomous Underwater Vehicles, June, pp. 1-6.
- Shkurti, F.; Anqi Xu; Meghjani, M.; Gamboa Higuera, J.C.; Girdhar, Y.; Giguere, P.; Dey, B.B.; Li, J.; Kalmbach, A.; Prahacs, C.; Turgeon, K.; Rekleitis, I.; Dudek, G., (2012)"Multi-domain monitoring of marine environments using a heterogeneous robot team", Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on, vol., no., pp.1747,1753, 7-12 Oct.
- Sujit, P. B.; Sousa, J.; Pereira, F.L., (2009) "UAV and AUVs coordination for ocean exploration", Oceans - EUROPE, vol., no., pp.1,7, 11-14 May
- Tanner H. G. (2007a) "Switched UAV-UGV cooperation scheme for target detection", IEEE International Conference on Robotics and Automation, Roma, Italy, April, pp. 3457-3462.
- Tanner H.G., D.K. Christodoulakis, (2007b) "Decentralized cooperative control of heterogeneous vehicle groups", Robotics and Autonomous Systems 55,pp 811–823
- Tether, T. (2009) "Darpa Strategic Plan", Technical Report DARPA, May
- Vail D. and M. Veloso, (2003) "Dynamic multi-robot coordination", Multi-Robot Systems: From Swarms to Intelligent Automata, Vol II, pp. 87-100.
- Zini A. (2012) "Virtual Ship: Dream or Reality", Keynote Speech at Summersim 2012, Genoa, July

Author's Index

AbouRizk	368	382	
Achkar	395		
Adel	46		
Adragna	228		
Affenzeller	222	289	307
Ahn	368	382	
Aizstrauta	36		
Aizstrauts	36		
Al-Hussein	234		
Alonso Lopez-Rojas	266		
Amini Khafri	368		
Atanasijević-Kunc	240	314	
Basán	166		
Bauer	430		
Beham	289	301	307
Bellamine Ben Saoud	103		
Bešter	314		
Bicher	430		
Bogner	159	207	
Bouanan	416		
Bouferguene	234		
Breitenecker	197	430	
Brožek	282	295	
Bruno	471		
Bruzzo	250	479	
Burešová	324		
Burlacu	222		
Capel Tuñón	213		
Caputi	471		
Cascaval	228		
Cayirci	445		
Champliaud	79		
Chen	403		
Chettupuzha	382		
Clavería	374		
Cóccola	166		
Corn	240		
D'Arienzo	228		
de Bock	172		
De Felice	250		
Di Bona	451		
di Matteo	479		
Di Pasquale	93		
Diviš	52		
Domínguez Mirón	213		
Duraccio	451		
Dvořáková	324		

Edésio Jungles	334		
El Boukili	389		
Elduque A.	359		
Elduque D.	359	374	
Falcone	451		
Fechter	301	307	
Feng	79		
Fernández A.	359		
Fernández J.B.	436		
Fikejz	191		
Flores	126	137	
Florio	471		
Forcina	451		
Forestier	416		
Frydman	340		
García	13		
García-Alcaraz	436	465	
Gebeshuber	222		
Ginters	36		
Glock	197		
Gómez	374		
Hague	382		
Hamri	340		
Hamza	46		
Hao	151		
Haspl	159		
Helo	151		
Hernández	126		
Holecek	222		
Iannone	93		
Jakeš	282	295	
Javierre	359	374	
Ji	368		
Jiménez Macías	349	436	465
Juuso	457		
Kaba Traoré	403		
Kaczorek	1		
Karapinar	445		
Karder	289		
Kartak	70		
Kavička	52	113	120
Kocyan	84		
Kommenda	222		
Körner	430		
Labba	103		
Latorre-Biel	349	465	
Lei	58		
León Samaniego	349		

Li F.	147	
Li X.	58	
Li W.	7	63
Liu Z.	7	
Liu H.	234	
Longo	250	471
Ma	20	
Maglione	479	
Maldonado-Macías	436	
Manzo	228	
Martinovič	84	
Massei	479	
Matschulat Ely	334	
McKenzie	410	
Méndez	166	395
Meng	63	
Merta	191	
Miranda	93	
Mujica Mota	126	172
Nicoletti	471	
Novotny	120	
Okolnishnikov	425	
Orlov	258	
Ozcakir	445	
Papić	314	
Pérez de la Parte	349	465
Petrillo	250	
Picech	395	
Pina	359	374
Pineda Torres	213	
Ponnambalam	13	
Popper	197	
Qian	63	
Qing	7	
Raggl	301	
Raska	181	
Řezanina	113	
Ribault	416	
Riemma	93	
Rossainz López	213	
Rudometov	425	
Sadiq Altaf	234	
Sáenz Díez Muro	465	
Santana Souza	334	
Sarasa	374	
Scala	172	
Scheibenpflug	289	
Segura Pérez	137	

Seo	340	
Shabrov	258	
Silvestri	451	
Slaninová	84	
Soares Freire	334	
Song	147	
Steps	207	
Stylios	382	
Szturcová	84	
Toshev	151	
Tozzi	272	
Tremori	479	
Ulrych	181	
Valdenazzi	272	
Vallespir	416	
Vetrano	471	
Vindel	137	
Voracek	32	
Wagner	289	307
Wang C.	147	
Wang H.	63	
Wang Y.	403	
Wiesinger	159	207
Winkler	430	
Xu	7	
Yacoub	340	
Yang	7	20
Youcef	46	
Zacharewicz	403	416
Zeigler	340	
Zhang Limao	382	
Zhang Lin	147	
Zhang Y.	234	
Zhou	20	
Zhu N.	58	
Zhu Y.	58	
Zhuravlev	425	
Zini	272	
Zupančič	25	314