

THE 26TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 10-12 2014

BORDEAUX, FRANCE



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

PRINTED IN RENDE (CS), ITALY, SEPTEMBER 2014

ISBN 978-88-97999-32-4 (Paperback)
ISBN 978-88-97999-38-6 (PDF)

© 2014 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 GENOVA. 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 GENOVA, PROF. AGOSTINO BRUZZONE, VIA OPERA PIA 15, 16145 GENOVA, ITALY. ADDITIONAL COPIES OF THE PROCEEDINGS OF THE EMSS ARE AVAILABLE FROM DIME UNIVERSITY OF GENOVA, PROF. AGOSTINO BRUZZONE, VIA OPERA PIA 15, 16145 GENOVA, ITALY.

ISBN 978-88-97999-32-4 (Paperback)
ISBN 978-88-97999-38-6 (PDF)

THE 26TH EUROPEAN MODELING & SIMULATION SYMPOSIUM

SEPTEMBER 10-12 2014, Bordeaux, France

ORGANIZED BY



DIME - UNIVERSITY OF GENOA



LIOPHANT SIMULATION



SIMULATION TEAM



IMCS - INTERNATIONAL MEDITERRANEAN & LATIN AMERICAN COUNCIL OF SIMULATION



DIMEG, UNIVERSITY OF CALABRIA



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



MODELING AND SIMULATION CENTER OF EXCELLENCE (MSCOE)



LATVIAN SIMULATION CENTER - RIGA TECHNICAL UNIVERSITY



LOGISIM



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



MIMOS - MOVIMENTO ITALIANO MODELLAZIONE E SIMULAZIONE



MITIM PERUGIA CENTER - UNIVERSITY OF PERUGIA



BRASILIAN SIMULATION CENTER, LAMCE-COPPE-UFRJ



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



M&SNET - MCLEOD MODELING AND SIMULATION NETWORK



LATVIAN SIMULATION SOCIETY



ECOLE SUPERIEURE D'INGENIERIE EN SCIENCES APPLIQUEES



UNIVERSITY OF LA LAGUNA



CIFASIS: CONICET-UNR-UPCAM



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



NATIONAL RUSSIAN SIMULATION SOCIETY



CEA - IFAC

LOCAL SPONSORS



UNIVERSITY OF BORDEAUX



IMS LAB



AQUITAINE REGION

I3M 2014 INDUSTRIAL SPONSORS



CAL-TEK SRL



LIOTECH LTD



MAST SRL



SIM-4-FUTURE

I3M 2014 MEDIA PARTNERS



INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SIMULATION AND PROCESS MODELING

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF CRITICAL INFRASTRUCTURES

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF ENGINEERING SYSTEMS MODELLING AND SIMULATION

INDERSCIENCE PUBLISHERS – INTERNATIONAL JOURNAL OF SERVICE AND COMPUTING ORIENTED MANUFACTURING



IGI GLOBAL – INTERNATIONAL JOURNAL OF PRIVACY AND HEALTH INFORMATION MANAGEMENT



HalldaleGroup



HALLDALE MEDIA GROUP: MILITARY SIMULATION AND TRAINING MAGAZINE

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



EUROMERCI

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 2014

GENERAL CO-CHAIRS

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

PROGRAM CHAIR

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

THE 26TH EUROPEAN MODELING & SIMULATION SYMPOSIUM, EMSS 2014

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 2014 INTERNATIONAL PROGRAM COMMITTEE

MICHAEL AFFENZELLER, *UPPER AUSTRIAN UNIV. OF AS, AUSTRIA*
JONI A. AMORIM, *UNIVERSITY OF SKOVDE, SWEDEN*
MAJA ATANASIJEVIC-KUNC, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
FELIX BREITENECKER, *TECHNICAL UNIVERSITY OF WIEN, AUSTRIA*
AGOSTINO BRUZZONE, *UNIVERSITY OF GENOA, ITALY*
HIPOLITO CARVAJAL, *UNIVERSITY OF ORIENTE, CUBA*
ALESSANDRO CHIURCO, *UNIVERSITY OF CALABRIA, ITALY*
JERRY COURETAS, *SOCIETY FOR MODELING AND SIMULATION, USA*
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 AUSTRIAN UNIV. OF AS, 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, MEXICO*
MIGUEL MÚJICA MOTA, *UAB, SPAIN*
TERESA MURINO, *UNIVERSITY OF NAPLES FEDERICO II, ITALY*
GASPER MUSIC, *UNIVERSITY OF LJUBLJANA, SLOVENIA*
LETIZIA NICOLETTI, *UNIVERSITY OF CALABRIA, ITALY*
TUNCER ÖREN, *M&SNET, UNIVERSITY OF OTTAWA, CANADA*
ELEONORA PANTANO, *UNIVERSITY OF CALABRIA, ITALY AND TECHNISCHE UNIVERSITEIT, EINDHOVEN, NETHERLANDS*
MERCEDES PEREZ, *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, FRANCE*
FEI TAO, *BEIHANG UNIVERSITY, CHINA*
ALBERTO TREMORI, *UNIVERSITY OF GENOA, 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*

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: LIN ZHANG, BEIHANG UNIVERSITY, BEIJING, CHINA; FEI TAO, BEIHANG UNIVERSITY, BEIJING, CHINA; SIMONLUCA POGGI, MAST SRL, ITALY

SIMULATION OPTIMIZATION APPROACHES IN INDUSTRY, AVIATION, SERVICES AND TRANSPORT

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

HUMAN-CENTRED AND HUMAN-FOCUSED MODELLING AND SIMULATION

CHAIRS: AGOSTINO BRUZZONE, DIME UNIVERSITY OF GENOA, ITALY; PEER OLAF SIEBERS, UNIVERSITY OF NOTTINGHAM, UNITED KINGDOM.

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, SIMULATION TEAM, 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, UNIVERSITY OF LJUBLJANA, SLOVENIA; ANDREAS KÖRNER, VIENNA UNIV. OF TECHNOLOGY, AUSTRIA

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

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

SIMULATION AND MODELLING FOR OCCUPATIONAL HEALTH AND SAFETY

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

ADVANCED MODELS AND APPLICATIONS OF LOGISTICS & MANUFACTURING

CHAIRS: TERESA MURINO, UNIVERSITY OF NAPLES FEDERICO II, ITALY; LIBERATINA SANTILLO, UNIVERSITY OF NAPLES FEDERICO II, ITALY

RETAIL-ORIENTED SIMULATION

CHAIR: ELEONORA PANTANO, UNIVERSITY OF CALABRIA, ITALY AND TECHNISCHE UNIVERSITEIT, EINDHOVEN, NETHERLANDS

COMPUTER SIMULATION FOR SECURITY

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

WORKSHOP ON MODEL ENGINEERING

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

GENERAL CO-CHAIRS' MESSAGE

WELCOME TO EMSS 2014!

The European Modeling & Simulation Symposium (EMSS) is the traditional simulation appointment in Europe; indeed EMSS is also known as the *Simulation in Industry* Symposium due to the large participation of researchers as well as practitioners working in the area of Modeling & Simulation (M&S) applied to Industry and Industrial Engineering. To this end, EMSS 2014 has strongly confirmed its own nature; tracks such as Industrial Engineering and Industrial Processes Modeling & Simulation have collected very high quality scientific papers proposing advanced simulation applications and methodology approaches in multiple domains. However EMSS 2014 has also confirmed its capability to be a point of attraction for researchers working in other areas. Indeed, this edition of EMSS includes papers in the area of Discrete, Combined, Distributed and Parallel Simulation, Simulation and Artificial Intelligence, Modeling & Simulation in Physiology and Medicine (joint track with the International Workshop on Innovative Simulation for Health Care, IWISH 2014), Modeling & Simulation Approaches In and For Education as well as Modeling & Simulation Methodologies, Techniques and Applications. As chair of EMSS 2014 we would like to take this opportunity to thank all the authors and the reviewers of EMSS; their invaluable work is the key of the symposium success. We would like also to inform authors that, as every year, the EMSS 2014 proceedings will be indexed by SCOPUS.

As every year, EMSS is co-located with the International Multidisciplinary Modeling & Simulation Multiconference (I3M 2014) that this year includes six International Conferences (EMSS, HMS, MAS, DHSS, IWISH, SESDE) therefore providing EMSS attendees with the possibility to spend the three conference days in an extended simulation framework where people working in different areas, such as Logistics and Supply Chain, Defense and Homeland Security, Healthcare, Energy, Sustainable Development and Environment, share their simulation experiences. Finally, this year EMSS is held in the wonderful framework of Bordeaux, France; nothing is better than working and meeting old and new friends while drinking red Bordeaux!



Francesco Longo
DIMEG, University of
Calabria, Italy



Emilio Jiménez
University of La Rioja,
Spain

ACKNOWLEDGEMENTS

The EMSS 2014 International Program Committee (IPC) has selected the papers for the Conference among many submissions; therefore, based on this effort, a very successful event is expected. The EMSS 2014 IPC would like to thank all the authors as well as the reviewers for their invaluable work.

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

I3M 2014 INTERNAL STAFF

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

I3M 2014 LOCAL ORGANIZATION COMMITTEE

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



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



Index

The model engineering for complex system simulation	1
Lin Zhang, Yuewei Shen, Xuesong Zhang, Xiao Song, Fei Tao, Ying Liu	
Teaching variance reduction in matlab	11
Jaroslav Sklenar	
Modeling and performance analysis of ammonia-water absorption refrigeration system for ocean-going fishing vessels	17
Li Ren, Zijian Guo, Hongxia Li	
Table top for scheduling and controlling constructions course	24
Caroline Cavalheiro, Ailton Soares Freire, Antônio Edesio Jungles	
Approach class library of high level parallel compositions to implements communication patterns using structured parallel programming	30
Mario Rossainz-López, Manuel I. Capel-Tuñón	
Data based identification of short term predictors for stock market trends using heterogeneous model ensembles	40
Stephan Winkler, Bonifacio Castaño, Sergio Luengo, Susanne Schaller, Gabriel Kronberger, Michael Affenzeller	
High level architecture virtual assistant framework	46
Josef Brozek, Bhakti Stephan Onggo, Antonin Kavicka	
Modelling and simulation of the non-ideal mixing behaviour of the crystallization stage in a sugar factory batch crystallizer	56
David Tejerina, Rogelio Mazaeda, César de Prada	
Modeling of lattice thermal expansion for simulation of thermal stress in 3D nano mosfets	65
Abderrazzak El Boukili	
The relationship between R&D expenses and performance: evidence from European manufacturing enterprises	72
Ondrej Zizlavsky, Michal Karas	
Identification of endothelial cell morphology in cornea using evolution strategies	79
Lisa Obritzberger, Susanne Schaller, Viktoria Dorfer, Claudia Loimayr, Simone Hennerbichler, Stephan Winkler	
Methodology for failure analysis in service processes through simulation	87
Oroselvia Sánchez, Idalia Flores	
Hierarchical feature selection for biological data	93
Witold Jacak, Karin Pröll	
Probabilistic behavioral model for the detection of criticalities when using morphine and fentanyl PCA	98
Henrikas Pranevicius, Mindaugas Snipas, Tadas Kraujalis, Mindaugas Pranevicius, Osvaldas Pranevicius, Vytautas Pilkauskas	
Modeling of mechatronic systems in Matlab (Simulink / Simmechanics)	106
Irina Kalapyshina, Anna D. Perechesova, Konstantin A. Nuzhdin, Victor M. Musalimov, Georgiy B. Zamoruev	
On the approximate solution for a fully coupled parabolic-hyperbolic PDES system applied in thermomechanical behaviour of hydrogel matrix used in drug delivery	112
Nirina Santatriniaina, Benjamain Boutin, Eric Darrigrand, Fabrice Mahé, Nicolas	

Crouseilles, Dominique Pioletti, Lalaonirina Rakotomanana	
Simulation optimization for a vaccine distribution strategy against the spread of a (H1N1) epidemic.	120
Javier Lara de Paz, Idalia Flores	
Kinematics and dynamics analysis of different mechanisms using simmechanics / simulink / matlab	128
Anna D. Perechesova, Irina I. Kalapyshina, Konstantin A. Nuzhdin	
A mono parameter analysis on a simulation model to support GALB heuristic optimization algorithms based on resource balancing	136
Sergio A. Gallo, Giovanni Davoli, Andrea Govoni, Francesco Longo, Riccardo Melloni, Teresa Murino	
BankSim: a bank payments simulator for fraud detection research	144
Edgar Alonso Lopez-Rojas, Stefan Axelsson	
A multi-agent model of typical competitive bidding process in construction	153
Ronald Ekyalimpa, Simaan AbouRizk	
Mathematical modeling of the diversity in human B and T cell receptors using machine learning	164
Susanne Schaller, Johannes Weinberger, Martin Danzer, Christian Gabriel, Rainer Oberbauer, Stephan M. Winkler	
A multi task assignment method in cloud-based simulation	171
Lei Ren, Hejian Ou, Jin Cui, Bowen Li, Baocun Hou	
Empirical modeling and simulation for discharge dynamics enabling catchment-scale water quality management	175
Huma Zia, Nick R. Harris, Geoff V. Merrett	
Event-oriented control functions for enhancing development process of war-game simulators	184
Se Jung Kwon, Changbeom Choi, Tag Gon Kim	
Simulation as a service in construction	191
Sean Newstead, Simaan AbouRizk, Stephen Hague, Yasser Mohamed, Junhao Zou	
Digital human models (DHM) to simulate meat processing to prevent work related Musculoskeletal disorders (WR-MSDS)	199
Sergio Amedeo Gallo, Riccardo Melloni, Simone Mosconi, Francesco Longo	
Hierarchical approach to developing a logistic discrete event simulation model using automated guided vehicles	205
Pavel Raska, Zdenek Ulrych	
A physics simulation tool for the container loading problem	212
António G. Ramos, João Jacob, Jorge Justo, José F. Oliveira, Rui Rodrigues, António Miguel Gomes	
Towards a methodology for human behaviour elicitation: preliminary results	220
João Emilio Almeida, Rosaldo J.F. Rossetti, Brígida Monica Faria, João Tiago Jacob, António Leça Coelho	
GPRS network monitoring of solar energy generation and supply in rural residences	229
Robson da Cunha Santos, Gerson Gomes da Cunha, Marcos António Cruz Moreira, Mónica Castelo Branco, David Douglas Nunes Oliveira, Matheus Muzitano Reis	
Approaching demand of cash transactions at bank branches	238
Miguel Aguilar Zaragoza, Idalia Flores de la Mota	
On the search for novel simulation application to support airport operations	245

management	
Olusola Theophilus Faboya, Peer-Olaf Siebers	
Analysis of obesity modelling	254
Maja Atanasijevic Kunc, Tina Sentocnik, Jože Drinovec	
Real time interaction and handling tool of historical data on timelapse record for construction management	263
Livia Fernandes, Gerson G. Cunha, Celia Lopes, Luiz Landau	
Mapping of wireless technologies to support real time location systems for tracking resources of large enterprises	269
Larissa P.M. Cruz, Luiz Landau, Gerson Gomes Cunha, Maria Célia S. Lopes	
Probabilistic approach as a support system for safety-critical environments by knowledge driven bayesian networks	280
Ciro D'Elia, Fabio De Felice, Paola Mariano, Antonella Petrillo, Simona Ruscino	
A multimodal optimization method for simulation systems	289
Zhizhao Liu, Wei Li, Ming Yang	
FEM analysis of radial-axial profile ring rolling process	295
Zhengkun Feng, Henri Champlaud	
Searching and indexing distorted data collections	301
Tomas Kocyan, Jan Martinovic, Michal Podhoranyi	
Train movement dynamics within AnyLogic tool	307
Roman Divis, Antonin Kavicka	
Self-optimizing real-time recommendation model design based on YARN	313
Tao Liu, Shuang Wang, Peng Wu	
Non HLA distributed simulation infrastructure	319
Jan Voracek, Jiri Penzes, Antonin Kavicka	
Predicting equipment availability using a high level architecture framework	325
Estacio Siemann Santos Pereira, Yasser Mohamed, Simaan AbouRizk	
Design and development of a distributed earthmoving simulation	334
Duanshun Li, Cristian Petre, Cayce Kerr, Tim Joseph, Simaan AbouRizk, Yasser Mohamed	
WebRTC technology as a solution for web-based distributed simulation	343
Stepan Kartak, Antonin Kavicka	
Combining DEVS and model-checking: using systems morphisms for integrating simulation and analysis in model engineering	350
Bernard P. Zeigler, James J. Nutaro	
Two stage simulation use in project verification and validation	357
Egils Ginters, Artis Aizstrauts, Miquel-Angel Piera Eroles, Roman Buil	
Migration among simulation paradigms and tools	364
Mairita Zake, Egils Ginters	
Utilization of computer simulation for the detection non-standard situations within the new data layer of railway network model	371
Jan Fikejz, Emil Rezanina	
Feedback DTC-SVM based a fractional PI controller: applied to rotor's speed of induction motor	378
Youcef Zennir, Lakhdar Bouras	
Influence of parked cars on smoke propagation during car park fire	384

Peter Weisenpacher, Ladislav Halada, Jan Glasa, Jan Aсталos	
ProOpter: production dynamics analysis and optimization tool	392
Gaspar Music, Miha Glavan, Dejan Gradisar, Stanko Strmcnik	
Evolutionary algorithms for hyperparameter tuning on neural networks models	402
David Orive, Gorka Sorrosal, Cruz E. Borges, Cristina Martin, Ainhoa Alonso-Vicario	
Modeling the multi-compartment vehicle routing problem with stochastic demands	410
Jan Melechovsky	
Integrated simulation and optimization in HeuristicLab	418
Andreas Beham, Gabriel Kronberger, Johannes Karder, Michael Kommenda, Andreas Scheibenpflug, Stefan Wagner, Michael Affenzeller	
Production line modeling and balancing: comparison of existing techniques and proposal of a new methodology	424
Alessandro Silvestri, Cristina Cerbaso, GianPaolo Di Bona, Antonio Forcina, Vincenzo Duraccio	
Maintenance critical analysis and priority index: a new model for maintenance policy	432
Alessandro Silvestri, Cristina Cerbaso, Domenico Falcone, Antonio Forcina, Vincenzo Duraccio	
Towards a metamodel for airport modeling and simulation	438
Deniz Çetinkaya, Ismet Camci	
Simulation based analysis and development of decision support system for virtual network bandwidth management	444
Julija Asmuss, Gunars Lauks	
Using tablets in distributed simulation	451
Josef Brozek, Martin Jakes, Lumir Gago	
Speedsim.net - an open.net based simulation system	457
Thomas Wiedemann, Karsten Wendt	
Applying an adaptive petri net to calculate the makespan in the job shop scheduling problem	463
Joselito Medina-Marin, Juan Carlos Seck-Tuoh- Mora, Norberto Hernandez-Romero, Nayeli Jazmin Escamilla-Serna	
Modeling and optimization of the extraction of lignosulfonate from barley straw by using artificial neural networks	469
Maria Guadalupe Serna-Diaz, Joselito Medina-Marin, Ainhoa Arana-Cuena, Juan Carlos Seck-Tuoh-Mora, Alejandro Tellez-Jurado, Yuridia Mercado-Flores, Angelica Jimenez-Gonzalez, Norberto Hernandez-Romero	
A friday 13th risk model for failure in cross - flow membrane filtration of passion fruit juice	474
Wei Zou, Kenneth Davey	
A Novel friday 13th risk analysis of a global food process – application to pasteurization of raw milk containing mycobacterium avium paratuberculosis	480
Kenneth R. Davey, Saravanan Chandrakash, Brian K. O'Neill	
A new transient predictive model to quantify taint as either geosmin (GSM) or 2-methylisoborneol (MIB) in rainbow trout (Oncorhynchus Mykiss) farmed in recirculating aquaculture systems (RAS)	490
Kenneth R. Davey, Priyantha I. Hathurusingha	
A modified beer game for simulation and optimization teaching	498

Diego Crespo-Pereira, Adolfo Lamas-Rodriguez, Rosa Rios Prado	
A simple queue model for an appointment system and applications in a hospital CT scan facility	504
Francesco Boenzi	
Simulation and experimental analysis of led weatherproof luminaire thermal performance	517
Carlos Javierre, Daniel Elduque, Víctor Camañes, David Franch	
Fupol simulators and advanced visualization framework integration	523
Egils Ginters, Artis Aizstrauts, Mikelis Baltruks, Kawa Nazemi, Dirk Burkhardt, Peter Sonntagbauer, Susane Sonntagbaure, Jorge Martin Gutierrez	
Visualization techniques for simulation models in virtual reality	530
Arnis Cirulis, Egils Ginters	
A coloured petri net model for the MAS simulation of urban economics in Yantai	536
Roman Buil, Miquel Angel Piera, Egils Ginters, Artis Aizstrauts	
Single sourcing vs. double sourcing: a simulation approach for supplier selection	543
Mattia Armenzoni, Marta Rinaldi, Roberto Montanari, Eleonora Bottani, Federico Solari	
A conjecture from learning simulations of series and parallel connections of components	550
Alexandre Muzy, Bernard P. Zeigler	
Skopje bicycle inter modality simulator – E-involvement through simulation and ticketing	557
Egils Ginters, Artis Aizstrauts, Girts Dreija, Maija Ablazevica, Sergey Stepucev, Inita Sakne, Mikelis Baltruks, Miquel Angel Piera Eroles, Roman Buil, Marjan Gusev, Goran Velkoski	
DEVS models design and test using AGILE-based methods with DEVSImPy	563
Timothée Ville, Laurent Capocchi, Jean-Francois Santucci	
CPN-simulation methodology for the boarding process of aircraft	570
Miguel Mujica Mota, Idalia Flores, Daniel Guimarans	
Tool profiles evaluation based on vibroacoustical signals generated by friction-stir welding	576
Julio Blanco-Fernández, Emilio Jiménez-Macías, Mario Sanchez-Orozco, Angel Sanchez-Roca, Hipolito Carvajal-Fals	
Evaluation of the erosive wear of the AA 6082 T6 alloy treated with friction stir Pprocessing (FSP)	582
E. Martínez-Cámara, J. Blanco-Fernández, Y.P. Chacón, A. Sanchez-Roca, H. Carvajal-Fals	
Design and operation of a dairy plant by means of a decision support tool based on the Petri nets paradigm	588
Juan Ignacio Latorre-Biel, Emilio Jiménez-Macías, Julio Blanco-Fernández, Eduardo Martínez-Cámara, Juan Carlos Sáenz-Díez, Mercedes Pérez-Parte	
Mixing formalisms based on Petri nets for improving the efficiency of modelling and simulation of DES	594
Juan Ignacio Latorre-Biel, Mercedes Pérez-Parte, Juan Carlos Sáenz-Díez, Jorge L. García-Alcaraz, Emilio Jiménez-Macías	
Disjunctive coloured Petri nets: a formalism for improving the applicability of CPN to the modeling of DES with alternative structural configurations	601
Juan Ignacio Latorre-Biel, Mercedes Pérez-Parte M, Emilio Jiménez-Macías	

Development of a Petri net model for a reconfigurable intelligent system based on experimental data	608
Juan Ignacio Latorre, Karim El-Laithy, Martin Bogdan, Emilio Jiménez	
Petri net representation with ciphered subnets: definition of PNML extensions for subnets representation and use of XML encryption for ciphering	613
Iñigo León-Samaniego, Juan-Carlos Sáenz-Díez, Jorge Luis García, Mercedes Pérez-Parte	
Production of compost for mushroom cultivation: a life cycle assessment study	620
Francisco J. Leiva-Lázaro, Julio Blanco-Fernández, Eduardo Martínez-Cámara, Emilio Jiménez-Macías	
Modeling of alternative for in situ unidirectional slabs	626
Esteban Fraile-García, Javier Ferreiro-Cabello, Eduardo Martínez-Cámara	
Model engineering for cyber complex adaptive systems	632
Saurabh Mittal	
Serious game at increased impact on culture and tourism	641
Francesco Longo, Letizia Nicoletti, Stefano Vena, Antonio Padovano	
Disaster and emergency management simulation in industrial plants	649
Agostino G. Bruzzone, Marco Frascio, Francesco Longo, Alessandro Chiurco, Simone Zanoni, Lucio Zavanella, Paolo Fadda, Gianfranco Fancello, Domenico Falcone, Fabio De Felice, Antonella Petrillo, Pasquale Carotenuto	
A system dynamics approach for improving container terminal operations	657
Giuseppe Converso, Mosè Gallo, Teresa Murino	
Innovation in hospitals: an e-procurement model in pharmacy operations in day surgery	662
Piera Centobelli, Giuseppe Converso, Antonio De Iasi, Teresa Murino	
Health worker monitoring: Kalman-based software design for fault isolation in human breathing	668
Improta Giovanni, Natale Pasquale, Santillo Liberatina Carmela, Triassi Maria	
The industrial plants relocation: issues, policies, procedures and algorithms for disassembly and reassembly phases	673
Massimo De Falco, Pellegrino Gaita, Mario IIsami, Liberatina Carmela Santillo	
Author's Index	679

The Model Engineering for Complex System Simulation

Lin Zhang^(a), Yuewei Shen^(b), Xuesong Zhang^(c), Xiao Song^(d), Fei Tao^(e), Ying Liu^(f)

^(a) ^(b) ^(c) ^(d) ^(e) ^(f) School of Automation Science and Electrical Engineering, Beihang University, Beijing, China

^(a) johnlin9999@163.com, ^(b) shenyuewei@gmail.com, ^(c) xs_zhang@126.com, ^(d) s761127@sina.com, ^(e) taofei999@gmail.com,
^(f) nervkidliu1103@126.com

ABSTRACT

The model is the foundation for simulation, especially for the complex system simulation. To meet the challenges in development and management of the complex system model, this paper proposes the concept of the model engineering, which aims at setting up a systematic, normalized, and quantifiable engineering methodology by exploring basic principles in model construction, management and maintenance to manage the data, processes and organizations/people involved in the full lifecycle of a model to guarantee the credibility of model lifecycle. The meaning and contents of the model engineering are given, a set of key technologies are introduced. A preliminary framework of the body of knowledge of the model engineering is given based on software engineering, system engineering and some other relative disciplines.

Keywords: simulation, model engineering, credibility, complex system, body of knowledge

1. INTRODUCTION

A model is an abstract expression of objects to study and embodies high intelligence of human beings in recognition of the world. With continuous development of science and technologies, the model is becoming more and more important. With a model, the key elements such as semantics and context in the information, system structure, functions and behaviors can be clearly displayed to exchange and share information among different fields, areas and platforms cross the obstacles in the culture, languages and technology level. The model has become one of the key contents in the modern science system. Human being has accumulated a large number of knowledge and experiences on model construction and use during long-term science research and practices, which play important role in the respective science and technology area. With continuous growth of complexity and diversity of objects to study, the models are becoming more complicated and diversified. Research on models is the key to affect science development.

Simulation technology has rapidly grown up and becomes a new interdisciplinary with the development of

computer technology, and has become one essential research method and means to almost all fields of science and technology (Li and Chai, 2002). Simulation is an activity based on the model, and its main research purpose is to create a system model and carry on the scientific research and experiments. In the simulation field, the research on the model is very representative.

The correct model can guarantee correct simulation results. Verification, Validation and Accreditation (VV&A) is the primary means of the credibility of the research simulation and simulation results (Pace 2004). VV&A for the model is to calibrate or validate the established model in order to determine whether it is credible. This post determination has important implications to discover model problems and defects, but it still cannot solve the problem that how to get a correct model, Especially for complex systems, due to the complexity and uncertainty of the system, the modeling process can be very complicated, which makes VV&A of a model become extremely difficult. Even if the defects are found via VV&A, the amendment of the model will be very difficult and high-cost.

More importantly, for a complex system, to construct a correct model is only the first step since the a complex system model generally experiences a long term of evolution and management. As a result, the key issue for a complex system model is to guarantee the credibility of the full model lifecycle with minimum cost.

In the current research on model, the theories and methods are generally aims at specific objects and are random and lack necessary fundamentals and rules. Most of the research is only for a certain phase and isolated from other phases of a model lifecycle, and seldom systematically consider the full lifecycle of a model as a whole. So it leads to many problems and difficulties in model development and management and cannot guarantee credibility of the model lifecycle. This case is caused due to three main reasons, first, no generic methodology guide for model lifecycle is available; Secondly, no effective process management mechanism for model lifecycle is established; thirdly, no uniform standard and technical support is available. Therefore, it is necessary to establish model-oriented

methodology system to restrain, guide and manage the model lifecycle process and guarantee correctness of the established models and credibility of the model lifecycle.

This paper aims to apply the engineering ideas and methods into the lifecycle of complex system model, get a standardized, normalized, quantified and systematic model engineering system by referring to and absorbing ideas and methods in the system engineering and software engineering, exploring, mining and summarizing ubiquitous theories, methods, technologies, standards and tools in the model related activities, serve for the model lifecycle including model construction, management, use, maintenance process, etc, and guarantee credibility of the model lifecycle.

This paper first analyzes the issues and challenges in modeling of the complex systems, then introduces existing research achievements in model development and management, gives concept and key technologies of the model engineering based on them, and proposes the body of knowledge framework of the model engineering for the complex system simulation by analyzing and arranging the software engineering, system engineering and body of knowledge in other related fields.

2. CHALLENGES IN DEVELOPMENT AND MANAGEMENT OF COMPLEX SYSTEM MODELS

No uniform definition for the complex systems is available now. The reference (Dai 1997) thinks that “a complex system features complicated composition relation, complex system mechanism, and complicated interaction relation and energy exchange between the system sub-systems and between system and environment” . Typical complex systems include social systems, complex environment systems, complex engineering systems, complex military systems and complex network systems, etc. (Li and Chai 2012). We can summarize main features of complex system models and challenges in construction and management of models.

1) High system complexity

First, a complex system is characterized by complicated composition, containing a large number of composition elements and very complicated relationship between elements. Secondly, generally a complex system is dynamic and variable and is very uncertain. Above complexities lead to very complicated verification, validation and accreditation (VV&A) of complex system models. Complexity of complex systems leads to complexity of models.

2) Long life cycle of a complex system.

With time elapse, the models should be continuously improved and changed. Different model versions are available. Each version may be applicable to different application phases. Different versions and application phases of multiple models compose complicated

network. How to keep consistence and credibility of different parts and versions of the model is the key for model maintenance.

3) Model heterogeneity

A complex system is composed of many heterogeneous component models. Heterogeneity of models generally comes from different development organizations, different platform and architecture, different development languages and databases, etc. Heterogeneity brings big challenges to integration and maintenance of the system models.

4) Complicated evolution of models

Generally a complex system is in continuous evolution, so the models will be continuously adjusted and changed. Changes of different relations are very complicated in evolution due to system complexity, so the model elements and its relation should be completely tracked and managed to guarantee correctness of the model evolution.

5) Difficult model reuse

With growth of complexity of the systems to study, the roles and values of model reuse is very remarkable in model development and use of a complex system (Liu, Tang, and Zheng 2008). Generally a complex system includes multiple combined systems. A huge number of models in past research and development practices have been accumulated. Correct and efficient reuse of the models will reduce model development cost, shorten development time much and effectively improve model credibility. Although some research on model reuse has been conducted, no efficient and practicable model reuse method is available now.

6) Massive processing data

Generally a complex system includes a large number of data to process, including required modeling data, data generated in modeling process, and data generated in model operation. Data processing includes data storage, inquiry, exchange, management, understanding, analysis and mining, which brings many challenges.

7) The multidisciplinary collaborative model development

Collaborative model development is associated with different steps in the whole model lifecycle. Collaboration is required on different phases, e.g. collaborative requirement analysis, collaborative design, and collaborative validation. All work compose a huge engineering and should be supported by management tools or platform.

8) Higher requirements for system performance

Compared to a simple system, a complex system requires higher performance, e.g. higher requirements for reliability, security, credibility, cost and energy saving. To guarantee that these performance requirements are met, special means should be required to analyze and process the models.

3. RELATED WORK

How to build a right model is the core issue in simulation. A large number of research achievements on models have been obtained in the past dozens of years. These achievements are related to different phases in a model lifecycle, e.g. modeling theory and method, model validation and model management.

For modeling theory and methods, as described in the above part, many researchers have developed different modeling methods for different systems. For modeling of complex systems, now issues in two aspects exist: (1) the confliction between model precision and complexity has not been well solved; (2) research on field-associated methodology for modeling process has seldom been conducted. Although many specific modeling methods are available, no enough generic guides and restraints are available for model development. For a complex system model requiring long-term use and maintenance, it is difficult to guarantee the model credibility and the aftermaths are catastrophic.

Research on the model validation, verification and accreditation (VV&A) is a hotspot in simulation field and many research achievements have been obtained (Wang and Lehmann 2007, Balci 1994). Related methods and technologies include model confidence rule, simulation credibility evaluation theory and method, VV&A framework and standard, quantified statistics and test method, data sequence comparison method, similarity method, and validity analysis method of simulation model acknowledge data (Yang and Fang 2011). Although researchers are plentiful, lot of work is required to establish a complete and mature VV&A theory system (Yang 2013, Liu 2007).

Compared to the research on modeling and VV&A of models, research on the model management, especially model lifecycle process management is very limited. Research on model management focuses on simulation model library design and management and model reuse, e.g. web-based simulation model library system framework (Wu, Qiu, and Liu 2006, Fensel and Bussler 2002), multi-model switching based on dynamic model library, object-oriented model library DSS reuse architecture and model library architecture management method. Research on model reuse includes meta-model (Alanen and Porres 2004, Wang and Lei 2007), component (Liang and Zhang 2008), modeling languages for model reuse, standardized modeling for model reuse, model driven architecture (MDA) (Kleppe, Warmer, and Bast 2003), and modeling environment and run environment for simulation model reuse (Jia and Zhang 2007, Brutzman 2002). Some preliminary achievements have been achieved in the fields such as combination of simulation models, model evolution and model configuration (Liu, Liu, and Wang 2012; Martens, Koziol, and Becker 2010; Dhungana, Grünbacher, and Rabiser 2010), but they are in the initial phase.

The lifecycle concept has not been emphasized enough in the simulation domain, and related research and applications are not sufficient (Balci 2012). The reference (Fishwick 1989) called the simulation model development process as simulation model engineering to emphasize engineering feature of the model development process, but no special explanation on its meaning was given and no systematic method system was established.

In recent years, the international simulation community is conscious of unfavorable influences of missing foundational theory of M&S on the development of simulation curriculum. As a result, research on the M&S lifecycle management is gradually attracting attention from the academic circle. The reference (Radeski and Parr 2002) proposed the modeling simulation lifecycle model framework, which defined organization mode and structure of the modeling simulation process, work products, quality assurance, project management, and described the features and requirements of the lifecycle phases such as development, use, maintenance and reuse of the modeling simulation system. The reference (Fishwick 1990, AbdouniKhayari 2010) achieved some valuable results in model lifecycle management and developed a model prototype management system, which provided valuable reference to model development for designers.

In a word, current research on models is generally focused on one phase in the model lifecycle and is separate and disperse. Although importance of the engineering idea is gradually recognized in applications of the full model lifecycle, now no complete theory and technology system and philosophy is available.

4. MEANING OF THE MODEL ENGINEERING

4.1. Concepts of The model engineering

Generally a model experiences requirement analysis, model design, model construction, model validation, model application and model maintenance. These processes compose a complete lifecycle of the model.

Based on state-of-the-art research on models, we propose a systematic methodology to cope with challenges in model lifecycle management of a complex system. The model development and management activities change from a spontaneous and random behavior to conscious, systematic, standardized and manageable behavior by constructing a model engineering theory and methodology system in order to guarantee credibility of different model phases.

Model engineering is defined as the general term of theories, methods, technologies, standards and tools relevant to a systematic, standardized, quantifiable engineering methodology that guarantees the credibility of the full lifecycle of a model with the minimum cost.

1) Model engineering is regarded the full lifecycle of a model as its object of study, which studies and establishes a

complete technology system at the methodology level based in order to guide and support the full model lifecycle process such as model construction, model management and model use of complex systems.

2) Model engineering aims to ensure credibility of the full model lifecycle, integrate different theories and methods of models, study and find the basic rules independent of specific fields in the model lifecycle, establish systematic theories, methods and technical systems, and develop corresponding standards and tools.

3) Model engineering manages the data, knowledge, activities, processes and organizations/people involved in the full lifecycle of a model, and takes into account time period, cost, and other metrics of development and maintenance of a model.

4) Here the model credibility is a comprehensive indicator and includes factors such as availability, accuracy, reliability and QoS.

4.2. Key Technologies of The model engineering

As the described in the above part, now the research on the technologies related to the full model lifecycle is preliminary and disperse. For comprehensive and systematic application and implementation of the model engineering, many key technologies should be studied.

1) Body of knowledge of the model engineering

The body of knowledge (BOK) includes the concepts and terminologies involved in a specific research field. This paper gives a preliminary knowledge system framework. The model engineering BOK identifies the research scope of the model engineering and its boundary and relationship with other related subjects. Establishment of systematic and complete BOK requires long-term accumulation and extraction.

2) Model engineering standards

Standards are the basis for the implementation of the model engineering. During the lifecycle process of a model, each activity requires corresponding standards, including model development process, model description, model component interface, model storage, model data exchange, model interoperation, model service, model maintenance, etc.

3) Modeling of model lifecycle process

The lifecycle model of the model engineering aims to identify the structural framework of activities involved in model construction and management (Zeigler, Kim, and Praehofer 2000), which is the methodology to guide the model engineering, and ensure improvement of model quality and development efficiency and reduction of full model lifecycle cost. Proper process models and corresponding implementation methods can be proposed by referring to the existing achievements in the system engineering, software engineering and other relevant fields

and combining the model development features of complex systems.

4) Model engineering process management

The data, knowledge, tools, persons/organization and technologies in the full model lifecycle should be effectively managed with the model lifecycle process model as the guide, with standards as the basis, and with the project management methods and means as reference in order to get the dependable model with the minimum cost. The model maturity definition and control, performance management, flow monitoring and optimization, risk control, and cost control are important in the model engineering process management.

5) Acquisition and management of model requirements

Accurate requirement acquisition is the key in modeling. Requirement acquisition and management is very challenging due to uncertainty and ambiguity of complex systems. Requirement acquisition studies to extract, describe, parse and validate requirement via automated or half-automated means. Requirement management studies how to reflect the changing requirements in model construction and maintenance accurately and timely.

6) Model description and modeling language

Complex systems generally contain multiple different systems with different properties, such as quantitative systems, qualitative systems, continuous systems, discrete event systems, deterministic systems, uncertain systems, etc. One of the core issues in model development of complex systems is how to take advantage of effective ways to describe the whole system. Therefore, it is required to study corresponding model description mechanism and structure and develop generic or specific description languages according to the characteristics of the various systems.

7) Quantitative analysis and evaluation of the model engineering

The quantitative analysis is one of main features of the model engineering. To ensure credibility of the full model lifecycle, many steps should be analyzed, evaluated and optimized in a quantitative manner, e.g. complexity analysis and evaluation of model development process, cost and benefit analysis and optimization, risk analysis and control, model availability and reliability analysis, and model service quality analysis.

8) Model validation, verification and accreditation (VV&A)

The model VV&A technology is one important part in the model engineering. Although some rich research achievements have been achieved, they cannot meet the actual requirements of modeling simulation of complex systems. Most research focus on qualitative analysis and quantitative and formalized analysis methods are lack, so VV&A technology, especially VV&A quantitative analysis

and formalized analysis technology are still main research content in the model engineering.

9) Model library

The model library is the foundational platform to carry out model management and perform standardized encapsulation, storage and query for the models (Ören and Zeigler 1979). The complicated applications such as model reuse, combination and configuration management can be based on the model library. Traditional database technology, service-oriented technology, and cloud computing technology can support construction and management of the model library.

10) Model composition and reuse

The model composition and reuse technology is an important means to improve model construction and maintenance efficiency and improve model credibility of complex systems. It mainly studies how to use the existing model components to quickly and correctly compose complicated models according to the system requirements and includes standardized encapsulation of model components, intelligent model matching, model relation management, dynamic model composition, model consistency validation and model service.

11) Model reconstruction and configuration management

The requirements for model functions and performances change due to diversified requirements and inside and environmental uncertainty, so the models should be quickly reconstructed or configured. The model reconstruction aims to adjust the internal structure without change of main external functions of models, further optimize the model performance, and ease its understanding, maintenance and transplant of models. Model configuration can adapt different requirements or changes of models in function and performance by adjusting and optimizing internal components and parameters. For complex system model engineering, model reconstruction and configuration management is very important and challenging.

12) Model data and knowledge management

Many complex system models contain many data to process. Some models are constructed based on massive data, even exist in the form of data and their relations. Data management aims to effectively organize and use the data, especially massive data, and plays a key role in quantitative analysis of the model engineering.

On the whole, the knowledge is divided into two classes. The class 1 indicates the knowledge in the model, e.g. some qualified models include massive knowledge rules. Another class indicates the knowledge on model development and management and generally includes experiences accumulated and extracted by developers and users in practices. Different knowledge should be managed and use in different manners to improve model quality and

intelligence and automation of model construction and maintenance.

13) Visualization of the model engineering

Visualization technology can be used on different phases of the model engineering, can realize transparent model development and management process, facilitate understanding and monitoring, and improve human-machine interaction efficiency. The visualization technology plays an important role in the model engineering.

14) Support environment and tools of the model engineering

Implementation of the model engineering requires an integrated support environment and corresponding support tool to support different activities of the model engineering, e.g. network collaboration, requirement management, process model construction and maintenance, model library management, qualitative and quantitative analysis and evaluation, data integration, knowledge management, model validation and simulation experiment.

5. PRELIMINARY FRAMEWORK OF THE BODY OF KNOWLEDGE

The model engineering is resultants of fusion of many crossing subjects including the software engineering, system engineering, computer science and engineering, mathematics, system modeling and simulation, knowledge engineering, project management, quality management and related application fields. Based on the body of knowledge (BOK) in these disciplines, specific BOK of the model engineering is formed according to the requirements and features of the model engineering.

To establish the model engineering BOK, it is necessary to tease out the involved knowledge in a systematic manner, extract features closely associated with the model-related activities from related fields, and summarize and condense those specific development technologies and management means of the model lifecycle process. Now we give a preliminary BOK framework of the model engineering. Two aspects are mainly considered in this process.

- 1) Identify the horizontal crossing relations between the model engineering and other closely associated subjects, properly tailor their overlapping parts, and make these overlapping parts reflect specific features of the model engineering.
- 2) Identify the modules in the model engineering system and vertical hierarchical relations and horizontal interface relations between modules, make the framework compose an organic whole, and serve for the full lifecycle process of the model.

By establishing the model engineering BOK framework, we hope to reach the following targets:

- 1) Promote consistent opinion of the academic circle on the meaning of model engineering;
- 2) Identify the research scope of model engineering;
- 3) Expound the position of model engineering to other subjects such as software engineering, system engineering, computer science and mathematics and set their boundaries.

We give the BOK framework of model engineering (Figure 1) by referring to knowledge system in the fields and subjects such as the software engineering, system engineering, modeling and simulation, project management and quality management (Mittal and Risco-Martín 2013, Feron 2013, Ludewig 2003).

Figure 1 gives the general logic structure of the BOK framework of the model engineering. Based on the full lifecycle process of model requirement, design, construction, VV&A, application and maintenance, the model engineering completes the model lifecycle process via implementation and guarantee means such as the model engineering process management, model engineering quality management and model configuration management. The model engineering tool provides assisted support for implementation of the model engineering. The model engineering standard provides the rules, protocols or specifications which should be complied with in implementation of the model engineering and development of related tools. The model engineering foundation includes the ideas, terms, theories and methods for the model engineering.

The BOK framework of the model engineering is divided into five parts:

Part I - foundation: including the basic concepts and terms, methodology, technology systems, etc. It provides the basic guidance for the implementation of the model engineering, and also is the foundation and guarantee of the model engineering independent of the other subjects.

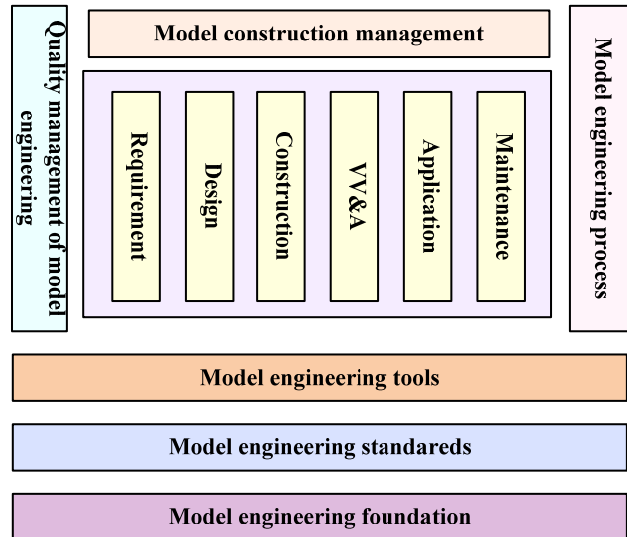


Figure 1: The BOK framework of the model engineering

Part II: model lifecycle: it describes different phases of the full model lifecycle at the technical level. This part is divided into six subjects, namely modeling requirement, model design, model construction, model VV&A, model application and model maintenance.

Part III: implementation and management: it includes the process management and quality management of the model engineering, and model configuration management. All activities in the full model lifecycle are managed and controlled in implementation, process and quality.

Part IV - tools: it provides the necessary software tools for the implementation and application of the full life cycle of the model engineering.

Part V - related standards: it includes rules, protocols or specifications which are necessary for implementation of the model engineering and development of related tools.

The detailed contents of each part are listed in Table 1.

Table 1: Details in the BOK framework of the model engineering

Theme		Contents	
Part I	Foundation of the model engineering	Outlines	The overviews, definition, function, body of knowledge of the model engineering, as well as necessary instructions of the brief introduction of other sections
		Concept of model	the basic contents, function, characteristics and meta-model
		History and status quo of modeling technology	Brief review of emergence and development of modeling technology, and the summary of status quo of the current modeling technology
		Model classification	The model classification is given from the consist, time characteristic, state characteristic, construction means, such as discrete, continuous, deterministic, random, qualitative, quantitative, mathematics, computer, concept, ontology model.
		Application field of model	According to the final application purpose of the model, and the macroscopic classification is given, such as communication,

	Theme	Contents	
		education, training, understanding, decision support, entertainment.	
	Life cycle of the model	Description of the main work of the model in every main phase from the full life cycle of production to the application, and each stage between join relationship	
	Modeling ideas	From the view of the cognitive theory, logic and methodology, to elaborate modeling ideas and its features	
Part II	Model requirement	Requirement foundation	Definition of requirements, process quality requirements, the functional and non-functional requirements, qualitative and quantitative requirements, system requirements, hardware and software requirements.
		Requirement acquisition	Requirement source, determination and negotiation technology.
		Requirement analysis	Requirement classification, requirement modeling, framework modeling and management.
		Requirement validation	Requirement assessment, requirement model verification, accreditation and revision.
		Requirement process	Process model, process support and management.
	Model design	Model design foundation	The design conception, model design context, model design process, enabling technologies
		The key issues of model design	Abstraction process, the interaction and performance, meta-model, data persistence
		The quality analysis and evaluation of model design	Quality attributes, quality analysis and evaluation technology, measurement
	Model construction	Modeling foundation	Complexity, related standards for validation and verification of modeling
		Modeling paradigm and modeling language	Modeling based on the graph, modeling based on the modeling language, description modeling, structure modeling, function modeling, relationship modeling, information modeling, process modeling, framework modeling and object-oriented modeling, formalization modeling, modeling based on component
		Modeling process	Modeling levels, refinement steps, implementation criterion
		Complex system modeling (hierarchy modeling)	Meta-model modeling, combined modeling, qualitative modeling, causal modeling, multi-resolution modeling, uncertainty modeling
		Emerging modeling	The multi-body modeling, modeling based on the self-organization theory, comprehensive integrated modeling
		Ontology modeling	ontology classification, candidate ontology, OWL
		Multidisciplinary collaborative modeling	Model with unified expression, model information exchange based on ontology, model abstraction level, model storage
	Model VV&A	VV&A foundation	Related termination of VV&A, key issues, relationship of VV&A and other activities
		VV&A technology	Non-formalization method, formalization method, index, steps
		Relevant measurement of VV&A	Model maturity, evaluation validation model, evaluation verification model, completed VV&A
Model application	Model deployment	The environment deployment, the model calibration, model sensitivity optimization	
	Model Integration	Model consistency, interface matching and transformation	
	Model reuse	Model splitting and combination, model integrity	

	Theme		Contents	
	Model maintenance	Model maintenance foundation	Definition and terminology, the essence, necessity, classification of maintenance	
		Key issues of model maintenance	Technical issues, management issues, maintenance cost estimation, model maintenance measurement	
		Model maintenance technology	Model meaning, model storage, model retrieval, models version control, the document generation technique based on the model	
Part III	Configuration management of the model	Process management of model configuration	Organization context, process scheme constraints and guidelines, plan making, management plan and supervision of model configuration.	
		Model configuration identification	Identifying item, model base	
		Model configuration control	Model change request, evaluation, approval, and implementation model change	
		Model configuration state bookkeeping	Model configuration state information, model configuration status report	
		Model Configuration verification	Model function configuration verification, model physical configuration verification, model base verification within the process	
	Process management of the model engineering	Definition of the process	The lifecycle model and processes of the model engineering, process definition notation, process revision and automation	
		Process evaluation	Process evaluation model and methods	
		Process realization and change	Process foundation structure, modeling process management cycle, process realization and change model	
		Engineering quality	Measurement establishment and maintenance, measurement process planning, measurement implementation process, measurement evaluation	
		Project management	Deliverables products determination, workload estimation, schedule, cost, resource allocation, risk management, quality management, plan management, process monitoring and control	
		Assessment and evaluation	Model consistency, the functional and non-functional demand determination, performance assessment and evaluation	
		Process and product measurement	Process measurement, model engineering product measurement, measurement results quality, information model, process measurement technology	
	Quality management of the model engineering	Model quality foundation	Model engineering culture and ethics, quality value and cost, model and quality characteristics, quality improvement	
		Model quality management process	Model quality validation, verification and accreditation and review	
		Model quality management technology	Model quality measurement, model comparison	
	Part IV	Tools of the model engineering		It mainly includes model requirement tools, design tools, construction tools, VV&A tools, maintenance tools, configuration management tools, engineering management tools, engineering process tools, model quality tools, etc.
	Part V	Related standards of the model engineering		It mainly includes related existing or intended standards supporting model engineering model, such as terminology standard, model description standard, packaging standard, interface standard, model data exchange standard, interoperability standards, validation, verification and accreditation (VV&A) standard, visualization standard, etc.

6. CONCLUSION

Model engineering (ME) is a systematic solution for the credibility of the full model lifecycle in modeling simulation of complex system based on the research achievements on model development and management in the simulation field. It can improve credibility of the full model lifecycle and reduce the model development and management cost by providing theories, technologies, methods, standards and tools on standardized, systematic and quantified engineering management and control in the model lifecycle.

It is a long-term and complicated work to establish model engineering and should extensively depend on collaboration of the academic circle and industry circle. It is expected that this paper can launch a discussion and provide a reference for final establishment of complete model engineering curriculum and true service for model development and management of complicated simulation systems.

Finally it should be pointed out that although the idea of model engineering is discussed for simulation in this paper, the application scope of model engineering is not limited to the simulation field and is valuable for any fields which requires modeling and model management.

Acknowledgements: This work is supported by the National Nature Science Foundation of China (61374199) and 863 program (No.2013AA041302) of China.

REFERENCE

- Li B., Chai X., et al., 2002. The New Development of Contemporary Modelling and Simulation Technology. *Proceedings of the 21st Chinese Control Conference*, pp. 1-4. August 12-14, Hangzhou (Zhejiang, China). (in Chinese)
- Pace D. K., 2004. Modeling and simulation verification and validation challenges. *Johns Hopkins APL Technical Digest*, 25(2): 163-172.
- Dai R., 1997. Complex Giant System Science—A 21st century science, *Chinese Journal of Nature*, (04):187-192. (in Chinese)
- Lewin R., 1999. *Complexity: Life at the edge of chaos*. Chicago: University of Chicago Press.
- Li B., Chai X., et al., 2012. Research on High-Efficiency Simulation Technology of Complex System. *Journal of china academy of electronics and information technology*, 7(3): 221-228. (in Chinese)
- Bonabeau E., 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3): 7280-7287.
- Chopard B., Droz M., 1998. *Cellular automata modeling of physical systems*. Cambridge: Cambridge University Press.
- Davis P. K., Bigelow J., 1997. Introduction to Multi-Resolution Model (MRM) with an Example Involving Precision Fires, Enable technology for simulation science (II). *Proceeding of SPIE AeoroSense*.
- Barth T. J., Chan T., H F., Haimes R., 2002. *Multiscale and multiresolution methods: theory and applications*. New York: Springer-Verlag Berlin and Heidelberg.
- Liu X., Tang Y., Zheng L., 2008. Survey of Complex System and Complex System Simulation. *Journal of System Simulation*, 20(23): 6303-6312. (in Chinese)
- Wang Z., Lehmann A., 2007. Verification and validation of simulation models and applications: a methodological approach. *Recent Advances in Modeling and Simulation Tools for Communication Networks and Services*. New York: Springer, 227-240.
- Balci O., 1994. Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of operations research*, 53(1): 121-173.
- Yang M., Fang K., 2011. VV&A solution for simulation systems in whole lifecycle. *Computer Simulation*, (z1): 20-25. (in Chinese)
- Yang M., Zhang Z., Fang K., 2013. Aggregation method for simulation credibility index within complete weight information. *Systems Engineering and Electronics*, 35(4): 885-888.
- Liu X., Liu L., Song K., et al., 2007. Important Considerations of Building Pattern to Complicated System. *Journal of System Simulation*, 19(13): 3073-3075. (in Chinese)
- Wu Y., Qiu X., Liu B., 2006. Overall Framework of Simulation Model System Based on Web. *Ordnance Industry Automation*, 25(1): 34-35. (in Chinese)
- Fensel D., Bussler C., 2002. The web service modeling framework WSMF. *Electronic Commerce Research and Applications*, 1(2): 113-137.
- van Der Vorst J. G. A. J., Tromp S. O., Zee D. J., 2009. Simulation modeling for food supply chain redesign; integrated decision making on product quality, sustainability and logistics. *International Journal of Production Research*, 47(23): 6611-6631.
- Van der Zee D J., 2006. Modeling decision making and control in manufacturing simulation. *International journal of production economics*, 100(1): 155-167.
- Li B., Chai X., et al., 2003. Supporting environment technology for collaborative manufacturing of complex product. *Computer Integrated Manufacturing Systems*, 9(8): 691-697. (in Chinese)
- Alanen M., Porres I., 2004. A relation between context-free grammars and meta object facility metamodels. Turku: Turku Centre for Computer Science.

- Wang W., Lei Y., Zhu Y., 2007. A Metamodel-based Approach for Reusable Simulation Model Representation. *Computer Simulation*, 24(8): 93-97. (in Chinese)
- Liang Y., Zhang W., et al., 2008. A Survey of Model Reuse Methods. *Computer Simulation*, 25(8): 1-5.
- Radeski A., Parr S., 2002. Towards a simulation component model for HLA. *Proceedings of the 2002 Fall Simulation Interoperability Workshop*.
- Liu C., Wei H., et al., 2006. A new simulation language for supporting model integration and reuse. *Computer Simulation*, 23(5): 254-258. (in Chinese)
- Reichenthal S. W., 2002. The Simulation Reference Markup Language (SRML): a foundation for representing BOMs and supporting reuse. *Proceedings of Fall 2002 Simulation Interoperability Workshop*, 1: 285-290.
- Robinson S., Nance R. E., Paul R. J., et al., 2004. Simulation model reuse: definitions, benefits and obstacles. *Simulation modeling practice and theory*, 12(7): 479-494.
- Kleppe A. G., Warmer J., Bast W., et al., 2003. *The model driven architecture: practice and promise*. London: Pearson Education
- Jia L., Zhang H., 2007. Research on Service Oriented Distributed M&S Framework. *Journal of System Simulation*, 19(20): 4680-4684. (in Chinese)
- Brutzman D., Zyda M., Pullen J. M., et al., 2002. Extensible modeling and simulation framework (XMSF): Challenges for Web-based modeling and simulation. *XMSF 2002 Findings and Recommendations Report: Technical Challenges Workshop and Strategic Opportunities Symposium*, Monterey, CA
- Liu W., Liu X., Wang L., 2012. Research on Exploratory Analysis Based on Model Combination. *Computer Simulation*, 29(2): 80-84.
- Martens A., Koziolok H., Becker S., et al., 2010. Automatically improve software architecture models for performance, reliability, and cost using evolutionary algorithms. *Proceedings of the first joint WOSP/SIPEW international conference on Performance engineering*, pp. 105-116. January 28- 30. San Jose: (CA, United states).
- Dhungana D., Grünbacher P., Rabiser R., et al., 2010. Structuring the modeling space and supporting evolution in software product line engineering. *Journal of Systems and Software*, 83(7): 1108-1122.
- Balci O., 2012. A life cycle for modeling and simulation. *Simulation*, 88(7): 870-883.
- Fishwick P. A., 1989. Qualitative methodology in simulation model engineering. *Simulation*, 52(3): 95-101.
- Fishwick P. A., 1990. Toward an Integrated Approach to Simulation Model Engineering. *International Journal of General System*, 17(1): 1-19.
- AbdouniKhayari R. E., Musovic A., Lehmann A, et al., 2010. A model validation study of hierarchical and distributed web caching model. *Proceedings of the 2010 Spring Simulation Multiconference*, pp. 98. April 11-15, Orlando: (FL, United states).
- Ören T. I., 2014. *Modeling and Simulation Body of Knowledge (M&S BoK)*. University of Ottawa. Available from: <http://www.site.uottawa.ca/~oren/pubsList/MSBOK.pdf>.
- Zeigler B. P., Kim T. G., Praehofer H., 2000. *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*. 2nd Revised edition. Waltham: Academic Press Inc.
- Ören, T. I., Zeigler, B.P., 1979. Concepts for Advanced Simulation Methodologies. *Simulation*, 32(3): 69-82
- Blanchard B. S., Fabrycky W. J., Fabrycky W. J., 1990. *Systems engineering and analysis*. Englewood Cliffs, New Jersey: Prentice Hall.
- Mittal S., Risco-Martín J. L., 2013. *Model-driven systems engineering for netcentric system of systems with DEVS unified process*. Boca Raton: CRC Press Inc.
- Feron H., Lehmann A., Hofmann M., 2013. Challenges of and Criteria for Validating a Physiology Model within a TCCC Serious Game, *Winter Simulation Conference 2013*, pp. 2807-2818. December 8-11, Washington DC: (USA).
- Ludewig J., 2003. Models in software engineering—an introduction. *Software and Systems Modeling*, 2(1): 5-14.

TEACHING VARIANCE REDUCTION IN MATLAB

Jaroslav Sklenar

Department of Statistics and Operations Research
University of Malta
Msida MSD 2080, Malta

jaroslav.sklenar@um.edu.mt

ABSTRACT

The paper describes an application of a new tool for programmed discrete event simulation in Matlab that became one of the most frequently taught languages for computations in mathematics, statistics, and operations research. To teach simulation related topics we need a simple to use and a fast to learn tool for creation of simple and medium-scale simulation models. We also need a tool where the code is visible and accessible and where all functions like generation of random numbers are directly under user's control. All this is caused by the need to incorporate simulation models into various algorithms based on repetitive experiments, variance reduction techniques, and simulation-based optimization. After a short introduction of the tool we present variance reduction examples.

Keywords: discrete event simulation, queuing systems, variance reduction, matlab.

1. INTRODUCTION

It is a well-known fact that for classical simulation applications like manufacturing, transportation, or similar described typically as queuing systems, programming is used less and less. For these systems the classical GPSS view of the world as represented by interactive tools like Arena, Simul8, Witness, and similar is satisfactory and programming simulation models of such systems is often considered as a waste of time and money. Fortunately there are still areas where simulation techniques are becoming more and more important and where the classical view of entities passing through a block diagram does not work. This is true in stochastic programming, finance, stochastic integration, reinforcement learning to mention just a few. In our situation there are two more arguments in favor of programmed simulation models where the user has full control over the model. In Statistics and OR courses we have recently introduced a study-unit called "Computational Methods in Statistics and OR" for students who know only basics of programming in Matlab. For the simulation part of this unit the obvious choice was an interactive simulation tool, in our case Arena (Kelton et al. 2006). Problems started with teaching Variance Reduction Techniques (L'Ecuyer

2007). Though some of these techniques are included in Arena and similar packages, we need to show their implementation. Another area where full control over the model is required is simulation-based optimization. There are optimization tools included in interactive simulation tools like OptQuest of Arena (Bradley 2007), but there is no feasible possibility to apply other than the built-in optimization algorithm and control over its working is very limited, leaving alone techniques like for example infeasibility detected by simulation.

So to summarize, we need a simple to use and a simple to learn tool for creating discrete event simulation models in Matlab. Simulation models should take a form of a function that given model specification and run control arguments provides the required results as outputs. Such function can then be incorporated into other algorithms, in our case algorithms used in variance reduction and simulation-based optimization.

2. SIMULATION IN MATLAB

Support is needed for simulation models with continuous time and discrete behavior. Simulation of discrete time or timeless models typical in finance and stochastic programming (often called Monte Carlo simulation) is from the time control point of view relatively easy and no special support in Matlab is needed. We are aware of two Matlab based discrete event simulation tools. SimEvents (Gray 2007) is a commercial interactive tool based on Simulink of Matlab. It belongs to the category of interactive tools with limited control over the model. MatlabDEVS2 (Deatcu 2003) is a tool created primarily as a support for research and education of abstract DEVS theory, so its use is not practical in our case either. That's why it has been decided to create a new tool with simplicity and transparency being the main objectives. The tool is definitely not supposed to be used for large-scale computationally demanding simulation studies.

3. TOOL DESCRIPTION

The paper (Sklenar 2013) describes the ideas and the implementation details of the tool. Here we just summarize its functions by categories.

3.1. Time Control Functions

The tool uses the classical event-oriented paradigm based on the sequencing set made of event notices. In our case the event notices are made of the event time, unique event notice identification, user event number, and user event notice data. The sequencing set is made of four arrays whose i -th items represent the event notice i . The set is not ordered, scheduling places the new items at the end, next event to be activated is found by the function `min` of Matlab in the array of event times. Removing notices is done in usual Matlab way by storing empty values [] in the four items. This approach is certainly not very fast, but it is simple and it works satisfactorily. As the system code is not protected, we start all system identifiers by “s_”.

`s_time` is the system variable that contains the current model time.

function `id = s_schedule(t,e,d)` schedules the event e at time t with user data d . It returns the event notice identification id assigned by the engine.

function `s_cancel(id)` removes the event notice id from the sequencing set.

function `s_simulation` starts the simulation run. It is assumed that at least one event has been scheduled.

function `s_terminate` ends the simulation run by clearing the sequencing set.

In addition to the above functions, the user has to write the common user event function:

function `event(e,d,id)` that starts the event e with data d and identification id . It typically tests the event number e and activates the particular event function. In addition to user events, there may be system events with negative numbers used by application-oriented additions to the basic tool - see later.

The simulation engine is the function `s_simulation` that repeatedly removes the next event notice from the sequencing set and activates either the user function `event` or a hidden system event function. The run ends when the empty sequencing set is detected.

3.2. Statistics

With respect to time there are two types of statistics. Time dependent statistics (using Arena’s terminology *time-persistent statistics*) is based on time integrals. We call such statistical objects *accumulators*, typical example is the statistics on a queue length. The other type is statistics based only on a collection of assigned values (using Arena’s terminology *counter statistics*). We call such statistical objects *tallies*, typical example is the statistics on waiting time in a queue. The following functions are available:

function `s_tupdate(t,x)` updates the tally t by the value x . The function keeps the minimum and the maximum values, the sum of assigned values, the sum of squared assigned values and the number of updates.

function `[mean,min,max,variance,updates] = s_tallystat(t)` returns the descriptive statistics on the tally t .

function `s_aupdate(a,x)` updates accumulator a to the value x . Call to this function replaces the assignment $a = x$.

function `s_aupdateby(a,x)` updates accumulator a by the value x . Call to this function replaces the assignment $a = a+x$. Both functions keep the minimum and the maximum values, the time integral and the time integral of squared assigned values.

function `[mean,min,max,variance,lastvalue] = s_accumstat(a)` returns the descriptive statistics on the accumulator a .

All statistical activities except assignment of accumulator values start after a user-defined warming up delay, for accumulators the user has to specify the initial values, mostly zeros.

3.3. Queues

Three usual types of queues (FIFO, LIFO, priority) with possibly limited capacity are implemented. Queues are represented by data structures with various fields used for statistics. Stored items are represented by the arrays of items structures, entry times, and priorities for priority queues. The following functions are available:

function `r = s_enqueue(q,i)` inserts the item i into the queue q . The output r specifies whether the insertion was successful (1) or not (0). Treatment of rejected arrivals is application dependent. Item data structure is specified by the user, the only compulsory field is *service* - the service duration when entering a queue. For all types of queues the item is placed at the end of an array.

function `[i,wt] = s_remove(q)` removes the next item from the queue q . The outputs are the item i and its waiting time wt . For priority queue the item is found by the Matlab function `min` in the array that contains the priorities. For all queues the item is physically removed by storing the empty values in the arrays.

function `s_nowait(q)` is used for statistics to record not waiting items in the queue q .

function `[...] = s_questat(q)` returns the statistics on queue q . The outputs are: mean queue length, mean waiting time, mean waiting of those who waited and left, maximum queue length, maximum waiting time, attempted arrival rate, effective arrival rate, rate of rejections, probability that the queue is full, number of attempted arrivals, number of rejected arrivals, number of not waiting arrivals and duration of statistics collection.

3.4. Model Function Structure

Simulation models are written as functions with a fixed structure. The input and output arguments are defined by the user. These are the parts of the model that have to be included in the following order:

- System functions
- User model initialization
- System model initialization
- User model functionality

The two system parts are the same for all models and they should not be modified. The two user parts can be any mixture of commands and local functions and of course any external functions can be called, typically functions for generation of random numbers. Anyway a very simple structure is suggested.

3.5. User Model Initialization

This part first tests the validity of model input arguments and initializes user model variables, if any. This optional code is application dependent. It is supposed to test the arguments of random number generators, array sizes, integrality, etc. Next some system variables have to be initialized by the user. This is in fact a part of the model specification. The following 8 system variables must be defined, the default initialization assumes a G/G/1 queuing model:

- Types of queues array. The items are 0/1/2 for FIFO, LIFO, and priority queues respectively.
- Maximum lengths of queues array. The items are non-negative integers or Inf for unlimited queues. Zeros for pure overflow models are accepted.
- Numbers of parallel channels array. It is assumed that each queue is served by several identical parallel channels. These three arrays must have the same length, but they can be empty.
- Data structure that represents entities (customers) stored in queues. In addition to already mentioned compulsory field *service*, there must be also the field *priority* if priority queues are part of the model. Other fields are user-defined, like for example attributes representing the history of the entity, types of entities, etc.
- Data structure that represents the user part of event notices. If stations are used (see later), the compulsory fields are *station* and *channel* used by the system event *end of service*.
- Warming-up delay for statistics collection.
- Number of tallies used in the model.
- Initial values of accumulators used in the model.

User code of the model is split into two parts because the values of the above system variables are needed for the system model initialization that prepares the sequencing set, the queues, and all statistics for the simulation run.

3.6. User Model Functionality

This part of user code follows the classical event-oriented paradigm. After scheduling at least one event, typically first arrival(s), breakdowns, etc., the simulation run is started by calling the function `s_simulation` followed by the simulation run evaluation, preferably implemented by another function. This function collects the statistics and assigns values to the model outputs.

3.7. Support for Queuing Systems

The tool is general; the only requirement is the possibility to express the model behavior in terms of events. Though the definition of the above 8 system variables has to be present, the values can be all empty, so there can be no queues in the models, no standard statistics, etc. Nevertheless typical application of discrete simulation is analysis and optimization of queuing systems, which is also our case. That's why we included a simple support that makes simulation of queuing networks simple and straightforward. We associate queues with a number of parallel channels serving the entities from the queue. This makes the so called *stations* supported by the following functions:

function `r = s_arrival(q,c)` is an arrival of the customer *c* to the station (queue) *q*. The result *r* specifies the outcome (0 = lost (rejected), 1 = enqueued, 2 = served without waiting). The user has to decide what to do in case of rejection due the limited capacity.

function `s_eos(ed)` is a system function activated by the engine `s_simulation` that is transparent to the user. It is an end of service specified by the data part *ed* of the corresponding event notice. For this purpose, if stations are used in the model, there are the two compulsory fields *station* and *channel* in the event notice data. After all necessary updates and statistics collection the following function is activated.

function `customer_leaving(s,c)` is the user's activity associated with the end of service to customer *c* in station *s*. Typically there is some decision about the next service, a call to `s_arrival`, or leaving the network.

function `r = s_stop(s,c)` stops the channel *c* in station *s*. The result *r* specifies the channel status (0 = idle, 1 = busy (the operation is completed), 2 = was already suspended).

function `r = s_resume(s,c)` re-activates the channel *c* in station *s*. The result *r* specifies the channel status (0 = idle, 1 = busy, 2 = suspended). Warning is given for the first two cases.

Additional statistics provided for stations by the function `s_questat` is the mean number of working channels and their utilization. Due to possible suspensions (failures) there is no simple relationship between these two figures.

The above mentioned functionality of stations is enabled by using the system events, so far only *end of service* was implemented. System events have negative numbers, are activated by the engine and for the user they are transparent. Also note that the functions

`s_arrival` and `customer_leaving` offer a sort of process-oriented view of the system dynamics. Call to `s_arrival` starts an internal process made of possible waiting in the queue and the service that ends when the customer appears as the argument in `customer_leaving`.

4. SIMULATION MODEL FOR EXAMPLES

The paper (Sklenar 2013) describes in detail how a model of a workshop made of two breaking down machines is built. For variance reduction examples we use a relatively simple model of a G/G/1 system given by the function:

```
function [LQ,WQ,LQmax,WQmax,duration,L,W,
rho,lambda,mu] = GG1(intervals,services)
```

The inputs are vectors of intervals between arrivals and service durations of the arriving customers. The outputs are self-explaining; `duration` is the total time of collecting statistics. For this model the default initialization of the system variables is satisfactory, so the first part of user code just tests validity of the input arguments (skipped here), initializes the counter of arrivals and computes some outputs because the traffic rate must be tested to be less than 1:

```
na = length(intervals);
    % number of attempted arrivals
arr = 1; % arrival number
lambda=1/mean(intervals); % arrival rate
mu = 1/mean(services); % service rate
rho = lambda/mu; % traffic rate
```

The model is in fact made of just one station, so the second part of user code is the following:

```
s_schedule(intervals(1),1,s_edata);
    % scheduling the first arrival
s_imulation; % starting the engine
evaluation; % experiment evaluation
```

```
function event(enumber,data,id)
    if enumber == 1 % arrival
        nextarrival;
    else
        error(['Unknown event']);
    end
end
```

```
function nextarrival % customer arrival
    itm = s_item; % creating item structure
    itm.service = services(arr);
    arr = arr + 1;
    if arr<=na % are there more arrivals ?
        s_schedule(s_time + intervals(arr),1,
            s_edata); % schedule next arrival
    end
    s_arrival(1,itm);% arrival to station 1
end
```

```
function customer_leaving(qn,itm)
    % here this function is not used
end
```

```
function evaluation % experiment eval.
    [LQ,WQ,...] = s_questat(1);
```

```
W = WQ + mean(services);
L = LQ + rho;
end
```

The function `evaluation` calls `s_questat` to get the first 5 output arguments and computes the remaining two outputs in obvious way. Note that the common event function `event` takes a trivial form; the function `customer_leaving` is not used by the model, but has to be present. The function `nextarrival` creates the entity structure, stores its service duration and schedules the next arrival (if any). The actual arrival is taken care of by the system function `s_arrival`.

5. VARIANCE REDUCTION EXAMPLES

The model described in the previous chapter can be used as a simulation “engine” in other demonstration models. So for example the following function runs repeatedly a given number of arrivals with intervals uniform in $[a, b]$ and services uniform in $[c, d]$:

```
function [meanLQ,...] = ...
    sim_gg1(a,b,c,d,arrivals,runs)
da = b-a;
ds = d-c;
for i=1:runs
    display(['Run number: ' num2str(i)]);
    rands = rand(1,arrivals);
    ints = a + rands*da; % intervals
    rands = rand(1,arrivals);
    servs = c + rands*ds; % services
    [LQ(i)] = GG1(ints,servs);
end
meanLQ = mean(LQ);
...

```

From each run only the mean queue length is taken, at the end some descriptive statistics of the sample is computed and a histogram is shown. The variance reduction techniques dealt with in this paper are well known, see for example (Gentle 1998) or (Pidd 2004).

5.1. Antithetic Variables

This example shows how the use of antithetic variables decreases the sample variance. Here we want to find the mean queue length of the G/G/1 system with intervals uniform in $[2,6]$ and services uniform in $[1,5]$. After taking a sample made of 500 independent runs, 100,000 arrivals each, we obtained the following results. The mean queue length $meanLQ = 0.1538$, the standard deviation $stdLQ = 0.0023$, so the coefficient of variation is $cvLQ = 0.0148$. As the measure of sample quality we are going to use the coefficient of variation (σ/μ) that gives the extent of variability independent of the actual value of the sample mean. We are dealing only with nonnegative values. It is well known that antithetic series can be created by replacing the underlying random numbers u by $1-u$ that are both uniformly distributed in $(0,1)$. So we performed 250 pairs of runs, 100,000 arrivals each (same computation price), from each pair an average queue length was computed. The sample of 250 values thus obtained gave the same mean

queue length (at 4 decimal places) with $cvLQ = 0.0083$ which is a decrease by 44%. In our case we have a single queue system, so there is another way how to obtain two antithetic series by swapping random numbers used to generate intervals and service durations respectively. So after performing another 250 antithetic pairs of runs of the same length (again the same computation price) we again obtained the same mean queue length (at 4 places) with $cvLQ = 0.0090$ which is a decrease by 39%.

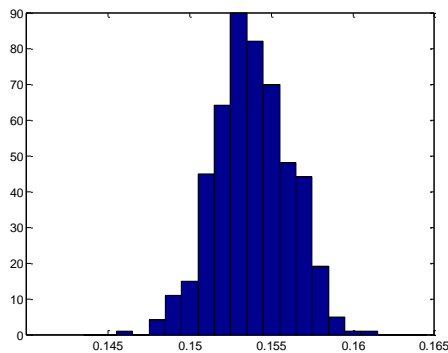


Figure 1: Sample histogram from 500 independent runs

Figures 1 and 2 show the histograms of the samples obtained by crude Monte Carlo and by using antithetic series (first case, the other one is very similar), both for the same 20 bins (0.144 : 0.001 : 0.164).

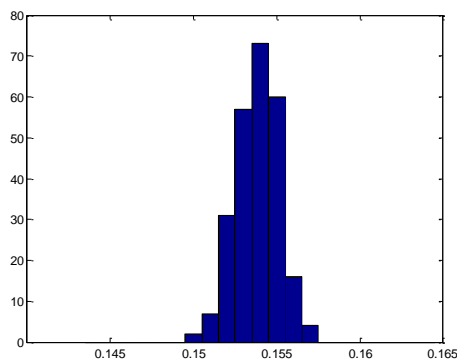


Figure 2: Sample histogram from 250 antithetic pairs of runs

Note that our very positive result is caused by the simplicity of the model. For more complicated models we would face the synchronization problem.

5.2. Common Random Numbers

This example shows how the use of common random numbers decreases the variance in situations where the purpose of the simulation is comparison of two systems. So let's compare the performance of the following two systems. The first is a G/G/1 system with intervals uniform in [5,15] and services uniform in [1,17] with traffic rate $\rho = \lambda/\mu = 0.9$. The other one is a G/M/1 system with same intervals uniform in [5,15] and exponentially distributed service with mean 8. Its traffic

rate is $\rho = \lambda/\mu = 0.8$. As the comparison criterion we use the difference of the mean queue lengths $L_{Q2} - L_{Q1}$. After taking a sample made from 500 independent pairs of runs of the two systems (500 values of the difference) made of 100,000 arrivals each, we obtained the following results. The mean difference of the two queue lengths $difLQ = 0.3047$, the standard deviation $stdifLQ = 0.0576$, so the coefficient of variation is $cvdifLQ = 0.1890$. Note the very negative effect of the service variability. Though the traffic rate of the G/M/1 system is smaller, its mean queue length was always bigger, so the use of the coefficient of variation is still justified. In order to decrease the sample variance we performed another 500 pairs of runs of the same length, but in each pair the same underlying sequences of random numbers uniform in (0,1) were used. The mean difference of the two queue lengths was $difLQ = 0.3068$, the standard deviation $stdifLQ = 0.0361$, so the coefficient of variation is $cvdifLQ = 0.1176$ which is a decrease by 38%. Figures 3 and 4 show the histograms of the samples obtained by crude Monte Carlo and by using common random numbers, both for the same 24 bins (0 : 0.025 : 0.6).

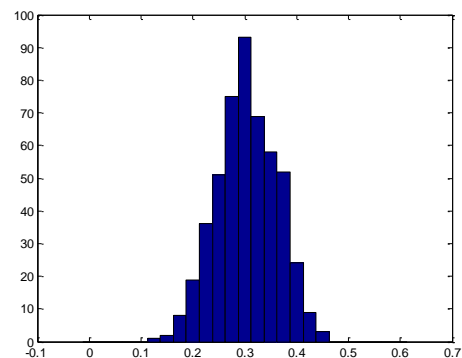


Figure 3: Sample histogram from 500 independent pairs of runs

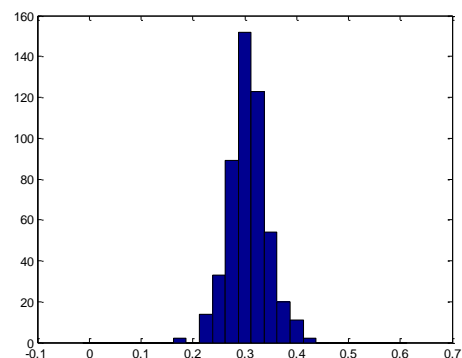


Figure 4: Sample histogram from 500 pairs of runs using common random numbers

The very positive result is again caused by the system simplicity, as synchronization in complex models is a complicated problem that may not be solved at all.

5.3. Control Variables

This example shows how the use of control variables decreases the sample variance. We want to find the mean queue length of the G/G/1 system with intervals uniform in [1,3] and services uniform in [1,2]. After taking a sample made of 500 independent runs, 50,000 arrivals each we obtained the following results. The mean queue length $meanLQ = 0.0746$, the standard deviation $stdLQ = 0.0014$, so the coefficient of variation is $cvLQ = 0.0193$. The technique of control variables is based on some knowledge about the internal working of the model. We first have to identify a variable such that its exact value is known and can be compared with the value obtained by simulation. Then we must know (at least approximately) the way how this value affects the simulation result, so we can perform a correction. In our example we select the control variable to be the mean service time T_S whose exact value is 1.5. The mean service time obtained by simulation is then $T_S + dT_S$. For simplicity we assume a linear relationship between the service time and the mean queue length, at least close to the actual values. So the change of the mean queue length caused by imperfect generation of service times is $c \times dT_S$. After simulation we perform the correction by subtracting this value from the mean queue length obtained by simulation. To find the value of the positive constant c , we may use a (simplified) analytical model or very long simulation runs. We performed 5 long simulation runs made of 10^6 arrivals each for the exact mean service 1.5 and another such 5 runs for the perturbed value 1.51 with services uniform in [1,2.02]. The difference of queue lengths (the two sample means) was then divided by 0.01 giving the value $c = 0.455643$. Using this constant we performed 500 corrected runs (50,000 arrivals each) with these results. The mean queue length $meanLQ = 0.0744$, the standard deviation $stdLQ = 0.0012$, so the coefficient of variation is $cvLQ = 0.0171$ which is a decrease by 11%.

Here the improvement is modest. Changing the perturbed value and increasing the numbers of long runs did not bring any considerable improvement. So we tried to use two control variables, the mean service time and the mean interval with the second one having obviously a negative effect on the queue length. So the correction subtracted from the simulated mean queue length was in this case $c_1 \times dT_S - c_2 \times dI_S$ where dI_S is the deviation between the exact mean interval 2 and the same value obtained by simulation. The value of c_1 is the same as the one used in the single control variable case, c_2 was obtained from two samples of long runs in similar way. Its value was $c_2 = 0.248055$. In this case the decrease of the coefficient of variation was 30%.

6. CONCLUSION

After using the Matlab based simulation tool to create various educational examples, see also two examples in (Sklenar 2013), we believe that the objective has been met. The user parts of the simulation models are very short, lucid, and all very similar. The differences are of cause given by the specific behavior of particular

models. Due to the choice of the simple event-oriented paradigm, the code resembles very much programs in simulation languages of this type, for example Simscript II. Similar to these languages the user is relieved from “background” functionality like time control, queues management, statistics, etc. and can concentrate on the model behavior as such. All this is achieved without a need to learn a special-purpose simulation language. Intermediate Matlab programming skills are enough to create simulation models of medium size and complexity. As already mentioned, the whole model code is a single function with local functions, so though not recommended, it is possible to modify the system functions in any way. Having all the functionality under control and directly visible was in fact the main objective. Single function simulation models can thus be incorporated in programs for various repetitive experiments, optimization algorithms, etc. Of course there is a lot of room for further improvements. Both sequencing set and queues are implemented in a very simple inefficient way. Also security might become an issue; so far there are no restrictions at all. Though we believe that these drawbacks are not serious because the tool is not supposed to be applied in large simulation studies, its further development will address these issues.

REFERENCES

- Bradley, A., 2007. *OptQuest for Arena - user's guide*. Rockwell Automation Technologies, Inc.
- Deatcu, C., 2003. An object-oriented solution to ARGESIM comparison C6 - Emergency Department with MATLAB-DEVS2, *Simulation News Europe*, 38/39, 56.
- Gentle, J.E., 1998. *Random Number Generation and Monte Carlo Methods*. Springer-Verlag.
- Gray, M.A., 2007. Discrete event simulation: a review of SimEvents. *Computing in Science and Engineering*, 9(6), 62-66.
- Kelton, W.D., Sadowski, R.P., Sadowski, D.A., 2006. *Simulation with Arena*. McGraw-Hill.
- L'Ecuyer, P., 2007. Variance reduction greatest hits. *Proceedings of European Simulation and Modelling Conference ESM'2007*, 5-12. October 22-24, Malta.
- Pidd, M., 2004. *Computer Simulation in Management Science*. 5th ed. John Wiley & Sons.
- Sklenar, J., 2013. Tool for Discrete Event Simulation in MATLAB. *Proceedings of the 27th European Conference on Modelling and Simulation ECMS-2013*, 110-116, May 27-30, Alesund.
- The MathWorks, Inc. 2005. *SimEvents user's guide*.
- The MathWorks, Inc. 2005. *Simulink: a program for simulating dynamic systems, user guide*.

MODELING AND PERFORMANCE ANALYSIS OF AMMONIA-WATER ABSORPTION REFRIGERATION SYSTEM FOR OCEAN-GOING FISHING VESSELS

Li Ren^(a), Zijian Guo^(b), Hongxia Li^(c)

^(a,c) School of Mechanical and Power Engineering, Dalian Ocean University

^(b) School of Civil and Hydraulic Engineering, Dalian University of Technology

^(a) renli@dlou.edu.cn, ^(b) zjguo@dlut.edu.cn, ^(c) li_hongxia2013@163.com

ABSTRACT

In this study, the modeling and performance analysis of ammonia-water absorption refrigeration system for ocean-going fishing vessels has been investigated. According to ammonia water solution saturated vapor and saturated liquid state equation, the thermodynamic properties of the refrigerant is studied. Based on the basic principles of the ammonia-water absorption refrigeration cycle, a mathematical model of the ammonia-water absorption refrigeration cycle is established so that the thermodynamic parameters of each state point and the heat load of each device can be calculated and then the coefficient of performance (COP), exergetic coefficient of performance (ECOP) for the system can also be calculated. Considering heat source temperature, cooling water temperature and refrigerating temperature, the effects of them on COP and ECOP are discussed by some simulation examples. After these calculations, some graphics indicating the change of the variables with the temperature parameters.

Keywords: ammonia-water absorption refrigeration system, ocean-going fishing vessels, modeling, exergy analysis

1. INTRODUCTION

For marine diesel engine, there is still about 50% of the fuel input energy not being put to productive use. The utilization of surplus heat rejected by the main propulsion engine provides substantial fuel savings, involving the reduction of pollutants. For the ocean-going fishing vessels, the catches' frozen preservation is very important, which related directly to the quality and economic value of the catches. Therefore, ammonia-water absorption refrigeration system has been paid an increased attention in which the surplus waste heat from the main engine is used to separate a refrigerant vapor from a binary solution. Sun[1] presented the thermodynamic design data and optimum design maps for absorption refrigeration systems. Ouadha[2] examined the feasibility of using waste heat from marine diesel engines to drive an ammonia-water absorption refrigeration system through a thermodynamic analysis. Sencan[3] analyzed

thermodynamic evaluation of ammonia-water absorption refrigeration system using linear regression and M5' rules models within data mining process and artificial neural network model. Sozen [4] investigated the effect of heat exchangers on the system performance in an aqua-ammonia absorption refrigeration system. Wang [5] investigated of the feasibility of applying aqua-ammonia absorption refrigeration device to fishing vessels. Ni[6] established the thermodynamic mathematical model and analyzed the theoretical operation characteristics for fishing boat ammonia-water absorption refrigeration system. These researches are concerned with the thermodynamic analysis of ammonia-water absorption refrigeration system. Here, following the above mentioned studies, the paper investigates the modeling and performance analysis of ammonia-water absorption refrigeration system for the ocean-going fishing vessels.

2. THERMODYNAMIC ANALYSIS OF AMMONIA-WATER ABSORPTION REFRIGERATION SYSTEM

The ammonia-water absorption refrigeration system consists mainly of a generator, a condenser, an evaporator, an absorber, a solution heat exchanger and a circulation pump. The schematic illustration of the ammonia-water ARS is presented in Fig.1. The generator uses waste heat from the marine Diesel engine to separate ammonia vapour from the concentrated ammonia solution. When ammonia is evaporated off the generator, it contains some water vapour. To remove as much water vapour as possible, the vapour driven off at the generator first flows counter currently to the incoming solution in the rectification column. In the condenser, high pressure ammonia vapours are cooled and condensed to liquid state. Liquid ammonia leaves the condenser and flows to the evaporator through an expansion valve. The refrigerant then enters the evaporator, where it receives heat from the cold source. Then, ammonia vapour enters the absorber, where a weak solution of water and low concentration ammonia absorbs the refrigerant and, at the same time, releases heat to the neighbourhood. The ammonia-water solution

flows back to the generator through a circulation pump to undergo a new cycle.

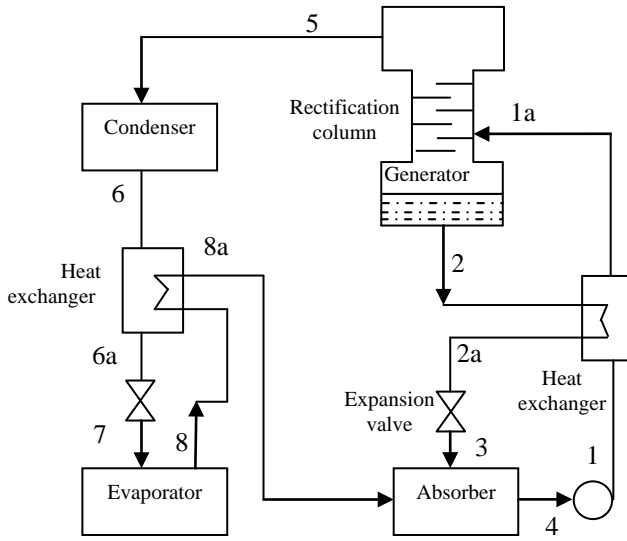


Figure 1: Schematic Diagram of an Ammonia-water Absorption Refrigeration System

2.1. Thermal-physical Properties of Ammonia-water Mixtures

For modeling and simulating an ammonia-water absorption refrigeration system, the thermal-physical properties of the ammonia-water have to be known as analytic function. Based on the state equation proposed by Schulz[7], using the thermodynamic relations, the expressions of thermodynamic parameters for temperature T , pressure P , enthalpy H , mass fraction x and y can be derived by mathematical transformation[8].

The enthalpy H_R^{gM} in vapor phase of ammonia-water mixture can be expressed:

$$\begin{aligned} H_R^{gM} &= (1-y)H_R^{gH_2O} + yH_R^{gNH_3} \\ &= (1-y)[G_R^{gH_2O} - \theta(\partial G_R^{gH_2O} / \partial \theta)_\pi] + \\ & \quad y[G_R^{gNH_3} - \theta(\partial G_R^{gNH_3} / \partial \theta)_\pi] \end{aligned} \quad (1)$$

Where, $G_R = \frac{G_m}{R_m T_b}$ is contrast-state Gibbs free energy

function, $\theta = \frac{T}{T_b}$ contrast-state temperature, $\pi = \frac{P}{P_b}$

contrast-state pressure. Superscript g is vapor phase, M mixture, e surplus. Subscript R is the contrast value.

The enthalpy H_R^{fM} in liquid phase of ammonia-water mixture can be defined as:

$$\begin{aligned} H_R^{fM} &= (1-x)H_R^{fH_2O} + xH_R^{fNH_3} + H_R^e \\ &= (1-x)[G_R^{fH_2O} - \theta(\partial G_R^{fH_2O} / \partial \theta)_\pi] + \\ & \quad x[G_R^{fNH_3} - \theta(\partial G_R^{fNH_3} / \partial \theta)_\pi] + G_R^e - \theta(\partial G_R^e / \partial \theta)_{\pi,x} \end{aligned} \quad (2)$$

Where, superscript f is liquid phase.

The relationship of T , P and x for vapor-liquid equilibrium is written as follows:

$$\begin{aligned} &\exp\left\{\frac{1}{\theta}\left[G_R^{fH_2O} + \theta \ln(1-x) + G_R^e - x(\partial G_R^e / \partial \theta)_{\pi,x} - G_R^{gH_2O}\right]\right\} + \\ &\exp\left\{\frac{1}{\theta}\left[G_R^{fNH_3} + \theta \ln x + G_R^e + (1-x)(\partial G_R^e / \partial \theta)_{\pi,x} - G_R^{gNH_3}\right]\right\} = 1 \end{aligned} \quad (3)$$

The relationship of T , P and y for vapor-liquid equilibrium is written as follows:

$$y = \exp\left\{\frac{1}{\theta}\left[G_R^{fNH_3} + \theta \ln x + G_R^e + (1-x)(\partial G_R^e / \partial \theta)_{\pi,x} - G_R^{gNH_3}\right]\right\} \quad (4)$$

The liquid equilibrium equation of pure ammonia is:

$$G_R^{fNH_3} - G_R^{gNH_3} = 1 \quad (5)$$

2.2. State Parameters Calculation of Ammonia-water Absorption Refrigeration System

Here, assuming the heat source temperature T_h , the cooling water temperature T_w and the refrigerating temperature T_e are known.

- point 6

The condensing temperature T_6 can be expressed:

$$T_6 = T_w + \Delta T_w + \Delta T_6 \quad (6)$$

Where, ΔT_w is the temperature rise of cooling water in the condenser, ΔT_6 the temperature difference between the hot side of the condenser.

The ammonia concentration $\xi_6 = 0.998$ and pressure P_6 , enthalpy h_6 can be calculated by Eq.(3) and Eq.(2).

- point 2

The termination temperature of the generator T_2 can be expressed:

$$T_2 = T_h - \Delta T_2 \quad (7)$$

Where, ΔT_2 is the temperature difference between the hot side of the generator.

The pressure P_2 is given by the following expression:

$$P_2 = P_6 + \Delta P_2 \quad (8)$$

Where, ΔP_2 is the pipeline pressure drop from generator to condenser.

Then, the ammonia concentration ξ_2 and enthalpy h_2 can be calculated by Eq.(3) and Eq.(2).

- point 8

The termination temperature of the evaporator T_8 can be expressed:

$$T_8 = T_c - \Delta T_8 \quad (9)$$

Where, ΔT_8 is the temperature difference of heat transfer in the evaporator.

The pressure P_8 can be calculated by Eq.(5) based on the initial temperature of the evaporator T_8^* and temperature T_8^* is defined as:

$$T_8^* = T_8 - \Delta T_8^* \quad (10)$$

Where, ΔT_8^* is the temperature rise of the evaporator.

The enthalpy h_8 can be calculated by the following expression:

$$h_8 = h'_8 + (h''_8 - h'_8) \frac{0.998 - \xi'_8}{1 - \xi'_8} \quad (11)$$

Where, h'_8 is the termination enthalpy in liquid phase of the evaporator, h''_8 the termination enthalpy in vapour phase of the evaporator, ξ'_8 the termination ammonia concentration in liquid phase of evaporator.

- point 4

The termination temperature of the absorber T_4 can be expressed:

$$T_4 = T_w + \Delta T_4 \quad (12)$$

Where, ΔT_4 is the the temperature difference between the cold side of the absorber.

The pressure P_4 can be expressed:

$$P_4 = P_8 - \Delta P_4 \quad (13)$$

Where, ΔP_4 is the pipeline pressure drop from evaporator to absorber.

Then, the ammonia concentration ξ_4 and enthalpy h_4 can be calculated by Eq.(3) and Eq.(2).

- point 3

The outlet temperature of the solution heat exchanger T_3 can be expressed:

$$T_3 = T_4 + \Delta T_3 \quad (14)$$

Where, ΔT_3 is the the temperature difference between the cold side of the solution heat exchanger.

Based on the temperature T_3 , ammonia concentration $\xi_3 (= \xi_1)$, pressure P_3 and enthalpy h_3 can be calculated by Eq.(3) and Eq.(2).

- point 1

Based on the thermal equilibrium of solution heat exchanger, the outlet enthalpy of concentrated ammonia solution h_{1a} can be calculated:

$$h_{1a} = h_4 + (h_2 - h_3) \frac{f - 1}{f} 0.95 \quad (15)$$

Where, f is the solution circulation ratio.

The outlet temperature of concentrated ammonia solution T_{1a} can be calculated:

$$T_{1a} = T_4 + (T_2 - T_3) \frac{f - 1}{f} \quad (16)$$

- point 5

The top ammonia concentration of rectification column ξ_5 should arrive 0.998. Based on pressure $P_5 (= P_2)$ and ammonia concentration ξ_5 , the temperature T_5 and enthalpy h_5 can be calculated by Eq.(3) and Eq.(1).

- point 8a, 6a

Based on the thermal equilibrium of supercooler, the outlet enthalpy in vapour phase of the supercooler h_{8a} can be calculated:

$$h_{8a} = h_8 + (h_8 - h_6) \nu \quad (17)$$

Where, ν is the load factor of supercooler.

the outlet temperature in vapour phase of the supercooler T_{8a} can be calculated:

$$T_{8a} = \frac{(h_{8a} - h_8)}{c''} + T_8 \quad (18)$$

Where, c'' is the average specific heat of ammonia between T_8 and T_{8a} .

The outlet enthalpy h_{6a} , temperature T_{6a} in liquid phase of the supercooler can be expressed:

$$\begin{aligned} h_{6a} &= h_6 + (h_{8a} - h_8) \\ T_{6a} &= T_6 - \frac{(h_{8a} - h_8)}{c'} \end{aligned} \quad (19)$$

Where, c' is the average specific heat of ammonia solution between T_6 and T_{6a} .

3. MODELING OF AMMONIA-WATER ABSORPTION REFRIGERATION SYSTEM

In the first law analysis, The cycle performance of ammonia-water absorption refrigeration system is measured by the coefficient of performance (COP). The larger COP is, the less energy will be consumed by the power unit in the same amount of refrigerating capacity.

In the second law analysis, the exergy is calculated for a system instead of energy and the difference between the energy qualities is also taken into account. The effectiveness of the ammonia-water absorption refrigeration system depends on the exergetic coefficient of performance (ECOP).

According to the mass conservation, the first and second laws of thermodynamics, the energy balance equations and exergy loss equations for some components of the ammonia-water absorption refrigeration system are expressed as follows:

Condenser:

$$\begin{aligned} q_C &= h_5 - h_6 \\ \Delta e_C &= e_5 - e_6 \end{aligned} \quad (20)$$

Evaporator:

$$\begin{aligned} q_E &= h_{8a} - h_6 \\ \Delta e_E &= e_6 - e_{8a} + q_E(T_0/T_e - 1) \end{aligned} \quad (21)$$

Absorber:

$$\begin{aligned} q_A &= h_{8a} - h_3 + f(h_3 - h_4) \\ \Delta e_A &= e_{8a} - e_3 + f(e_3 - e_4) \end{aligned} \quad (22)$$

Generator :

$$\begin{aligned} q_G &= h_5 - h_2 + f(h_2 - h_{1a}) + q_R \\ \Delta e_G &= f(e_{1a} - e_2) + e_2 - e_e + q_G(1 - T_0/T_h) \end{aligned} \quad (23)$$

Where, T_0 is the reference temperature. q is the unit thermal load, e the unit exergy. Subscripts A, C, E and G are absorber, condenser, evaporator and generator, respectively. q_R the unit thermal load of reflux condenser.

Then, ignoring the work required for the pumping process, COP is defined as:

$$COP = \frac{q_E}{q_G} = \frac{h_{8a} - h_6}{h_5 - h_2 + f(h_2 - h_{1a}) + q_R} \quad (24)$$

ECOP is defined as:

$$ECOP = \frac{e_E}{e_G} = COP \frac{\frac{T_0}{T_e} - 1}{1 - \frac{T_0}{T_g}} \quad (25)$$

4. PERFORMANCE ANALYSIS OF AMMONIA-WATER ABSORPTION REFRIGERATION SYSTEM

Based on the modeling of ammonia-water absorption refrigeration system, the effects of heat source temperature T_h , cooling water temperature T_w and refrigerating temperature T_e on the performance of it are discussed. The relevant temperature and pressure parameters are listed in Tab.1.

Table 1: Relevant temperature and pressure parameters

ΔT_w ($^{\circ}C$)	5.00	ΔT_4 ($^{\circ}C$)	7.00
ΔT_6 ($^{\circ}C$)	3.00	ΔT_3 ($^{\circ}C$)	8.00
ΔT_2 ($^{\circ}C$)	13.00	ΔP_2 (MPa)	0.01
ΔT_8 ($^{\circ}C$)	5.00	ΔP_4 (MPa)	0.03

4.1. Effect of Heat Source Temperature

The heat source temperature T_h are taken as 125-170 $^{\circ}C$. The cooling water temperature T_w is taken as 30 $^{\circ}C$ and the refrigerating temperature T_e is 5 $^{\circ}C$. The variation of circulation ratio f with different heat source temperatures is shown in Fig.2. The circulation ratio f decreases with T_h increasing.

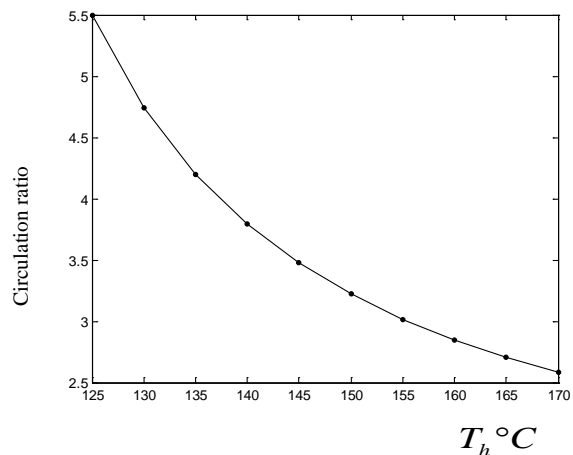


Figure 2: Variation of the circulation ratio with different heat source temperatures

The variations of COP and ECOP with different heat source temperatures are shown in Fig.3 and Fig.4. With the heat source temperature increasing, the COP value

increases firstly and then decrease. The high COP value is obtained about at 145°C . However, compared with COP, ECOP value decrease and the variation of ECOP value is different greatly.

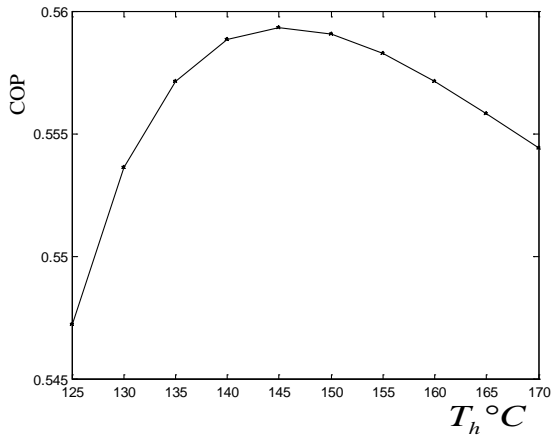


Figure 3: Variation of COP with different heat source temperatures

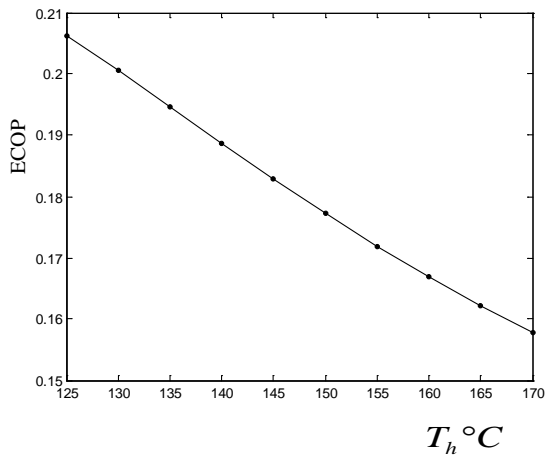


Figure 4: Variation of ECOP with different heat source temperatures

4.2. Effect of Cooling Water Temperature

The heat source temperature T_h is taken as 150°C and the refrigerating temperature T_e is 5°C . The cooling water temperatures T_w are taken as $24-42^{\circ}\text{C}$. The variation of circulation ratio f with different cooling water temperatures is shown in Fig.5. The circulation ratio f increases with the cooling water temperature T_w increasing.

The variations of COP and ECOP with different cooling water temperatures are shown in Fig.6 and Fig.7. With the cooling water temperature increasing, the COP value increases and the ECOP value decreases.

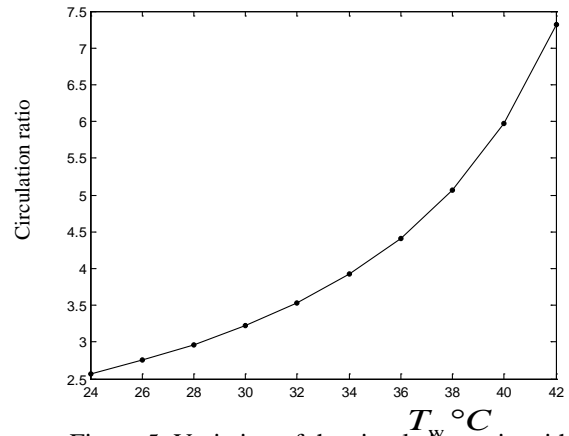


Figure 5: Variation of the circulation ratio with different cooling water temperatures

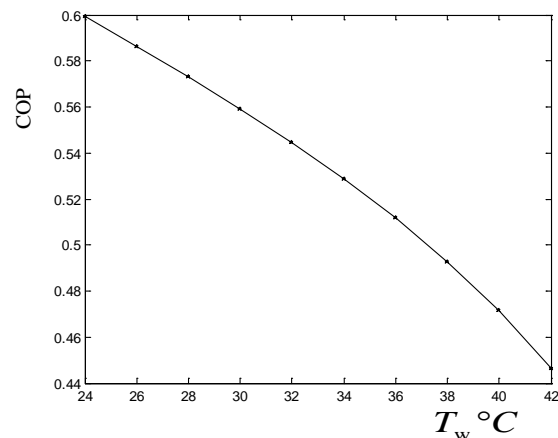


Figure 6: Variation of COP with different cooling water temperatures

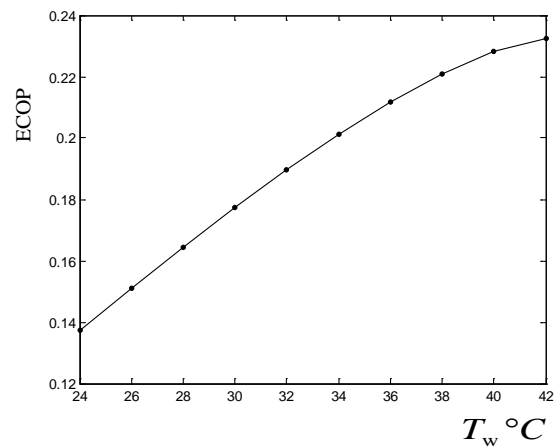


Figure 7: Variation of ECOP with different cooling water temperatures

4.3. Effect of Refrigerating Temperature

The heat source temperature T_h is taken as 150°C and the cooling water temperature T_w is 30°C . The refrigerating temperatures T_e are taken as $-15-15^{\circ}\text{C}$.

The variation of circulation ratio f with different refrigerating temperatures is shown in Fig.8. The circulation ratio f decreases with the refrigerating temperature T_e increasing.

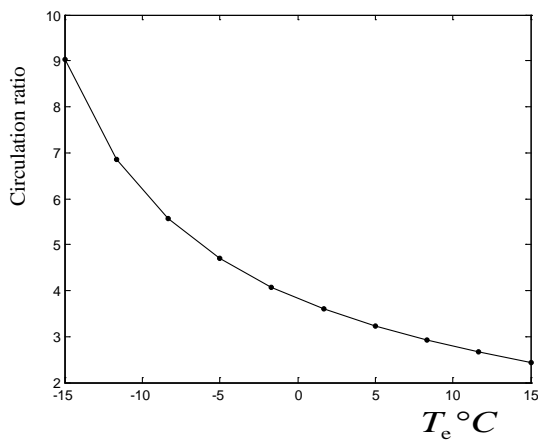


Figure 8: Variation of the circulation ratio with different refrigerating temperatures

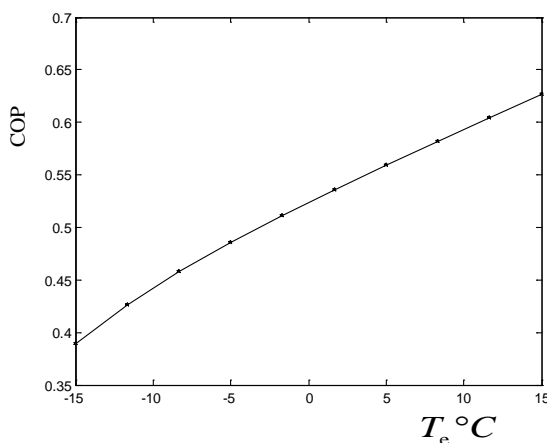


Figure 9: Variation of COP with different refrigerating temperatures

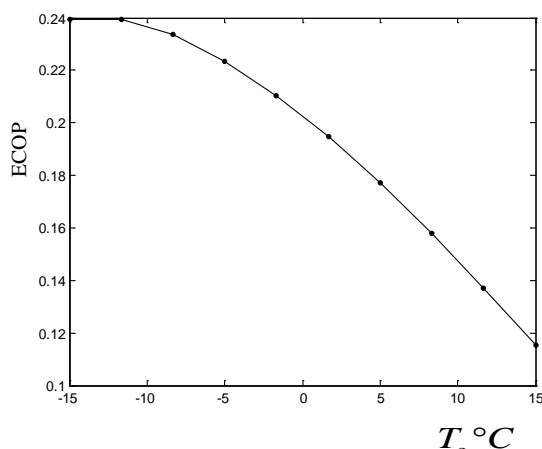


Figure 10: Variation of ECOP with different refrigerating temperatures

The variations of COP and ECOP with different cooling water temperatures are shown in Fig.9 and Fig.10. With the refrigerating temperature increasing, the COP value increases, however, the ECOP value decreases.

5. CONCLUSIONS

An modeling and performance analysis of ammonia-water absorption refrigeration system for ocean-going fishing vessels has been investigated in this work. Based on the Schulz state equation, the thermodynamic parameters analysis is carried on. Considering heat source temperature, cooling water temperature and refrigerating temperature, a mathematics model is constructed in which COP and ECOP are taken as the evaluation parameter for ammonia-water absorption refrigeration system. The simulation results showed that the effects of heat source temperature, cooling water temperature and refrigerating temperature on COP and ECOP are different greatly, for COP is built on the first law and ECOP is on the second law of thermodynamics. Thus can be used for the optimization design of ammonia-water absorption refrigeration system.

ACKNOWLEDGMENTS

This research was financially supported by the National Natural Science Foundation of China under grant No. 51109023.

REFERENCES

- Sun, D. W., 1997. Thermodynamic design data and optimum design maps for absorption refrigeration systems. *Applied Thermal Engineering*, 17, 211-221.
- Ouahda, A., El-Gotni, Y., 2013. Integration of an ammonia-water absorption refrigeration system with a marine diesel engine: A thermodynamic study. *Procedia Computer Science*, 19, 754-761.
- Sencan, A., 2006. Artificial intelligent methods for thermodynamic evaluation of ammonia-water refrigeration systems. *Energy Conversion and Management*, 47, 3319-3332.
- Sozen, A., 2001. Effect of heat exchangers on performance of absorption refrigeration systems. *Energy Conversion and Management*, Vol. 42, 1699-1716.
- Wang, W. W., Pan, X. X., Shen, B., 2012. Analysis to the feasibility of applying aqua-ammonia absorption refrigeration device to fishing vessels. *Energy Conservation Technology*, 30, 397-399.
- Ni, J., Gu, J. H., Shen, J., 2012. Simulation and theoretical operation characteristics analysis of ammonia-water absorption refrigeration system for fishing boat. *Fishery Modernization*, 39, 54-58.
- Schultz, S. C. G., 1971. Equations of state for the system ammonia water for use with computers. *Progr.*

Refrigeration Science and Technology, Proc. 13th Int. Congr. Refrigeration, 2:231-236.

Ma, Q., 2009. Study on Transportation of Low-Grade Thermal Energy over Long Distance by Ammonia-water Absorption Technology. PhD Thesis, Shanghai Jiao Tong University. (in Chinese)

AUTHORS BIOGRAPHY

In the post-graduate stage, the applicant participated in several scientific researches supported by national and provincial funds. Based on traditional topology optimization and element free Galerkin method (EFGM), the topology optimization design for multi-stiff structure, multi-compliant mechanism and heterogeneous heat conduction structure with specific heat characteristics, are investigated, respectively. On this basis, the concurrent design method for structure with material was proposed. And the digital design framework of Ideal Functional Material Components (IFMC) was constructed. Involved in these projects, the applicant has accumulated rich experience in various aspects, such as numerical modeling and simulation, structural topology optimization, optimization algorithm design and so on.

Since starting work at School of Mechanical Engineering, Dalian Ocean University in April 2008, the applicant has launched a series of exploratory researches with the support of DOU's Startup Fund for Doctor and the National Natural Science Foundation of China under grant No. 51109023, which mainly include numerical simulation for fishing vessels power plant, fishing vessels' structure and material optimization, fishing vessels' energy structure optimization, and processing complex data.

TABLE TOP FOR SCHEDULING AND CONTROLLING CONSTRUCTIONS COURSE

Caroline Cavalheiro^(a), Ailton Soares Freire^(b), Antonio Edesio Jungles^(c)

^(a)Universidade Federal de Santa Catarina

^(b)Instituto Federal de Educação, Ciência e Tecnologia do Piauí

^(c)Universidade Federal de Santa Catarina

^(a)carolenecavalheiro@gmail.com, ^(b)ailton99@gmail.com, ^(c)ajungles@ceped-ufsc.br

ABSTRACT

This paper brings the report of using the table simulation method as a tool to reach required competences to Civil Engineering students. To obtain results in this project it was used questionnaires, pictures, movies and individual perceptions of the researchers. The theories used are directly linked to teaching by competence, the methodology of table simulation games and techniques of Scheduling and Controlling Constructions. The reached results are: a) identifying and development of competences of civil engineer specialized in Scheduling and Controlling Constructions; and b) ascertainment that learning is more effective when applied setting exercises and table simulation techniques. In the classes where this methodology was applied it was verified that the students were able to develop the required competences to the proposed course and their reaction showed their preference and approval for that kind of classes. In addition, when compared, it reaches levels of attendance and knowledge higher than lecture classes.

Keywords: table simulation, competence, construction scheduling.

1. INTRODUCTION

Many projects show the different way of using simulation in the educational environment, from finding solutions to particular problems to creating an environment that allows qualification for the activity (Larréché 1987; Wolfe 1993; Curland and Fawcett 2001; Khaled 2001; Santos 2002; Van Der Zee and Slomp 2009; Pasin and Giroux 2011).

The simulation here understood as a technique that allows studying, previewing and understanding the behavior of a system in a simplified way is the “permanently representation of reality” (Ellington et al 1981).

According to Daneshfar (2006); Salas, Wildman, and Piccolo (2009); Mahboubian (2010), learning from simulation takes the learner to construct knowledge in a safe environment where it is possible to make mistakes,

understanding it and reaching correct results with no damage in the equipment or changing productive flow of the activity to be learned.

Learning through experimentation and simulation balanced with the application of theoretical classes results in a solid basis for students who needs knowledge for taking decisions in real scenarios (Latorre and Jiménez 2012).

Developing competence in courses of under graduation programs based on teaching methods that goes beyond the traditional teaching method may be the solution to achieve better results in the students’ education.

According to Magee (2006) an approach to learning through the use of simulation and games in adults education has been working quite well in training paradigms, where the aim of the educational experience is the creation of a homogeneous labor force with consistent and predictable levels of competence.

The aim of this project is to prove the acquisition of the required competences to the students of Civil Engineering’s “Scheduling and Controlling Constructions” course from the use of table simulation techniques (serious games) as a tool to add value to the methodology of teaching by competence.

2. SIMULATION IN UNDERGRADUATE EDUCATION

According to Larréché (1987); Thavikulwat (2004); Faria and Wellington (2004), it was in the 1950s that the first simulators were used in manager, economics and business area.

In the end of 1960s, in the USA, every business school was using some kind of games and simulation in their teaching programs.

In Brazil, the use in large scale began only the 1980s (Pantaleão, Oliveira, and Antunes 2003) based on the use of business games in undergraduate programs of business area.

Thereafter emerged games applications and simulations in undergraduate programs in other knowledge field, and some application in Engineering teaching may be emphasized, as the papers of Paxton (2003); Ross, Victor, and Stlather (2004); Costa, Bogado, and Jungles (2006).

In other knowledge areas, aligning learning and competence, from the focus of simulation and educational games, we can find some studies that bring the students' learner training in undergraduate programs.

In Thavikulwat, Chang, and Sanford (2013), it is shown to a group of 180 university students the analyses and the comprehension of the characteristics of merging and purchasing companies through a business game, simulating this experience in the academic 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.

Reviewing the literature of the current status of simulation in adults' education, there is Magee's studies (2006) that go further this review reporting the main problems of the implementation of this model, the main trends of research, and the main educational issues by adopting this model of education.

The paper of Siddiqui, Khan, and Akhtar (2008); Salas, Wildman, and Piccolo (2009) discuss the benefits and effectiveness that the application of games and simulation in the context of academic education, though many cases the adoption of this teaching strategies may have a high cost, which is the main challenge to be overcome (Pasin and Giroux 2011).

Thus, the use of simulation from table games, a technique with low cost, was the one chosen to mediate the acquisition of the required competences to turn the Civil Engineer competent in the area of scheduling and controlling constructions.

The expected result was to verify how was the progress of competences development comparing to groups that were exposed to the traditional teaching and the competences methodology (Knowledge – Skills – Attitude).

3. METHODOLOGY

It is an applied research, in the view of the kind of research, once it "aims to obtain knowledge to the practical application headed to solving specific problems" (Silva and Menezes 2005) which is the qualifications of "scheduling and controlling constructions" students in Civil Engineering.

The moment built, described and analyzed in this paper is part of the Master thesis of one of the authors of this paper, exemplified below:

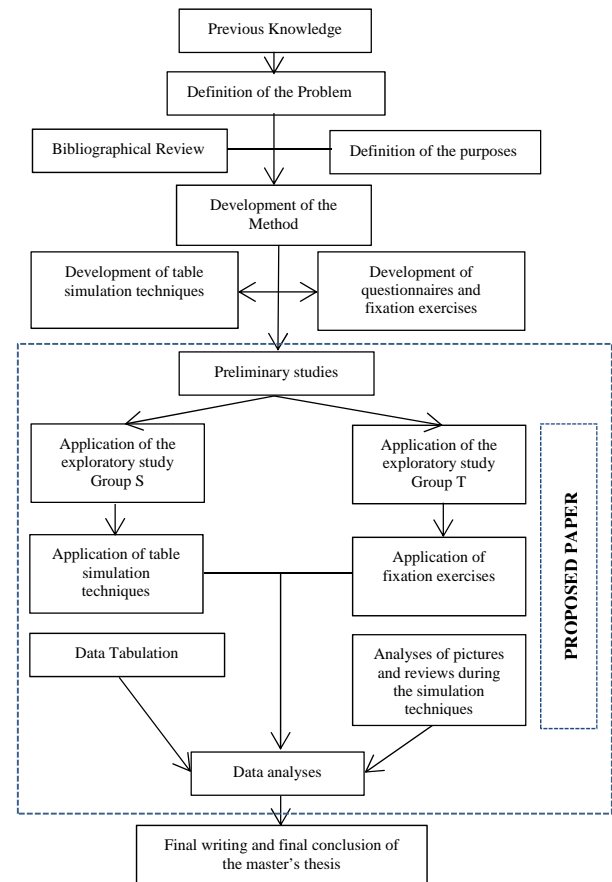


Figure 1 – Flowchart of master's thesis

For the moment of knowledge construction in classroom, it was chosen the strategy of Constructive Research, which is for Kasanen, Lukka and Siitonen (1996) a research procedure to produce "innovative constructions", to solve real problems and to supply theoretical framework to the course that it is applied.

The constructive research has as characteristics (Kasanen, Lukka, and Siitonen 1996): a) focus in real world problems to be solved functionally; b) production of an innovative construction aiming to solve the problem proposed; c) implementation to the developed construction, testing the practical applicability; d) cooperation among researchers to allow learning based on experiments; e) a powerful connection to a prior theoretical knowledge; and f) a reflection about empirical evidence based on a theory. Thus the main characteristic of constructive research is that the "constructions" are created and developed, where it is not possible to it be discovered.

Based in the Constructive Research, follows the structure of the work done which led the results presented in this paper. Figure 2 brings the logical system of this research.

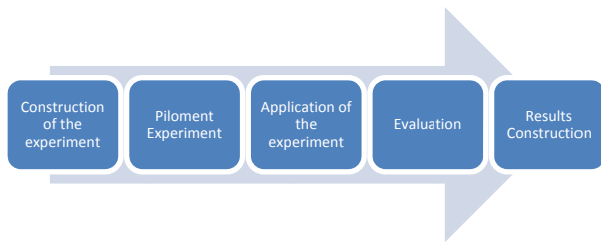


Figure 2: Logical scheme of research

The construction of this experiment came from the evaluation of the profile of the students enrolled at “scheduling and controlling constructions” course, with the sample of 64 students.

The experiment consisted in the application of the traditional teaching methodology and the methodology by competences (Knowledge – Skills – Attitudes) for both groups of Scheduling and Controlling Constructions course of University of Santa Catarina.

The experimentation was applied in two academic semesters (9 months) in two groups, with the amount of 64 students, nominated as traditional group (Group T) and group with simulation (Group S).

Group T: thirty students under lecture classes;

Group S: thirty-four students under KSA classes with simulator applications.

The application of the traditional methodology consisted in lecture classes with application of theoretical classes and practice of exercises, closing with the application of two writing tests.

The KSA methodology was built from the required competences for the academic student of Civil Engineering of the proposed course, using for it the application of table simulation.

At this point was defined the competences – KSA (knowledge, skills, abilities) (Durand 2000; Nisembaum 2000; Le Boterf 2002) required to become a Civil Engineer competent in scheduling and controlling constructions’ field, achieving the following Table 01.

Table 1: Required competences for Scheduling and Controlling Constructions course

Knowledge	Skills	Attitude
<ul style="list-style-type: none"> - PERT/CPM Method Development of work plain; Sequencing of implementation activities; Scheduling resources. - Line of balance Repetitive tasks; Usage of Buffer in scheduling resources; Teamwork allocation. - Last Planner: Specific and long-term vision of planning; Long, medium and short term planning. 	<ul style="list-style-type: none"> - Know how to plan; - Work as a team; - Communication; - Organization; - To delegate activities; - To solve problems; - To motivate the team; - To manage and control people; - Flexibility. 	<ul style="list-style-type: none"> - Own-initiative; - Leadership; - Dynamism; - To take decisions; - Bargaining power.

Source: Adjusted from Odusami (2002); Moore, Cheng, and Dainty (2005); Cripe et al (2007); Dudon and Marchigto (2007); Navarro (2007).

Identified the students’ profile and the required competences for the construction of learning in this course, the next step was to define the course leading using KSA technique, the instructional design.

At this point was chosen the table simulation techniques that would simulate the required practical to achieve the competences defined in Table 01, these techniques are: “Jogo das casinhas”; “Miru Woli” and “Jogo do Carteiro”.



Figure 3: Example of the used techniques of table simulation

According to Vargas (1998), the “Jogo das casinhas” is a simulation of the execution of a set of houses using a reduced material design. Its purposes is to apply techniques of evaluation of labor force productivity and measurement of losses, showing the variety of modern project management techniques, and the gains that may come with the technological innovations adopted in the building site.

The table simulation technique “Miru Woli” works with the communication between the participants, where they have to choose, from many information, which ones fits to the purposed.

For Sugiura et al (1999) the technique of “Jogo do Carteiro” consists in walking between two P.O. Boxes, not talking to each other, communicating only through written messages. This simulation technique highlights the leaders who have the fundamental task of knowing how to express through the written communication.

These techniques of table simulation were applied during the analyzed term time, interspersed with setting exercises solved by the students which were collected and assessed statistically.

In order to validate the instructional design, of the materials prepared to the instruction and the chosen learning situation, it was made a pilot trial with 16 students for 4.5 months, where it was possible to verify mistakes, adjust time and activities, resulting the definition of the learning instruction based on KSA and supported by table simulation techniques.

The evaluation of the acquisition of the competences was made from the method of Kirkpatrick (1988), which developed the technique of evaluation and classification of training in four different levels, as showed below:

Table 2: Evaluation levels according to Kirkpatrick

EVALUATION LEVELS		
Nível 01	Reaction/Satisfaction	Evaluates the satisfaction of training after undertaking.
Nível 02	Learning	Evaluates before and after training, measuring the learning.
Nível 03	Change in behavior	Evaluates the different behavior before and after training.
Nível 04	Results	Evaluates the impact caused in the organization, from the training.

Source: Adapted from Kirkpatrick (1988)

This project was focused on levels 01, 02 and 03 in the evaluation of acquisition of competences from the use of lecture classes and the classes that used the table simulation techniques. Level 04 was observed by the researchers, assessed and reported to Analyses of Results.

4. ANALYSES OF RESULTS

During the reaction/satisfaction evaluation process (Level 01) during Scheduling and Controlling Constructions course was applied two questionnaires, developed and tested for this purpose, applied in the beginning and in the end of the course.

Table 3: Evaluation of Reaction/Satisfaction in "Scheduling and Controlling Constructions" course.

Reaction/Satisfaction Evaluation (Level 01)		
	Before	After
Group T	62.4%	70.5%
Group S	65.1%	88.7%

Table 03 shows similar results as the questionnaires applied in the beginning of the course in both groups; however the results of the end of the course are quite different, a small increase in Group T, and a substantial difference in Group S.

Observation and communication were the two points identified by the students of Group S as

important to filter the knowledge taught and to use it in the working atmosphere, allowing:

- Make the relation with work and professional practice;
- Develop interaction and team work;
- Relate the developed content with other parts of the course and also with the under-graduation program.

In Group T, only with lecture classes, it was verified little perception of the importance of the course for the development of education in the undergraduate program and the integration of education and entrenchment of future working life of the student.

In the learning evaluation (Level 02) were applied eight fixing exercises, four evaluation tests and a final project in both groups, and in Group S it was interspersed with table simulation planned and scheduled for the course. It was taken for analyses, corrected and measured by the researchers group.

The following Table 04 shows the summary of results reached by the average of these activities for Group T and S, and also the average of attendance in both groups.

Table 4: Learning Evaluation of Scheduling and Controlling Constructions course

Learning Evaluation (Level 02)		
	Final Grade	Attendance
Group T	7.32	83.1%
Group S	8.06	89.5%

The difference between the average of final grades of the group who had lecture classes and the group who had KSA classes with table simulations was 0.74 points, what represents an increasing of about 10% in the final grade.

With the final result of attendance and the analyses of the course's offering records, it can be stated that there is lower truancy when it the methodology of KSA is applied in classes of Scheduling and Controlling Constructions course.

Another insight is the increasing of knowledge acquirement during this experiment (Chart 01) of Group S, since the student did not have a sufficient knowledge of constructive techniques or professional experience which could have affect in a higher final grade when faced with the use of a table simulation technique in the construction of this knowledge.

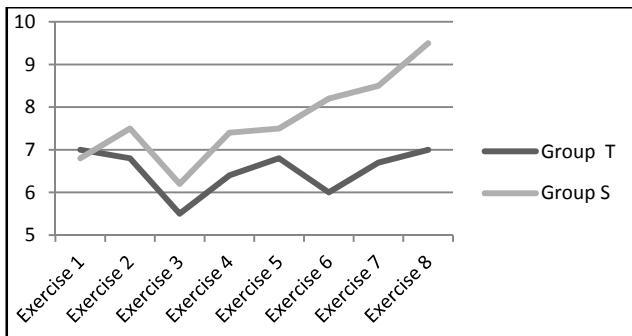


Chart 01: Average of the fixing exercises applied in Scheduling and Controlling Constructions' course.

At the moment to analyze the acquisition of skills and attitudes, it was observed the development of the skills mentioned in the study (Table 01), while the skill lower developed was team motivation, since it was not included situations that require this skill in the simulations, only during the technique of table simulation – *Jogo do Carteiro* – it was observed in some instance, when the group was delayed the leader acted to motivate.

The attitudes observed the own-initiative and dynamism were the most frequent, the leadership role, the decision-making and the observation power were also observed, however it were more frequent in table simulation of “*jogo das casinhas*”.

Observing the changes of behavior was more complex, the majority of students have some of the skills mentioned, but in the study it is targeted to the professional field and since the students did not have a professional experience, these skills have never been used to scheduling and managing works.

Considering that it was the first time that they acted as Engineers and played as planners, it can be observed that there is a change of behavior between before and after the course (Level 04).

It can be stated that the Kirkpatrick (1988) training evaluation tool was successfully used in this study, proving that the students were, in the end of the course, competent in scheduling and controlling constructions from using table simulation as a tool to the construction of competences (knowledge, skills and attitudes).

5. CONCLUSION

For a student to be considered competent when finishing a program at a university, it should be given to them tools that develop their knowledge, skills and attitudes. It can be stated that the aim of this study was confirmed by the results obtained.

The skills and attitudes were identified in the national and international bibliographical researches and it is based on the Brazilian legislation, aiming to diagnose which would be necessary to turn them

competent in Civil Engineering field and, consequently, in Scheduling and Controlling Constructions field.

Therefore, it can be observed that the competences, defined in the study, were constantly produced by the students during the application of table simulation techniques.

The motion of the methodology of KSA classes' proved that in the beginning of the course the students were totally lacking of competences referring to Scheduling and Controlling Constructions and in the end they had an excellent performance in knowledge, skills and attitudes.

Besides that, when applied the KSA classes' method it is noticed that the attendance to classes and the grade achieved by the students are higher than the achieved in lecture classes.

REFERENCES

- Cavalheiro, C. *Simulações e técnica de vivência na disciplina de planejamento e controle e empreendimentos*. Dissertação - Pós-graduação em Engenharia Civil - Universidade Federal de Santa Catarina (UFSC) 233p. Brasil - Florianópolis, SC. 2010.
- Costa, A. C. F.; Bogado, J. G. M; Jungles, A. E.; Heineck, L. F. M. *Apresentação dos resultados da simulação de uma fábrica de montagem de canetas a luz dos conceitos da mentalidade enxuta*. In: Encontro Nacional de Tecnologia do Ambiente Construído, Artigo Técnico. Brasil – Florianópolis, SC. 2006.
- Cripe, E. J.; Gerlach, R. R.; Lurie, R.; Masfield, R. S. - *Competency Model*, <http://www.opm.gov/studies/transapp.pdf>. Available: 10/ abril/2007.
- Durand, T. *L'alchimie de já compétence*. Revue Française de Gestion. Paris, 127, p. 84 -102, Janvier-Février 2000.
- Ellington, H.; Addinall, E.; Percival, F. *Games and simulations in science education*. London: Kogan Page Limited. 1981.
- Faria, A. J.; Wellington, W. J. *A survey of simulation games users, former-users and never-user*. Simulation & Gaming, v. 35, Issue 2, pp. 178 – 207. 2004.
- Khaled, N. *Managing construction equipment buy and sell decisions replacement: A simulation game*. In: ASC proceedings of the 37th annual conference, pp. 187–198. 2001.
- Kirkpatrick, D. *Evaluating Training Programs – The Four Levels*. Berret-Koehler Publisher, Inc – San Francisco. 1988.
- Koops, M. Hoevenaar, M. *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. 2013.

- Larréché, J.C. *On Simulations in Business Education and Research*. Journal of Business Research, v. 15, Issue 6, pp 559 – 571, 1987.
- Latorre, J.I., Jiménez, E. Simulation for Education in Business decision-Making. Magazine of the Society for Modeling & Simulation International (SCS M&S Magazine), (2), 59–65. 2012.
- Le Boterf, G. *Developper la competence des professionnels*. 4ème édition. Paris: Les éditions d'organisations, 2002.
- Magee, M., 2006. State of the Field Review: Simulation in Education. Internal report. Alberta Online Learning Consortium. Calgary AB.
- Mahboubian, M. *Educational aspects of business simulation softwares*. Procedia Social and Behavioral Sciences, v.2, pp. 5403–5407. 2010.
- Moore, D.R.; Cheng, M.; Dainty, A.R.F. *Competence, competency and competencies: performance assessment in organizations*. Work Study, Bingley, v.51, n.6, pp.314-319, 2005.
- Nisembaum, H. A. *A competência essencial*. São Paulo: Infinito, 2000.
- Odusami, K. T. *Perceptions of Construction Professionals Concerning Important Skills of Effective Project Leaders*. Revista de Gestão em Engenharia, v. 18, nº 2, pp. 61 – 67. Abril/ 2002.
- Pantaleão, L. H.; Oliveira, R. M.; Antunes J. A. V. *Utilização de um jogo de produção como ferramenta de aprendizagem de conceitos de engenharia de produção: O jogo do barco*. In: Encontro Nacional de Engenharia de Produção (XXII Enegep). Ouro Preto, MG, 2003.
- Pasin, F.; Giroux, H. *The impact of a simulation game on operations management education*. Computers & Education 57: 1240–1254. 2011.
- Paxton, J. *Teaching brief: a short, simple learning curve classroom exercise*. Decision Sciences Journal of Innovative Education, v. 1, n. 2, p. 303-307, 2003.
- Roos, J.; Victor, B.; Statler, M. *Playing seriously with strategy*. Long Range Planning, v. 37, p.549-568, 2004.
- Salas, E.; Wildman, J. L.; Piccolo, R. F. *Using simulation-based training to enhance management education*. Academy of Management Learning and Education 8: 559–573. 2009.
- Santos, J. 2002. *Development and implementing an internet-based financial system simulation game*. Journal of Economic Education 33: 31–39.
- Siddiqui, A., Khan, M., Akhtar, S. 2008. *Supply chain simulator: A scenario-based educational tool to enhance student learning*. Computers & Education 51: 252–261.
- Silva, E. L.; Menezes, E. M. *Metodologia da Pesquisa e elaboração da dissertação*. 4 Ed. Florianópolis: Edufsc, 2005.
- Sugiura, T.; Kaneko, N.; Yamada, Y.; Oda, T. *Introdução a jogos de treinamento para equipes*. Rio de Janeiro: Qualitymark, 1999.
- Thavikulwat, P. *The architecture of computerized business gaming simulations*. Simulation & Gaming, V. 35, Issue 2, pp. 242 – 269. 2004.
- Thavikulwat, P. Chang, J. Sanford, D. *Mergers and Acquisitions in a Business Game..* Simulation & Gaming, V. 44, Issue 5, pp. 706 – 731. 2013.
- Van Der Zee, D. J.; Slomp, J. 2009. *Simulation as a tool for gaming and training in operations management a case study*. Journal of Simulation 3: 17–28.
- Vargas, C. L. S. *Desenvolvimento de modelos físicos reduzidos como simuladores para a aplicação de conceitos produtividade, perdas, programação e controle de obras de construção civil*. Dissertação - Pós-graduação em Engenharia de Produção - Universidade Federal de Santa Catarina (UFSC) 115p. Brasil - Florianópolis, SC. 1998.
- Wolfe, J. *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. 1993.

AUTHORS BIOGRAPHY

Caroline Cavalheiro – Master in Civil Engineering – Universidade Federal de Santa Catarina in Brazil.

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

Antônio Edésio Jungles – Associate professor at Universidade Federal de Santa Catarina, with 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.

APPROACH CLASS LIBRARY OF HIGH LEVEL PARALLEL COMPOSITIONS TO IMPLEMENTS COMMUNICATION PATTERNS USING STRUCTURED PARALLEL PROGRAMMING

M. Rossainz-López^(a), M. I. Capel-Tuñón^(b)

^(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

ABSTRACT

This article presents through an environment of Parallel Objects, an approach to Structured Parallel Programming and the Object-Oriented paradigm, a programming methodology based on High Level Parallel Compositions (HLPC). By means of the method application, the parallelization of commonly used communication patterns among processes is presented, which is initially constituted by the HLPCs Farm, Pipe and TreeDV that represent, respectively, the patterns of communication Farm, Pipeline and Binary Tree, the latter one used within a parallel version of the design technique known as Divide and Conquer.

Keywords: Parallel Objects, Communication Patterns, Structured Parallel Programming, High Performance Computing.

1. INTRODUCTION

At the moment the construction of concurrent and parallel systems has less conditioning than one decade ago, since there currently exist, within the realms of HPC or Grid computing, high performance parallel computation systems that are becoming more and more affordable, being therefore possible to obtain a great efficiency today in parallel computing without having to invest a huge amount of money in purchasing a state-of-the-art multiprocessor. Nevertheless, to obtain efficiency in parallel programs is not only a problem of acquiring processor speed; on the contrary, it is rather about how to program efficient interaction/communication patterns among the processes (Brinch Hansen 1993; Capel and Palma 1992; Darlington 1993), which will allow us to achieve the maximum possible speed-up of a given parallel application. These patterns are aimed at encapsulating parallel code within programs, so that an inexperienced parallel applications programmer can produce efficient code by only programming the sequential parts of the applications processes (Brinch Hansen 1993; Brinch Hansen 1994; Capel and Troya 1994). Parallel

Programming based on the use of communication patterns is known as Structured Parallel Programming (SPP) (Corradi and Zambonelli 1995; Danelutto and Orlando 1995). The widespread adoption of SPP methods by programmers and system analysts currently presents a series of open problems, which motivate us to carry out research in this area; we are particularly interested on those that have to do with parallel applications that use predetermined communication patterns among their processes. In regarding an improvement of the use of SPP methods, several open problems have currently been identified. It is worth mentioning, among the most important, the following ones:

- The lack of acceptance of structured parallel environments applicable to developing a wider range of software applications.
- Determination of a complete set of communication patterns as well as their concrete semantics.
- The necessity to make available predefined communication patterns or high level parallel compositions aimed to encapsulate parallel code within programs.
- Adoption without anomalies of a programming approach merging concurrent primitives and Object-Oriented (O-O) features, thereby fulfilling the requirements of uniformity, genericity and reusability of software components (Corradi and Zambonelli 1995).

This work study the Methods of Structured Parallel Programming, proposing a new implementation of a library classes of High Level Parallel Compositions-HLPC or CPAN according to its Spanish acronym (Corradi and Zambonelli 1995; Danelutto and Orlando 1995; Capel and Rossainz 2004; Rossainz 2005), which provide the programmer with the communication patterns more commonly used in Parallel Programming. At the moment, the library includes the following ones:

Farm, Pipeline and Binary Tree (Rossainz and Capel 2008), the latter one being used in a parallel version of Divide and Conquer algorithmic design technique and implementation of the solution of the traveling salesman problem using the HPLC Farm using branch & bound algorithmic design technique for solving NP-complete problems (Rossainz and Capel 2006).

2. HIGH LEVEL PARALLEL COMPOSITIONS (HLPC)

The basic idea of the programming method consists in the implementation of any type of communication patterns between parallel processes of an application or distributed/parallel algorithm as classes, following the O-O paradigm. Starting from these classes, an object can be instantiated and its methods can be invoked, and then be executed, on a client-server basis, thereby hiding parallelism or communication protocols to the application processes. A HLPC is made up of the composition of a set of objects of three types, (see Figure 1).

An object manager that represents the HLPC itself, making of it an encapsulated abstraction that hides its internal structure. The manager controls the references of a set of objects (a denominated Collector object and several denominated Stage objects) that represent the components of the HLPC and whose execution is carried out in parallel and should be coordinated by the manager itself.

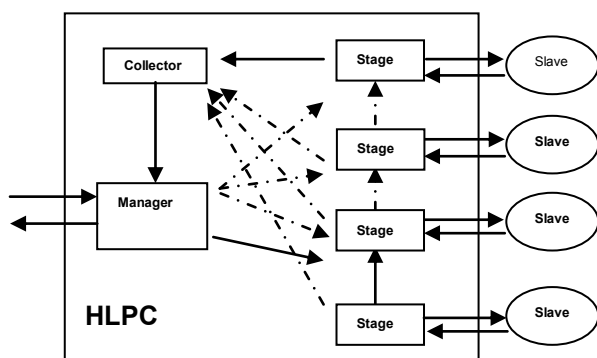


Figure 1: Internal Structure of a HLPC

The stage objects are objects of specific purpose responsible for encapsulating a client-server type interface between the manager and the object slaves (objects that are not actively participative in the composition of the HLPC, but rather, are considered external entities that contain the sequential algorithm constituting the solution of a given problem) as well as providing the necessary connection among them to implement the communication pattern semantics, whose definition is being sought. In other words, each stage should act in parallel as a node of the graph that represents the communication pattern and should be capable of executing its methods as an active object. A stage can be directly connected to the manager and/or to another component stage depending on the particular pattern of the HLPC.

An object collector which is an object in charge of storing in parallel the results that it receives from the stage objects connected to it, i.e., during the service of a petition. The control flow within the stages of a HLPC depends on the communication pattern implemented between these. When the HLPC concludes its execution, the result does not return to the manager directly, but rather to an instance of the class Collector, which takes charge of storing these results and of sending them to the manager, which then sends them to the exterior as they arrive, i.e., without being necessary to wait for all the results to be obtained at the end of the computation.

2.1. The HLPC as composition of parallel objects

The objects manager, collector and stages are included within the definition of Parallel Object - PO (Corradi and Zambonelli 1995). The Parallel Objects are active objects, that is to say, objects that have execution capacity in them. 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 (Rossainz and Capel 2012). 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). 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 inheritance.

2.2. Types of communication between parallel objects

1. The synchronous way stops the client's activity until the object's active server gives back the answer to the petition.
2. The asynchronous way does not force any waiting in the client's activity; the client simply sends its petition to the active server and then it continues.
3. The asynchronous future way makes only to wait the client's activity when the result of the invoked method is needed to evaluate an expression during its code execution.

2.3. Semantic and Syntactic definition of the classes bases of a HLPC anyone

The abstract class ComponentManager defines the generic structure of the component manager of a HLPC, from which all the concrete manager classes are derived,

depending on the parallel behaviour which is needed to create a specific HLPC.

```

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 abstract class ComponentStage defines the generic structure of the component stage of a HLPC as well as its interconnections, so that all the concrete stages needed to provide a HLPC with a given parallel behaviour can be obtained by class instantiation.

```

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);
};

```

The concrete class ComponentCollector defines the concrete structure of the component collector of any HLPC. It implements a multi-item buffer, which permits the storage of the results from stages that make reference to this collector.

```

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);
};

```

2.4. The synchronization restrictions MaxPar, Mutex and Sync

Synchronization mechanisms are needed when several petitions of service take place in parallel in a HLPC, being capable its constituting parallel objects of interleaving their concurrent executions while, and at the same time, they preserve the consistency of the data being processed. Within the code of any HLPC, execution constraints are automatically included when the reserved words MAXPAR, MUTEX and SYNC of the library are found. The latter ones must be used to obtain a correct programming of object methods and to guarantee data consistency in applications.

3. THE HLPC PIPE, FARM AND TREEDV

The parallel patterns worked until now have been the *pipeline*, the *farm* and the *TreeDVD*, since they constitute a significant set of reusable communication patterns in multiple parallel applications and algorithms.

3.1. The HLPC PipeLine

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

3.1.1. 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 (Roosta 1999). The processes of the pipeline are sometimes called stages of the pipeline. 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.

3.1.2. Representation of the Pipeline as a HLPC

The figure 2, represent the parallel pattern of communication Pipeline as a HLPC.

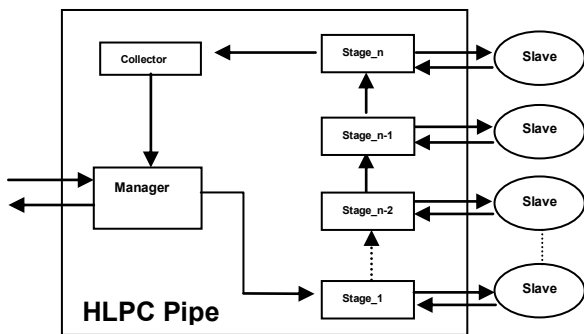


Figure 2: The HLPC of a Pipeline

The objects *stage i* and *Manager* of the graphic pattern of the HLPC-Pipe are instances of concrete classes that inherit the characteristics of the classes *ComponentManager* and *ComponentStage*.

3.1.3. Semantic and Syntactic Definition of the HLPC-Pipe

The HLPC-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. 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);
    }
};
```

The operations *execution()* and *request()* are inherited of their respective super class and the operations private *commandStages()* (of the class *PipeManager*) and *commandOtherStages()* (of the class *PipeStage*), together with the operation *init()* are redefined. However there are not synchronization problems since in their definitions the synchronization restrictions they are inherited of their super class.

3.2. The HLPC Farm

It is shown the technique of the parallel processing of the FARM as a High Level Parallel Composition or HLPC.

3.2.1. The technique of the Farm

The parallel pattern of interaction Farm, is formed to each other with a set of independent processes, called worker processes, and a process that controls them, call the process controller (Roosta 1999). The worker processes are executed in parallel until reaching a common objective for all them. The process controller is in charge of distributing the work and of controlling the progress of the farm until finding the solution of the problem. With this model it could be interesting to observe the performance of the execution in parallel of several sorting algorithms with oneself set of data for all them.

3.2.2. Representation of the Farm as a HLPC

The representation of parallel pattern FARM as a HLPC is show in figure 3.

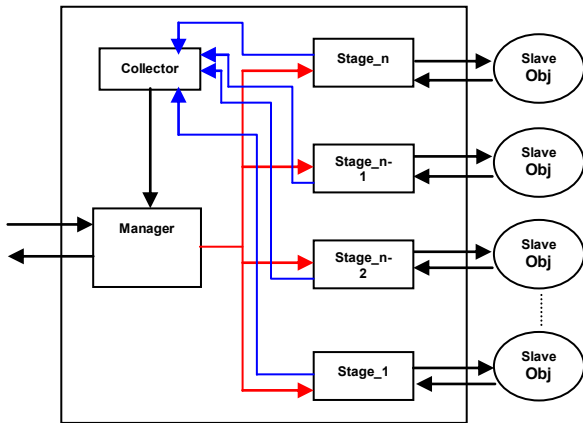


Figure 3: The HLPC of a Farm

The same as in the previous pattern, the objects *Manager* and *stage_i* are respectively instances of the classes that inherit of the classes base denominated *ComponenManager* and *ComponentStage*.

3.2.3. Semantic and Syntactic Definition of the HLPC Farm

A first policy for the composition of the FARM is that the manager only waits the first available result given by anyone of the stages, which respond to a petition of service in an asynchronous way.

```
CLASS CONCRETE FarmManager EXTENDS OF
ComponentManager
```

```
{
VAR
  INT nWorker;

PUBLIC VOID init (ASOCIACION[] list)
VAR
  asociacion[] newList, INT i = 0;
{
  WHILE (!(newlist = TAIL(list)))
  {
    stages[i++] = FarmStage CREATE
      (CONS(HEAD(list), NULL));
    list = newList;
  }
  nWorker = i;
}

PRIVATE VOID commandStages(ANYTYPE
  datain, ComponentCollector res)
VAR INT i;
{
  FOR i =(0,nWorker)
  {
    THREAD stages[i].request(datain, res);
  }
}
};
```

The concrete class *FarmManager* inherits of *ComponentManager*. The operation *init ()* create all the

necessary stages, while the operation execution () it is thrown in parallel in an asynchronous way, distributing data to all the stages, waiting the first available result in the object collector. As in the case of the pipeline, the synchronization restrictions are inherited of the abstract class *ComponentManager* without any problem. The stages of the farm are objects of *FarmStage* that inherits of *ComponentStage*:

```
CLASS CONCRETE FarmStage EXTENDS OF
ComponentStage
{
};
```

The stages of the Farm are not connected some with others (*am_i_last* is always *true* and the *CommandOtherStages* operation, as in the abstract class this empty). The manager commands them to all in his execution and the result of each one is a correspondent to the object collector and position to the manager's disposition. As the stages are executed in parallel in an asynchronous way, the policies of scheduling inherited of the super class guarantees that the access to the collector, necessary to return the results, it will be made in a way synchronized on the part of this objects.

3.3. The HLPC TreeDV

Finally, the programming technique is presented it Divide and Conquer as a HLPC, applicable to a wide range of problems that can be parallelizable within this scheme.

3.3.1. The technique of the Divide and Conquer

The technique of it Divide and Conquer it is characterized by the division of a problem in sub-problems that have the same form that the complete problem. The division of the problem in smaller sub-problems is carried out using the recursion. The method recursive continues dividing the problem until the parts divided can no longer follow dividing itself, then they combine the partial results of each sub-problem to obtain at the end the solution to the initial problem (Blelloch 1996). In this technique the division of the problem is always made in two parts, therefore a formulation recursive of the method Divide and Conquer form a binary tree whose nodes will be processors, processes or threads. The node root of the tree receives as input a complete problem that is divided in two parts. It is sent to the node left son, while the other is sent to the node that represents the right son.

This division process is repeated of recursive form until the lowest levels in the tree. Lapsed a certain time, all the nodes leaf receives as input a problem given by its node father; they solve it and the solutions (that are the exit of the node leaf) are again correspondents to its progenitor. Any node father in the tree will obtain his children's two partial solutions and it will combine them to provide an only solution that will be the node father's exit. Finally the node root will give as exit the complete solution of the problem, (Brinch Ansen 1993). This

way, while in a sequential implementation a single node of the tree can be executed or visited at the same time, in a parallel implementation, more than a node it can be executed at the same time in the different levels, it is, when dividing the problem in two sub-problems, both can be processed in a simultaneous way.

3.3.2. Representation of the TreeDV as a HLPC

The representation of the patron tree that defines the technique of it Divide and Conquer as HLPC has their model represented in figure 4.

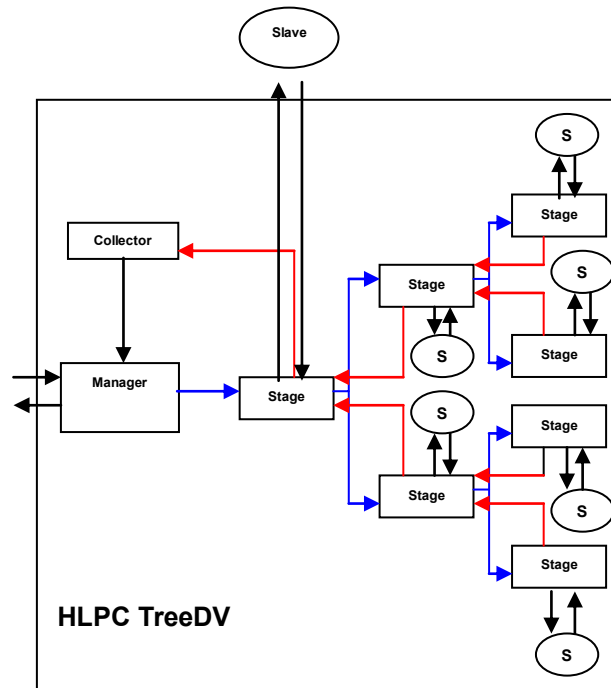


Figure 4: The HLPC of a TreeDV

Contrary to the previous models, where the objects slaves were predetermined outside of the pattern HLPC, in this model an object slave is only predefined statically and associated to the first stage of the tree.

The following objects slaves will be created internally by the own stages in a dynamic way, because the levels of the tree depend from the problem to solve and a priori the number of nodes that can have the tree is not known, neither its level of depth.

3.3.3. Semantic and Syntactic Definition of the HLPC TreeDV

The class TreeDVManager is created, which inherits of ComponentManager, and a pattern of communication of a binary tree implements within the programming technique it Divide and Conquer. The nodes of the binary tree are represented by the stages that are objects of the class TreeDVStage that inherits of ComponentStage. Any instance of the class TreeDVManager only takes charge of the first stage or node root of the tree in its initialization. During the execution of a petition of service, the root of the binary tree, that is to say, the first stage of the structure is

commanded by the manager and internally each node father created in the lowest levels in the tree will command his respective nodes children in the solution of the problem.

CLASS CONCRETE TreeDVManager EXTENDS OF ComponentManager

```

{
  PUBLIC VOID init (ASOCIACION[] list)
  { stages[0] = TreeDVStage CREATE (list); }

  PRIVATE VOID commandStages(ANYTYPE
    datain, ComponentCollector res)
  {
    THREAD stages[0].request(datain,res);
    THREAD res.put(datain);
  }
};

```

Any object TreeDVStage will take charge of creating a node of the binary tree (left or right). When the node root or initial stage execute the operation *request()* in parallel, the problem is evaluated with the method of the slave object associate, returning the division of the problem in two parts. Later on, will call himself to the method *commandOtherStages()* who will take this sub-problems, it will create two nodes stages associated to their node father, associating these last their respective objects slaves that will be created dynamically conform to they go creating the nodes of the binary tree, and will send to each one a part of the problem to solve. Of recursive form the nodes stages children will receive the sub-problem and they will execute their method *request()* in parallel, carrying out the same process until the sub-problem can no longer be divided more. The last objects TreeDVStage of the tree, that is to say, the leaves, send the result from their calculation process to their progenitor and this combines the solutions to pass them, in turn, to their progenitor and so on, in the return of the recursion, until arriving to the node root who will send the final result to an object collector who in turn will pass the result to the manager of the composition.

CLASS CONCRETE TreeDVStage EXTENDS OF ComponentStage

```

{
  VAR
    ASOCIACION list;

  PUBLIC VOID init(ASOCIACION[] *list)
  {
    this.list=list;
    stage.init(list);
    am_i_last=FALSE;
  }
}

```

```

PRIVATE VOID commandOtherStages
  (ANYTYPE datain, ComponentCollector res)

```

```

VAR
  ANYTYPE data_izq, data_der;
{
  IF(datain.inicio<datain.fin)
  {
    otherStages[0]=TreeDVStage CREATE(list);
    otherStages[1]=TreeStage CREATE(list);

    THREAD otherStages[0].request
      (dv(datain,datain.inicio, datain.medio),res);

    THREAD otherStages[1].request
      (dv (datain,datain.medio+1, datain.tam-1),res);
  }
}

```

```

PRIVATE ANYTYPE dv(ANYTYPE dataout,
                    int inicio, int fin)
{
  datain.inicio=inicio;
  datain.fin=fin;
  RETURN datain;
}
};

```

4. USE OF A HLPC WITHIN AN APPLICATION

Once implemented the HLPCs of interest, the way in that you/they are used in user's application is the following one:

1. It will be necessary to create an instance of the class manager of interest, that is to say, one that implements the required parallel behavior in agreement with the following steps:
 - 1.1. To initialize the instance with the reference to the objects slaves that will be controlled by each stage and the name of the method requested as an association of even (slave_obj, associated_method).
 - 1.2. The internal stages is created (using the operation *init()*) and they are passed each one an association (slave_obj, associated_method) that will use invoking the associated_method on their slave object.
2. The user asks the manager to begin a calculation through the execution within the HLPC of the method execution(). This execution is carried out as it continues:
 - 2.1. The object collector is created with respect to the petition.
 - 2.2. They are passed to the stages the input data (without verification of types) and the reference to the collector.
 - 2.3. The results are obtained from the object collector.
 - 2.4. The collector returns the results again to the exterior without verification of types.

3. An object manager has been created and initialized and some execution petitions can be dispatched in parallel.

The speedup analysis of the Farm, PipeLine and TreeDV HLPCs appears in (Rossainz and Capel 2008)

4.1. Example of the use of a HLPC

Supposing that we want process an array that contains data that must be ordered. It is well-known that, depending on the data, different sorting algorithms can show different performances. It can be interesting to order the same data by means of the execution of different algorithms in parallel and to wait the results. The parallel pattern identified is FarmManager.

1. Objects of different sorting classes are created that represent the slaves objects and are stored in an array of references to object:


```

Object obj[] = {
  Qsort CREATE(. . .),
  BubleSort CREATE(. . .),
  Isort CREATE(. . .)
};

```
2. The objects are created that they represent the methods associated to the objects slaves and they are stored in an array of references to method.


```

method meth[] = {
  method CREATE (. . . resolve),
  method CREATE (. . . resolve),
  method CREATE (. . . resolve)
};

```
3. A list of associations is created (slave_obj, asociación_method)


```

Asociación pareja =
crea_asociacion(obj, meth, 3);

```
4. An instance of the class FarmManager is created initializing it with the previous list


```

FarmManager cpanfarmInt =
FarmManager CREATE(pareja);

```
5. The initial data are specified to process by means of the specific creation of the type and of the defined data as an object for the user, in this example this object it is denominated MyInt. Each object MyInt represents a problem to solve


```

int nums[t1] = {-11, -14, -6, -1, ...};
int nums2[t2] = {5, 1, 9, 11, 4, 8, ...};
ANYTYPE data[] = {
  MyInt CREATE(t1, nums),
  MyInt CREATE(t2, nums2)
};

```
6. The initial data are printed in screen


```

FOR i = (0, num_problems)
  {imprime_datos(data[i]);}

```
7. The instance FarmManager is ready to work, that is to say, the execution of the operation *execution()* is requested that sort the array of data in parallel:

```

FUTURETYPE resul[num_problems];
FOR i =(0,num_problems)
{
  resul[i]=THREAD
  cpanfarmInt.execution(data[i]);
}

```

8. The final results are printed in screen
- ```

ANYTYPE resultados[num_problems];
FOR i =(0,num_problems)
{
 resultados[i]=resul[i];
 imprime_datos(resultados[i]);
}

```

## 5. THE WORK OF PRESENT PROGRAMMING

The work of programming made until this moment is made up of six sets of classes that constitute the implementation of the parallel patterns farm, pipe and treeDV as HLPCs that are part of what will be the library of High Level Parallel Compositions. These sets of classes are the following ones:

The set of the classes base (Table.1), necessary to build a HLPC, or said in another way, the classes that implement the Parallel Objects of a HLPC.

Table 1: Classes base for the construction of a HLPC

|                         |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Class Object</i>     | It represents slave objects generic, that is to say, it is an abstract class inherited by the slave objects concrete (for example ISort, QSort, Invierte, etc.) to implement the virtual method <i>resuelve()</i> that will be the method to be executed by the slave object within the HLPC.                                                                                                                                                                                                                                                                                                                          |
| <i>Class FutureType</i> | It defines the type FutureType referred as a Future of type void * that will be the type of the value of return of a function. This class denotes instances of FutureType to indicate that but this available one the return value in a given moment, if it can be he in some time in the future.                                                                                                                                                                                                                                                                                                                      |
| <i>Header Util.h</i>    | It contains the definition of "primitive function" used in the code of the HLPC (the primitive CONS for to append elements in a list, the primitive HEAD that obtains the head of a list, the primitive TAIL that obtains the rest of a list, the primitive EVAL that evaluates a function), as well as the definition of several types of abstract data (the type association that defines a pair (method,slave_obj), the type <i>method</i> that defines the execution method associated to an slave object, the type vector_sol that is used as container of results) necessary in the implementation of the HLPCs. |

|                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Class ComponentCollector</i>     | It is used to create instances or objects Collector. An object collector will to set the solutions from the connected stages to him and it will store them in a variable of type vector_sol.                                                                                                                                                                                                                                                                                      |
| <i>Class Sched_GetPut</i>           | It defines the synchronization restrictions for the functions member <i>get()</i> and <i>put()</i> within the class ComponentCollector: Mutex for the processes <i>put()</i> , Mutex for the processes <i>get()</i> and Sync for the communication of processes <i>put()</i> and <i>get()</i> . All the members of this class are static.                                                                                                                                         |
| <i>Class ComponentStage</i>         | It defines the generic structure of a STAGE in the construction of a HLPC and has to be inherited by the specific stages that can be built in the implementation of the parallel composition for to concretize a particular stage. It is constituted of two parts: the part of initialization of the "stage" and the execution part, in parallel.                                                                                                                                 |
| <i>Class ComponentManager</i>       | It defines the generic structure of a MANAGER in the construction of a HLPC that has to be inherited by the specific manager that can be built for the implementation of the HLPC. It is constituted of two parts: the part of the "manager's" initialization and the execution part, in parallel.                                                                                                                                                                                |
| <i>Class Sched_RequestExecution</i> | It implements the restrictions of maximum parallelism and use of futures for the functions member <i>execution()</i> and <i>request()</i> of the classes ComponentManager and ComponentStage respectively. Their methods <i>request_t()</i> and <i>execution_t()</i> implement the maxpar restrictions for the functions <i>request()</i> and <i>execution()</i> and the use of futures in the assignment of the value of return of <i>execution()</i> to an instance FutureType. |

The set of the classes that define the types of own data (Table.2), used for the creation of types of abstract data to be worked by the HLPCs in form of objects:

Table 2: Classes that define own data types

|                      |                                                                                                                                                                                                                                                 |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Class MyInt</i>   | It defines the specific data and their type (in this case integer) that the user wants to provide to the HLPC for his processing as an instance of this class. In particular the objects of this class are used within the HLPCs Farm and Pipe. |
| <i>Class MyIntDV</i> | It defines the specific data and their type (also integer) that the user wants to provide to the HLPC for his processing as a object. In particular instances of this class are used within the HLPC TreeDV.                                    |

The set of classes that define the slave objects (Table.3), they should inherit the characteristics of Object to be worked in a generic way by the HLPCs, implementing the virtual method *resuelve()* that is where the code of solution of the problem will be.

Table 3: Classes that define slave objects

|                        |                                                                                                                                                                                                                                                                  |
|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Class QsortDV</i>   | It is used for instance slave objects type QSortDV and it solves the problem partially of sorting a set of data in disorder using the algorithm of QuickSort (implementation of the part of “conquer” of the technique divide and conquer) for HLPC type TreeDV. |
| <i>Class Qsort</i>     | It is used for instance slave objects type QSort and it solves the one problem of sorting a set of data in disorder using the algorithm of QuickSort within the paradigm of it divide and conquer, for HLPC type Farm and type Pipe.                             |
| <i>Class BubleSort</i> | Used for instance slave objects type BubleSort and it solves the problem of sorting a set of data in disorder using the algorithm of the Bubble for HLPCs type Farm or type Pipe.                                                                                |
| <i>Class ISort</i>     | It is used for instance slave objects type ISort and it solves the problem of sorting a set of data in disorder using the algorithm of Sorting for Insertion for HLPCs type Farm or type Pipe.                                                                   |
| <i>Class Invierte</i>  | It is used for instance slave objects type Invierte and it solves the problem of investing the sequence of a set of data. Instances of this type were used in the implementation of the HLPC Pipe.                                                               |

The set of classes that define the HLPC Farm (Table.4) that is formed of:

Table 4: Classes that define the HLPC Farm

|                                  |                                                                                                                                                                                                                                  |
|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Main Program cpanfarm.cpp</i> | It proves the execution of the HLPC Farm. The problem that is solved is the sorting in parallel of two arrays of integer numbers using three different algorithms of sorting (QuickSort, BubleSort and ISort) sent concurrently. |
| <i>Class FarmStage</i>           | It defines the component stage relative to the parallel pattern FARM and it inherits of ComponentStage.                                                                                                                          |
| <i>Class FarmManager</i>         | It defines a concrete instance of a manager for a HLPC type Farm. The class inherits of ComponentManager.                                                                                                                        |
| <i>Class CpanFarm</i>            | It is used in the main program to create an active object that represents The High Level Parallel Composition Farm and to solve "n" problems in parallel through this implemented communication pattern.                         |

The set of classes that define the HLPC Pipe (Table.5), formed by the following classes.

Table 5: Classes that define the HLPC Pipe

|                                  |                                                                                                                                                                                  |
|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Main Program cpanpipe.cpp</i> | It proves the execution of the HLPC Pipe. The problem that is solved is the execution of a sequence in parallel of 3 algorithms: the algorithm Invierte, the algorithm Qsort and |
|----------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                          |                                                                                                                                                                                                          |
|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                          | again the algorithm Invierte to invest a sequence of data in disorder, to order the sequence change and to invest the sorted sequence of two arrays of integer numbers sent concurrently again.          |
| <i>Class PipeStage</i>   | It defines the component stage relative to the parallel pattern PIPE and it inherits of ComponentStage.                                                                                                  |
| <i>Class PipeManager</i> | It defines a concrete instance of a manager for a HLPC type Pipe. The class inherits of ComponentManager.                                                                                                |
| <i>Class CpanPipe</i>    | It is used in the main program to create an active object that represents The High Level Parallel Composition Pipe and to solve "n" problems in parallel through this implemented communication pattern. |

The set of classes that define the HLPC TreeDV (Table 6), formed by the classes:

Table 6: Classes that define the HLPC TreeDV

|                                    |                                                                                                                                                                                                                                                                                                 |
|------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Main Program cpantreeDV.cpp</i> | It proves the execution of the HLPC TreeDV. The problem that is solved is the sorting in parallel of two arrays of integer numbers using the QuickSort algorithm so many times as nodes stage of the tree leave creating in the solution process which are sent concurrently for its execution. |
| <i>Class TreeDVStage</i>           | It defines the component stage relative to the parallel pattern TreeDV and it inherits of ComponentStage.                                                                                                                                                                                       |
| <i>Class TreeDVManager</i>         | It defines a concrete instance of a manager for a HLPC type TreeDV. The class inherits of ComponentManager.                                                                                                                                                                                     |
| <i>Class CpanTreeDV</i>            | It is used in the main program to create an active object that represents The High Level Parallel Composition TreeDV and to solve "n" problems in parallel through this implemented communication pattern.                                                                                      |

## 6. CONCLUSIONS

Method of original programming has been developed based on HLPCs.

Patterns of communication/interaction have implemented themselves within the model of the HLPC commonly used in the parallel and distributed programming: the HLPC Pipe, the HLPC Farm and the HLPC TreeDV.

The implemented HLPCs can be exploited, thanks to the adoption of the approach oriented to objects using the different mechanisms of reusability of the paradigm to define new patterns already using those built.

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.

The HLPCs Pipe, Farm and TreeDV conform the library of classes that intends in this work.

The restrictions of synchronization have been programmed of original form suggested by the model of

the HLPC for their parallel and concurrent operation: the maximum parallelism (MaxPar), the mutual exclusion (Mutex) and the synchronization of communication of processes readers/writers (Sync).

Of equal it forms the programming in the asynchronous future communication way for results “futures” within the HLPCs it has been carried out in an original way by means of classes.

## REFERENCES

- Blelloch, G.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.
- Capel, M.I. and Palma A., 1992. A Programming tool for Distributed Implementation of Branch-and-Bound Algorithms. *Parallel Computing and Transputer Applications*. IOS Press/CIMNE. Barcelona.
- Capel, M.I. and Troya J.M., 1994. An Object-Based Tool and Methodological Approach for Distributed Programming. *Software Concepts and Tools*.
- Capel, M.I. and 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.
- Corradi A., Leonardi L., 1991. PO Constraints as tools to synchronize active objects. *Journal Object Oriented Programming*. Volume10: 42-53.
- Corradi, A., Leonardo, L., Zambonelli, F., 1995. *Experiences toward an Object-Oriented Approach to Structured Parallel Programming*. DEIS technical report. 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).
- 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)*. Thesis (PhD). Universidad de Granada.
- Rossainz, M., Capel, M.I., 2006. Design and Implementation of the Branch & Bound Algorithmic Design Technique as an High Level Parallel Composition. *Proceedings of*

*International Mediterranean Modelling Multiconference*. Barcelona, Spain.

- Rossainz, M., Capel M.I., 2008. A Parallel Programming Methodology using Communication Patterns named CPANS or Composition of Parallel Object. *Proceedings of 20TH European Modeling & Simulation Symposium*. Campora S. Giovanni. Italy.
- Rossainz M., Capel M.I., 2012. Compositions of Parallel Object to Implement Communication Patterns. *Proceedings of XXIII Jornadas de Paralelismo*, pp.8-13. September 19-21. Elche, Spain.



# DATA-BASED IDENTIFICATION OF SHORT TERM PREDICTORS FOR STOCK MARKET TRENDS USING HETEROGENEOUS MODEL ENSEMBLES

Stephan Winkler<sup>(a)</sup>, Bonifacio Castaño<sup>(b)</sup>, Sergio Luengo<sup>(b)</sup>,  
Susanne Schaller<sup>(a)</sup>, Gabriel Kronberger<sup>(a)</sup>, Michael Affenzeller<sup>(a,c)</sup>

<sup>(a)</sup> Heuristic and Evolutionary Algorithms Laboratory  
University of Applied Sciences Upper Austria, Softwarepark 11, 4232 Hagenberg, Austria

<sup>(b)</sup> Department of Physics and Mathematics, University of Alcalá, Madrid (Spain)

<sup>(c)</sup> Institute for Formal Models and Verification, Johannes Kepler University Linz  
Altenberger Straße 69, 4040 Linz, Austria

<sup>(a)</sup> [stephan.winkler@fh-hagenberg.at](mailto:stephan.winkler@fh-hagenberg.at), [michael.affenzeller@fh-hagenberg.at](mailto:michael.affenzeller@fh-hagenberg.at), [susanne.schaller@fh-hagenberg.at](mailto:susanne.schaller@fh-hagenberg.at),  
[gabriel.kronberger@fh-hagenberg.at](mailto:gabriel.kronberger@fh-hagenberg.at)

<sup>(b)</sup> [bonifacio.castano@uah.es](mailto:bonifacio.castano@uah.es), [sergio.luengo@edu.uah.es](mailto:sergio.luengo@edu.uah.es)

## ABSTRACT

We here show the application of heterogeneous ensemble modeling for training short term predictors of trends in stock markets. A sliding window approach is used; model ensembles are iteratively learned and tested on subsequent data points. The goal is to predict trends (positive, neutral, or negative stock changes) for the next day, the next week, and the next month.

Several machine learning approaches implemented in HeuristicLab and WEKA have been applied; the models produced using these methods have been combined to heterogeneous model ensembles. We calculate the final estimation for each sample via majority voting, and the relative ratio of a sample's majority vote is used for calculating the confidence in the final estimation; we use a confidence threshold that specifies the minimum confidence level that has to be reached.

We show results of empirical tests performed using data of the Spanish stock market recorded from 2003 to 2013.

Keywords: financial data analysis, ensemble modeling, trend classification, machine learning

## 1. INTRODUCTION AND OVERVIEW

A lot of research on using machine learning for predicting stock market trends has already been discussed in the literature, see for example (Kaboudan et al. 2002), (Potvina et al. 2004), (Summers et al. 2004), (Mallick et al. 2008), and (Bodas-Sagi et al. 2012).

The aim of the research discussed in this paper is to learn classifiers for short-term trends in stock markets. We use several machine learning techniques for training sets of heterogeneous models that are able to estimate trends in stock market data.

We particularly focus on short term prognosis, i.e., models are trained that are designed to predict trends for a short future period: The goal is to predict the trend for the next trading day, the next 5 days, and the next 20

days. Trends are classified as positive, neutral, or negative, so that the goal is to learn classification models that are able to correctly predict the classification of future trends.

For evaluating this approach we apply a sliding window strategy: The given data are split into partitions, and each partition is used for training models that shall explain the following test data points.

In the following section we describe the sliding window short term training and testing approach used here, and in Section 3 we give an overview of the modeling methods applied. In Section 4 we summarize the results achieved using this approach on data of the Spanish stock market (recorded 2003 – 2013), and Section 5 concludes the paper.

## 2. SHORT TERM PROGNOSIS AND SLIDING WINDOW TREND CLASSIFICATION

We simulate real world applications for stock market trend estimations by applying training and test on the data using a sliding window approach. As depicted in Figure 1, we split the given time series data in partitions; the training of models (as explained in the following section) is done for each partition separately, and all models are tested only on data samples that are directly subsequent to the respective training samples. This simulates short term prediction for stock market data that are (in real life) updated periodically.



Figure 1: Sliding window short term prognosis training and testing of models.



For the analyzed target variables we are in this research work interested in the classification of their trends. I.e., the goal is to train classification models that estimate the trend of a variable  $V$  for the next  $n$  samples (days) as positive (rising by at least  $e\%$ ), negative (falling by at least  $e\%$ ), or neutral.

### 3. HETEROGENEOUS ENSEMBLE MODELING AND CONFIDENCE ESTIMATION

The following techniques for training classifiers have been used in this research study:

- Neural networks (Nelles 2001),
- k-nearest-neighbor classification (Duda et al. 2000),
- support vector machines (Vapnik 1998; Chang & Lin 2001),
- genetic programming (Koza 1992; Affenzeller et al. 2009; Winkler 2009),
- decision trees (Kotsiantis 2007), and
- random forests (Breiman 2001).

Most of these machine learning methods have been implemented using the HeuristicLab framework (Wagner et al. 2014), a framework for prototyping and analyzing optimization techniques for which both generic concepts of evolutionary algorithms and many functions to evaluate and analyze them are available; for learning decision trees the WEKA framework (Hall et al. 2009) was used. Details of these methods are summarized in the appendix of this paper.

For all methods we have used parameter settings that were identified as the best settings by the approach described in (Winkler et al. 2011). There we have described the use of evolutionary optimization for identifying optimal modeling parameters for the here presented data collections as well as the modeling methods applied here.

Each modeling method was executed five times using 10-fold cross validation. All so created models were then applied on their respective test partitions; the final classification for each sample was calculated in the following way (as described in detail in (Winkler et al. 2014(a)) and (Winkler et al. 2014(b))):

For each sample  $s$  various models are trained; models are here referred to as  $m_i$  where  $i$  represents the current model index. Subsequently, a voting over all models for each sample  $s$  is performed:

$$vote(c,s) = | \{ m_i : m_i(s) = c \} | \quad (1)$$

where  $c$  represents an arbitrary, but fix class  $c$ .

The final classification  $fc$  for a sample  $s$  is calculated by using a majority voting of votes  $vote(c,s)$ :

$$fc(s) = argmax_c (vote(c,s)) \quad (2)$$

Finally, we define the confidence of classification  $c$  for a sample  $s$  (which represents the confidence towards this classification) on the basis of a set of models  $m$ :

$$conf(c,s,m) = | \{ m_i : m_i(s) = c \} | / | m | \quad (3)$$

As we are here facing ternary classification tasks, the classification confidence for any final classification  $fc$  (that is a majority vote winner and thus must have more than 1/3 of the votes) will always be in the interval  $]1/3, 1]$ .

The overall work flow followed here is graphically shown in Figure 2:

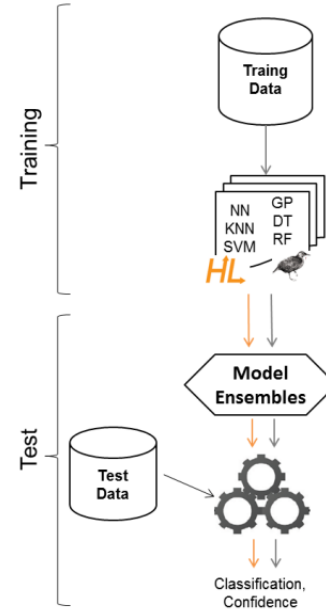


Figure 2: Hybrid ensemble modeling with confidence estimation.

In the empirical tests documented in the following sections we use a confidence threshold  $\theta \in [0 \dots 1]$ . If the confidence for a sample's classification is smaller than this threshold, then there is no estimation statement for this sample. As a consequence, the ratio of samples for which a classification statement is given will be below 100%; we expect this samples coverage ratio to decrease for increasing values of  $\theta$ .

### 4. EMPIRICAL STUDY

For this testing this approach we have used quotes from 10 of the most representative shares in the Spanish market for the years 2003 - 2013. In that period we have identified three types of different market developments: bull market (a long upward price movement where prices in the end of the period are higher than in the beginning), bear market (a long downward price movement where prices in the end are lower than in the beginning), and sideways market (prices in the end and the beginning are almost the same) (Kirkpatrick & Dahlquist 2006).

In the following we describe the executed data preprocessing steps, the target variables for which we trained estimators, and show statistics on the results achieved using the sliding window short term approach with heterogeneous model ensembles.

#### 4.1. Raw Data

We have obtained data from <http://www.Yahoo.com/finances> using the *quantmod* R library (Ryan 2008). This data source contains most of the Spanish stock market data recorded on a daily basis since 2000. For our experiments we have selected a set of relevant stocks listed on the IBEX35 index.

For all stocks we use the following daily available information:

- The *adjusted close price* represents the price of a share at the end of a trading day and takes into account the dividends and splits of that specific share.
- The *volume* represents the number of trading actions that were executed for a share on a trading day.

The companies we have chosen are 10 of the most representative ones in the Spanish stock market for which reliable data can be obtained in a period of 10 years from January 2003 to January 2013. These are:

- Telecommunication: **TEF** (Telefónica)
- Electricity: **ELE** (Endesa)
- Banks: **POP** (Popular), **SAN** (Santander) and **BBV** (Bilbao Vizcaya)
- Oil industry: **REP** (Repsol)
- Construction: **FCC** (Fomento de Construcciones y Contratas) and **SYV** (Sacyr Vallehermoso)
- Insurance: **MAP** (Maphre)
- Engineering: **IDR** (Indra)
- Index: **IBEX35** (summarizes the most capitalized companies in the Spanish stock market; shown in Figure 3)

For these variables the data from January 02, 2003 until January 02, 2013 have been collected; in total, 2548 samples (i.e., data of 2548 days) are available.



Figure 3: The IBEX35 index from 2003.01 to 2013.01

#### 4.2. Data Preprocessing, Definition of Features

For each share  $S$  at time  $t$  we calculate the following features that are used as input for training classifiers:

- $VolStd10(S,t)$ : Standard deviation of the volume of  $S$  in time period  $[t-10, \dots, t]$
- $VolStd25(S,t)$ : Standard deviation of the volume of  $S$  in time period  $[t-25, \dots, t]$
- $PriceAvg10(S,t)$ : Average price of  $S$  in time period  $[t-10, \dots, t]$
- $PriceAvg25(S,t)$ : Average price of  $S$  in time period  $[t-25, \dots, t]$

- $PriceStaq(S,t)$ : Short term averages quotient, calculated as  $PriceAvg25(S,t) / PriceAvg10(S,t)$
- $PriceStd10(S,t)$ : Standard deviation of the price of  $S$  in time period  $[t-10, \dots, t]$
- $PriceStd25(S,t)$ : Standard deviation of the price of  $S$  in time period  $[t-25, \dots, t]$

#### 4.3. Target Variables

For a share  $S$  the goal in this research work is to train classifiers for the following targets:

- $PriceFutureTrend1(S,t)$ : Trend (positive, neutral, or negative) for the price of  $S$  at time  $t+1$  relative to the price of  $S$  at time  $t$
- $PriceFutureTrend5(S,t)$ : Trend (positive, neutral, or negative) for the price of  $S$  at time  $t+5$  relative to the price of  $S$  at time  $t$
- $PriceFutureTrend20(S,t)$ : Trend (positive, neutral, or negative) for the price of  $S$  at time  $t+20$  relative to the price of  $S$  at time  $t$

A trend is classified positive if the difference is at least +1% and negative if the difference is at least -1%.

#### 4.4. Results

For five selected stocks (BBV, ELE, TEF, MAP, and POP) as well as the IBEX35 index we have executed the heterogeneous short term learning and prognosis approach described in Sections 2 and 3; we trained models for the classification of the trend for the next day, the next week (i.e. in five trading days), and the next month (i.e. in 20 trading days). The number of training and test samples in each training / test cycle was set to 100 and 10, respectively. The modeling methods listed in Section 3 were used 20 times independently with varying parameter settings leading to a set of test classification for each sample.

The results for these test series are summarized in Table 1. For varying confidence thresholds  $\theta$  we report classification accuracies and sample coverage ratios.

We see that for all target variables increasing the confidence threshold  $\theta$  leads to increased classification accuracies, but also to decreased sample coverage ratios.

Furthermore, we also see that for all target variables the classification accuracy for the close future's trend ranges from approx. 0.5 (for  $\theta=0$ ) to approx. 0.7 (for  $\theta=1$ ). As expected, the classification accuracies for the trend for the next day are higher than those for the next week. Surprisingly, the accuracies for the next month's trend are for all target variables higher than those for the next week.

We exemplarily show classification rates vs. samples coverage ratios for IBEX in Figure 4. In Figures 5 and 6 we show the classification accuracies and confidences for BBV for the trends for the next day and the next week, respectively. In Figure 7 we show the 30 day moving average of classification confidence, accuracy, and samples coverage for IBEXA prognosis for the next day and  $\theta=0.8$ .

| Target variable | $\theta$ | classification accuracy | samples coverage |
|-----------------|----------|-------------------------|------------------|
| BBV 1d          | 0.0      | 0.490                   | 1.000            |
|                 | 0.5      | 0.493                   | 0.906            |
|                 | 0.8      | 0.543                   | 0.553            |
|                 | 1.0      | 0.603                   | 0.136            |
| BBV 5d          | 0.0      | 0.484                   | 1.000            |
|                 | 0.5      | 0.491                   | 0.936            |
|                 | 0.8      | 0.509                   | 0.694            |
|                 | 1.0      | 0.517                   | 0.227            |
| BBV 20d         | 0.0      | 0.574                   | 1.000            |
|                 | 0.5      | 0.582                   | 0.939            |
|                 | 0.8      | 0.605                   | 0.766            |
|                 | 1.0      | 0.629                   | 0.413            |
| ELE 1d          | 0.0      | 0.494                   | 1.000            |
|                 | 0.5      | 0.514                   | 0.931            |
|                 | 0.8      | 0.561                   | 0.676            |
|                 | 1.0      | 0.668                   | 0.311            |
| ELE 5d          | 0.0      | 0.447                   | 1.000            |
|                 | 0.5      | 0.454                   | 0.944            |
|                 | 0.8      | 0.479                   | 0.672            |
|                 | 1.0      | 0.494                   | 0.242            |
| ELE 20d         | 0.0      | 0.504                   | 1.000            |
|                 | 0.5      | 0.508                   | 0.942            |
|                 | 0.8      | 0.523                   | 0.756            |
|                 | 1.0      | 0.568                   | 0.394            |
| TEF 1d          | 0.0      | 0.507                   | 1.000            |
|                 | 0.5      | 0.521                   | 0.929            |
|                 | 0.8      | 0.562                   | 0.713            |
|                 | 1.0      | 0.633                   | 0.306            |
| TEF 5d          | 0.0      | 0.449                   | 1.000            |
|                 | 0.5      | 0.447                   | 0.931            |
|                 | 0.8      | 0.480                   | 0.660            |
|                 | 1.0      | 0.512                   | 0.188            |
| TEF 20d         | 0.0      | 0.542                   | 1.000            |
|                 | 0.5      | 0.549                   | 0.940            |
|                 | 0.8      | 0.568                   | 0.759            |
|                 | 1.0      | 0.587                   | 0.416            |
| IBEXX 1d        | 0.0      | 0.545                   | 1.000            |
|                 | 0.5      | 0.551                   | 0.928            |
|                 | 0.8      | 0.601                   | 0.707            |
|                 | 1.0      | 0.690                   | 0.388            |
| IBEX 5d         | 0.0      | 0.458                   | 1.000            |
|                 | 0.5      | 0.458                   | 0.934            |
|                 | 0.8      | 0.481                   | 0.653            |
|                 | 1.0      | 0.506                   | 0.190            |
| IBEX 20d        | 0.0      | 0.564                   | 1.000            |
|                 | 0.5      | 0.576                   | 0.937            |
|                 | 0.8      | 0.595                   | 0.753            |
|                 | 1.0      | 0.613                   | 0.396            |
| MAP 1d          | 0.0      | 0.456                   | 1.000            |
|                 | 0.5      | 0.474                   | 0.919            |
|                 | 0.8      | 0.521                   | 0.625            |
|                 | 1.0      | 0.628                   | 0.237            |
| MAP 5d          | 0.0      | 0.484                   | 1.000            |
|                 | 0.5      | 0.489                   | 0.938            |
|                 | 0.8      | 0.509                   | 0.685            |
|                 | 1.0      | 0.575                   | 0.245            |
| MAP 20d         | 0.0      | 0.550                   | 1.000            |
|                 | 0.5      | 0.561                   | 0.943            |
|                 | 0.8      | 0.579                   | 0.793            |
|                 | 1.0      | 0.594                   | 0.461            |
| POP 1d          | 0.0      | 0.460                   | 1.000            |
|                 | 0.5      | 0.472                   | 0.909            |
|                 | 0.8      | 0.540                   | 0.627            |
|                 | 1.0      | 0.672                   | 0.283            |
| POP 5d          | 0.0      | 0.467                   | 1.000            |
|                 | 0.5      | 0.474                   | 0.938            |
|                 | 0.8      | 0.509                   | 0.690            |
|                 | 1.0      | 0.531                   | 0.267            |
| POP 20d         | 0.0      | 0.593                   | 1.000            |
|                 | 0.5      | 0.605                   | 0.942            |
|                 | 0.8      | 0.622                   | 0.776            |
|                 | 1.0      | 0.649                   | 0.445            |

Table 1: Heterogeneous short-term prognosis modeling test results. The confidence threshold  $\theta$  is varied leading to varying classification accuracies and sample coverage ratios.

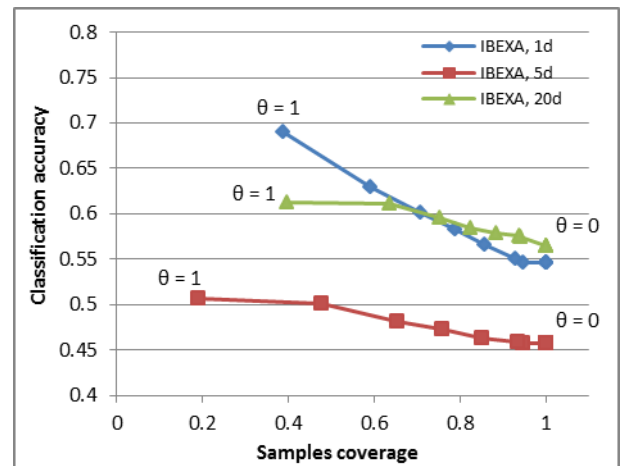


Figure 4: Classification accuracy vs. samples coverage for IBEX trend for the next day, next week, and next month for  $\theta$  varied from 0.0 to 1.0.

## 5. CONCLUSION, OUTLOOK

In this paper we have described the application of heterogeneous short time modeling and prognosis to stock market data. In the test series results we have shown that the use of heterogeneous model ensembles leads to increased classification accuracy for the near future's trend of stocks and indices; this can be increased even more by increasing the threshold for the classification confidence, bearing in mind that this decreases the ratio of samples for which a classification statement can be given.

Future work shall focus on the application of this approach to data of other stock markets as well as the identification of variable impacts.

## REFERENCES

- 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.
- Bodas-Sagi, D.J., Fernández-Blanco, P., Hidalgo, J. I., & Soltero-Domingo F.J. (2012) A parallel evolutionary algorithm for technical market indicators optimization. *Natural Computing*, Vol. 12 (2), pp 195-207. Springer.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45 (1): 5-32.
- Chang, C.-C., & Lin, C.-J. (2001). *LIBSVM: a library for support vector machines*. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification*. Wiley Interscience, 2nd edition.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten I.H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, 11(1).
- Kaboudan, M. (2002) GP forecasts of stock prices for profitable trading. *Evolutionary computation in economics and finance*, pp. 359-382. Springer.

- Kirkpatrick, C.D. & Dahlquist, J. (2006) *Technical Analysis: The Complete Resource for Financial Market Technicians*. Financial Times Press.
- Kotsiantis, S. (2007) Supervised machine learning: A review of classification techniques. *Informatica*, 31:249–268.
- Koza, J. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. University of Michigan Press.
- Ljung, L. (1999). *System Identification – Theory For the User*, 2nd edition. PTR Prentice Hall, Upper Saddle River, N.J.
- Mallick, D., Lee, V.C., Yew Soon Ong (2008). An empirical study of Genetic Programming generated trading rules in computerized stock trading service system. *Proceedings of the 2008 International Conference on Service Systems and Service Management*, pp. 1-6.
- Nelles, O. (2001). *Nonlinear System Identification*. Springer Verlag, Berlin Heidelberg New York.
- Potvina, J.-Y., Soriano, P., Vallee, M. (2004) Generating trading rules on the stock markets with genetic programming. *Computers & Operations Research* 31(7), 1033-1047. Elsevier.
- Ryan, J. A. (2008) Quantmod: Quantitative Financial Modeling Framework. *Google academy*.
- Summers, B., Griffithsz, E., & Hudson, R. (2004) Back to the future: an empirical investigation into the validity of stock index models over time. *Appl. Fin. Econ.* 14, 209-214.
- Vapnik, V. (1998). *Statistical Learning Theory*. Wiley, New York.
- Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., Dorfer, V., & Affenzeller, M. (2014) Architecture and Design of the HeuristicLab Optimization Environment. In *Advanced Methods and Applications in Computational Intelligence, Topics in Intelligent Engineering and Informatics Series*, pp. 197-261. Springer
- Winkler, S. (2009). *Evolutionary System Identification - Modern Concepts and Practical Applications*. Schriften der Johannes Kepler Universität Linz, Reihe C: Technik und Naturwissenschaften. Universitätsverlag Rudolf Trauner. ISBN 978-3-85499-569-2.
- Winkler, S.M., Schaller, S., Dorfer, V., Affenzeller, M., Petz, G., & Karpowicz, M. (2014) Data based prediction of sentiments using heterogeneous model ensembles. Submitted to *Soft Computing*.
- Winkler, S.M., Affenzeller, M., Schaller, S., & Stekel, H. (2014) Data Based Prediction of Cancer Diagnoses Using Heterogeneous Model Ensembles - A Case Study for Breast Cancer, Melanoma, and Cancer in the Respiratory System. In *Proceedings of the GECCO 2014 Workshop on Medical Applications of Genetic and Evolutionary Computation (MedGEC 2014)*, 2014.

## AUTHORS BIOGRAPHIES

**STEPHAN M. WINKLER** received his PhD in engineering sciences in 2008 from Johannes Kepler University (JKU) Linz, Austria. His research interests include genetic programming, nonlinear model identification and machine learning. Since 2009, Dr. Winkler is professor at the Department for Medical and Bioinformatics at the University of Applied Sciences (UAS) Upper Austria at Hagenberg Campus; since 2010, Dr. Winkler is head of the Bioinformatics Research Group at UAS, Hagenberg.

**BONIFACIO CASTAÑO** received his PhD in Physics from Complutense University of Madrid (UCM) in 1985. He has collaborated with multiple international research organizations, for example with CERN in Geneva, where he worked as a research fellow in High Energy Physics from 1985 to 1986. Since 1989 he is staff professor at the Department of Physics and Mathematics at the Alcalá University of Madrid (UAH). His research interests include computer science, wireless communications, robotics, and genetic programming applied to stocks markets analysis.

**SERGIO LUENGO**'s received a degree in Computer Science from the University of Alcalá de Henares (UAH), Spain, in 2013. Currently he is M.Sc. student in Artificial Intelligence at the Technical University of Madrid (UPM) and is a researcher at the Computational Intelligence Group in UPM.

**SUSANNE SCHALLER** received her Masters of Science in Biomedical Informatics in 2009 and 2010 from the University of Skoevde and Upper Austrian University of Applied Sciences. She is a research and teaching associate at the Research & Development GmbH, in Hagenberg. Her research interests include system biology, medical image analysis, tissue engineering, data mining and machine learning.

**GABRIEL KRONBERGER** received his PhD in engineering sciences in 2010 from JKU Linz, Austria, and is a research associate at the UAS Research Center Hagenberg. His research interests include genetic programming, machine learning, and data mining and knowledge discovery.

**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.

Figure 5: Classification accuracies for the next day's and next week's trend for BBV over the analyzed time period. The confidence threshold  $\theta$  set to 0.8. Classification accuracies for the next day's trend decrease after 2007.

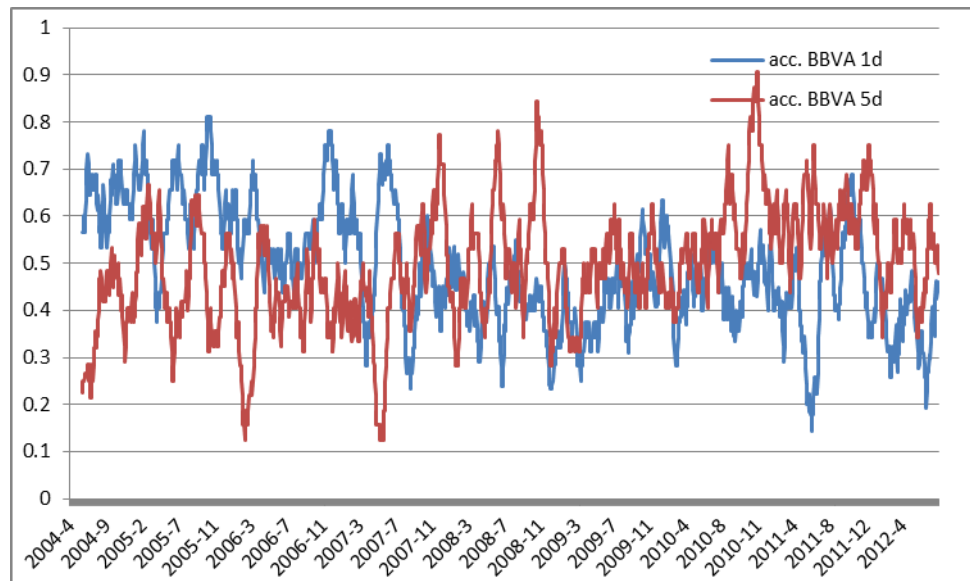


Figure 6: Classification confidence for the next day's and next week's trend for BBV over the analyzed time period. We see that the confidences for both the next day's and the next weeks trends are rather high until 2007; after 2007 the confidence for the next day's classification decreases which can be interpreted as an indicator for changes in the analyzed system.

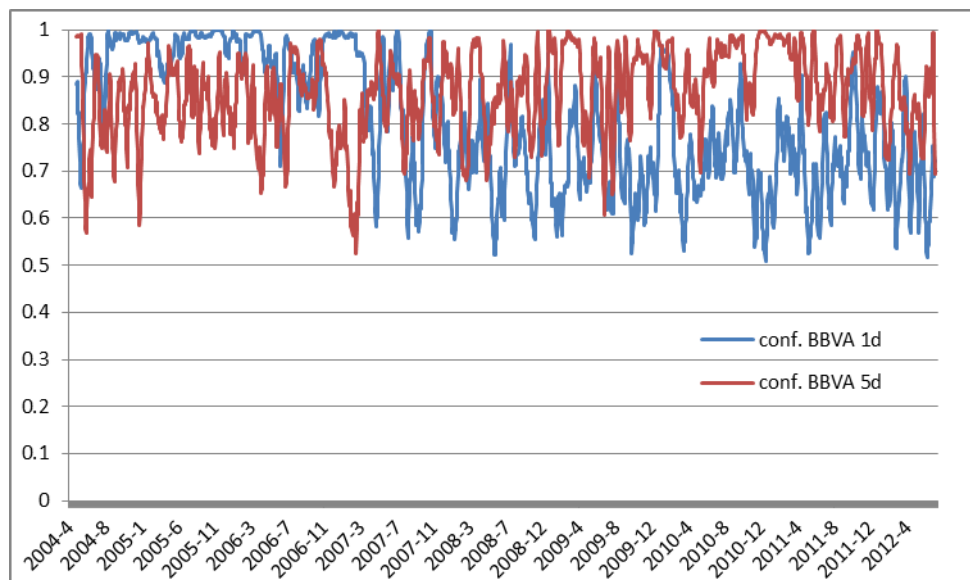
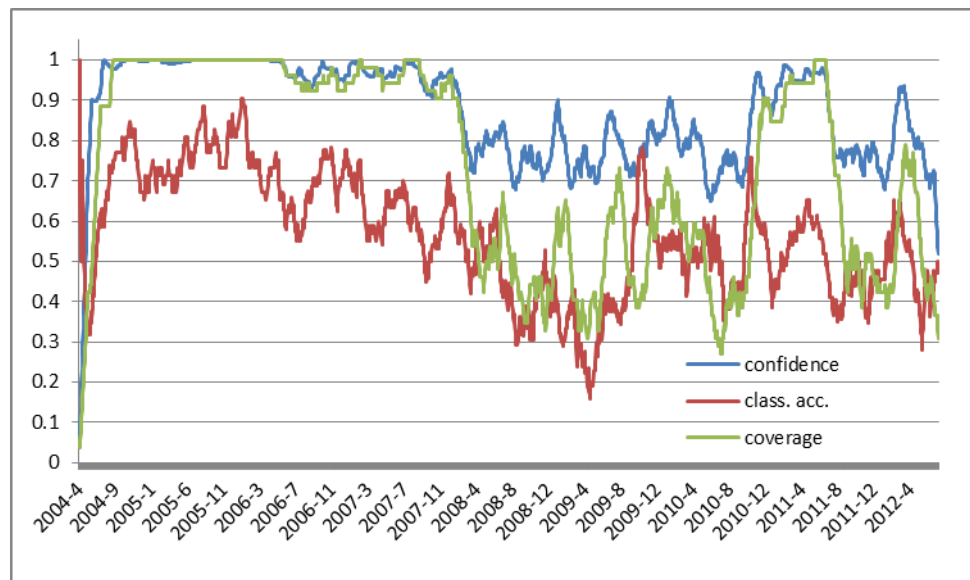


Figure 7: Classification confidence, accuracy, and samples coverage for IBEXA prognosis for the next day ( $\theta=0.8$ ). We here again see that classification confidence and samples coverage decrease after 2007 which can be interpreted as an indicator for changes in the analyzed system.





# HIGH LEVEL ARCHITECTURE VIRTUAL ASSISTANT FRAMEWORK

Josef Brožek<sup>(a)</sup>, Bhakti Stephan Onggo<sup>(b)</sup>, Antonín Kavička<sup>(c)</sup>

<sup>(a)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice, Czech republic

<sup>(b)</sup> Lancaster University Management School, University of Lancaster, United Kingdom

<sup>(c)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice, Czech republic

<sup>(a)</sup> [mail@jobro.cz](mailto:mail@jobro.cz), <sup>(b)</sup> [s.onggo@lancaster.ac.uk](mailto:s.onggo@lancaster.ac.uk), <sup>(c)</sup> [antonin.kavicka@upce.cz](mailto:antonin.kavicka@upce.cz)

## ABSTRACT

High Level Architecture provides a standard for abstraction, design, construction, development and operation of distributed computer simulation systems. This paper focuses on the latest version of the standard, i.e. IEEE1516:2010. Users who are not familiar with distributed computer programming may find it difficult to create HLA-compliant models using a tool that assumes some familiarities with distributed computer programming. Hence, a tool that can help such users is useful, especially in encouraging more people to develop HLA-compliant models.

We have developed HLAVA Framework that encapsulates the detailed steps of main HLA interfaces such as the methods for creating logical processes, and management thereof, synchronization methods, and communication protocols - in fact, the framework is a simulation kernel for the distributed simulation logical process, which is compliant with (among others) the HLA standard, but due to a simple interface has only about 10 methods, which have done creation of (potentially) much simpler.

Keywords: HLA, High Level Architecture, Framework, simplification interface, distributed simulation, traffic simulation

## 1. MOTIVATION

High Level Architecture (HLA) is a standard for the creation and operation of distributed computer simulation systems. The standard does not limit the application domain and simulation modelling method. Hence, HLA has been applied to domains such as military, healthcare, supply chain and many more. HLA has also been used to link models developed using various simulation modelling methods such as discrete-event and agent-based. At a lower level, the standard specifies the method of communication between the nodes (or federates) of a distributed simulation. The standard sets the requirements for the actual transmission format using XML. This enables us to link simulation models written using different programming languages or simulation software. It also enables us to

write a wrapper for legacy simulation models (non HLA compliant) by converting the input-output data into an XML format.

The openness, flexibility and many functionalities of the standard may overwhelm novice users or users who are not familiar with distributed computer programming. The main contribution of this paper is to propose a framework that helps users to develop HLA-compliant models more easily so that they can use the full potential of HLA with minimal effort (at the cost of moderate restrictions which will be discussed later). This research is partly motivated by the lack of adoption of HLA in the Czech Republic. We hope this work may attract more people to develop HLA-compliant models. The remainder of this paper is organised as follows.

## 2. CHARACTERISTICS OF HLA

This section provides an overview of HLA. The explanation is based on Fujimoto (2000), Kuhl, Dahmann, Weatherly (2000) and standards IEEE1516:2010 (2010). Readers who are familiar with HLA may skip this section.

### 2.1. History of HLA

HLA was originally created for the military needs of the United States of America Department of Defense (more specifically, it was developed at the Defense Modeling and Simulation Office), where there was a motivation for linking various simulation models into a bigger and more complex simulator. HLA, which was handed over to the army in 1995, met all the military requirements. It enabled the interconnection between different models implemented in different programming languages. It also enabled the simulation to be used in a human-in-the-loop training exercises using the simulation models.

In 1997, the Army released the standard for public use; HLA 1.0 standard was thus issued, which began to be used in universities and space research. Gradually, the standard began to be applied in the private sector too. In 1998, HLA was included in the NATO military simulation standard (STANAG 4603 standard, which has since been updated). The standard was approved in 2000 (IEEE 1516). Another important milestone was a

major revision that was already based on real experience, not only in the military but also in the private sector and the university sector, taking place in 2010. HLA-Evolved (IEEE 1516:2010) was thus created. Example of solution on Roberto, Sala-Diakanda, Pastrana and col (2013).

## 2.2. Basic Terminology

### 2.2.1. Federation

A federation in HLA refers to an entire distributed simulation (or a complex distributed simulation model). There is one federation during one distributed simulation experiment.

### 2.2.2. Federate

A federate is an HLA-compliant application that can participate in a distributed simulation experiment. A federate may be in a form of a simulation model, software application, software agent of any type, an input sensor or panel, a display unit, a system for retrieving historical data, etc.

### 2.2.3. Objects

An object is the information to be exchanged during a distributed simulation experiment. In HLA, an object represents an entity with persistent states.

### 2.2.4. Attributes

An object has a set of attributes that are relevant to an entity represented by the object. An attribute is basically a data field of a defined data type. Basic data types and attributes are defined in the Object Model Template (OMT). We will explain OMT in Section 2.2.6.

### 2.2.5. Interactions

An interaction represents an event that may be of interest to two or more federates. Similar to objects, interactions must be pre-defined. Each interaction can have a set of parameters. Interaction can be performed directly through interaction classes, e.g. passing an object to another federate etc.

Federates cannot send interactions to each other directly, but they must make a request to RTI (runtime infrastructure) as shown in Figure 1. We will explain RTI in Section 2.2.7.

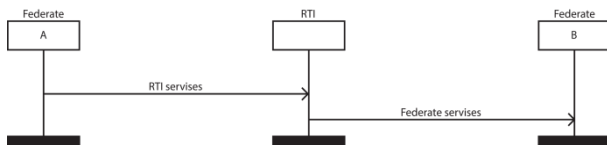


Figure 1: Interactions in HLA

### 2.2.6. Object Model Template (OMT)

Objects and their attributes as well as interactions and their parameters need to be defined for the whole federation based on the standard set in the OMT (IEEE1516.2:2010). The definitions for Objects and Interactions must be provided in a Federation Object Model (FOM) or Simulation Object Model (SOM).

Both are XML documents and are relatively easy to read. In practice, FOM is often built from one or more SOMs.

### 2.2.7. RTI

RTI (run-time infrastructure) is a middleware that provides communication services to all federates in a federation. It is defined directly by the standard (IEEE1516.3:2010), independent of the platform and language. RTI provides APIs that can be called to perform certain functions such as creating a federation, creating a federate, connecting a federate to the federation, etc. RTI needs FOM to understand the objects and interactions that will be exchanged between federates. Common terminology and the logic of RTI are shown in Figure 2.

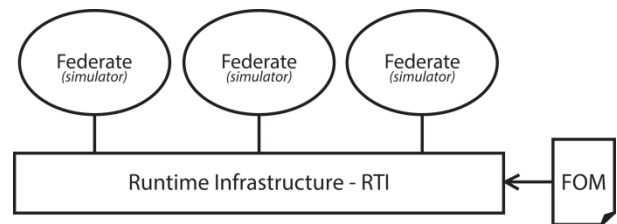


Figure 2: HLA RTI Diagram

A number of vendors have implemented RTI products, for example Pitch RTI, MAK RTI, OpenHLA and many more. In principle, the implementation approach is almost the same as the one shown in Figure 3. A federate usually runs on one computing node (but one node can be shared by multiple federates). For each federate, we need to include the local RTI component. It is a library that needs to be linked to the actual software solution (federates) or to the application wrapper (for legacy or non-compliant software). This component actively manages communication between a federate and a central RTI component. The central RTI component is the busiest part in a federation<sup>1</sup>. It is responsible for managing the communication within a federation, for example, connecting a federate to a federation, managing objects, etc.

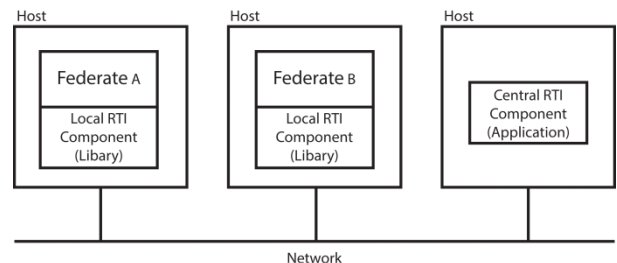


Figure 3: RTI physical implementation

<sup>1</sup> However, it is necessary to note that it is possible to create other than simply linear non-hierarchical simulation models, which makes it possible to use a number of central RTI components in a single simulation.

### 2.3. Federation Rules

HLA defines 5 rules for federations and 5 rules for federates. These rules must be met even in a situation where we want to declare a model to be partially HLA compliant. In other words, compliance with all ten rules is a necessary condition for us to refer to a simulation model as HLA compliant.

The rules for federations are as follows:

- Federations shall have an HLA federation object model (FOM), documented in accordance with the HLA object model template (OMT).
- All representation of objects in the FOM shall be in the federates, not in the run-time infrastructure (RTI)<sup>2</sup>.
- All exchange of FOM data among federates shall occur via the RTI<sup>3</sup>.
- Federates shall interact with the run-time infrastructure (RTI) in accordance with the HLA interface specification.
- An attribute of an instance of an object shall be owned by only one federate at any given time. Only this federate is authorized to modify the values of the owned element.

The rules for federates are as follows:

- Federates shall have an HLA simulation object model (SOM), documented in accordance with the HLA object model template (OMT).
- Federates shall be able to update and reflect any attributes of objects in their SOM and send and receive SOM object interactions externally, as specified in their SOM.
- Federates shall be able to transfer and/or accept ownership of an attribute dynamically during a federation execution, as specified in their SOM.
- Federates shall be able to vary the conditions under which they provide updates of attributes of objects, as specified in their SOM.
- Federates shall be able to manage local time in coordination with the RTI requirements (in effect, in coordination with other federates).

<sup>2</sup> The consequence of this rule is the fact that the transfer of the object from one federate to another is performed so that federate1 requests RTI for placing its object in federate2. RTI locks the object in federate1 before the changes and requests federate2 for placing the new object (while passing the object itself). If the location is in federate2, the object is removed from federate1. If not, the object is unlocked in federate1.

<sup>3</sup> The consequence of this rule is that RTI can be equipped with a software upgrade that allows you to track all interactions and potentially current states of most objects.

### 2.4. Synchronization Methods

Synchronization is provided through RTI, it is therefore not necessary to create additional mechanisms that are normally used in the architecture for distributed simulation models.

The actual synchronization and its algorithms are transparent to the user, we can simply use the services provided. If we want a federate to be synchronized with other federate, we can set which federate is time-regulating and which federate is time-constrained. Alternatively, we can run all federates in a synchronous mode.

### 2.5. HLA Software Solution

Therefore, to easily understand the benefits of an own framework, it is necessary to specify how the application issue is dealt with.

#### 2.5.1. General Principle

We can design a federate by creating a class that implements (or inherits) one of the federate type classes (for example, NullFederate). This will automatically give the federate access to a number of RTI services such as creating a federation or joining a federation.

Figure 4 shows the details on what happen if a class federate is instantiated during a distributed simulation experiment. First, a pair RTI Ambassador – Federate ambassador will be instantiated for each federate instance. These two components manage the communication between a federate and an RTI. RTI Ambassador is used by the federate to communicate with RTI. The Federate Ambassador is used by the RTI to call any of the callback methods defined in the federate.

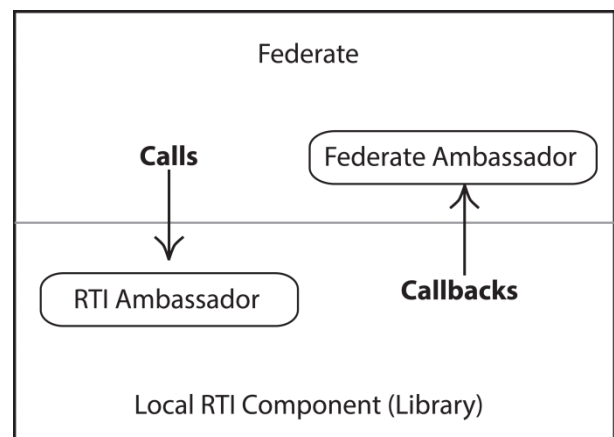


Figure 4: Functioning of local RTI

#### 2.5.2. Federate Ambassador

Federate Ambassador is located in same place as the instance of a running federate. It is generated by default directly when the federate is instantiated. Federate Ambassador allows RTI to invoke a callback method defined in the federate. Since a federate will have access to FOM, it allows the federate to implement appropriate behaviour or to appropriately respond to any callback.



### 2.5.3. RTI Ambassador

RTI Ambassador is the implementation of a local RTI component discussed previously in Figure 3. During a distributed simulation experiment, each RTI Ambassador is physically located in the same compute node as the corresponding federate. We can use RTI Ambassador to create a federation or to join a federation. It also provides services to allow a federate to communicate with the RTI (i.e., the global RTI component). HLA provides standard for interfaces that have to be implemented in RTI Ambassador and Federate Ambassador which is independent of any RTI software implementation. This standard ensures platform independence<sup>4</sup>.

### 2.5.4. Global RTI Component

The second essential part of RTI that provides communication between RTI ambassadors is the global RTI component. In other words, RTI ambassadors do not communicate directly but via a global RTI component (i.e. the principle described in Section 2.5.3 is valid for some implementations of RTI that do not contain a global RTI component). The global RTI component provides forwarding to all federates that have requested specific information such as an attribute or an interaction. In addition to this forwarding function, this component also provides synchronization, monitoring, administration of the federation and many other supporting functions.

## 2.6. Overview of Existing RTI software

This section provides an introduction to a number of existing RTI software. Generally, it is possible to characterize various implementations of RTI by specifying whether it is an open or a closed source product; commercial or freely available product; and then according to the HLA standard that the RTI complies with.

### 2.6.1. Pitch RTI

The commercial product of the Swedish company Pitch Technologies (belonging to the BAE Systems Group) is one of the most powerful variants of RTI, meeting all the HLA standards and is developed with a lot of utilities, more at Pitch websites (2014).

It is very easy to use. It is supplied with libraries for Java and C++ languages (also applicable in C#). To use it, we just need to import a relevant library and then create a sequence of desired objects and serve a quantity of functions.

---

<sup>4</sup> If both federates have knowledge of OMT, they are able to exchange information with each other through XML. This information can then be forwarded to a federate and RTI (it is worth noting that during the decomposition of the communication process RTI is treated as an active software component, not only as a communication bus - even though it may initially appear so).

Pitch RTI also provides a trial version in which only two federates can run. There may be other limitations that we are not aware of. Currently, this product is used at the University of Pardubice; the proposed software library HLA-VA is designed for Pitch RTI (although the job to write it for a different RTI is trivial).

### 2.6.2. MÄK RTI

Similar to Pitch RTI, MÄK RTI (more information at MAK websites (2014)) is one of the most widely used commercial RTIs. This software also comes with a trial version in which the maximum number of federates that can run is two. It is a very good and very powerful product that meets the latest standards that is commonly used in the private sector and universities. It supports programs written in Java and C++.

### 2.6.3. Open HLA

Open HLA is an open source software distributed under the Apache license more at Geeknet (2005). It is a relatively good product for basic experimentation with HLA.

At the time of writing, the development has stopped (i.e. there are no new updates). The latest version does not support all the functionalities that are required by the standard IEEE1516:2010 (e.g. models with optimistic synchronization methods cannot run in Open HLA).

Open HLA launching and compiling scripts are written with the support of the library Ant. It is therefore necessary to have this library installed in the computer (this fact also makes it different from other software where there is an RTI executable .exe file).

### 2.6.4. CERTI

CERTI is an open source software which is distributed under the GPL (or LGPL) license more at Free software foundation web(2002). Its functionalities are better than that of Open HLA; It fully supports the HLA 1.3 standard for Java and C++. However, the support for IEEE 1516:2000 is only partially supported and only supports C++.

### 2.6.5. Other Products

There is a large amount of RTI implementations including products of the U.S. Army and the Chinese People's Liberation Army. Reportedly, RTIs, developed by the Chinese public sector are very good but due to language barriers the author was not able to get acquainted with them. A larger, though not entirely up-to-date, overview of individual RTIs can be found in older resources such as Knight, Corder, Liedel (2001) and report from The simulation interoperability standards organization (2001). A more up-to-date review can be found at Wikipedia (2014), as this resource is updated relatively more regularly to provide the most comprehensive overview of the existing solutions.

## 3. HLA Virtual Assistant (HLA-VA) Framework

The main motivation for the development of the "virtual assistant" framework is to facilitate easy access to HLA. In this section, we will explain the design and components of the HLA-VA framework, its benefits and limitations.

### 3.1. Framework Architecture

The framework is built over the libraries that are supplied as a standard to RTI (different RTIs from different vendors have different control libraries). The block diagram of the architecture is shown in Figure 5. The benefit of the design of this framework is that it is possible to replace the relevant control library (the bottom layer in Figure 5) without the need to change the upper layers. The explanation for each layer is as follows.

#### 3.1.1. Isolation of the RTI Library

The actual RTI library is isolated from the rest of the code in the framework via an interface. Parts of the RTI component that are not supplied as a standard in the RTI library are implemented separately. In theory, this design allows us to use an RTI from any provider<sup>5</sup>.

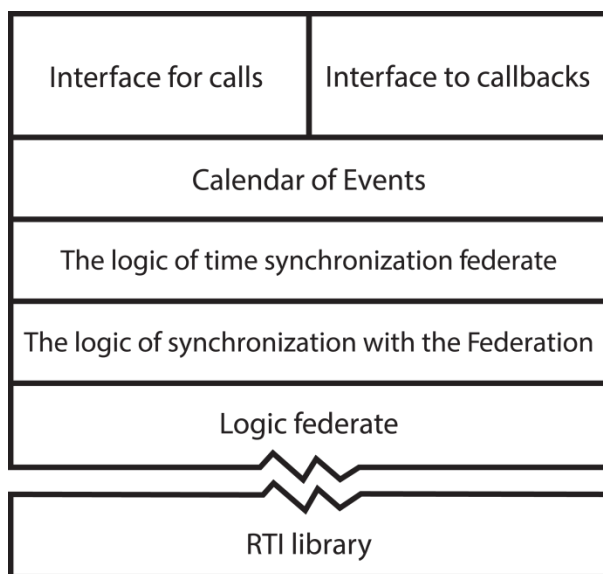


Figure 5: Block diagram of the framework

#### 3.1.2. Logic of the Federate

This layer deals with the communication between the global RTI component and the local federate. One of the main role of this layer is to make the definition of the information exchanged during distributed simulation execution easier.

<sup>5</sup> We have tested our framework with MAK RTI, open HLA and Pitch RTI. Hence, we are confident that our framework can be used with any compliant RTI solution.

When we write an HLA-compliant model, we often need to deal dynamics arguments<sup>6</sup>. Each message (Objects or Interactions) in HLA is strongly typed and controlled by RTI (as specified in FOM or SOM). If we would like a federate to react upon an incoming message, we need to write a callback method.

The problem is solved by a framework at the program level. Immediately after running a local instance of the framework, it passes through FOM and SOM and selects the interaction classes and types. Consequently, the framework stores them in the hash table.

When a message from RTI arrives, the framework compares its hash fingerprint with those of all message types in the Federation (stored in the hash table) and finds the relevant type. When we identify the type of message, we can start reading - that is, determining the number and type of arguments, their translation from the stream, and select individual data components.

As the number of the existing message types compared to the number of individual sendings is negligible, this method is computationally efficient (the approximated complexity of the search is smaller than the subsequent translation of the data from the stream).

Since we have to assume that the individual Federates can be implemented in different programming languages, it is impossible to rely on programming techniques, such as Reflection, and problems need to be solved by implementation as such.

Hence, we can edit the FOM/SOM and send / receive messages with dynamic arguments without the need to change the code for our federate.

#### 3.1.3. Logic of Synchronization with the Federation

This layer and layer indicated in section 3.1.4 are responsible for time synchronization. Central component of cooperation between both the time synchronizations and also the part available to the user is called Calendar of events (see section 3.1.5.).

Specifically, Logic of Synchronization with the Federation ensures proper functioning of the time synchronization federate (on which it runs) with the Federation. It encapsulate self-programmed and also standard methods specified by HLA enabling the framework to fully control the synchronization (using

<sup>6</sup> Standard communications (or more precisely in the HLA language "Interaction Objects") which takes place within HLA, is defined in the XML format, wherein it differs for different types of communication primarily in the number and types of arguments (however, it is not impossible that messages will have the same types and the number of arguments). However, programming languages often have a problem with coding sufficiently general programs so that we could control incoming communications regardless of the type and number of arguments they have - at the same time, a successful control requirement is also the decoding of messages by type and passing thereof on. This problem is possible to partially solve by using of Lambda expressions, but they are not available in a lot of programming languages.

only the knowledge of the desired synchronization procedure.). This layer passes all the messages (incoming and outgoing) marked as timed and solves the outwards federate requirements for time changes. It is the responsibility of the layer to respond to the changes in GST to be enacted by RTI.

When conservative synchronization is used, each federation makes a request for the local virtual time and after all federates are ready, RTI orders to go to the next lowest required LVT (this is how the easiest global synchronization solved by the methodology with zero look ahead works – more about conservative synchronization with zero look ahead in Fujimoto (2000)). It should be noted that, as standard, the information about the transition to LVT is sent only to those federates that have requested it. The logic of time synchronization of federations is thus responsible not only for forwarding the requests for shifting the time and information about the federate being ready to shift the time, but it also monitors time changes taking place in other federates. In addition, if the mode with look ahead is activated, it provides automatic adjustment of look ahead so that it is as advantageous as possible for the federate and the federation.

### **3.1.4. Logic of Time Synchronization of the Federate**

This layer is responsible for setting up the mode of which a federate runs, for example, real time run, run without time synchronization (e.g., for modelling and use of Monte Carlo simulation), standard mode (i.e., maximum speed run) or speed limits (e.g. for visualization).

This layer is responsible for setting up the mode in which a federate runs. For example, the real time run, run synchronization free run (e.g. for modelling and use of Monte Carlo method), standard mode (i.e. maximum speed run) or speed limits (e.g. for visualization).

The layer consists of two major parts: the application logic running in a standard fibre framework and the time base running in its thread.

The time base may not be used in all the simulation types. However, for some runs (e.g. the real time) it is required because it is the only part that cannot be influenced computationally and the time is the most accurate.

The application part of the layer provides for the synchronization of the local level - monitoring requirements from the user (as inserted into the Calendar of events, and if needed, passes the requirements for the Synchronization Layer with the Federation) serving the requirements of the Federation (by entering them into the Calendar of events) and provides the synchronization with independent time base.

### **3.1.5. Calendar of Events**

It is a standard calendar of events represented by the priority queue that is primarily solved by secured transactions to avoid the risk of inserting and removing

across the threads. It is the same as in a monolithic simulator.

Due to the similarity with a monolithic simulator and because there exists only a set of basic user operations such as "Add Event", "Take event" and "Cancel the scheduled event" that makes the entire interface framework very simple.

The behaviour of the other two layers that provide for the synchronization (see above) is transparent to the user. Despite the simple interface, it is possible to reach the synchronization with any time-base derived from real time and with the Federation.

### **3.1.6. Interface for Calls**

It creates a space for calling basic methods necessary for simulation. These include insertion of events into the calendar of events, change of registration in the calendar of events, sending an object or a message to another federate.

Messages are always defined generically to make it possible to insert any type of object into an argument. This object is then the return type of callback.

### **3.1.7. Interface for Callbacks**

It contains Callbacks that must be removed. By default, all Callbacks are implemented the design pattern Observer, it is thus possible to connect to them any number of classes that will process the data.

Callbacks can be viewed as events. The only difference is in the way that leads to evocation. Callbacks are always dependent on the call and they do not occur unbidden.

Basic Callbacks include time change call, incoming message call, incoming object call and information call.

## **3.2. Services Provided**

Basic services that are provided by the framework can thus be summarized as services of the simulation kernel, which includes the time calendar and is able to synchronize the local simulation by the event scanning method.

In addition, support is included for synchronization within distributed simulations. It is possible to forward messages, synchronize the local virtual time with other federates as the global virtual time of the whole federation.

The first major consequence of the creation of the simulation kernel is facilitation of the creation of simulation models for developers who would opt to use this kernel.

Moreover, it is possible to use only a few other methods, and the whole principle of the framework allows developers to create more federates that will communicate with each other and synchronize their time. Therefore, it is possible to create distributed simulation models. At the same time, all models created comply with the HLA standard. The consequence of this fact is that it is possible to connect other simulators implemented without the use of the HLA-VA technology.

### 3.3. Comparison of Demands of the Development

When comparing individual development methodologies, we assume a distributed simulation model with several federates, while creation of the application is assumed from the very beginning.

#### 3.3.1. Development for HLA without the Framework

Even if we use the absolutely simplest applications developed to comply with the HLA recommendations, namely to operate on the simplest RTI, we need to perform the following steps:

1. Create OMT using a specialized tool, or using a text editor. It is created in XML.
2. Connect all the necessary libraries to the program.
3. Create objects that will implement all methods of the general definition Federate (or NullFederate).
4. Create methods for connecting to the federation, identification of federates and other utility methods.
5. Create methods for operating each individual interaction.
6. Create synchronization methods for the actual simulation logic. While it is indeed possible to use synchronization mechanisms provided by RTI between individual federates, it is still necessary to create a mechanism that will ensure the synchronization within the federate. (The advantage of this step is again the potential reuse of the solution, because if the internal synchronization service is implemented sufficiently generally, it can be used also in other solutions). Set rules for synchronization and solution of the actual synchronization within a particular node.
7. Program the application of the specific model (or logical process).

#### 3.3.2. Development with the Framework

Since the framework has an implemented simulation kernel, there is no need to address synchronization, and it is designed to work as an intermediate layer between the application logic and the federate, so the development process changes to:

1. Create OMT specifications through the XML editor or specialized software.
2. Connect the HLA-VA library to the application.
3. Create the logic of own simulation model.

It is very important to find that all the steps that require knowledge of programming associated with RTI have thus been omitted. At the present moment, it is only the first step that requires some knowledge of HLA (nevertheless, this can be circumvented by using a

specialized editor). Still, the development process (especially for a layman) becomes much easier to grasp and implement.

It is this facilitating of the access to HLA that is the greatest advantage of HLA-VA Framework, as it is possible to create HLA-compliant models more quickly. Furthermore, because the RTI library is only connected to framework code in the form of a dynamic link (instead of using static compilation), it is possible to use the framework with (virtually) any RTIs.

### 3.4. Limitations of the Framework

#### 3.4.1. Limitations on the Choice of Synchronization Methods

Currently, from the user's perspective, we can choose from several synchronization methods in the framework. However, physical implementation is limited to the use of RTI services in the field of conservative synchronization, which is sufficient for the class of tasks that the software is focused on, but it is still a certain limitation for the programmer.

#### 3.4.2. Limitations on the Connection to Federates Implemented without the Framework

By default, HLA assumes that any two federates will be able to cooperate provided that they have a relevant operating method implemented. However, as is clear from the preceding paragraph, if we have an existing federate that contains methods for optimistic synchronization, it is not possible to create a federate through the framework that could cooperate.

In other cases, federates implemented through the framework and without it can cooperate.

#### 3.4.3. Limitations on the Choice of Programming Languages

Since the Framework was implemented only in the Java programming language, it enables writing more software components only in Java, C# and C++.

Java support is implicit, and it is possible to link some parts of the framework as libraries.

Support for C# and C++ has been tested in the Windows environment where it was necessary to use Wrapper for the possibility of using the Java code in other programming languages, which can have negative effects on the computational complexity.<sup>7</sup>

---

<sup>7</sup> Experiments confirmed the extension of the computational time if we compare an identical task implemented solely in one programming language and the same solution implemented using Wrapper and combination of Java and C#. The computational complexity, however, did not increase dramatically enough to make it possible not to recommend such a solution. In the demonstration tests, the solution through Wrapper was useful even if we ran a distributed interactive simulation with real-time synchronization.

#### 4. Prototype

In the early stages of the development of the framework, basic functionalities were tested through direct calls. Using these methodologies, validation of the framework was performed, first in the local environment (more federates running on one local node) and subsequently also in a distributed way within the local network. After basic validation tests were completed, it was decided to build several prototypes.

##### 4.1. Prototype: Motorway Rest Areas

The first prototype designed to test synchronization, connection to the visualization components and testing of the development of the framework in various programming languages (Java, C #) was a simple example of motorway rest areas. The demonstration example contains three federates of two kinds. The federate of motorway is used to generate input streams and its logic determines in which direction the vehicles are going and what percentage of them is heading to an exit with a rest area; the physical situation is shown in Figure No. 6.

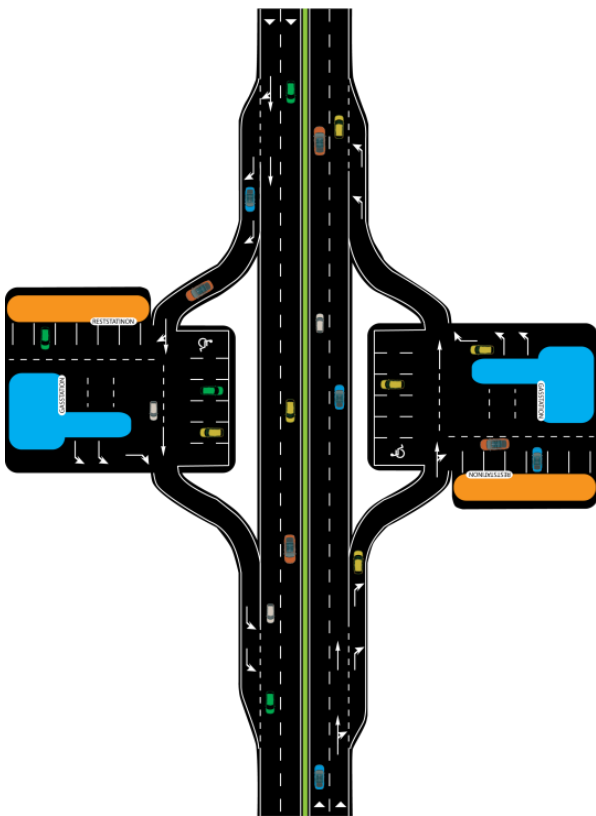


Figure 6: Physical situation

The second type of the federate is a rest area, its internal logic contains only an input from the federate of the highway (similarly, the output is only to the federates of the motorway, it cannot disturb objects by its own) and it is de facto a process that only determines the time during which a given object (in this case, a vehicle) will be delayed. The actual design of the model, including the dividing into individual federates can be seen in Figure 7.

While this may be a relatively trivial demonstration case, by using it, we managed to prove that synchronization solved within the federation at the RTI level (or by calling RTI services) and synchronization solved by default by the framework are functional and the results of their (federates) cooperation are fully in line with expectations..

##### 4.1.1. Testing Methodology

The actual simulation model was programed over the framework, and also in the simulation tool Arena. The outputs of the simulation tool Arena were then taken as the standard.

As for the motorway federate, the selected monitored indicators included verification of the functionality of the generators by comparing individual input streams and also verification of the dividing of the number of vehicles that just pass the section and the number of vehicles that turn off to some of the rest areas (in fact, the number of objects that are passed to the other federates).

As for the system of the rest areas, the selected monitored indicators were the numbers of vehicles and monitoring of their forwarding in time, travel time from the turning to the rest area, up to the arrival and stopping at the rest area (the same for departures) and service/stop time at the rest area.

Results of the solutions within monolithic synchronization in the commercial product Arena and within own programming solutions over the framework reached a high degree of conformity.

Figure 7: Dividing into federates

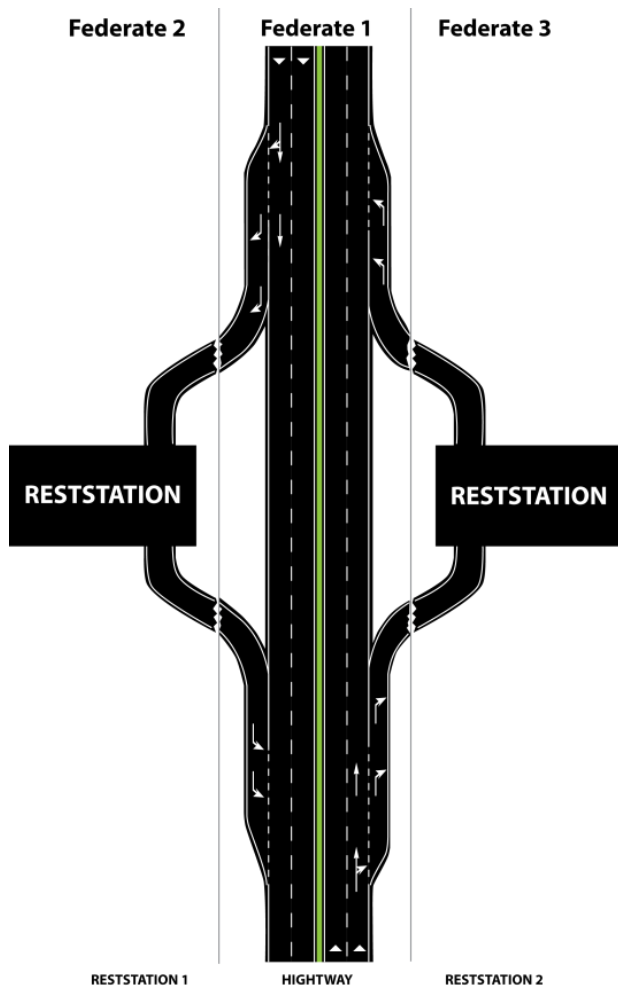


Figure 7: Dividing into federates

#### 4.2. Other Prototypes

For validation and verification of the framework, other prototypes were created reflecting bulk service systems OR queuing systems. The basic principles are, however, very similar to the first demonstration model - they differ only in the area of implementation.

Another suitable prototype model chosen was a multi-storey building whose floors are connected both by a lift and by two staircases. The model is decomposed into three heterogeneous types of logical processes: the operator of the floor (it is a logical process that addresses the movement of people within the individual floors - each floor is operated by one common LP), the operator of vehicles (it is a logical process that addresses the movement between the floors where the objects are passed to the operator of the floor the moment they reach the destination floor) and the operator of doors (addresses the time individual people stay in a particular door - possible to replace with less general terms "surgery", if it was a hospital model, or "shop" if we declared the model to be a department store model. All doors are controlled by only one federate. The specific type of internal doors is entrance doors that serve as a generator).

Another area, for which it was desirable to create a demonstration example, was the possibilities of real-time synchronization and the possibility of interactive interventions into the course of the simulation during its course. Based on these requirements, a third demonstration model was created that reflects the operation in a railway transport system. Federates in this model represent individual stations (these may not be homogeneous) and a special federate (or a set thereof), which is responsible for railway sections between the stations. Operators of each station are enabled to perform interactive interventions into the control of individual signaling devices and railway switches.

For the purpose of this model, it is necessary to say that it is subject to a higher level of abstraction, specifically where it deals with motion characteristics of trains (acceleration, deceleration, driving dynamics). Although the model is designed so that individual characteristics have the potential for future extension so that the system corresponds to reality as much as possible, but currently it uses only the medium speed of the train with an step onset of the speed and step braking distance.

In addition to these examples, the framework is tested in the workplace on a number of demonstration examples ranging from communication testing of communication protocols through an analogue of chat), modelling without time synchronization (testing in turn-based strategy games) and application in the context of larger, potentially real-time simulations.

#### 5. Conclusion

The framework makes it easier for the users to access HLA at the cost of certain restrictions. Prototyping of standard models is easier and faster and, moreover, the user needs no special knowledge of the HLA architecture.

The toll for easy access to the HLA simulations is the limited amount of synchronization methods and some handover protocols.

The framework does not aim to be applicable for large and complex models that consist of high-trained teams - those are assumed to have mastered HLA brilliantly. On the other hand, if a team begins with HLA, or just looks for a methodology through which they would connect existing simulators into the distributed simulation (without trying to specialize purely HLA), or just experiments with simulation technologies (when studying and teaching), this own framework is very useful tool for them as it will help bridge the initial problems with the implementation of distributed simulation models.

Prospects of further development are aimed at the implementation of optimistic synchronization methods and application in practice.



## Acknowledgement

This article was created within the project Support for Internships and Professional Activities in Innovation in Tertiary Education at Jan Perner Transport Faculty and Faculty of Electrical Engineering and Informatics of the University of Pardubice, Reg. No.: CZ.1.07/2.4.00/17.0107.

## References

Free software foundation, Inc, 2002, *CERTI - Summary [Savannah]*. Available from: <http://savannah.nongnu.org/projects/certi> [accessed 22 April 2014].

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].

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].

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.

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].

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.

Vtmäk, 2014, *HLA RTI – Run Time Infrastructure – MÄK RTI*. Available from: <http://www.mak.com/products/link/mak-rti.html> [accessed 22 April 2014].

Wikipedia: the free encyclopedia, 2014, *Run-time infrastructure (simulation)*. Available from: [http://en.wikipedia.org/wiki/Run-time\\_infrastructure\\_%28simulation%29](http://en.wikipedia.org/wiki/Run-time_infrastructure_%28simulation%29) [accessed 4 April 2014].

# MODELLING AND SIMULATION OF THE NON-IDEAL MIXING BEHAVIOUR OF THE CRYSTALLIZATION STAGE IN A SUGAR FACTORY BATCH CRYSTALLIZER

David Tejerina, Rogelio Mazaeda, César de Prada

Department of Systems Engineering and Automatic Control, University of Valladolid  
Address: C/ Real de Burgos s/n, Sede Mergelina EII  
47011 Valladolid, Spain

[david@autom.uva.es](mailto:david@autom.uva.es), [rogelio@cta.uva.es](mailto:rogelio@cta.uva.es), [prada@autom.uva.es](mailto:prada@autom.uva.es)

## ABSTRACT

A first principles model of a semi-batch crystallizer is proposed, taking into account the non-ideal mixing behaviour of the process. For this purpose, a two-compartment model is developed in order to predict the evolution of the variables related to the sugar crystallization in a more realistic way. The model is based on mass and energy balance equations. It also includes population balance equations to keep track of the sugar crystal size distribution (CSD). Finally, the model is adjusted and validated with real factory data.

Keywords: Sugar crystallization; Dynamic modelling; Crystal Size Distribution (CSD); Process simulation

## 1. INTRODUCTION

The development of a suitable process simulation model is a difficult engineering task, due to the complexity of industrial processes, the uncertainty related to them and, in some cases, the lack of information about some important variables.

In the case of crystallization processes, one of the challenges is associated with sensor limitations, considering that super-saturation, for instance, is not directly measured. In this context, the design of model-based soft-sensors plays an important role (Damour et al, 2010).

This paper presents an extension of a previously developed model of a beet sugar crystallizer (Mazaeda, 2010). The model presented in this work considered perfect mixing, but this assumption is rarely true for an industrial-scale crystallizer. In fact, a significant source of uncertainty in industrial crystallizers is associated with mixing (Braatz, 2002). For this reason, the new approach in this contribution is oriented to capture the non-ideal mixing behaviour of the massecuite that takes place in actual industrial practice, in order to improve the fidelity of the model.

Besides, this model is adjusted and validated, in a practical operation range, from data obtained from a sugar factory in Spain. The aim of this model will be to perform an optimal control of the process.

This paper is organized as follows: Section 2 focuses on the description of the process. Then, the

improved model is presented in Section 3. The results obtained using the model comparing them against industrial data are presented in Section 4, ending with some conclusions in Section 5.

## 2. SUGAR CRYSTALLIZATION PROCESS

Crystallization process can be considered as the key stage of sugar production (Van der Poel, 1998). It takes place in evaporative crystallizers, called “tachas”, operating in semi-batch mode, where each cycle is divided into several sequential stages.

As many batch processes, sugar crystallization is controlled by a recipe. In this paper, we are going to make reference to the specific way of controlling the strike of a particular factory in Spain.

Initially, a reference pressure of 0.20-0.25 bar is established in the syrup chamber in order to reduce the boiling point, thus preventing the thermal decomposition of the sucrose.

Once the vacuum pressure is established, the pan's chamber is filled with the sucrose technical solution, up to a pre-specified level. This initial charge of liquor is concentrated in the next stage, by evaporating the amount of water needed to achieve the required sucrose super-saturation, which is the driving force of the crystallization process. The energy required for evaporation is delivered by means of a heat exchanger or calandria, which uses, as heating source, saturated steam from other sections of the factory.

Once the specified concentration or Brix is reached, it is assumed that the correct super-saturation has been obtained, so that it is safe to introduce a small amount of seed crystals, with an average size of about 10 micrometres.

Then, the dissolved sucrose begins to crystallize around this initial population of crystals, resulting in their growth in the subsequent boiling stage. The crystallization will then proceed smoothly along this latter stage, the most important and lengthiest of the batch cycle, provided that the super-saturation of the solution is kept at correct values. Super-saturation, which is defined as the ratio of the actual concentration of sucrose to the one defining the solubility, is the key variable governing the growing of the grains. It



depends, obviously, on the concentration and temperature, but also on the purity of the mother liquor. The objective of the boiling stage is to deliver a uniform population of sugar grains with an average size of around 550  $\mu\text{m}$  and a coefficient of variation (CV) that should, typically, be below 30%. This can be accomplished only if the mass, size and uniformity of the initial seed is carefully considered and the super-saturation of the mother liquor is kept in the meta-stable zone of the Ostwald-Miers Diagram (see Fig. 5), avoiding the spontaneous nucleation (formation of new crystals) or the dissolution of the existing grains that would occur otherwise.

But super-saturation is not a directly measured variable since it has to be inferred from other measurements. In the industrial crystallizer here taken as reference, the crystal growing process is conducted by carefully balancing the rate of water evaporation with the feeding of the right amount of syrup to compensate the sucrose that migrates from the mother liquor to the crystal phase (Mazaeda, 2010). This is achieved by controlling the Brix or concentration of the massecuite, implementing a so called boiling curve that relates the desired evolution of Brix to the level attained by the mass inside the pan, which is thus taken as a measure of the evolution of the process. The manipulated variable is the standard liquor inlet flow, through the valve that regulates this stream (see Fig. 1).

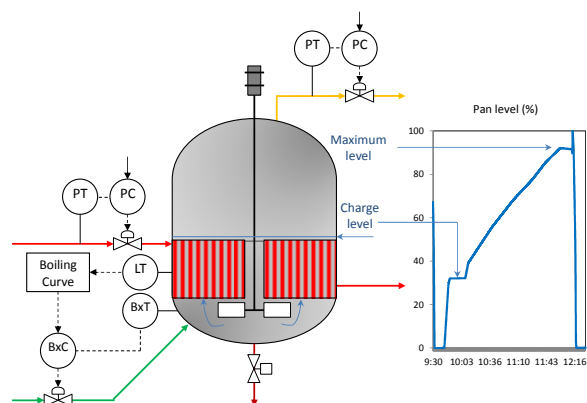


Figure 1: Crystallizer control scheme

The Brix vs. level curve is completely parameterized by specifying its endpoints and a parameter  $\alpha$  that specifies its shape as shown in Figure 2. The curve is adjusted by the operator taking in consideration the purity of the feeding syrup.

The boiling phase is concluded when the level of the strike in the pan reaches the value that determines its maximum capacity.

After the growth phase, a tightening stage follows, with the purpose of increasing the concentration of the massecuite in preparation for its discharge.

Finally, the vacuum is broken and the massecuite is evacuated to the downstream centrifuges, where the sugar grain, which is the commercial product, gets separated from the mother liquor in which it was embedded.

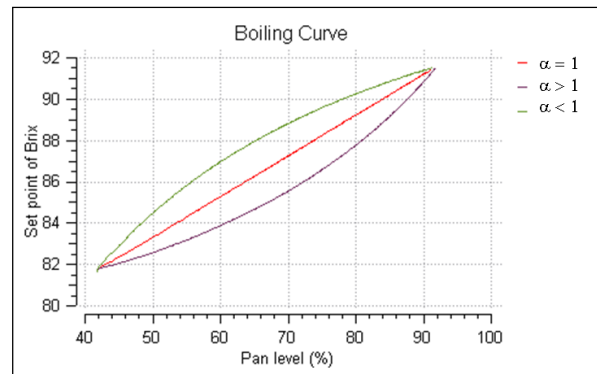


Figure 2: Boiling Curve

### 3. THE MODEL

The concept of a mixed-suspension, mixed-product removal (MSMPR) reactor has been widely adopted for the modelling of many industrial processes. However, there are some cases where deviation from the ideal mixing can be significant (Lie-Ding and Tsan-Sheng, 2005). Frequently the MSMPR supposition is more difficult to justify for continuous processes: it is the one of two limiting simplifying cases: perfectly mixed or plug flow.

In general, it is common to assume a perfectly mixed volume for batch processes: the discharge is delayed to the end of the process and the contents of the agitated reactor have the same residence time. However in our case, for its particular characteristics, we are assuming that a two volume representation would best capture the problem.

In the sugar crystallization batch process, the assumption of a perfectly mixed mass is quite reasonable in the first minutes of the boiling stage, when the agitation produced by the mechanical stirrer and by the bubbles of the evaporated water flowing up through the volume of massecuite, is high enough to ensure a high degree of homogeneity. However, the progression of the process causes a considerable decrease of the agitation effect (see Fig. 3). Some studies have been carried out about the influence of the circulation in vacuum pans (Echeverri, 2005). The causes are multiple: the level of the mass keeps increasing and with it the pressure and thus the temperature in the neighborhood of the calandria (it is no longer realistic to assume the temperature is everywhere the same). A higher temperature of the mass in direct contact with the calandria implies a smaller temperature difference between the two sides, and thus a lesser heat flux and with it, less agitation provoked by the bubbles. Furthermore, we must take into account the increased viscosity of the massecuite due to the crystals growth, one of the main facts that explain the deterioration of the circulation intensity.

To tackle this issue, a two compartment model is proposed. The total volume of massecuite is divided into two regions:

- The volume when the boiling phase is started ( $V_1$ ), which remains constant, essentially

corresponds to the region around the calandria where the circulation is somehow better due to the presence of the mechanical stirrer.

- The variable volume up to the final level (V2, as shown in Figure 3) is going to increase to absorb the syrup which is introduced after the initial charge.

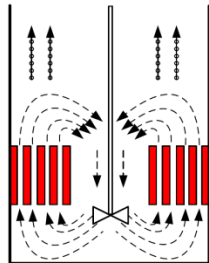


Figure 3. Scheme of the circulation inside the pan

The V1 compartment is in direct contact with the calandria and so it receives the heat contribution from the heat exchanger. The V2 volume, on its turn, is heated only as the result of the convective interchange of mass between the two compartments (F1 and F2 as shown in Fig. 4).

This interchange flow ( $F1 = F2$ ) decreases as the strike proceeds, and this behaviour models the aforementioned deterioration of massecuite circulation. The flow between compartments is then parameterized by a factor  $\beta_{NI}$ , which is defined as the ratio between the interchanged flow and the flow of standard liquor as follows,

$$\beta_{NI} = \frac{F_i}{F_{input}} = b_0 - b_1 \cdot h_2 \quad (1)$$

where  $h_2$  denotes the height of massecuite in V2, while  $b_0$  and  $b_1$  are adjustable coefficients.

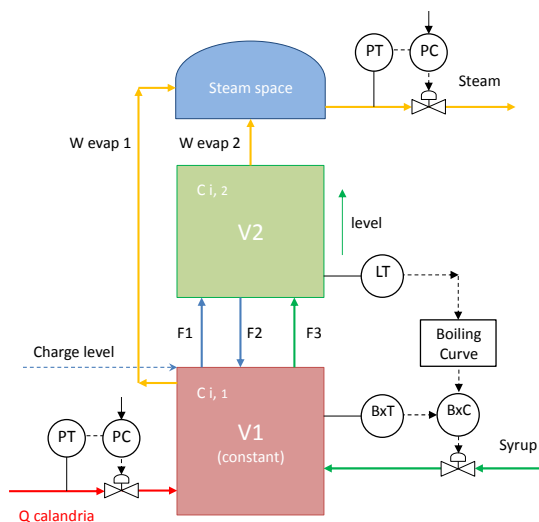


Figure 4: Schematic diagram of the model

The dynamic mass and energy balances are formulated for the two considered crystallization volumes, as well as the growth and nucleation rates. The population balance equation (PBE) is solved by discretising the original partial differential equation which allows to keep track of the shape of the crystal size distribution (CSD). For the sake of comparison, the PBE is also solved with the more computationally efficient method of moments (Hulburt and Katz, 1964). The model equations can be reviewed in the works of Mazaeda (2010) for a single compartment.

### 3.1. Concepts related to the sugar production

Some of the most common concepts used in the field of sugar industry are:

- The dry substance content or Brix
- The purity
- The super-saturation
- Nucleation and growth kinetics

The Brix is a measure of total soluble solids content in sugar juice or syrup, defined as follows:

$$Brix = \frac{m_{sucrose} + m_{imp}}{m_{sucrose} + m_{imp} + m_{water}} \cdot 100 \quad (2)$$

The purity is the sucrose content as a percent of the dry substance or dissolved solids content:

$$Purity = \frac{m_{sucrose}}{m_{sucrose} + m_{imp}} \cdot 100 \quad (3)$$

Note that when these concepts are referred to the massecuite, the mass of crystals must be included in the equations (2) and (3):

$$Brix_{mc} = \frac{m_{sucrose} + m_{imp} + m_{crystal}}{m_{sucrose} + m_{imp} + m_{water} + m_{crystal}} \cdot 100 \quad (4)$$

$$Purity_{mc} = \frac{m_{sucrose} + m_{crystal}}{m_{sucrose} + m_{imp} + m_{crystal}} \cdot 100 \quad (5)$$

The super-saturation can be defined as the degree to which the sucrose content in the solution is greater than the sucrose content in a saturated solution, expressed as

$$s = \frac{q_{suc/W}}{q_{suc/W,sat} \cdot y_{sat}} \quad (6)$$

where  $y_{sat}$  denotes the super-saturation coefficient, which takes into account the presence of impurities in the syrup. Its value is estimated as follows:

$$y_{sat} = A \cdot q_{ns/W} + B + (1 - B) \cdot \exp(-C \cdot q_{ns/W}) \quad (7)$$

where  $q_{ns/w}$  is the concentration of non-sucrose substances, while A, B and C are adjusted parameters according to the relative amount and the kind of specific impurities whose values are quite uncertain. Nevertheless, the values suggested by Van der Poel (1998) are A = 0.178; B = 0.820; C = 2.1.

As shown in Figure 5, over the solubility line, where  $s = 1$ , crystallization takes place. For high values of super-saturation, the labile region is reached and the generation of new small crystals happens. This phenomenon is known as secondary nucleation.

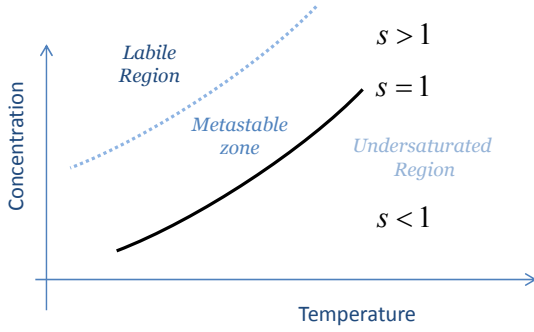


Figure 5: Ostwald-Miers diagram

However, the frontier between the metastable region and the labile zone is not clear. In other words, nucleation is not modelled to be completely non-existent all along the metastable zone width, but it will increase exponentially once super-saturation reaches the labile zone.

Several models for growth and nucleation kinetics have been developed (Garside, 1984; Nyvlt et al., 1985; Rawlings et al., 1993). The expression used for nucleation here is the one that follows:

$$B = \frac{k_b}{\rho_{ml}} \cdot (S - 1)^{n_b} \quad (8)$$

where  $k_b$  and  $n_b$  are adjustable parameters and  $\rho_{ml}$  the density of the mother liquor. Anyway, as said before, because of the high uncertainty related to this phenomenon, it is a complex task to model it accurately.

With regard to the growth rate, the expression used is presented below

$$R_g = \frac{\Delta C_{eff}}{\frac{1}{\beta} + \frac{1}{\Gamma_{kd0}} \frac{C_{sat,i}}{\Delta C_{eff}} + \frac{1}{k_R}} \quad (9)$$

where  $\Delta C_{eff}$  denotes the effective difference of sucrose concentration, while  $\beta$  and  $k_R$  are mass transfer coefficients (Ekelhof & Schliephake, 1995).

### 3.2. Mass balances

The dynamic mass and energy balances applicable to the boiling phase are formulated as usual, taking into

account, in this case, the two volumes, as shown previously in Figure 4.

The mass balances are shown below.

- Compartment V1

$$\frac{dM_1}{dt} = M_{syrup} - W_{evap\ 1} + F_2 - F_1 - F_3 \quad (10)$$

$$\frac{dM_{suc1}}{dt} = M_{syrup} \cdot w_{syrup}^{suc} + F_2 \cdot w_2^{suc} - [F_1 + F_3] \cdot w_1^{suc} - W_{c1} \quad (11)$$

$$\frac{dM_{c1}}{dt} = W_{c1} + F_2 \cdot w_2^c - [F_1 + F_3] \cdot w_1^c \quad (12)$$

$$\frac{dM_{imp1}}{dt} = M_{syrup} \cdot w_{syrup}^{imp} + F_2 \cdot w_2^{imp} - [F_1 + F_3] \cdot w_1^{imp} \quad (13)$$

$$M_{w1} = M_1 - M_{suc1} - M_{c1} - M_{imp1} \quad (14)$$

- Compartment V2

$$\frac{dM_2}{dt} = F_3 - W_{evap\ 2} + F_1 - F_2 \quad (15)$$

$$\frac{dM_{suc2}}{dt} = [F_3 + F_1] \cdot w_1^{suc} - F_2 \cdot w_2^{suc} - W_{c2} \quad (16)$$

$$\frac{dM_{c2}}{dt} = W_{c2} + [F_3 + F_1] \cdot w_1^c - F_2 \cdot w_2^c \quad (17)$$

$$\frac{dM_{imp2}}{dt} = [F_1 + F_3] \cdot w_1^{imp} - F_2 \cdot w_2^{imp} \quad (18)$$

$$M_{w2} = M_2 - M_{suc2} - M_{c2} - M_{imp2} \quad (19)$$

The global mass balances to the total amount of massecuite in V1 and V2 are performed in the equations (10) and (15) respectively, while the individual balances are shown in the equations (11) and (16) to the sucrose, (12) and (17) to the crystals and (13) and (18) to the impurities. The mass of water is calculated algebraically by means of (14) and (19) for both considered volumes.

The terms  $W_{evap\ 1}$  and  $W_{evap\ 2}$  (from the equations 10 and 15 respectively) denote the flow of evaporating water that leaves the solution phase. The mass balance to the steam space is shown in (20):

$$\frac{dM_{steam}}{dt} = W_{evap\ 1} + W_{evap\ 2} - W_{steam\ out} \quad (20)$$

On the other hand, the terms  $W_{c\ i}$  represent the mass flow of sucrose that leaves the mother liquor to be incorporated to the growing crystals:

$$W_{crisi} = R_{g_i} \cdot A_{c_i}; i \in (1, 2) \quad (21)$$

where  $A_c$  represents the total surface of the crystals and  $R_g$  the growth rate, previously mentioned (equation 9).

### 3.3. Energy balances

To formulate the dynamic energy balances to the massecuite and to the steam space, it is assumed the thermal equilibrium between the two phases, which implies that the existing pressure in the steam space  $P_{cham}$  is equal to the equilibrium pressure. Nevertheless, for the compartment V1, the hydrostatic head of the column on top (V2) is taken into account as follows:

$$P_{eq1} = P_{cham} + \rho \cdot g \cdot h_2 \quad (22)$$

The pressure in the steam space of the crystallizer is calculated using the ideal gas law:

$$P_{cham} = \frac{\left[ \frac{M_{steam}}{MM_{H2O}} \right] \cdot R \cdot T_{cham}}{V_{cham}} \quad (23)$$

As  $V_{cham}$  decreases along the process, it is estimated as follows:

$$V_{cham} = V_{tot} - V_{mc} = V_{tot} - \frac{M_{mc}}{\rho_{mc}} \quad (24)$$

where  $V_{mc}$  denotes the volume occupied by the massecuite and  $V_{tot}$  is the total volume of the crystallizer.

The energy balances to both considered volumes are expressed in (25) and (26), where H denotes the enthalpy:

$$\frac{dH_1}{dt} = \frac{\left[ \begin{array}{l} M_{syrup} \cdot H_{syrup} + F_2 \cdot H_2 - [F_1 + F_3] \cdot H_1 \\ - W_{evap1} \cdot H_{vap1} + Q_{cal} - H_1 \cdot \frac{dM_1}{dt} \end{array} \right]}{M_1} \quad (25)$$

$$\frac{dH_2}{dt} = \frac{\left[ \begin{array}{l} [F_1 + F_3] \cdot H_1 - F_2 \cdot H_2 \\ - W_{evap2} \cdot H_{vap2} - H_2 \cdot \frac{dM_2}{dt} \end{array} \right]}{M_2} \quad (26)$$

The temperature of the massecuite is calculated as a function of the enthalpy, the Brix of the massecuite and the Brix and purity of the mother liquor (Bubnik et al., 1995).

Note that, as represented in Figure 4, it is assumed that only the compartment V1 is heated from the calandria ( $Q_{cal}$  in the equation 25) and heat losses are neglected. This flow of heat is calculated using the expression that follows

$$Q_{cal} = U \cdot A_{cal} \cdot (T_{cal} - T_{mc1}) \quad (27)$$

where  $T_{cal}$  is the temperature of the saturated steam circulating through the shell of the calandria,  $T_{mc1}$  is the temperature of the massecuite in V1,  $A_{cal}$  denotes the heating surface and U represents the overall heat transfer coefficient, estimated as:

$$U = \frac{K_U}{Brix_{mc1}} \quad (28)$$

The heat transfer coefficient is modelled to drop along the boiling phase, due to the increase of the Brix of the massecuite and, therefore, of the viscosity.

### 3.4. Population Balance Equation (PBE)

Considering a single growth direction with one characteristic length L, the Crystal Size Distribution (CSD) is given by the population balance equation (Nagy, 2009) with the form

$$\frac{\partial n}{\partial t} = \frac{Q_e n_e - Q_s n - n \frac{dV}{dt}}{V} - \frac{\partial(G(L, x) \cdot n)}{\partial L} + B \quad (29)$$

where the first term of the right hand side of (29) represents the convective term, the second is the growth rate of the characteristic size, and B and D denote the spontaneous nucleation and the dissolution. The growth rate G can be put into correspondence with the mass transfer rate  $R_g$  by means of (30), where  $k_s$  and  $k_v$  are the surface and the volume shape factors of the particles.

$$G = \frac{k_s}{3 \cdot k_v \cdot \rho_c} \cdot R_g \quad (30)$$

As (29) is a partial differential equation, it is necessary to convert it to a set of ODEs by the method of moments (Hulburt and Katz, 1964):

$$\frac{d\mu_k}{dt} = \frac{Q_e \mu_k^e - Q_s \mu_k - \frac{dV}{dt} \mu_k}{V} + k \cdot G \cdot \mu_{k-1} + B \cdot L_{nuc}^k \quad (31)$$

Note that the zero moment is proportional to the total number of particles per unit of volume, while the first, second and third moments are proportional to the aggregate values of length, area and volume of the particles per unit of volume, respectively (Mazaeda, 2007). From the moments, the mass based average size and the coefficient of variation of the crystals are estimated as follows (Randolph & Larson, 1988).

$$\bar{L} = \frac{\mu_4}{\mu_3} \quad (32)$$

$$CV = \sqrt{\frac{\mu_5 \mu_3}{(\mu_4)^2} - 1} \quad (33)$$

Note that the equations (29)-(33) are solved for both volumes.

#### 4. MODEL ADJUSTMENT TO REAL DATA

The model has been adjusted and validated from data obtained from a sugar factory in Spain.

An optimisation problem is formulated to minimise the error between the real measured data and those predicted by the model, selecting the most sensitive parameters as decision variables. Among the large number of parameters, the ones that are adjusted, bounded in a reasonable interval, are:

- The super-saturation coefficients (A, B and C from the equation 7), used to take into account the effect of the relative amount of impurities on the super-saturation value (Bubnik et al, 1995).
- The spontaneous nucleation parameters ( $k_b$  and  $n_b$  from the equation 8), which relate the super-saturation to the nucleation rate.
- A factor used to estimate the value of the overall heat transfer coefficient along the process ( $k_U$  from the equation 28).
- The mass and the population characteristics of the seed, which are important sources of uncertainty.

The less sensitive parameters, that are not adjusted, have been set according to the typical values compiled in technical literature.

Measurements obtained directly from factory floor instrumentation include pressure in the pan, pressure of the steam in the calandria, temperature, level and Brix of the massecuite. There are also laboratory results, gathered at a much lower frequency, for the purity of the feeding syrup and the characteristics of the sugar grain population: average size and CV.

The set point of the pressure controllers are considered as inputs of the system, as well as the Brix and the purity of the standard liquor; while the temperature, level and Brix of the massecuite, as well as the final conditions of the crystals (average size, coefficient of variation and crystal production) are considered as the outputs to adjust.

Thus, the optimisation problem is defined in (34) and solved using a reduced SQP algorithm (Gill et al., 2002).

$$\begin{aligned} \min_p J = & \sum \gamma_l \cdot (\text{level}(t)_{\text{mod}} - \text{level}(t)_{\text{plant}})^2 + \\ & \gamma_B \cdot (\text{Brix}_{mc}(t)_{\text{mod}} - \text{Brix}_{mc}(t)_{\text{plant}})^2 + \\ & \gamma_T \cdot (T_{mc}(t)_{\text{mod}} - T_{mc}(t)_{\text{plant}})^2 + \\ & \gamma_L \cdot (\bar{L}_{\text{mod}} - \bar{L}_{\text{plant}})^2 + \\ & \gamma_{CV} \cdot (CV_{\text{mod}} - CV_{\text{plant}})^2 + \\ & \gamma_{CC} \cdot (CC_{\text{mod}} - CC_{\text{plant}})^2 \\ \text{s.t.} : & \text{Measured inputs: } P_{cal}(t), P_{cham}(t) \\ & \text{Brix, Purity} \\ & \text{Crystallizer geometry} \\ & P_{low} \leq p \leq P_{upp} \end{aligned} \quad (34)$$

The cost function is the sum of the errors between the measured and the model-predicted data weighted with a factor  $\gamma$ , which is greater for the less uncertain variables. The vector of parameters to be estimated is denoted as  $p$ , bounded in a reasonable interval.

The modelling and simulation software used to perform the simulation of the model is EcosimPro, capable of modelling dynamic systems represented by differential-algebraic equations (DAE) or ordinary-differential equations (ODE) and discrete events (EcosimPro / PROOSIS by EA Internacional).

Table 1: Values of the parameters after the optimisation

| Parameters                                          |                                         |                     |
|-----------------------------------------------------|-----------------------------------------|---------------------|
| Super-saturation coefficient parameters             | A                                       | 0.12                |
|                                                     | B                                       | 1.07                |
|                                                     | C                                       | 2.08                |
| Nucleation parameters                               | $n_b$                                   | 5.0                 |
|                                                     | $k_b$                                   | $2.0 \cdot 10^{12}$ |
| Gain of the overall heat transfer coefficient       | $K_U$                                   | $5.45 \cdot 10^4$   |
| Mass and the population characteristics of the seed | $M_{c \text{ seed}} \text{ (kg)}$       | 0.18                |
|                                                     | $L_{\text{seed}} \text{ (}\mu\text{m)}$ | 15                  |
|                                                     | $CV_{\text{seed}} \text{ (\%)}$         | 25                  |

The Figures 6 and 7 show the crystallization stage model identification and validation respectively, for one of the factory crystallizers' taken as reference. Due to the high uncertainty and the inherent process nonlinearities, not forgetting the unmeasured disturbances and the limited available data, a perfect fit is impossible. Nevertheless, a robust model is validated in a practical operation range.

Table 2: Model inputs from the factory

| Brix and purity of standard liquor |                 |
|------------------------------------|-----------------|
| Brix = 69.6                        | Purity = 95.3 % |

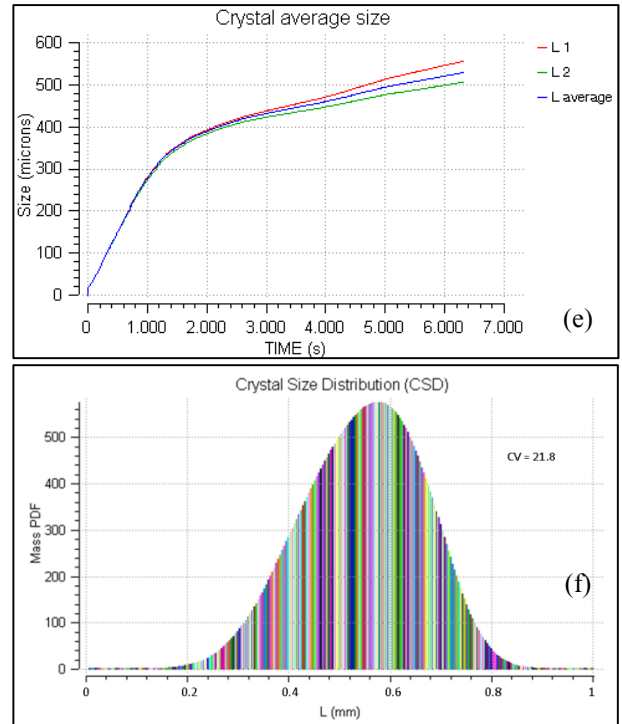
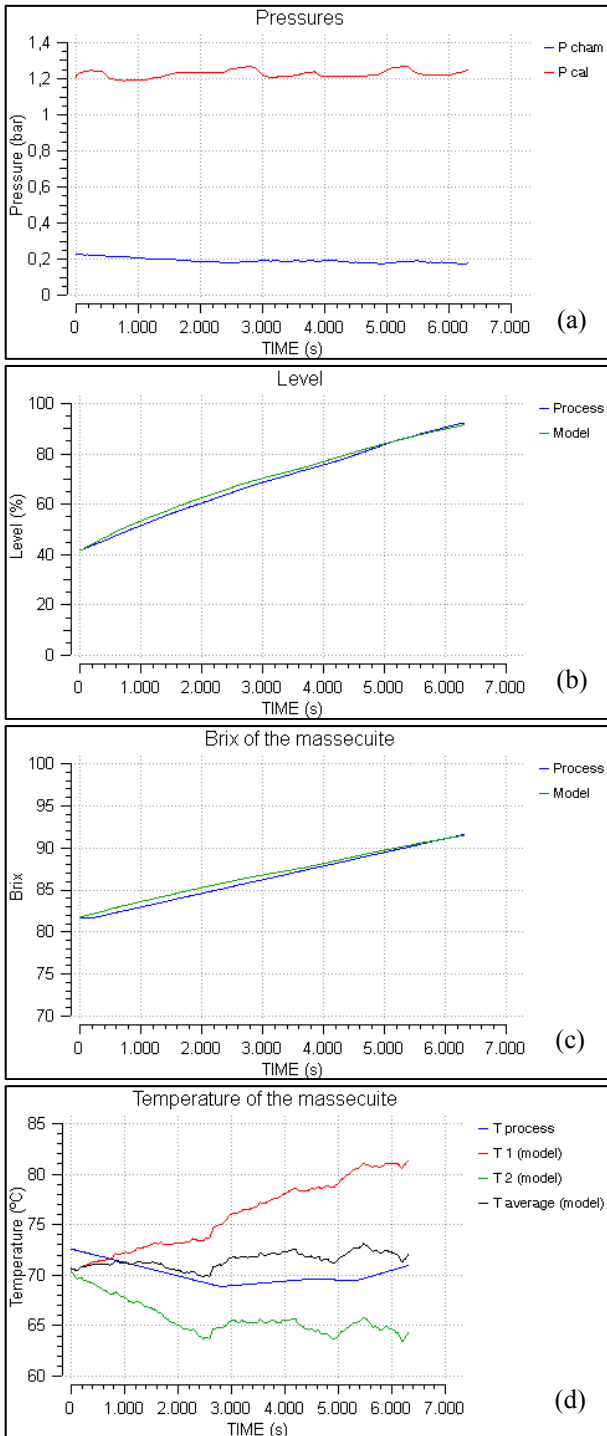


Figure 6: Model adjustment. (a) Pressures (calandria and vacuum); (b) Pan level; (c) Brix of the massecuite; (d) Temperature of the massecuite; (e) Crystal average size; (f) CSD.

Table 3: Final average size, CV and crystal content

| Crystal average size and CV |        |       |
|-----------------------------|--------|-------|
| L average                   | CV     | CC    |
| 529 $\mu\text{m}$           | 21.8 % | 50.9% |

The Figure 6-(a) shows the evolution of the pressure in the calandria (red line) and the vacuum established in the crystallizer (blue line), which are inputs to the model. These pressures are controlled by means of PIDs. The profile of set points is determined by the recipe based sequential controller, as parameter that can be modified by the operators.

The evolution of level and Brix of massecuite are shown in the Figures 6-(b) and 6-(c) respectively, where the blue line represents the measured data and the green one the values predicted by the model. The parameter of the boiling curve  $\alpha$  was set to 1.2.

In Figure 6-(d) the dispersion between the temperature of the massecuite in the compartment V1 (where heat is directly received from the calandria) and the temperature in V2 can be observed. In this way, the average temperature (black line) fits well to the measured temperature data (blue line), which depend on the particular position where it is placed.

The evolution of the crystal size in V1, V2, as well as the average size, is shown in Figure 6-(e), where the dispersion between the grain size in V1 and V2 grows as the boiling stage proceeds.

Finally, the crystal size distribution is represented in Figure 6-(f), where we can see that the final average



crystal size is near 500 – 550  $\mu\text{m}$  and coefficient of variation is between 20 and 30% (Table 3).

Table 4: Model inputs from the factory (validation)

| Brix and purity of standard liquor |                 |
|------------------------------------|-----------------|
| Brix = 67.8                        | Purity = 95.2 % |

In this case, as the Brix of the standard liquor is lower than before (see Tables 3 and 4), logically, the duration of the boiling stage is larger (almost 2 hours). Concerning the rest of the inputs, their values are very similar to the ones used in the previous batch.

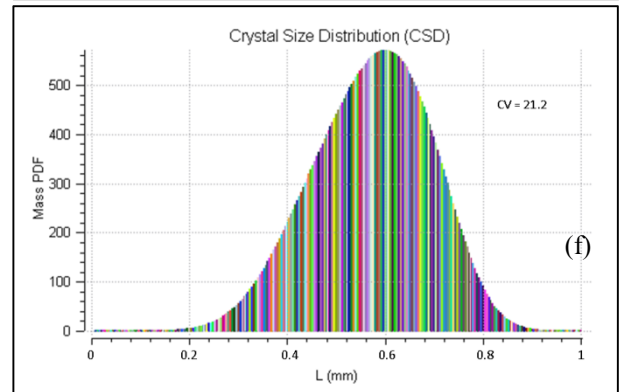
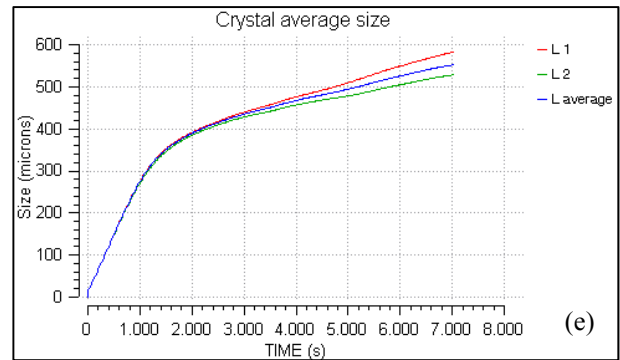
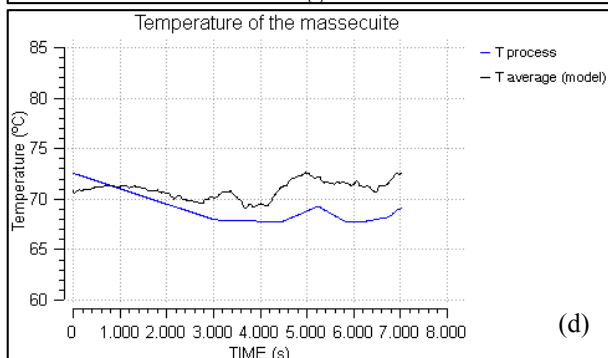
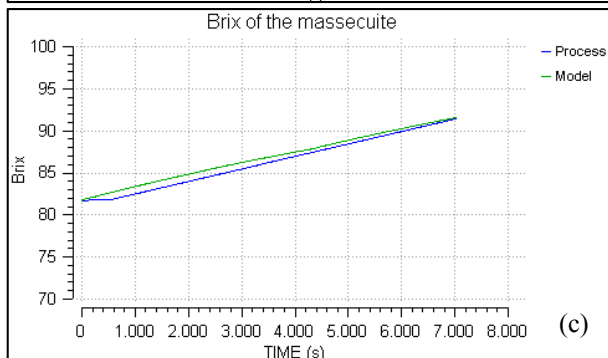
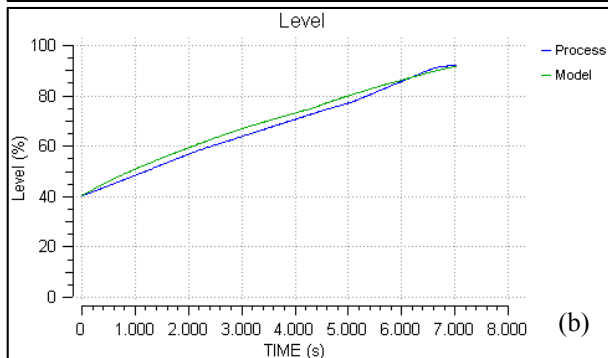
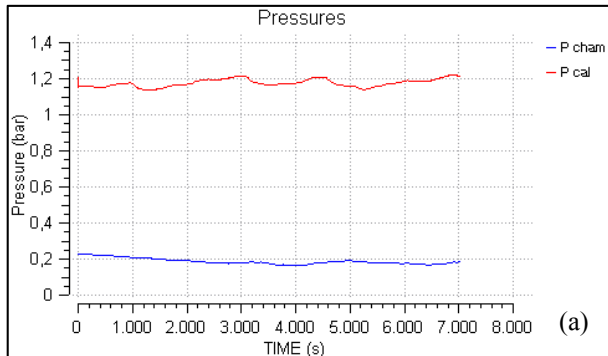


Figure 7: Model validation. (a) Pressures (calandria and vacuum); (b) Pan level; (c) Brix of the massecuite; (d) Temperature of the massecuite; (e) Crystal average size; (f) CSD.

Table 5: Final average size, CV and CC (validation)

| Crystal average size and CV |        |       |
|-----------------------------|--------|-------|
| L average                   | CV     | CC    |
| 552 $\mu\text{m}$           | 21.2 % | 51.1% |

## 5. CONCLUSIONS

A two-MSMPR-crystallizers-with-interchange model has been developed to capture the deviations from ideal behaviour caused by the existence of non-homogeneous agitated regions in a sugar batch crystallizer, in order to improve the fidelity of the previously developed model (Mazaeda, 2010). The model adjustment and validation are presented in the previous section, where a good fit is achieved, despite the unmeasured disturbances and the limited available data.

## ACKNOWLEDGMENTS

The authors would like to express their gratitude to ACOR (Spain) for their financial support and technical assistance, as well as projects DPI2012-37859 of Spanish MINECO and the EU FP7 NoE HYCON2.

## REFERENCES

- Braatz, R. D. (2002). Advanced control of crystallization processes. *Annual Reviews in Control*, 26, 87–99.
- Bubnik, Z., Kadlec, P., Urban, D., M. Bruhns (1995). *Sugar Technologist Manual*. Chemical and Physical Data Manufacturers and Users. Ed. Verlag Dr. Albert Bartens.

- Damour, C., Benne, M., Grondin-Perez, B., Chabriat, J. P. (2010) Soft-sensor for industrial sugar crystallization: On-line mass of crystals, concentration and purity measurement. *Control Engineering Practice*, 18 (2010) 839-844. Elsevier.
- Echeverri, L. (2005). *Measurements and Computations of the flow in Full-scale Sugar Evaporative-Crystallizers and in Lab-scale Models*. PhD Thesis, Louisiana State University.
- EcosimPro / PROOSIS by EA Internacional, System Modeling & Simulation Software, [On line] <http://www.ecosimpro.com/> Last access: June 20, 2014.
- Ekelhof, B., Schliephake, D. (1995). Description of Crystal Growth Rate in pure and impure sucrose solution by means of a single kinetic expression. *In Proceedings of the 20th General Assembly of the Comission Internationale Technique Sucrierie (CITS), Munich, Germany*, 216-235.
- Garside, J. (1984). Advances in characterization of crystal growth. In G. R. Youngquist (Ed.), *Advances in crystallization from solutions, AIChE symposium series*, Vol. 80, N° 240 (23–38). New York: AIChE Press.
- Gill, P. E., Murray, W., Saunders, M. A. (2002). SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM J. Optim.*, Vol. 12 No 4, 979–1006.
- Hulburt, H. M., & Katz, S. (1964). Some problems in particle technology: A statistical mechanical formulation. *Chemical Engineering Science*, 19, 555–574.
- Nyvtl, J., Sohnel, O., Matuchova, M., & Broul, M. (1985). *The kinetics of industrial crystallization*. Elsevier.
- Lie-Ding, S., Tsan-Sheng, L. (2005). Modeling the nonideal mixing behaviour in a continuous-stirred crystallizer. *Computers & Chemical Engineering*, 30 (2006) 970-977. Elsevier.
- Ma, D. L., Braatz, R. D. (2003). Robust identification and control of batch processes. *Computers & Chemical Engineering*, 27 (2003) 1175-1184. Elsevier.
- Mazaeda, R., Prada, C. (2007). Dynamic simulation of a sucrose batch evaporative crystallizer for operators training. *19th European Modeling and Simulation Symposium, EMSS07*.
- Mazaeda, R. (2010). *Librería de modelos del cuarto de azúcar de la industria azucarera para simuladores de entrenamiento de operarios*. Thesis (PhD). University of Valladolid.
- Nagy, Z. K. (2009). Model based robust control approach for batch crystallization product design. *Computers & Chemical Engineering*, 33 (2009) 1685-1691. Elsevier.
- Randolph, A. D. & Larson, M. A. (1988). *Theory of particulate processes* (2nd ed.). New York: Academic Press.
- Rawlings, J., Miller, S., & Witkowski, W. (1993). Model identification and control of solution crystallization processes: A review. *Industrial & Engineering Chemistry Research*, 32, 1275–1296.
- Sugar Engineers [On line] <http://www.sugartech.co.za/> Last access: June 20, 2014.
- Van der Poel, P. W., Schiweck, H., Schwartz, T. (1998). *Sugar Technology. Beet and Cane Sugar Manufacture*. Ed. Bartens.



# MODELING OF LATTICE THERMAL EXPANSION FOR SIMULATION OF THERMAL STRESS IN 3D NANO MOSFETS

Abderrazzak El Boukili

Al Akhawayn University in Ifrane, Morocco

[Email: a.elboukili@aui.ma](mailto:a.elboukili@aui.ma)

## ABSTRACT

We are developing new model to calculate the lattice thermal expansion coefficient of Silicon Germanium (Si(1-x)Ge(x)) thin films used as performance boosters for nanoscale PMOSFETs. These models take into account the dependence of thermal expansion coefficient on Germanium mole fraction  $x$  and temperature. This thermal expansion coefficient is used to calculate the thermal induced intrinsic stress in Si(1-x)Ge(x) thin film after deposition. And this intrinsic stress is used to calculate and simulate numerically the resulting extrinsic stress in the channel of a nano PMOS based on Intel technology. 3D simulation results of channel stress will be presented, analyzed, and compared with literature.

Keywords: lattice thermal expansion coefficients, 3D modeling and simulation, thermal induced stress, Intel nano PMOSFETs

## 1. INTRODUCTION

Most of nano semiconductor device manufacturers as Intel, IBM and TSMC are intentionally using the intrinsic stress  $\sigma_0$  to produce uniaxial extrinsic stress  $\sigma$  in the Silicon channel of nano MOSFETs. And, it is now admitted that the channel stress enhances carrier mobilities for both nano PMOS and NMOS transistors (Chui 2007; Bera 2006; Krivokapic 2003). This mobility enhancement fundamentally results from alteration of electronic band structure of silicon due to extrinsic stress. The extrinsic stress is the stress that exists in the whole transistor to balance the intrinsic stress that exists in different materials (or films) that make up the transistor. The intrinsic stress is either introduced intentionally or unintentionally or both. The unintentional intrinsic stress is induced by the processing steps as: implantation, etching, deposition of thin films, oxidation, shallow trench isolation, doping, diffusion, dislocation loops, or capping layers. The intentional intrinsic stress is introduced intentionally by the manufacturers to increase performance. Intel technology is using compressed Si(1-x)Ge(x) pockets in source and drain of nano PMOSFETs to introduce a

compressive uniaxial extrinsic stress  $\sigma$  in the channel. This compressive stress  $\sigma$  has the advantage of enhancing the mobility of holes. Consequently, this will enhance the electrical performances of the nano PMOS devices by 30% or more (Chui 2007, Ghani 2003). This extrinsic stress  $\sigma$  is due to an intrinsic stress  $\sigma_0$  that is generated in Si(1-x)Ge(x) films after deposition. And, this intrinsic stress  $\sigma_0$  is due mainly to material mismatch or thermal mismatch. In this paper, we are focusing mainly on developing new analytical models for the thermal expansion coefficients  $\alpha_{Si(1-x)Ge(x)}(T)$  of Si(1-x)Ge(x) film and  $\alpha_{Si}(T)$  of the Silicon (Si) substrate. Here  $T$  represents the deposition temperature. It could also represent the diffusion temperature. These models take into account the dependence of thermal expansion coefficients (TECs) on the Germanium mole fraction  $x$  and the temperature. These coefficients are used to calculate accurately the intrinsic stress  $\sigma_0$  due to thermal mismatch between the Si substrate and the Si(1-x)Ge(x) film. The 3D simulations and analysis of the resulting extrinsic stress  $\sigma$  will also be presented. The equations relating  $\sigma_0$  and  $\sigma$  will be presented in Section 3. The intrinsic stress  $\sigma_0$  due to thermal mismatch is caused by the different thermal expansion coefficients of the thin film Si(1-x)Ge(x) and the Silicon substrate. In fact, the temperature is high during deposition of Si(1-x)Ge(x) films. And, the temperature will cool down to room temperature after deposition. Then, a compressive intrinsic stress  $\sigma_0$  will develop in the thin film and in the substrate during the cooling down to room temperature. We are going to neglect the intrinsic stress in the Si substrate since Si substrate is thick. Thermal expansion coefficients  $\alpha_{Si(1-x)Ge(x)}(T)$  or  $\alpha_{Si}(T)$  are, in general, extremely important for the performance of semiconductor devices as MOSFETs, bipolar transistors, laser diodes, solar cells, or micro-electromechanical systems (MEMS). This is because they create stress at the interface between the film and

the substrate during deposition, diffusion, etching, oxidation, doping, or any other processing step. And, this stress can affect significantly the performance. Thermal expansion coefficients of different semiconductor materials have been the focus of many researchers in the last decades. They have been studied both experimentally (Reeber 2000; Toshimaro 2002; Giles 2005; Okada 1984) and theoretically (Berrardi 1996; Xu 1992; Talwer 1995; Xie 1999). However, even for the same material such as Diamond (Toshimaro 2002, Giles 2005), Silicon (Okada 1984) and Gallium Nitride (Reeber 2000, Roder 2005) the available experimental data scattered significantly depending on the measuring methods. The main objective of this paper is to develop new analytical models based on polynomials for the temperature dependent TECs. These models are simple, and easy to fit with experiments compared to those found in literature (Garai 2007; Reeber 2000; Toshimaro 2002; Okada 1984; Xu 1992; Bruls 2001; Reeber 1975; Debernardi 2001). The scientific evidence of the simplicity of the proposed polynomial models is shown in the equations (2) and (3). The four unknowns in the equations (2) and (3) are easily and exactly calculated by using the equations (4) or (5) and a simple Gauss-Jordan elimination method for a dense matrix. Some researchers have used the *ab initio* method to calculate the temperature dependent TECs for Al (Debernardi 2001). But, this method involves the bulk modulus, mode Grüneisen constant, and the concave parameter. These parameters are based on many physical assumptions that are not always valid and they are not easy to fit to experiments. This paper is organized as follows. Section 2 will outline the main sources of intrinsic stress in Si(1-x)Ge(x) films after deposition. Section 3 will present the new physically based models for thermal expansion coefficients of Si(1-x)Ge(x) films and Si substrate. It will also present the mathematical models we are using to calculate the initial stress  $\sigma_0$  due to thermal mismatch, and the resulting extrinsic stress  $\sigma_e$ . Section 4 will present the 3D simulation results of the channel extrinsic stress  $\sigma_e$ . This section will also analyze qualitatively and quantitatively the numerical results and provide some comparisons with the results found in the literature. Section 5 presents the concluding thoughts and future work.

## 2. INITIAL STRESS SOURCES

The main processing step in determining the initial stress in Si(1-x)Ge(x) films is deposition. This processing step takes place at elevated temperatures. When the temperature is decreased, the volumes of the grains of Si(1-x)Ge(x) film shrink and the stresses in the material increase. The stress gradient and the average stress in the film depend mainly on the Germanium ratio x, the substrate temperature and orientation, and the deposition technique which is usually LPCVD (low pressure chemical vapor deposition) or PECVD (plasma enhanced chemical

vapor deposition), or Epitaxy. The initial stress  $\sigma_0$  existing in thin films after deposition is caused mainly by lattice mismatch and thermal mismatch.

### 2.1. Intrinsic stress due to lattice mismatch

During deposition, thin films are either stretched or compressed to fit the substrate on which they are deposited. After deposition, the film wants to be smaller if it was stretched earlier, thus creating tensile intrinsic stress. And similarly, it creates a compressive intrinsic stress if it was compressed during deposition. The intrinsic stress generated due to this phenomenon can be quantified by Stoney's equation which relates the stress to the substrate curvature or by other advanced models (El Boukili 2010).

### 2.2. Intrinsic stress due to thermal mismatch

Thermal mismatch stress occurs when two materials with different coefficients of thermal expansion are heated and 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 intrinsic strain and stress in the film and also in the substrate. The thermal expansion coefficient is defined as the rate of change of strain with respect to temperature.

## 3. MODELING OF LATTICE THERMAL EXPANSION COEFFICIENTS AND THERMAL INDUCED INTRINSIC STRESS

From the point of view of classical thermodynamics and lattice quantum vibrations, the temperature dependent TECs are found analogous to the temperature dependent specific heat (Xie 1999; Roder 2005; Bruls 2001). Many researchers have used the model of specific heat based on exponential functions to derive different models for temperature dependence of TECs (Reeber 2000; Toshimaro 2002; Okada 1984; Xu 1992; Bruls 2001; Garai 2007; Reeber 1975; Debernardi 2001).

These models could reproduce the profile of measured temperature dependent TECs. But, they are based on many physical assumptions that are not always valid. And, they involve many physical parameters that should be fitted. Which make these models difficult to use in practice.

Some researchers have used the *ab initio* method to calculate the temperature dependent TECs for Al (Debernardi 2001). But, this method involves the bulk modulus, mode Grüneisen constant, and the concave parameter. These parameters are also not easy to fit to experiments.

Some researchers have used the phenomenological lattice dynamical theory in quasi-harmonic approximation to describe the temperature dependence of TECs (Xie 1999).

A deep understanding of the atomistic origin of the thermal expansion and the accurate modeling of TECs by an easy and physically based formula is still a

challenge. Experimentally, the temperature dependent TECs have been described by polynomials (Toshimaro 2002; Bruls 2001; Garai 2007).

The objective of this paper is to develop and analyze new analytical models based polynomials for the temperature dependent TECs. We are proposing the following models:

$$\alpha_{Si(1-x)Ge(x)}(T) = (1-x)\alpha_{Si}(T) + x\alpha_{Ge}(T), \quad (1)$$

where  $x$  is the Germanium mole fraction.

To take into account the temperature effects, we are proposing the following polynomials of degree two since most of observed TECs are nonlinear:

$$\alpha_{Si}(T) = T(a_{Si}T + b_{Si}) \quad (2)$$

$$\alpha_{Ge}(T) = T(a_{Ge}T + b_{Ge}) \quad (3)$$

The constants  $a_{Si}, b_{Si}, a_{Ge}, b_{Ge}$  are calculated exactly using the following experimental models given for  $T = 300$  K (Schaffler 2001):

If ( $x < 0.85$ ) then:

$$\alpha_{Si(1-x)Ge(x)}(300) = (2.6 + 2.55x) \times 10^{-6} \quad (4)$$

If ( $x > 0.85$ ) then:

$$\alpha_{Si(1-x)Ge(x)}(300) = (-0.89 + 7.53x) \times 10^{-6} \quad (5)$$

Now let's use the thermal expansion coefficients  $\alpha_{Si(1-x)Ge(x)}(T)$  and  $\alpha_{Si}(T)$  to model analytically the thermal induced intrinsic strain  $\epsilon_0$  and intrinsic stress  $\sigma_0$ .

The intrinsic strain tensor,  $\epsilon_0$ , is defined by:

$$\epsilon_0 = (\epsilon_0^{xx}, \epsilon_0^{yy}, \epsilon_0^{zz}, \epsilon_0^{xy}, \epsilon_0^{yz}, \epsilon_0^{xz}), \quad (6)$$

where  $\epsilon_0^{xx}, \epsilon_0^{yy}, \epsilon_0^{zz}$  are the intrinsic normal strain components, and  $\epsilon_0^{xy}, \epsilon_0^{yz}, \epsilon_0^{xz}$  are the intrinsic shear strain components. We assume that in Si(1-x)Ge(x):

$$\epsilon_0^{xx} = \epsilon_0^{yy} = \epsilon_0^{zz} = \epsilon_{SiGe}^0(T). \quad (7)$$

For simplicity, let:

$$\alpha_{Si(1-x)Ge(x)}(T) = \alpha_{SiGe}(T)$$

The thermal expansion coefficient of Si(1-x)Ge(x) is defined as the rate of change of the intrinsic strain component,  $\epsilon_{SiGe}^0(T)$ , with respect to

temperature  $T$ . Its unit is micro strain/Kelvin ( $\mu\epsilon/K$ ) and it is given by:

$$\alpha_{SiGe}(T) = \frac{d\epsilon_{SiGe}^0(T)}{dT} \quad (8)$$

Then, the intrinsic strain component will be given by:

$$\epsilon_{SiGe}^0(T) = - \int_{T_0}^T \alpha_{SiGe}(t) dt \quad (9)$$

where  $T_0$  is the ambient temperature and  $T$  is the processing temperature. For example,  $T$  could be the temperature during deposition of Si(1-x)Ge(x) thin film. The Si(1-x)Ge(x) film will cool down after deposition from  $T$  to  $T_0$ . Then, it will contract. This explains the use the sign minus in the equation (9). We also assume that there is no thermal strain at room temperature  $T_0$ .

On the other hand, to include the effects of the thermal expansion coefficient of the Si substrate,  $\alpha_{Si}(T)$ , on the intrinsic strain in Si(1-x)Ge(x) thin film, we add the term  $\Delta\alpha(T)\Delta T$  to the model given by the equation (9). Then, the proposed model for the intrinsic strain component  $\epsilon_{SiGe}^0(T)$  is given by:

$$\epsilon_{SiGe}^0(T) = \int_{T_0}^T \alpha_{SiGe}(t) dt + \Delta\alpha(T)\Delta T \quad (10)$$

where

$$\Delta\alpha(T) = \alpha_{SiGe}(T) - \alpha_{Si}(T), \quad (11)$$

$$\Delta T = T - T_0. \quad (12)$$

The component form of the intrinsic stress,  $\sigma_0$ , is defined by:

$$\sigma_0 = (\sigma_0^{xx}, \sigma_0^{yy}, \sigma_0^{zz}, \sigma_0^{xy}, \sigma_0^{yz}, \sigma_0^{xz}),$$

where  $\sigma_0^{xx}, \sigma_0^{yy}, \sigma_0^{zz}$  are the intrinsic normal stress components and  $\sigma_0^{xy}, \sigma_0^{yz}, \sigma_0^{xz}$  are the intrinsic shear stress components that are also assumed to be zero. We also assume that in SiGe:

$$\sigma_0^{xx} = \sigma_0^{yy} = \sigma_0^{zz} = \sigma_{SiGe}^0(T). \quad (13)$$

On the other hand, we are using a physically based model to define the intrinsic stress tensor  $\sigma_0$  in the Si(1-x)Ge(x) thin film, since, we are using the Hooke's elastic law to express the relation between strain and stress as follows:

$$\sigma_0 = D \cdot \epsilon_0, \quad (14)$$

In three dimensions, the stiffness tensor  $D$ , used in our model, for isotropic elastic materials as SiGe is given by:

$$D = \begin{bmatrix} c11 & c12 & c12 & 0 & 0 & 0 \\ c12 & c11 & c12 & 0 & 0 & 0 \\ c12 & c12 & c11 & 0 & 0 & 0 \\ 0 & 0 & 0 & c44 & 0 & 0 \\ 0 & 0 & 0 & 0 & c44 & 0 \\ 0 & 0 & 0 & 0 & 0 & c44 \end{bmatrix} \quad (15)$$

where the elastic constants  $c11, c12$ , and  $c44$  for each material are depending on the Young modulus ( $E$ ) and Poisson's ratio ( $\gamma$ ). In this paper, we are taking into account the dependence of  $\gamma$  and  $E$  on the Germanium mole fraction and substrate orientations.

In our simulation program, the intrinsic stress tensor  $\sigma_0$  is used as a source term to calculate, in the whole 3D nano MOSFET structure, the extrinsic stress tensor  $\sigma = (\sigma^{xx}, \sigma^{yy}, \sigma^{zz}, \sigma^{xy}, \sigma^{yz}, \sigma^{zx})$ . We note that  $\sigma^{xx}$ ,  $\sigma^{yy}$ , and  $\sigma^{zz}$  represent the extrinsic stress along the channel, vertical to the channel, and across the channel. This channel stress is used to enhance the mobility of holes in 3D nano PMOSFETs based on Intel technology (Ghani 2003). We assume that Silicon and Silicon Germanium are elastic materials. And, to calculate the stress tensor  $\sigma$  we use the elastic stress model based on Newton's second law of motion, and the following Hooke's law relating stress to strain:

$$\sigma = D (\varepsilon - \varepsilon_0) + \sigma_{00} \quad (16)$$

Here  $\sigma_{00}$  is taken to be zero for simplicity. And, the components of  $\varepsilon_0$  are given by the equation (6). A detailed description and how to use finite volume method to solve numerically the elastic model (16) in 3D is given in (El Boukili 2010).

#### 4. 3D NUMERICAL RESULTS AND ANALYSIS

The third order model of intrinsic stress  $\sigma_0$  is based on the proposed second order model of the thermal expansion coefficients  $\alpha_{Si(1-x)Ge(x)}(T)$  and  $\alpha_{Si}(T)$ .

##### 4.1 Validation of the proposed models for TECs using experimental values

For 300°K, the values of thermal expansion coefficients obtained from the proposed models for Si, Ge, and Si(1-x)Ge(x) are in excellent agreement with the values obtained from experiments. The developed models for Si and Ge are given by:

$$\alpha_{Si}(T) = (3.02 e^{-10} T^2 - 8.19 e^{-8} T)$$

$$\alpha_{Ge}(T) = (5.72 e^{-12} T^2 + 3.13 e^{-23} T)$$

If we take  $T=300^\circ\text{K}$ , we find out that:

$$\alpha_{Si}(300) = 2.6 e^{-6},$$

$$\alpha_{Ge}(300) = 5.15 e^{-6},$$

The experimental values (Sze 1981) for  $T=300^\circ\text{K}$  are:

$$\alpha_{Si}(300) = 2.6 e^{-6},$$

$$\alpha_{Ge}(300) = 5.8 e^{-6}.$$

We could then see that the values we get from the proposed models are exactly the same for Silicon and very close for Germanium. The following table (Table 1) shows the values of the TECs for different Germanium model fractions for Si(1-x)Ge(x) at  $T=300^\circ\text{K}$  that we get from the proposed models and from experiments (Sze 1981).

Table 1: Comparison of Si(1-x)Ge(x) TECs

| Germanium mole fraction | Proposed TECs | Experimental TECs |
|-------------------------|---------------|-------------------|
| 0.25                    | 3.23e-6       | 3.4e-6            |
| 0.50                    | 3.87e-6       | 4.2e-6            |
| 0.75                    | 4.51e-6       | 5e-6              |

From this table, we should notice that the values of the proposed TECs are in good agreement with those given by experiments (Sze 1981). Another important point from Table 1 is that the values of the proposed TECs are increasing with Ge mole fraction as those obtained from experiments. The following table (Table 2) shows the behavior of the proposed TEC of Si(1-x)Ge(x)(T) for different temperatures at  $x=0.17$ .

Table 2: Proposed Si(.83)Ge(.17) TEC as function of temperature

| Temperature T in °K | $\alpha_{Si(.83)Ge(.17)}(T)$ |
|---------------------|------------------------------|
| 300                 | 3.16e-6                      |
| 400                 | 1.44e-5                      |
| 500                 | 3.11e-5                      |
| 600                 | 5.29e-5                      |
| 700                 | 8e-5                         |
| 800                 | 1.12e-4                      |
| 900                 | 1.49e-4                      |

From the Table 2, the TEC of Si(0.83)Ge(0.17)(T) is increasing with temperature. The same thing is valid for the TECs of Si and Ge since they are polynomials. Then, we could use the proposed models to predict the TECs of Si, Ge, and SiGe for high temperatures. Some models found in literature can predict reasonably the TECs of these materials only at low temperatures. And, for temperatures greater than Debye temperature ( $\theta_D$ ) these models fail.

#### 4.2 Validation of the proposed TECs models using the numerically calculated channel stress

The intrinsic stress  $\sigma_0$  is used to simulate numerically the 3D extrinsic stress  $\sigma$  in the channel of an Intel 45 nm gate length PMOSFET shown in Figure 1. For the following numerical results, we used (001) for the substrate orientation and 17% as the Germanium mole fraction. The results in Figures 2 and 3 show the 3D distribution of stress components along channel for 300°C and 1000°C respectively. A similar stress distribution has been reported in (Victor 2004) for a similar structure. The values of the calculated 3D extrinsic stress are also qualitatively and quantitatively in good agreement with those calculated in (Victor 2004). Figure 4 shows the contour lines of the stress component in x direction at 300°C. All these results did show that the processing temperatures have significant effects on intrinsic and extrinsic stress profiles. These temperature effects will also affect the performances of the MOSFETs devices. On the other hand, these numerical results confirm that the proposed models for thermal expansion coefficients  $\alpha_{Si(1-x)Ge(x)}(T)$  and  $\alpha_{Si}(T)$  provide valid and correct results for the extrinsic stress  $\sigma$  in 3D. We also believe that these results are of great interest to the semiconductor community including manufacturers as Intel, IBM, TSMC, and academia.

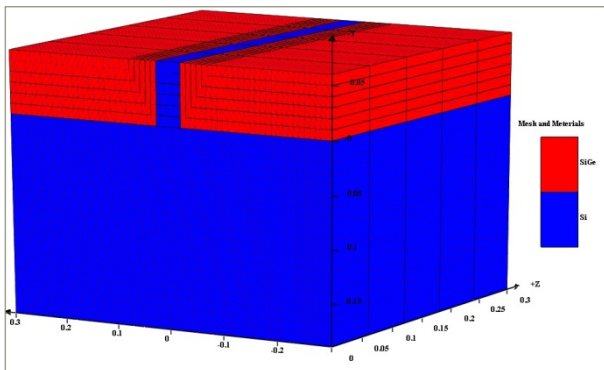


Figure 1: Materials and Mesh of The Simulated Structure

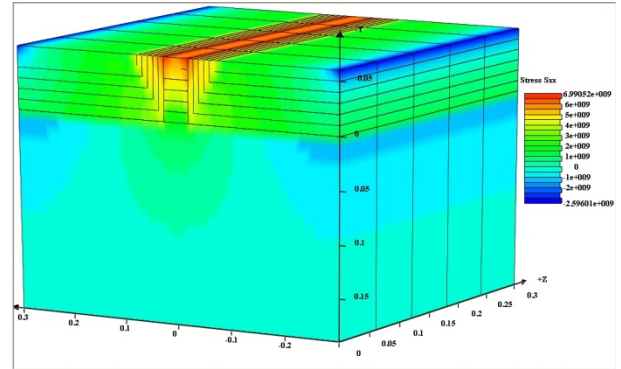


Figure 2: 3D Distribution of Stress Component Along Channel at 300°C

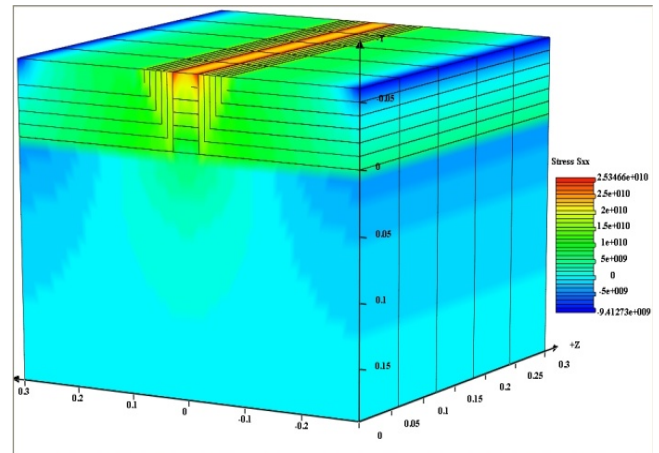


Figure 3: 3D Distribution of Stress Component Along Channel at 1000°C

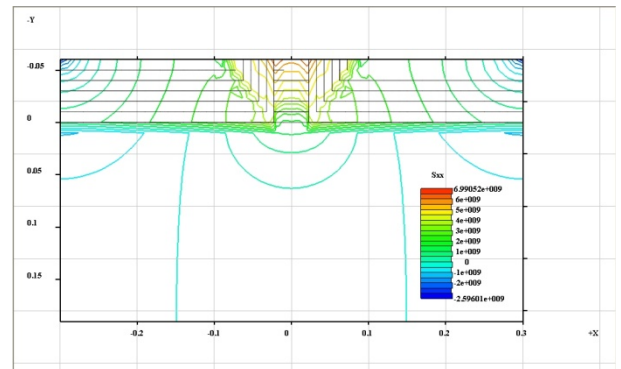


Figure 4: Contour Lines of Stress Component Along Channel at 300°C

## 5. CONCLUSIONS

In this paper, we have developed new second order models for the thermal expansion coefficients  $\alpha_{Si(1-x)Ge(x)}(T)$  and  $\alpha_{Si}(T)$ . The results we got from these models for TECs of Si, Ge, and SiGe are in excellent agreement with experiments. And, we have used these models to calculate accurately and analytically the intrinsic stress  $\sigma_0$  due to thermal mismatch after deposition. We have used the intrinsic

stress  $\sigma_0$  to calculate numerically the resulting extrinsic stress  $\sigma$  in the channel of the Intel 45 nm PMOSFET shown in Figure 1. The numerical results we have got for  $\sigma$  are also in good agreement, qualitatively and quantitatively, with those found in the literature for a similar structure (Victor 2004). In the future, we will apply these models to predict the TECs of other semiconductor materials and metals.

## REFERENCES

- Bera, L.K., 2006. The impact of uniform strain applied via bonding onto plastic substrate on MOSFET performance. *IEEE Electron Device Lett.*, Vol. 27, pp. 58- 61.
- Bernardi, M.,C., 1996. Isotope effects on the lattice constant by perturbation theory: ab initio calculations, *Phys. Rev.*, B 54, 11305-11310.
- Brash, B., 2004. Mobility enhancement in compressively strained SiGe surface channel PMOS transistors with HFO2/TIN gate stack. *Electrochemical society proceedings*, Vol. 7, pp. 12-30, San Antonio, (California, USA).
- Bruls, R.,J., 2001. The temperature dependence of the Grüneisen parameters of MgSiN, *J. Phy. Chem. Sol.* 62 (4), 783-792.
- Chui, K.J., 2007. n-MOSFET with silicon-carbon source/drain for enhancement of carrier transport, *IEEE Trans. Electron Devices*, Vol. 54, pp. 249-256.
- Debernardi, A., 2001, Ab-initio thermodynamics of metals: Al and W, *Phys. Rev.* B 63, 064305.
- El Boukili, A., 2010. 3D Stress Simulations of NanoTransistors. *Progress in Industrial Mathematics at 85-91.ECMI,85-91.*
- Fischetti, M., Laux, S., 1996. Band Structures Deformation Potentials, and Carrier Mobility in Strained Si, Ge, and SiGe Alloys. *J. Appl. Phys.*, Vol. 80, 2234-2240.
- El Boukili, A., 2010. 3D Stress Simulations of NanoTransistors. *Progress in Industrial Mathematics at 85-91.ECMI,85-91.*
- Garai, J., Laugier, A., 2007, The temperature dependence of the isothermal bulk modulus at 1 bar pressure, *J. Appl. Phys.* **101**, 023514–023514.
- Freund, L.B. and Suresh, S., 2003. *Thin Film: Materials:Stress Defect Formation and Surface Evolution.* Cambridge, United Kindom, Cambridge Unievrstity Press.
- Garai, J., Laugier, A., 2007, The temperature dependence of the isothermal bulk modulus at 1 bar pressure, *J. Appl. Phys.* **101**, 023514–023514.
- Giles, C., et al., 2005, Diamond thermal expansion measurement using transmitted X-ray back-diffraction, *J Synchrotron Radiat*, 349-53.
- 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* Washington, DC, USA.
- Horn, F., 1955. Densitometric and Electrical Investigation of Boron in Silicon. *Physical Review*, Vol. 97, 1521-1525.
- Hollauer, C., 2007. *Modeling of thermal oxidation and stress effects.* Thesis (PhD), Technical University of Wien.
- 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, Washington, DC, USA.
- Nirav, S., 2005. *Stress modeling of nanoscale MOSFETs.* Thesis (PhD). Florida University.
- Okada, Y., Tokumaru, Y., 1984, Precise determination of lattice parameter and thermal expansion coefficient of silicon between 300 and 1500 K, *J. Appl. Phys.* 56(2), 314–320.
- Randell, H., 2005. *Applications Of Stress From Boron Doping And Other Challenges in Silicon Technology.* Thesis (Master), University of Florida.
- Reeber, R.,R., 1975, Thermal expansion of some group IV elements and ZnS, *Physica Status Solidi*, Volume 32, Issue 1, 321-331.
- Reeber , R.R, Wang, K., 2000, Lattice parameters and thermal expansion of GaN, *Journal of Materials Research*, 15(1), 40-44.
- Rieger, M Vogl, P., 1993. Electronic-band parameters in strained Si(1-x)Ge(x) alloys on Si(1-y)Ge(y) substrate. *Phy. Rev. B*, Vol. 48, No 19, 14276-14287.
- Rim, E. et al., 2000. Fabrication and Analysis of Deep Submicron Strained-Si N-MOSFET's. *IEEE Transactions on Electron Devices*, Vol.47, No 7, 1406-1415.
- Roder, C., et al., 2005, Temperature dependence of the thermal expansion of GaN, *Phys. Rev. B* 72, 085218.
- Schaffler F., 2001, *Properties of Advanced Semiconductor Materials GaN, AlN, InN, BN, SiC, SiGe* . Eds. Levinshtein M.E., Rumyantsev S.L., Shur M.S., John Wiley & Sons, Inc., New York, 149-188.
- Takagi, S. et al., 2003. Channel Structure Design, Fabrication and Carrier Transport Properties of Strained-Si/SiGe-On-Insulator (Strained-SOI) MOSFETs. *IEDM Technical Digest*, pp. 57-60, 10 Dec., Washington, DC, USA.
- Talwar, D.,N., Sherbondy, J.C., 1995, Thermal expansion coefficient of 3C-SiC, *App. Phys. Lett*, 3301-3303.
- Toshimaro, S., et al., 2002, Thermal expansion of a high purity synthetic diamond single crystal at low temperatures, *Phys. Rev. B* 65, 092102.
- Van de Walle, C., Martin, R., 1986. Lattice constants of unstrained bulk Si(1-x)Ge(x). *Phy. Rev. B.*, Vol. 34, 5621-5630.
- Victor, M., 2004. *Analyzing strained-silicon options for stress-engineering transistors.* July Edition of Solid State Technology Magazine.

- Xie, J., et al., 1999, First-principles calculation of the thermal properties of silver. *Physical Review B*, 59, 965-969.
- Xu, C.H., et al., 1992, Theory of the thermal expansion of Si and diamond, *Phys. Rev. B* 43, 5024.
- Sze, S.M., 1981. *Physics of semiconductor devices*. New York, Johnson and Sons.

#### **AUTHOR'S BIOGRAPHY**

**Abderrazzak El Boukili** received both the PhD degree in Applied Mathematics in 1995, and the MSc degree in Numerical Analysis, Scientific Computing and Nonlinear Analysis in 1991 at Pierre et Marie Curie University in Paris-France. He received the BSc degree in Applied Mathematics and Computer Science at Picardie University in Amiens-France. In 1996 he had an industrial Post-Doctoral position at Thomson-LCR company in Orsay-France where he worked as software engineer on Drift-Diffusion model to simulate hetero junction bipolar transistors for radar applications. In 1997, he had European Post-Doctoral position at University of Pavia-Italy where he worked as research engineer on software development for simulation and modeling of quantum effects in hetero junction bipolar transistors for mobile phones and high frequency applications. In 2000, he was Assistant Professor and Research Engineer at the University of Ottawa-Canada. Through 2001-2002 he was working at Silvaco Software Inc. in Santa Clara, California-USA as Senior Software Developer on mathematical modeling and simulations of vertical cavity surface emitting lasers. Between 2002-2008, he was working at Crosslight Software Inc. in Vancouver-Canada as Senior Software Developer on 3D Process simulation and Modeling. Since Fall 2008, he is working as Assistant Professor of Applied Mathematics at Al Akhawayn University in Ifrane-Morocco. His main research interests are in industrial TCAD software development for simulations and modeling of opto-electronic devices and processes.

<http://www.aui.ma/personal/~A.Elboukili>.



# THE RELATIONSHIP BETWEEN R&D EXPENSES AND PERFORMANCE: EVIDENCE FROM EUROPEAN MANUFACTURING ENTERPRISES

Ondrej Zizlavsky<sup>(a)</sup>, Michal Karas<sup>(b)</sup>

<sup>(a)</sup> Faculty of Business and Management Brno University of Technology

<sup>(b)</sup> Faculty of Business and Management Brno University of Technology

<sup>(a)</sup> zizlavsky@fbm.vutbr.cz, <sup>(b)</sup> karas@fbm.vutbr.cz

## ABSTRACT

Measuring performance and contribution of R&D to performance has become a critical concern for managers and executives. This paper illustrates a link between R&D expenses and performance through a statistical model. We test whether and how innovation influences performance. Therefore, we consider R&D expenses (the independent variable) and other financial indicators of the enterprise's performance (the dependent variables). Enterprises from manufacturing industries have been chosen as an examined sample. The data was obtained from the Amadeus database in the period 2007 to 2012. From a managerial point of view, such a model should be useful in predicting how enterprises may invest in new R&D capabilities in the future. We found significant relationship between the return on assets and past R&D expenses. The main limitation of our analysis is, that data provide only a view into medium and large-sized companies due to focus on top R&D investing companies in the EU.

Keywords: R&D expenses, performance measurement and management, statistical model in measurement, European manufacturing companies.

## 1. INTRODUCTION

Innovation is generally considered as a major cause of enterprise's performance growth. In addition, the examination of the impact of innovation on enterprise's survival has shown that the ability to innovate increases survival probabilities for all enterprises across most manufacturing sectors (Cefis and Ciccarelli, 2005).

However, it is not clear that innovation has actually a positive impact on an enterprise's profits. Innovation may be considered as a largely random and unpredictable phenomenon (Brusoni et al., 2006). The relationship between innovation and performance (measured by R&D expenses and return on assets) is a priori unclear and it is by no means clear that innovative activities really lead to higher returns at the microeconomic level. Furthermore, a large part of all R&D expenses may no return at all and whether the innovative activities have a positive return on average is a matter that is not clear at the outset.

Many researches are conducting studies to determine the degree to which R&D really improves an

enterprise's performance (Bae and Kim, 2003; Chauvin and Hirschey, 1993; Smith, 2006; Szewczyk et al., 1996; Youndt et al., 2004). Czarnitzki and Kraft [2010] proved that innovation has a strong a robust impact on profitability. An innovating enterprise realizes an about 0.67 percentage points higher return on sales than an enterprise not performing innovations, on average. Therefore, enterprises invest into R&D in order to maximize their individual profits (Czarnitzki and Kraft, 2010).

There seems to be general agreement that the accounting definition of R&D is incredibly loose. It is also the case that output from various R&D processes may differ fundamentally. Research ranges, where the goal is to advance the state of the art along a predetermined dimension (Marshall, 1980).

More likely is that the R&D process is affected on the input side. Labor, capital, materials, and energy are combined to produce knowledge. Prominent Czech expert in innovation Valenta (1969, 2001) defined chain activity-innovation-effect. The activity presents human creativity, which lead into invention. Invention is knowledge basis for innovation. Well-managed and successfully-introduced innovation into the market gives rise to final effect.

In spite of the abundance of books and publication written over past few years in the field of performance measurement, the problem of defining a rigorous model for measuring innovation and its impact on company financial performance has not been solved yet (Lazarotti et al., 2011; Neely, 2005), although some notable and interesting attempts have been recently published (Apergis et al., 2013; Carayannis and Provan, 2008; Smith, 2006; Tohumcu and Karasakal, 2010). The most typical indicator used is R&D expenses (Gault, 2013; OECD, 2002; OECD, 2005).

In general, we can understand R&D expenses as an expenses associated with the R&D of an enterprise's goods or services. R&D expenses are a type of operating expense that can be deducted as such on the business tax return. This type of expense is incurred in the process of finding and creating new products or services.

R&D expenses can be relatively minor, or they can easily run into the billions of dollars for large enterprises. R&D expenses are usually the highest for industrial,



technological, healthcare and pharmaceutical enterprises. Some companies reinvest a significant portion of their profits back into R&D, as they see this as an investment in their continued growth.

World industrial R&D spending has reached a level of € 373 billion and is expected to grow continuously, in spite of the financial crisis and restructuring of the world economy after 2008 (Gerybadze, 2010).

We consider R&D expenses as an accounting item for which measurement under International Financial Reporting Standards (IFRS) are likely to differ considerably from measurements under domestic accounting systems across the EU countries prior to the mandatory introduction of the IFRS.

The scientific aim of this paper is to analyze the relationship between the current performance of an enterprise measured by the return on assets indicator and past R&D expenses. We focus on manufacturing enterprises, since this is the sector that undertakes the majority of total business R&D. In this paper, R&D expenses is expressed relatively, as a ratio of R&D expenses and other items in the same period (for example, number of employees, added value, value of fixed assets, etc.).

The paper is organized as follows. After introduction to basics and essential definition in view of interpreting them in the case of R&D managers and executives, the next section describes research method and basis for models. Section 3 presents the results and models of relationship between R&D expenses and performance of enterprise. Section 4 synthesis the highlights of our work, outlines some of its limitation, any suggests some direction for future research.

## 2. RESEARCH METHODS

The purpose of our paper, as previously stated, is to contribute to the knowledge of innovation by creating a link between innovation and performance from a managerial point of view. Considering the European manufacturing enterprises, we stated the following research question:

*Is there a relationship between performance (measured by return on assets) and R&D expenses?*

From the theoretical point of view, the goal of creating such a link is consistent with the data. The sample under our evaluation comprises 2,666 private enterprises active in the processing industry (NACE rev. 2 Main section C: Manufacturing) from the EU 28 countries in the period 2007 to 2012. The data was obtained from the Amadeus database provided to the company Bureau Van Dijk.

The performance of an enterprise is measured by the return on assets indicator or, in other words, the EBIT and total assets (ROA) ratio. The reason we use ROA a measure of performance is represented by characteristic of this ration. It allows the enterprise financing strategy to be neutral with respect to the performance (Ferrari and Rocca, 2010) and comparison among companies in the sample to be favored.

As regards R&D expenses, we take into consideration the total amount of expenses on research and development activities reported in Amadeus database according to IFRS. The relative R&D expenses has been defined in the following ways:

1. R&D expenses and sales ratio (RD/S) – expresses the proportion of R&D expenses per unit of sales in the given year, in other words the relative amount of R&D expenses where sales represent the size of the enterprise.
2. R&D expenses and number of employees ratio (RD/Empl.) – expresses the proportion of R&D expenses per employee in the given year,
3. R&D expenses and fixed assets ratio (RD/FA) – expresses the proportion of R&D expenses per unit of fixed assets in the given year.
4. Employee cost and R&D expenses ratio (EC/RD) – expresses the proportion of employee cost per unit of R&D expenses in the given year,
5. EBIT and R&D expenses ratio (EBIT/RD) – represents the profits (EBIT) per unit of R&D expenses,
6. Added value per R&D expenses ratio (AV/RD) – represents the degree of covering R&D expenses from added value.

Values beyond the interval ( $\mu-3\sigma$ ;  $\mu+3\sigma$ ) were removed from the sample of data under examination.

The analysis of the relationship between performance and R&D expenses was carried out in two stages; in the first stage the correlation was examined between the ROA values for 2012 and the values of relative R&D expenses in the years 2012-2007. To this end, Spearman's correlation coefficient was applied, particularly due to its non-parametric assumptions:

$$\rho = \frac{6 \sum_{i=1}^n (p_i - q_i)^2}{n(n^2 - 1)} \quad (1)$$

where

$p_i$  or  $q_i$  refers to the ranking of  $x_i$  or  $y_i$  values for the random quantity X or Y,

$n$  refers to the number of observations/number of values  $x_i$  or  $y_i$

In the second stage, a regression model was set up with only those variables (relatively defined R&D expenses) that showed a statistically significant correlation to the 2012 ROA, at least at a 5% significance level. The purpose of the model is to determine as to what degree the current performance (ROA) is influenced by past R&D expenses.

We assume that there is a quadratic dependence between the performance of enterprises and R&D expenses. The reason is that on the one hand, growing R&D expenses suggests an innovative activity which, if successful, leads to lower cost or higher revenues. On the

other hand, these costs decrease profits or EBIT no matter whether the innovation is successful. In other words, we assume that there is a theoretical value of relative R&D expenses which maximizes the current ROA.

To describe this dependence, we shall apply the following generalized model of linear regression (generalized linear model – GLM):

$$ROA_t = \alpha_{t-i} + \beta X_{t-i} + \gamma \cdot X_{t-i}^2 + \varepsilon = \alpha_{t-i} + \beta_1 x_{1(t-i)} + \beta_2 x_{2(t-i)} + \gamma_1 x_{1(t-i)}^2 + \gamma_2 x_{2(t-i)}^2 + \dots + \beta_n x_{n(t-i)} + \gamma_n x_{n(t-i)}^2 + \varepsilon \quad (2)$$

where

$ROA_t$  refers to the return on assets in year  $t$  where  $t$  stands for the year 2012,

$i$  refers to the number of periods prior to year  $t$  where  $i = 0, 1, \dots, 5$

$\alpha, \beta$  refer to regression coefficients,

$X_{t-i}$  represents a vector of independent variables defined in time,  $t-i$

### 3. EMPIRICAL FINDINGS

The following table shows the values of correlation between ROA in 2012 and the given relatively defined R&D expenses in the period  $t-i$ .

Table 1: Values of Spearman's correlation coefficient between ROA in 2012 and the given relative R&D expenses in the given period

|              | No. | Spearman | t (N-2) | p-value  |               | No. | Spearman | t(N-2)  | p-value  |
|--------------|-----|----------|---------|----------|---------------|-----|----------|---------|----------|
| RD/S_12*     | 143 | 0,1447   | 1,7370  | 0,084572 | EC/RD_12      | 297 | -0,0761  | -1,3109 | 0,190923 |
| RD/S_11      | 147 | 0,1282   | 1,5571  | 0,12162  | EC/RD_11      | 292 | -0,0368  | -0,6265 | 0,531455 |
| RD/S_10      | 146 | 0,1193   | 1,4418  | 0,151528 | EC/RD_10      | 288 | -0,0331  | -0,5607 | 0,57542  |
| RD/S_09      | 147 | 0,1097   | 1,3296  | 0,185747 | EC/RD_09      | 153 | -0,0533  | -0,6556 | 0,513102 |
| RD/S_08      | 146 | 0,1127   | 1,3612  | 0,175572 | EC/RD_08      | 27  | 0,1319   | 0,6652  | 0,512045 |
| RD/S_07      | 154 | 0,1028   | 1,2747  | 0,204358 | EC/RD_07      | 24  | 0,1670   | 0,7942  | 0,435538 |
| RD/Empl_12** | 419 | 0,1493   | 3,0841  | 0,002177 | EBIT/RD_12*** | 368 | 0,6011   | 14,3894 | 0,000000 |
| RD/Empl_11** | 412 | 0,1072   | 2,1838  | 0,029542 | EBIT/RD_11*** | 365 | 0,3757   | 7,7243  | 0,000000 |
| RD/Empl_10   | 411 | 0,0725   | 1,4697  | 0,142417 | EBIT/RD_10*** | 356 | 0,3142   | 6,2274  | 0,000000 |
| RD/Empl_09   | 276 | 0,0834   | 1,3857  | 0,166977 | EBIT/RD_09*** | 221 | 0,2786   | 4,2921  | 0,000027 |
| RD/Empl_08*  | 146 | 0,1379   | 1,6705  | 0,096999 | EBIT/RD_08*** | 99  | 0,3784   | 4,0259  | 0,000113 |
| RD/Empl_07   | 154 | 0,1213   | 1,5065  | 0,134021 | EBIT/RD_07*   | 106 | 0,1724   | 1,7853  | 0,077127 |
| RD/FA_12***  | 422 | 0,1940   | 4,0528  | 0,000060 | AV/RD_12      | 259 | 0,0312   | 0,5002  | 0,617335 |
| RD/FA_11***  | 419 | 0,1761   | 3,6539  | 0,000291 | AV/RD_11      | 254 | 0,0059   | 0,0931  | 0,925908 |
| RD/FA_10***  | 404 | 0,1577   | 3,2017  | 0,001475 | AV/RD_10      | 252 | 0,0069   | 0,1095  | 0,912896 |
| RD/FA_09***  | 274 | 0,1730   | 2,8969  | 0,004076 | AV/RD_09      | 135 | -0,0609  | -0,7037 | 0,482878 |
| RD/FA_08**   | 147 | 0,1647   | 2,0111  | 0,046166 | AV/RD_08      | 22  | 0,0322   | 0,1440  | 0,886933 |
| RD/FA_07**   | 155 | 0,1652   | 2,0714  | 0,040003 | AV/RD_07*     | 15  | 0,4857   | 2,0035  | 0,066426 |

Note: \*significant at a 10% level, \*\*significant at a 5% level, \*\*\*significant at a 1% level

Source: Own processing of data from the Amadeus database.

It is obvious from the table 1 that the R&D expenses and sales ratio (RD/S) does not have a statistically significant effect in any manner, at the 5% significance level, the enterprise performance values evaluated using ROA in any of the past periods under survey.

However, a statistically significant relationship exists in the same year, 2012, although only at the 10% significance level. This can be explained by the fact that R&D expenses reduces EBIT.

The employee cost/R&D expenses ratio or added value/R&D expenses ratio does not have a significant effect on the ROA values in 2012 at any standard significance level in any of the periods under survey.

In contrast, the R&D expenses per employee (RD/Empl.) values in the same year, as well as in the previous year, have an effect on ROA in 2012 at a statistically significant level, namely 5%. However, the correlation values reported are relatively low.

Stronger correlations can be identified between the 2012 ROA and the R&D expenses/fixed assets ratio defined for the years 2007 to 2012. In other words, such relatively defined R&D expenses has an effect on performance measured by the ROA relationship up to 5 years in advance.

The strongest correlation of all was identified between ROA in 2012 and the EBIT/R&D expenses ratio defined for the years 2008 to 2012. In other words, such a relatively defined R&D expenses has an effect on performance measured by the ROA relationship up to 4 years in advance.

Indicators that show a statistically significant correlation to ROA, at least at a 5% significance level, were used to set up a model or as an independent model variable in the (2) form. These indicators were RD/FA and EBIT/RD defined for the years 2008 to 2012. The role of dependent variable is given to the ROA indicator

defined for 2012. The following table shows the descriptive statistics of these variables.

Table 2: Descriptive statistics of ROA indicators (2012), RD/FA (2012-2018), EBIT/RD (2012-2008)

| Variable     | N.   | Mean   | Min.     | Max     | quant. (25%) | quant. (50%) | quant. (75%) | quant. (90%) | quant. (95%) | std. dev |
|--------------|------|--------|----------|---------|--------------|--------------|--------------|--------------|--------------|----------|
| ROA_2012     | 2579 | 0,078  | -0,541   | 0,711   | 0,023        | 0,068136     | 0,126        | 0,207        | 0,262        | 0,11239  |
| RD/FA_2012   | 424  | 0,196  | 0,000    | 2,296   | 0,008        | 0,042552     | 0,197        | 0,663        | 0,960        | 0,35886  |
| RD/FA_2011   | 421  | 0,202  | 0,000    | 2,389   | 0,008        | 0,043668     | 0,184        | 0,606        | 0,957        | 0,38208  |
| RD/FA_2010   | 406  | 0,154  | 0,000    | 1,578   | 0,011        | 0,04704      | 0,169        | 0,475        | 0,711        | 0,25914  |
| RD/FA_2009   | 275  | 0,154  | 0,000    | 1,719   | 0,004        | 0,043054     | 0,165        | 0,466        | 0,716        | 0,27004  |
| RD/FA_2008   | 148  | 0,165  | 0,000    | 1,362   | 0,000        | 0,042526     | 0,221        | 0,491        | 0,755        | 0,26106  |
| EBIT/RD_2012 | 370  | 10,419 | -143,333 | 158,276 | 0,544        | 3,561111     | 11,459       | 33,890       | 65,729       | 28,2159  |
| EBIT/RD_2011 | 367  | 12,494 | -157,118 | 202,778 | 0,863        | 3,286011     | 13,018       | 42,750       | 61,333       | 33,5472  |
| EBIT/RD_2010 | 359  | 9,107  | -55,430  | 113,394 | 0,700        | 3,585818     | 10,793       | 26,438       | 47,037       | 18,2143  |
| EBIT/RD_2009 | 222  | 5,847  | -90,114  | 127,641 | -0,462       | 1,761475     | 7,121        | 16,457       | 29,725       | 20,8587  |
| EBIT/RD_2008 | 100  | 4,838  | -20,524  | 46,810  | 0,417        | 2,373614     | 6,410        | 13,933       | 23,291       | 9,2822   |

Source: Own processing of data from the Amadeus database

The model for these applied variables can be transformed into the following formula:

$$ROA_t = \alpha_{t-i} + \beta_1(RD/FA)_{t-i} + \beta_2(EBIT/RD)_{t-i} + \gamma_1(RD/FA)_{t-i}^2 + \gamma_2(EBIT/RD)_{t-i}^2 + \varepsilon \quad (3)$$

Assuming that the parameters  $\gamma_1$ ,  $\beta_1$ ,  $\gamma_2$ ,  $\beta_2$  are statistically significant in the given model, it is reasonable to look for the values of indicators RD/FA<sub>t-i</sub> or EBIT/RD<sub>t-i</sub> that maximize the ROA<sub>t</sub> values. These values can be found when placing the first partial derivative of the ROA<sub>t</sub> function under RD/FA or EBIT/RD at equal to zero, as follows:

$$\frac{dROA_t}{d(RD/FA)_{t-i}} = \beta_1 + 2\gamma_1(RD/FA)_{t-i} = 0 \quad (4)$$

Then the RD/FA<sub>t-i</sub> value maximizing ROA<sub>t</sub>

$$\beta_1 + 2\gamma_1(RD/FA)_{t-i} = 0 \rightarrow (RD/FA)_{t-i} = -\frac{\beta_1}{2\gamma_1} \quad (5)$$

This value will maximize ROA<sub>t</sub> if  $\gamma_1 < 0$ .

Or

$$\frac{dROA_t}{d(EBIT/RD)_{t-i}} = \beta_2 + 2\gamma_2(EBIT/RD)_{t-i} = 0 \quad (6)$$

Then the EBIT/RD<sub>t-i</sub> value maximizing ROA<sub>t</sub>

$$\beta_1 + 2\gamma_1(EBIT/RD)_{t-i} = 0 \rightarrow (EBIT/RD)_{t-i} = -\frac{\beta_2}{2\gamma_2} \quad (7)$$

This value will maximize ROA<sub>t</sub> if  $\gamma_2 < 0$ .

### 3.1. Regression model set up

A total of five models were set up, depending on the time span between the independent variables and the dependent variable.

The first model, referred to as Model 12, explains the ROA values in 2012 using the values of the RD/FA and EBIT/RD variables defined in 2012. The second model, referred to as Model 11, explains the ROA values in 2012 using the values of the RD/FA and EBIT/RD variables defined in 2011. Analogously, only the last, fifth model (Model 08) explains the ROA values in 2012 using the values of the RD/FA and EBIT/RD variables defined in 2008.

All the thus-created models are, as a whole, statistically significant at a 1% level; see the following table.

Table 3: Overall characteristics of the models set up

|                         | Model 2012 | Model 2011 | Model 2010 | Model 2009 | Model 2008 |
|-------------------------|------------|------------|------------|------------|------------|
| Multiple R              | 0,4124     | 0,3305     | 0,3162     | 0,2648     | 0,3761     |
| Multiple R <sup>2</sup> | 0,1701     | 0,1092     | 0,1000     | 0,0701     | 0,1415     |
| Adjust R <sup>2</sup>   | 0,1608     | 0,0991     | 0,0893     | 0,0527     | 0,1046     |
| F-stat.                 | 18,2937    | 10,8229    | 9,3889     | 4,0160     | 3,8318     |
| p-value                 | 0,000000   | 0,000000   | 0,000000   | 0,003671   | 0,006292   |
| Std. Error              | 0,12642    | 0,13220    | 0,13383    | 0,14389    | 0,15582    |

Source: Own processing of data from the Amadeus database.

The values of multiple  $R^2$  in the individual models vary from 7.01 to 17.01%; in other words, the models can explain from 7.01 to 17.01% of ROA values using the values of relative R&D expenses with various time spans.

The following table provides details of the individual models. The parameters of these models are distinguished in the same way as the models themselves, e.g. EBIT/RD<sub>12</sub> is a variable of model 12, EBIT/RD<sub>11</sub> a variable of model 11, etc.

Table 4: Details of individual models

| Variable                               | Parameter | t-stat. | p-value  | Variable                             | Parameter | t-stat. | p-value  |
|----------------------------------------|-----------|---------|----------|--------------------------------------|-----------|---------|----------|
| Constant <sub>12</sub> ***             | 0,040800  | 4,2183  | 0,000031 | Constant <sub>09</sub> ***           | 0,050716  | 3,4954  | 0,000576 |
| EBIT/RD <sub>12</sub> ***              | 0,001951  | 7,4900  | 0,000000 | RD/FA <sub>09</sub> **               | 0,215699  | 2,4795  | 0,013931 |
| RD/FA <sub>12</sub> ***                | 0,222569  | 4,7733  | 0,000003 | EBIT/RD <sub>09</sub> ***            | 0,001767  | 2,6411  | 0,008876 |
| EBIT/RD <sup>2</sup> <sub>12</sub> *   | -0,000004 | -1,8619 | 0,063434 | RD/FA <sup>2</sup> <sub>09</sub>     | -0,10116  | -1,5014 | 0,134738 |
| RD/FA <sup>2</sup> <sub>12</sub> ***   | -0,107245 | -3,7203 | 0,000231 | EBIT/RD <sup>2</sup> <sub>09</sub> * | -1,2E-05  | -1,6611 | 0,098168 |
| Constant <sub>11</sub> ***             | 0,051003  | 5,0696  | 0,000001 | Constant <sub>08</sub>               | 0,012942  | 0,4326  | 0,666323 |
| RD/FA <sub>11</sub> ***                | 0,190982  | 4,0820  | 0,000055 | EBIT/RD <sub>08</sub> **             | 0,0074    | 2,4574  | 0,015848 |
| EBIT/RD <sub>11</sub> ***              | 0,001451  | 5,4218  | 0,000000 | RD/FA <sub>08</sub> **               | 0,425819  | 2,6103  | 0,010545 |
| RD/FA <sup>2</sup> <sub>11</sub> ***   | -0,08504  | -3,2782 | 0,001149 | EBIT/RD <sup>2</sup> <sub>08</sub>   | -7,8E-05  | -0,9161 | 0,361997 |
| EBIT/RD <sup>2</sup> <sub>11</sub> *** | -0,000006 | -3,5363 | 0,000460 | RD/FA <sup>2</sup> <sub>08</sub> *   | -0,28002  | -1,8984 | 0,060744 |
| Constant <sub>10</sub> ***             | 0,037308  | 3,1892  | 0,001560 |                                      |           |         |          |
| RD/FA <sub>10</sub> ***                | 0,229331  | 3,2004  | 0,001503 |                                      |           |         |          |
| EBIT/RD <sub>10</sub> ***              | 0,003511  | 4,9684  | 0,000001 |                                      |           |         |          |
| RD/FA <sup>2</sup> <sub>10</sub> **    | -0,121305 | -1,9701 | 0,049644 |                                      |           |         |          |
| EBIT/RD <sup>2</sup> <sub>10</sub> *** | -0,000026 | -2,7965 | 0,005462 |                                      |           |         |          |

Note: \*significant at a 10% level, \*\*significant at a 5% level, \*\*\*significant at a 1% level

Source: Own processing of data from the Amadeus database.

The parameter for the RD/FA variable, i.e. parameter  $\beta_1$  (see formula 3), is statistically significant at a 1% level in models 12, 11 and 10, and at a 5% level in models 09 and 08.

The parameter for the RD/FA<sup>2</sup> variable, i.e. parameter  $\gamma_1$  (see formula 3), is statistically significant at a 1% level in models 12 and 11, at a 5% level in model 10, at a 10% level in model 08 and is not statistically significant at any standard level of significance in model 09.

The parameter for the EBIT/RD variable, i.e. parameter  $\beta_2$  (see formula 3), is statistically significant at a 1% level in models with the exception of Model 08 where it is significant only at the 5% level.

The parameter for the EBIT/RD<sup>2</sup> variable, i.e. the  $\gamma_2$  parameter (see formula 3), is statistically significant at a 1% level in models 11 and 10, at a 10% level in models 12 and 09, and it is not statistically significant at any standard level of significance in model 08.

The constant, i.e. parameter  $\alpha_{t-i}$ , is statistically significant at a 1% level in all models, except for model 08 where it is not statistically significant at any standard level of significance. The value of the constant varies from 16.51 to 65.05% of the average ROA value in the sample.

### 3.2. Theoretical values of ROA-maximizing R&D expenses

Using relationships (5) and (7), it is possible to deduce the theoretical values of RD/FA<sub>t-i</sub> and EBIT/RD<sub>t-i</sub> that result in maximization of the ROA<sub>t</sub> value. They can provide a theoretical maximum value of ROA<sub>t</sub> corresponding to these indicators RD/FA<sub>t-i</sub> and EBIT/RD<sub>t-i</sub> (see Table 5).

Table 5: Theoretical values of ROA-maximizing EBIT/RD and RD/FA under evaluation.

| Variable              | Value | Variable            | Value | ROA**  |
|-----------------------|-------|---------------------|-------|--------|
| EBIT/RD <sub>12</sub> | 243,8 | RD/FA <sub>12</sub> | 1,04  | 0,3942 |
| EBIT/RD <sub>11</sub> | 120,9 | RD/FA <sub>11</sub> | 1,12  | 0,2460 |
| EBIT/RD <sub>10</sub> | 67,52 | RD/FA <sub>10</sub> | 0,95  | 0,2642 |
| EBIT/RD <sub>09</sub> | 73,63 | RD/FA <sub>09</sub> | 1,07* | 0,2307 |
| EBIT/RD <sub>08</sub> | 47,44 | RD/FA <sub>08</sub> | 0,76* | 0,3503 |

\*calculated using parameters that were not statistically significant

\*\*theoretical values of ROA<sub>t</sub> in application of ROA maximizing EBIT/RD and RD/FA

The actual value of ROA in 2012 ranged from 0.541 to 0.711. The median value was 0.0681. Only the top 10, resp. 5% of the surveyed enterprises reached values higher than 0.207, resp. 0.262.

Under optimal R&D expenses (measured by EBIT/RD and RD/FA), it is theoretically possible to achieve ROA values in the range of 0.230 to 0.394. However, it should be noted that further growth in R&D expenses would have led to a decrease in ROA.

A comparison of the actual and theoretical values shows that there are enterprises (approximately 5% of enterprises in our sample) that take advantage of this potential. Their R&D expenses are, in terms of the impact on ROA, optimal. Remaining 95% of the companies below this theoretical limit.

## 4. DISCUSSION

The performance of an enterprise is influenced by a wide range of factors, including innovation activity. One of the methods of its evaluation is to use the expenses expended on that activity in the period in question. Innovation

intensity can be evaluated by referring these expenses to some other variable. The potential benefits of innovation activity can be expected particularly in future periods.

The research performed examined the effect of R&D expenses in the period 2007 to 2012 on the values of the return on assets (ROA) of enterprises for the year 2012. It was ascertained that a statistically significant correlation can be found between the values of R&D expenses/fixed assets ratio, or EBIT/R&D expenses, defined in the period 2007 to 2012 and the return on assets value for 2012.

The fact that the current return on assets (ROA 2012) in the period 2012 to 2010 can be described using quadratic dependence where the coefficients in the quadratic member (see  $\gamma_1$  and  $\gamma_2$ ) are statistically significant and the negative values suggest that there are R&D expenses degrees that maximize this current ROA value. Thus, our assumption that there is a R&D expenses degree which optimally contributes to performance was confirmed for the above period. In other words, in the period concerned there is an optimum degree of R&D expenses definable for up to 2 years back (i.e. back to 2010). Lower R&D expenses values (compared with the optimum value) represent a reserve in increase in the performance of the enterprise, while higher R&D expenses values decrease this performance and it would be appropriate to reduce them.

Only the coefficients of those linear members of the models (see the  $\beta_1$  and  $\beta_2$  coefficients) that have positive values are statistically significant in the period 2009 to 2008. This can be interpreted in that the R&D expenses values in a period dating 3 to 5 years back only increase the current performance while higher R&D expenses values do not decrease current performance in the same period.

A statistically significant constant also exists in all the models, which can be interpreted in that the relevant part of the current ROA under examination (16.51 to 65.05% of the average value from 2012) can be attributed to other factors.

## 5. CONCLUSIONS

In our paper, we presents the results of an empirical study on the effects of R&D expenses on enterprise's performance.

In order to explain this performance of company (measured by return on assets ROA, one of the typical enterprise performance indicator) in terms of innovation strategies, we considered the quadratic model. Our representative sample of European manufacturing companies contains mostly medium and large-sized companies.

With the help of statistical model we found evidence on the positive impact of past R&D expenses on the return on assets.

Even if the results do not satisfy our expectation, due to relatively weak relationship between R&D expenses and performance, they are in some way consistent with previous studies (Ferrari and La Rocca,

2010; Gerybadze, 2010; Griffith et al., 2006; Barrett and Musso, 2010; Smith, 2006).

Most of previous studies (e.g. Bae and Kim, 2003; Smith, 2006; Zhu and Huang, 2012) utilized for evaluation of relationship between R&D expenses and enterprise performance linear relationship. In other words, they are based on an assumption that higher R&D expenses lead to higher performance. They have proved that innovation activities (measured by R&D expenses) is consistent within an enterprise from one year to another. They can be applied in planning within own enterprise. Moreover, they can be used to analyze and predict the innovation levels of competitors.

Ferrari and La Rocca [2010] found the relationship between performance (measured by ROA) and types of innovation and industry structure. Through the linear model they proved that Pavitt's (Pavitt, 1984) taxonomy and the different typologies of innovation are capable of influencing ROA.

Gerybadze [2010] set three types of strategy for large enterprises in mixed combinations: incremental innovation, dynamic growth strategy and industry creation.

Our paper extends previous studies by using quadratic dependence that allows the derivation of the optimal R&D expenses, which maximize the performance on the ROA level. By comparison of the actual and theoretical values we found that more than 95% of companies not fully exploit their innovation potential to performance increasing.

Although, we set theoretical models for innovative activities performance measurement, these models represent just one managerial tool and they should be combined with other techniques to contribute to the optimal decision making. Moreover, further research has to focus on stability of derived values in long term period and also the possibility of application of this approach for small enterprises.

## ACKNOWLEDGMENTS

The authors would like to thank Czech Science Foundation for its funding support within postdoc project No. 13-20123P "Innovation Process Performance Assessment: a Management Control System Approach in the Czech Small and Medium-sized Enterprises".

## REFERENCES

- Apergis, N, Eleftheriou, S. and Payne J.E. 2013. The relationship between international financial reporting standards, carbon emissions, and R&D expenditures: Evidence from European manufacturing firms. *Ecological Economics*, 88, 57-66.
- Bae, S. and Kim, D. 2003. The effect of R&D investments on market value of firms: evidence from the U.S., Germany, and Japan. *Multinational Business Review*, 11(3), 51-75.
- Barrett, Ch. and Musso, Ch. 2010. *R&D after the crisis: McKinsey Global Survey results*. McKinsey. Available from:

[http://www.mckinsey.com/insights/operations/r\\_and\\_nd\\_38d\\_after\\_the\\_crisis\\_mckinsey\\_global\\_survey\\_results](http://www.mckinsey.com/insights/operations/r_and_nd_38d_after_the_crisis_mckinsey_global_survey_results) [accessed 28 June 2014].

- Brusoni, S., Cefis, E. and Orsenigo, L. 2006. Innovate or die? A critical review of the literature on innovation and performance. *KITeS Working Paper No. 179*, 1-30.
- Carayannis, E.G. and Provan, M. 2008. Measuring firm effectiveness: towards a composite innovation index built on firm innovative posture, propensity and performance attributes. *International Journal Innovation and Regional development*, 1(1), 90-107.
- Cefis, E. and Ciccareli, M. 2005. Profits differentials and innovation. *Economics of Innovation and New Technology*, 14(1/2), 43-61.
- Chauvin, K. and Hirschey, M. 1993. Advertising, R&D expenditures and the market value of the firm. *Financial Management*, 22(1), 128-140.
- Czarnitzki, D and Kraft, K. 2010. On the profitability of innovative assets. *Applied Economics*, 42, 1941-1953.
- Ferrari, M. and Rocca, L.L. 2010. Innovation and performance: Some evidence from Italian firms. In: Epstein, M. et al., eds. *Performance measurement and management control: Innovative concepts and practices*. Bradford: Emerald Group Publishing, 115-141.
- Gault, F. 2013. *Handbook of innovation indicators and measurement*. Cheltenham: Edward Elgar Publishing.
- Gerybadze, A. 2010. R&D, innovation and growth: Performance of the world's leading technology corporations. In: Gerybadze, A. et al., eds. *Innovation and international corporate growth*. Berlin: Springer, 11-30.
- Griffith, R., Huergo, E., Mairesse, J. And Peters, B. 2006. Innovation and productivity across four European countries. *Oxford Review of Economic Policy*, 22(4), 483-498.
- Lazzarotti, V., Manzini, R. and Mari, L. 2011. A model for R&D performance measurement. *International Journal of Production Economics*, 134, 212-223.
- Marshall, W. 1980. Discussion of the economic effects of involuntary uniformity in the financial reporting of R&D expenditures and accounting for research and development costs: The impact on research and development expenditures. *Journal of Accounting Research*, 18, 84-90.
- Neely, A. 2005. The evolution of performance measurement research. Development in the last decade and a research agenda for the next. *International Journal of Operations & Production Management*, 25(12), 1264-1277.
- OECD. 2002. Frascati manual. Paris: OECD.
- OECD. 2005. Oslo manual, guidelines for collecting and interpreting innovation data. Paris: OECD.
- Pavitt, K. 1984. Sectoral patterns of technical change: Towards a taxonomy and a theory. *Research Policy*, 13, 343-373.
- Smith, R. 2006. Modelling R&D investments. *Research Technology Management*, 49(6), 16-22.
- Szewczyk, S., Tsetsekos, G. and Zantout, Z. 1996. The valuation of corporate R&D expenditures: evidence from investment opportunities and free cash flow. *Financial Management*, 25(2), 105-110.
- Tohumcu, Z. and Karasakal, E. 2010. R&D project performance evaluation with multiple and interdependent criteria. *IEEE Transactions on Engineering Management*, 99, 1-14.
- Valenta, F. 1969. *Tvurci aktivita-inovace-efekty*. Prague: Velryba.
- Valenta, F. 2001. *Inovace v manazerske praxi*. Prague: Velryba.
- Youndt, M., Subramaniam, M. and Snell, S. 2004. Intellectual capital profiles: an examination of investments and returns. *Journal of Management Studies*, 41(2), 335-361.
- Zhu, Z. H. and Huang, F. 2012. The Effect of R&D Investment on Firms Financial Performance: Evidence from Chinese Listed IT Firms. *Modern Economy*, 3 (1), 915-919.

#### AUTHORS BIOGRAPHY

**Ondrej Zizlavsky** is a Professor Assistant in the Institute of finances at Faculty of Business and Management Brno University of Technology, where he specialized in innovation, R&D management control, performance measurement systems and finances. His research primarily aims to understand what drives the success of innovation, how to measure and develop an innovative performance in the company. Part of his active research consists on studying the innovation process in a company and its performance measurement. He combines financial and non-financial metrics to set up a complex innovative performance measurement system for the Czech enterprises. His work has been published in two monographs, in a number of scientific peer-viewed journals, book chapters, as well as in handbook of innovative performance assessment. He has been a speaker at conferences on the subjects of innovation, performance management and innovative potential development. He is a member of advisory board of two scholarly and research centres and serves on the editorial board of five international scientific journals.

**Michal Karas** is a Professor Assistant in the Institute of finances at Faculty of Business and Management Brno University of Technology, where he specialized in possibilities of creating bankruptcy prediction models. His research primarily aims to understand why companies bankrupt and how far ahead the risk of bankruptcy could be detected. Moreover, what factors limits the robustness of bankruptcy prediction models and how the prediction accuracy of these models could be increased. His work has been published in a number of scientific peer-viewed journals and conference proceedings.

# IDENTIFICATION OF ENDOTHELIAL CELL MORPHOLOGY IN CORNEA USING EVOLUTION STRATEGIES

Lisa Obritzberger <sup>(a)</sup>, Susanne Schaller <sup>(a)</sup>, Viktoria Dorfer <sup>(a)</sup>,  
Claudia Loimayr <sup>(b)</sup>, Simone Hennerbichler <sup>(b)</sup>, Stephan Winkler <sup>(a)</sup>

<sup>(a)</sup> University of Applied Sciences Upper Austria  
Bioinformatics Research Group  
Softwarepark 13, 4232 Hagenberg, Austria

<sup>(b)</sup> Red Cross Blood Transfusion Service for Upper Austria  
Austrian Cluster for Tissue Regeneration  
Krankenhausstraße 9, 4020 Linz, Austria

<sup>(a)</sup> [lisa.obritzberger@fh-hagenberg.at](mailto:lisa.obritzberger@fh-hagenberg.at), [susanne.schaller@fh-hagenberg.at](mailto:susanne.schaller@fh-hagenberg.at), [viktoria.dorfer@fh-hagenberg.at](mailto:viktoria.dorfer@fh-hagenberg.at),  
<sup>(b)</sup> [claudia.loimayr@o.ropeskreuz.at](mailto:claudia.loimayr@o.ropeskreuz.at), [simone.hennerbichler@o.ropeskreuz.at](mailto:simone.hennerbichler@o.ropeskreuz.at), <sup>(a)</sup> [stephan.winkler@fh-hagenberg.at](mailto:stephan.winkler@fh-hagenberg.at)

## ABSTRACT

In this paper we describe an algorithm for identifying shapes of endothelial cells in images of human cornea. In order to analyze the quality of cornea samples that shall be transplanted, endothelial cells have to be counted and the morphology of cells has to be identified.

The algorithm for shape detection presented in this paper is based on evolution strategies (ES): After several pre-processing steps an initial shape is fitted into the detected cells and repeatedly evaluated and mutated in order to create new candidates from which the best ones are promoted to the next generation. In the experimental section of this paper we analyze the performance of this approach for identifying cell morphologies in images of human cornea samples using different fitness functions.

Keywords: structure identification, image analysis evolution strategies, human cornea analysis

## 1. INTRODUCTION

Cornea transplantation may be the only recovery chance for diseases such as Fuchs' Dystrophy, Keratoconus or similar disorders (Kampik and Grehn 2002). Here the cornea of the donor is taken and fixed at the right position of the recipient eye with few sutures. The post mortem enucleation and cornea processing according to European Community Directives 2004/23/EC, 2006/17/EC, and 2006/86/EC can be performed within 48 hours without loss of quality. To minimize risks of cornea transplantations, the tissue donation has to pass strict quality checks. Besides tests to exclude the possibility of HIV, hepatitis, and syphilis, corneas are also evaluated concerning their quality.

The quality of corneas can be defined by two criteria: the endothelial cell count and the morphology

of consisting cells. As cornea endothelial cells are not able to undergo cell division and consequently cannot be replicated, the number of endothelial cells decreases over the years. Therefore, higher donor age leads to a reduced endothelial cell count (Loimayr et al. 2012).

The minimum of endothelial cell count should be 2000 cells per mm<sup>2</sup> to be adequate for transplantation use (Dichtl et al. 2010). Quality determination through cell counting is in many tissue banks still a manual process. Usually, cornea endothelial cells have a hexagonal cell form as apparent in Figure 1.

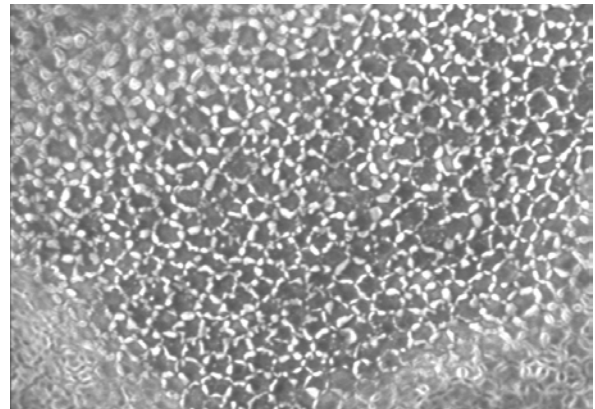


Figure 1: Endothelial cells of which most have hexagonal shapes.

The typical form of endothelial cells is lost during apoptosis. In this process, cells lose their original form and size and cannot be counted as hexagonal. The higher the number of cells with hexagonal shape, the better is the quality of a cornea and the higher is the possibility to be used for transplantation.

Thus it is not only important to identify the endothelial cell count, but also to analyze the shape of cells on the cornea that are to be used for transplantation. The goal of the research work presented in this paper is to develop an algorithm that is able to automatically identify the number of cells and the classification according to their shape.



## 2. STATE OF THE ART

Currently, there are only a view tools for analyzing images of corneas. There is for example the Endothel Analysis System of the German company Rhine-Tec, which rather concentrates on counting the cells and makes it not clear how the identification of the cell morphology is calculated. Another software specialized on endothelial cell count is the NAVIS Cell Count-Advanced Vision Information System of the German company Nidek Technologies, which offers also functionalities to identify the cells shape. Researchers have compared these two software tools against manual endothelial cell counts and their studies have shown that the fully automated analysis of cornea cells remains problematic (Hirneiss et al. 2007).

There are also several cell counting tools that are not geared towards endothelial cell count and hence not optimally appropriate for this purpose.

Due to the lack of appropriate cell counting tools, in many tissue banks the cell count and cell morphology identification is still a manual progress, for which an alternative shall be shown in this paper.

## 3. DATA PRE-PROCESSING

In this section we describe the pre-processing steps that have been performed before applying evolution strategy on processed images. The following pre-processing steps are applied: identification of sharp areas in images, dilation, transformation to binary images, and cell detection using a flood fill algorithm.

### 3.1. Identification of Sharp Areas

Microscopy images of cornea samples are out of focus at the edges, which is caused by curvature of the cornea. In these areas a proper analysis of endothelial cells is not possible and would falsify the results. Therefore, the focused area of cornea images has to be detected. This is achieved by using the change of intensity in the images: Areas that are out of focus have smoother transitions than focused areas hence change of intensity is higher in focused areas of the image.

First the intensity of each pixel of the image is calculated by using the RGB-data of the pixel  $p$ . The thereby used definition can be seen in Equation 1, where  $p.R$  represents the red component,  $p.G$  the green component, and  $p.B$  the blue component of pixel  $p$  (Nagabhushana, 2005).

$$intensity = 0.3 \times p.R + 0.59 \times p.G + 0.11 \times p.B \quad (1)$$

In the next step a radius is defined in which each pixel is compared to its neighbors with regard to the intensity value. If the difference between the maximum and the minimum intensity value in the neighborhood exceeds a pre-defined threshold, the centered pixel is marked as focused.

To identify not only sharp pixels, but also whole areas as focused, a second radius is defined, in which a minimum number of pixels that have been marked as sharp has to be reached. If the number of sharp pixels in the neighborhood is at least as high as the pre-defined minimum, then the pixel belongs to a focused area. An example for this is shown in Figure 2.

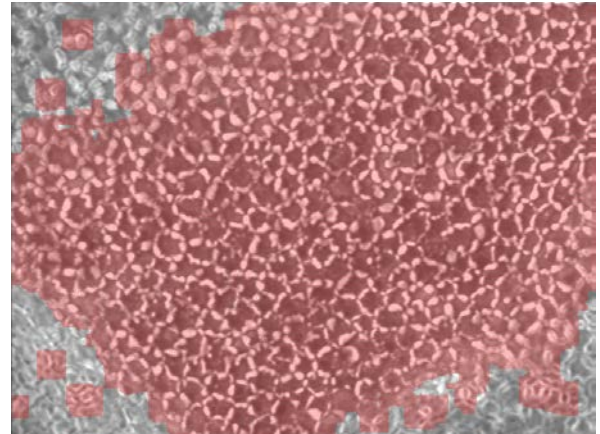


Figure 2: Focused area identified in a cornea image.

### 3.2. Dilation for Cell Margin

In order to identify pixels belonging to a cell area proper cell margin has to be performed. Therefore, the holes between the cell membrane of endothelial cells have to be closed. To achieve the closing of cell membrane dilation is processed on the cornea cell image. Dilation uses a mask that causes objects to increase and thereby fills small holes between objects (Jähne 2012) as depicted in Figure 3.

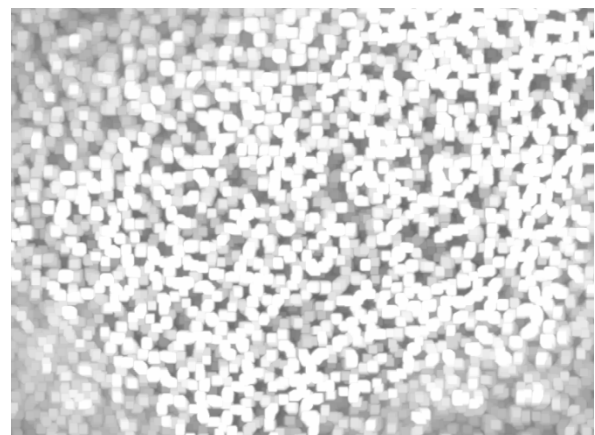


Figure 3: Cornea image after performing dilation.

### 3.3. Transformation to Binary Images

To distinguish cell areas from cell membrane, we consider the different intensity levels in the images. A transformation to a binary image is achieved using a pre-defined threshold; intensity levels below the



threshold are transformed to black representing cell pixels, others to white ones.

### 3.4. Flood Fill Algorithm for Cell Separation

To identify pixels that belong to the same cell a recursive flood fill algorithm is applied. A point within the cell area is selected as seed from where all 4-connected pixels, if they are also part of the cell, are colored. The process is repeated recursively for each of the connected pixels until there is no neighboring pixel left that can be colored (Pachghare 2008). This leads to images as the one shown in in Figure 4:

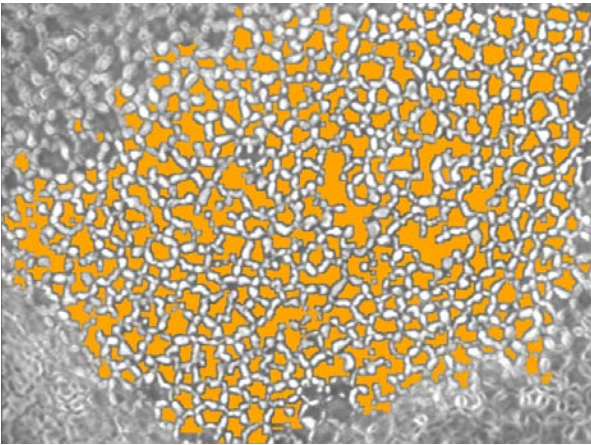


Figure 4: Detected cells after performing a flood fill algorithm.

## 4. IDENTIFICATION OF ENDOTHELIAL CELL MORPHOLOGY IN IMAGES OF CORNEA SAMPLES BASED ON EVOLUTION STRATEGIES

In this section we describe an approach for identifying cell forms in images using evolution strategies: The fitness of several created shapes is repeatedly evaluated and mutated in order to generate new solution candidates.

The initial shapes are created with a randomly chosen number of corners in the range [4, 7]. This form is fitted into the cell and evaluated. Then, mutation is executed by changing the corners of the initial shape, followed by re-evolution; solution candidates with the best fitness are chosen as new parents for the next generation.

### 4.1. Evolution Strategies

Evolutionary algorithms are inspired by the Darwinian paradigm of evolution, and use the principle of variation and selection combined with generational changes (Beyer 2001). The two main types of evolutionary algorithms are genetic algorithms and evolution strategies.

Evolution strategies were developed since the 1960s, primarily by a German research community around Rechenberg and Schwefel at the Technical University of Berlin, and have been extensively studied in Europe (see for example (Rechenberg 1973) and (Schwefel 1994)).

As one of the most representative of evolutionary computation ES executes the optimization process by applying operators in a loop. Each step of the algorithm's execution is known as a generation. Over generations main operations are applied on the solution candidates repeatedly until a given termination criterion is met. One of the termination criteria can for example be the reaching of a pre-defined number of generations.

An ES works with a population of individuals which are also called solution candidates and are characterized by their parameter vector. The vector is used to calculate the individual's fitness value.

In each generation the old population is replaced by successful child individuals. Individuals are represented as real-valued vectors in which each position corresponds to a feature of the individual.

In contrast to genetic algorithms, where mutation is mainly used for avoiding stagnation, mutation is the main reproduction operator in evolution strategies. In each generation each component of the vector is mutated individually, where small mutations are more likely than big ones.

As the algorithm converges with increasing quality of the solution candidates, the mutation width  $\delta$  (convergence rate) should be adapted. For the purpose of self-adaptation, the 1/5 success rule proposed by Rechenberg (for Details see Rechenberg 1973) is processed. Here the percentage of successful mutations  $s$  is recorded over  $n$  generations, to calculate the convergence rate as seen in Equation 2, whereas  $c_d = 0.82$  and  $c_i = 1 - 0.82$ .

$$\delta(t) = \begin{cases} c_d \times \delta(t - n) & \text{if } s < \frac{1}{5} \\ c_i \times \delta(t - n) & \text{if } s > \frac{1}{5} \\ \delta(t - n) & \text{if } s = \frac{1}{5} \end{cases} \quad (2)$$

Another possibility to create new individuals can be recombination where a child is created by recombining two solution candidates (parents). This can for example be processed by calculating their geometric average. Strategies for choosing candidates for recombination and recombination techniques can be very different. Typically, parent selection in ES is performed uniformly and random, with no regard to fitness.

The algorithm described in this paper does not use recombination.

Examples for mutation and recombination in the context of ESs are shown in Figure 5.

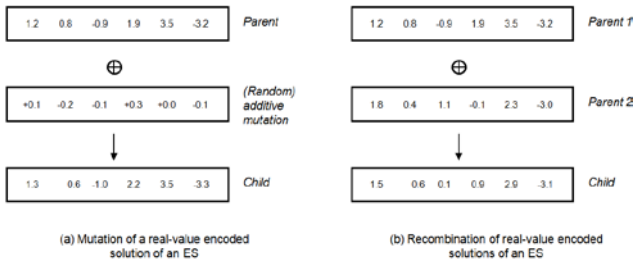


Figure 5: Exemplary solution candidates and the effect of genetic operations in ES (Winkler 2009).

The selection of individuals for the next generation is a deterministic method that uses the survival of the fittest principle: only the best individuals remain.

In each generation of an ES algorithm,  $\lambda$  children are produced by  $\mu$  parent individuals. By selection, the best children are chosen and become the parents of the next generation. The  $\mu$  best individuals are based on the relative ordering of fitness values.

Basically, there are two selection strategies for ESs: the  $(\mu, \lambda)$ -strategy and the  $(\mu + \lambda)$ -strategy. We have focused on the  $(\mu + \lambda)$ -strategy, which is also called plus-selection and allows not only the  $\lambda$  produced children, but also the  $\mu$  parents to be included in the pool of potential new parents. The best individuals of parents and children are chosen as the next generation's parents.

The main procedure steps of the execution of ES is summarized and graphically shown in Figure 6.

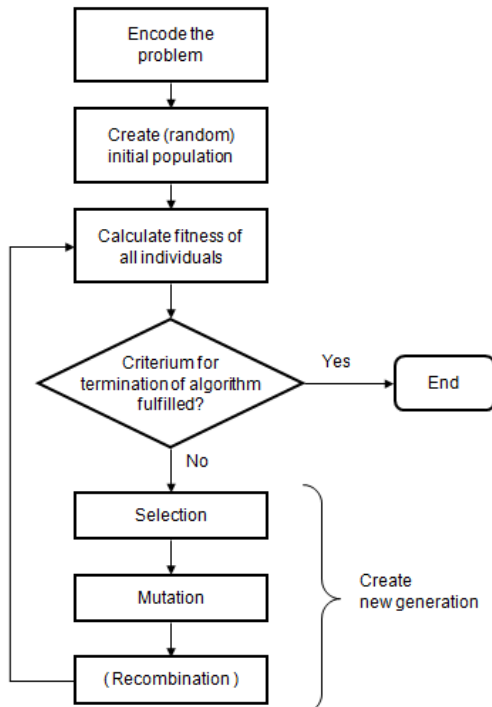


Figure 6: Workflow of the standard evolution strategy (ES) algorithm (Winkler 2009).

## 4.2. Cell Morphology Identification using Evolution Strategies

### Solution Candidates

A solution candidate for the ES performed to detect the correct cell morphologies is represented as a list of points which describe the corners of a shape. Figure 7 shows an example:

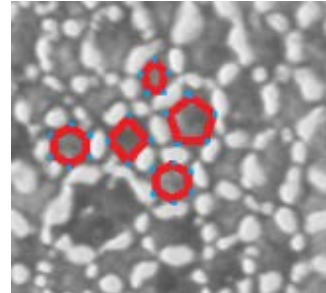


Figure 7: Idea of fitting shapes into cells; shapes are defined by their corners.

### Initial Candidates

Initial solution candidates have a uniformly distributed possibility of being quadrangular, pentagonal, hexagonal or heptagonal. As the best fit can be created by using pixels that are part of the cell border, the cell border pixel with the minimum distance from the original corner of the initial solution candidate is calculated for each corner and assigned as the new corner. The so created solution candidates are the first individuals of the ES' population.

### Mutation

After calculating the fitness of each parent they are mutated in order to create children as new candidates from whom the best ones are promoted to the next generation. Mutation is here accomplished in two ways:

- Mutation of points: Each corner of a parent solution candidate has a pre-defined possibility of being mutated. This mutation causes a small movement of the edges and thereby changes the shape and area of the original solution candidate.
- Change of the number of points: There can either be a new corner added to the shape by mutating an existing corner and adding it as an additional one, or a randomly chosen corner can be deleted.

### Fitness Function

The fitness function has to return a value that reflects how well the shape fits into the original cell area. To get the fitness  $f$  of a solution candidate  $s$ , the area of the created shape is calculated and compared with the area of the cell. Overlapping areas lead to

improvement of the fitness value, whereas area that does not fit leads to fall off in fitness quality. Thereby the overlapping areas are weighted with a value  $\alpha_1$  and points outside the shape are weighted with a value  $\alpha_2$  as defined in Equation 3:

$$f(s) = (\sum_{p \in ca(s)} I(p)) \times \alpha_1 - (\sum_{p \notin ca(s)} I(p)) \times \alpha_2 \quad (3)$$

Where  $ca(s)$  is the area covered by the shape  $s$ . The aim of this approach is to maximize the values returned by the fitness function. The degree of reward / punishment depends on the grayscale intensity  $I$  of the pixels. In the experiments summarized in this paper  $\alpha_1$  was always greater than  $\alpha_2$ .

As it is easier for shapes with a higher number of corners to fit the original cell, fitness has to be weighted by the number of corners.

## 5. EXPERIMENTAL RESULTS

Several variants of this solution evaluation approach have been evaluated and are described in the following section.

### 5.1. Using Fitness Function with Non-Weighted Corners

Initial ES runs have been performed without weighting the fitness by the number of corners of a shape. As it is easier for shapes with higher number of corners to fit a cell, a non-corner weighted fitness function favored shapes with higher corner number which is shown in Figure 8.

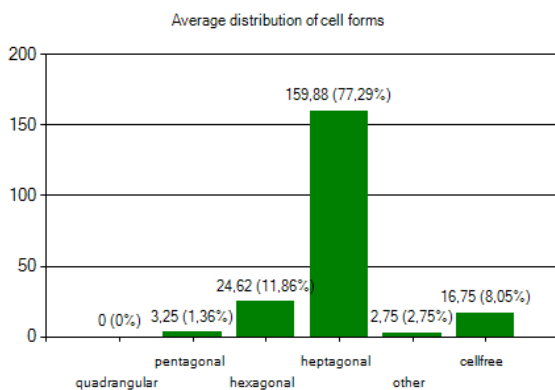


Figure 8: Distribution of identified cells using the non-corner weighted fitness function.

To further evaluate the quality of the results a manual analysis has been executed where the shape of the cells (originally defined by a human expert) was

compared to the result of the algorithm. We have used 8 images of corneas provided by the Red Cross Blood Transfusion Service for Upper Austria. The evaluation of the results of five independent identification runs per image (ES with 100 generations and 1 parent with 10 offspring) are summarized in Table 1. The overall classification of calculated forms compared to the original shape classification can be seen in the confusion matrix shown in Table 2.

| Image Sample        | Classified correctly |                    |
|---------------------|----------------------|--------------------|
|                     | Average              | Standard deviation |
| 1                   | 32.50%               | 0.07               |
| 2                   | 12.86%               | 0.03               |
| 3                   | 14.67%               | 0.03               |
| 4                   | 13.33%               | 0.05               |
| 5                   | 9.09%                | 0.00               |
| 6                   | 4.62%                | 0.04               |
| 7                   | 13.85%               | 0.08               |
| 8                   | 1.33%                | 0.03               |
| <b>Mean Quality</b> | <b>12.78%</b>        | <b>0.10</b>        |

Table 1: Average results achieved with five runs, each with 100 ES iterations per cell, using a fitness function with non-weighted corners.

| expected / identified | 4     | 5      | 6      | 7     |        |
|-----------------------|-------|--------|--------|-------|--------|
| 4                     | 0.00% | 0.00%  | 0.00%  | 0.00% |        |
| 5                     | 0.00% | 0.20%  | 0.20%  | 0.00% |        |
| 6                     | 0.41% | 2.24%  | 2.86%  | 0.61% |        |
| 7                     | 2.65% | 25.10% | 57.14% | 8.57% |        |
|                       |       |        |        |       | 11.63% |

Table 2: Confusion matrix with average classification ratios for cells in 8 cornea images. Classifications computed in 5 independent test runs per cell are compared to original cell shapes classifications defined by a human expert. The results were achieved by using a fitness function with non-weighted corners.

### 5.2. Using Fitness Function with Non-Linear Functions

To avoid drifting to shapes with higher numbers of corners, tests with a weighted fitness function have been performed: The calculated fitness values have been weighted by dividing them by the number of corners, by the logarithm of number of corners and by the square root of number of corners.

The results obtained using these functions are similar, they all perform better than the non-corner weighted fitness function; still, shapes with higher number of corners are favored which can be seen in Figure 9. Though results improved they were still not satisfying which can be seen in Table 3 and Table 4.

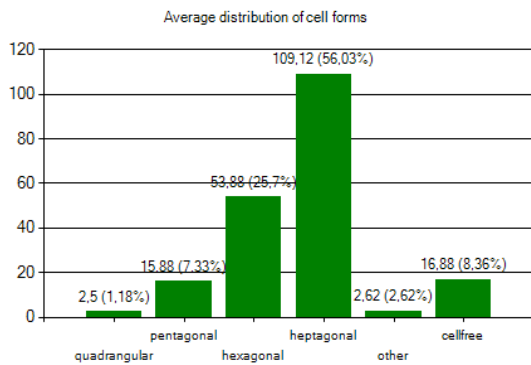


Figure 9: Distribution of identified cells using the fitness function weighted by logarithm of corner number.

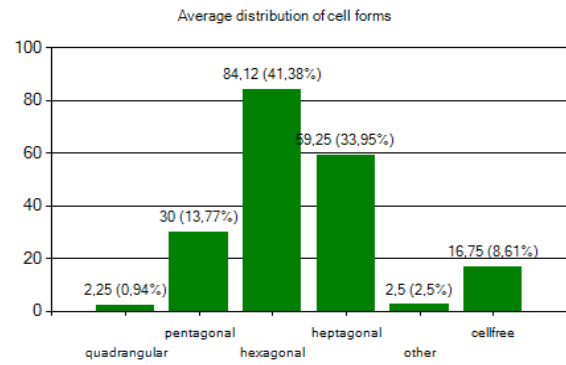


Figure 10: Distribution of identified cells using fitness function weighted by pre-defined values.

| Image Sample        | Classified correctly |                    |
|---------------------|----------------------|--------------------|
|                     | Average              | Standard deviation |
| 1                   | 40.00%               | 0.21               |
| 2                   | 24.29%               | 0.10               |
| 3                   | 34.67%               | 0.16               |
| 4                   | 26.67%               | 0.13               |
| 5                   | 21.82%               | 0.08               |
| 6                   | 16.92%               | 0.03               |
| 7                   | 23.08%               | 0.05               |
| 8                   | 20.00%               | 0.11               |
| <b>Mean Quality</b> | <b>25.93%</b>        | <b>0.13</b>        |

Table 3: Average results achieved with five runs, each with 100 ES iterations per cell, with a fitness function weighted by logarithm of corner number.

| expected / identified | 4     | 5      | 6      | 7     |               |
|-----------------------|-------|--------|--------|-------|---------------|
| 4                     | 0.00% | 0.20%  | 0.82%  | 0.20% |               |
| 5                     | 0.20% | 1.63%  | 6.73%  | 0.82% |               |
| 6                     | 1.22% | 10.61% | 18.57% | 3.06% |               |
| 7                     | 1.63% | 15.10% | 34.08% | 5.10% |               |
|                       |       |        |        |       | <b>25.31%</b> |

Table 4: Confusion matrix with average classification ratios for cells in 8 cornea images. Classifications computed in 5 independent test runs per cell are compared to original cell shapes classifications defined by a human expert. The results were achieved by using a fitness function weighted by logarithm of corner number.

### 5.3. Using Fitness Function with Linear Weighting

As the previous methods did not show the expected success, the fitness was weighted by a pre-defined value depending on corner numbers which are depicted in Table 5. This approach has led to better results with mainly hexagonal cell forms, which can be seen in Figure 10.

| Numbers of corners | Weight |
|--------------------|--------|
| 4                  | 1.6    |
| 5                  | 1.4    |
| 6                  | 1.2    |
| 7                  | 1.0    |

Table 5: Weights for shapes depending on corners.

To evaluate the performance of this approach in more detail, the change of fitness over generations was observed and showed that the shapes' fitness improves over time. This can be seen in Figure 11:

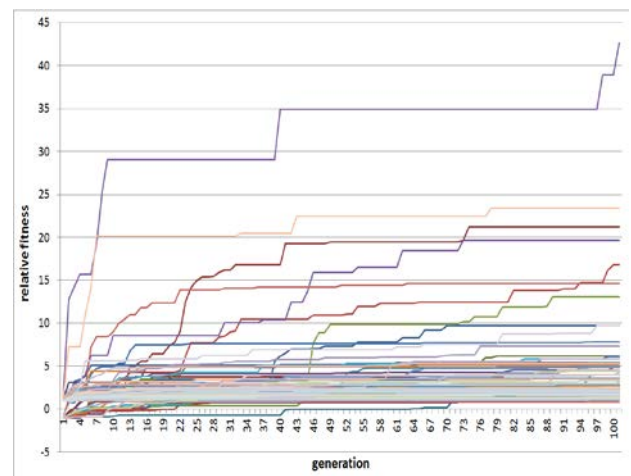


Figure 11: Improvement of fitness of shapes over generations for randomly selected cells.

The shape with the best fitness is selected to represent the morphology of the cell and plotted into the original image which is shown in Figure 12:

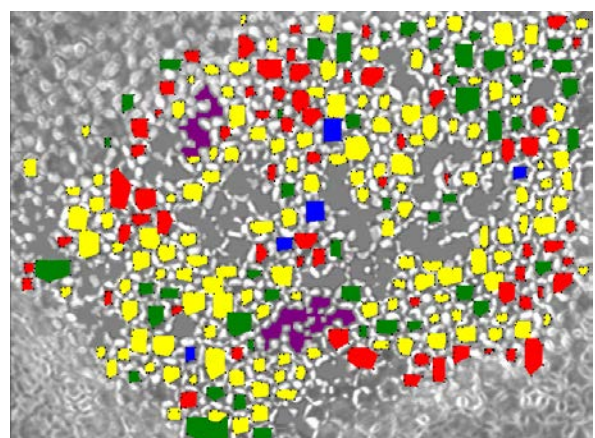


Figure 12: Identified shapes with the best fitness fitted into endothelial cells.



The evaluation of the results of five independent identification runs per image (ES with 100 generations) are summarized in Table 6. The overall classification of calculated forms compared to the original (human) shape classification can be seen in the confusion matrix shown in Table 7.

With this approach 70.03% of the samples were classified correctly. By accepting an error of + / - one corner the number of correctly classified samples is even 92.80%.

| Image Sample        | Classified correctly |                    |
|---------------------|----------------------|--------------------|
|                     | Average              | Standard deviation |
| 1                   | 62.10%               | 0.13               |
| 2                   | 62.90%               | 0.04               |
| 3                   | 72.17%               | 0.06               |
| 4                   | 67.94%               | 0.08               |
| 5                   | 74.07%               | 0.12               |
| 6                   | 74.05%               | 0.11               |
| 7                   | 70.98%               | 0.11               |
| 8                   | 69.92%               | 0.07               |
| <b>Mean Quality</b> | <b>69.27%</b>        | <b>0.09</b>        |

Table 6: Average results achieved with five runs, each with 100 ES iterations per cell, with a fitness function with linear weighted corners.

| expected / identified | 4     | 5     | 6      | 7     |        |
|-----------------------|-------|-------|--------|-------|--------|
| 4                     | 0.31% | 0.32% | 1.47%  | 0.30% |        |
| 5                     | 0.60% | 8.77% | 7.67%  | 1.32% |        |
| 6                     | 1.03% | 6.86% | 51.62% | 3.64% |        |
| 7                     | 0.14% | 2.93% | 3.68%  | 9.33% |        |
|                       |       |       |        |       | 70.03% |

Table 7: Confusion matrix with average classification ratios for cells in 8 cornea images. Classifications computed in 5 independent test runs per cell are compared to original cell shapes classifications defined by a human expert. The results were achieved by using a fitness function with linear weighted corners.

## 6. CONCLUSION

In this paper we have described an evolutionary algorithm that is able to automatically identify the morphology of endothelial cells of human cornea in microscopy images. In future work this approach shall be included in a fully automated image analysis framework that is designed for biomedical research and production.

## ACKNOWLEDGMENTS

The work described in this paper was done within the FIT-IT project number 835918 "NanoDetect: A Bioinformatics Image Processing Framework for Automated Analysis of Cellular Macro and Nano Structures" sponsored by the Austrian Research Promotion Agency (FFG).

## REFERENCES

- Beyer, H., 2001. *The Theory of Evolution Strategies*. Heidelberg: Springer Verlag.
- Dichtl M., Gabriel C., Hennerbichler S., Seitz B., Priglinger SG., *EU konformes Hornhautbanking – eine Bestandsaufnahme Hornhautbank Linz*, Spektrum der Augenheilkunde, 2010;24(3): 166-173
- Hirneiss C., Schumann RG., Grüterich M., Welge-Luessen UC., Kampik A., Neubauer AS., *Endothelial cell density in donor corneas: a comparison of automatic software programs with manual counting*. PubMed, 02/2007; 26(1):80-3
- Jähne, B., 2012. *Digitale Bildverarbeitung und Bildgewinnung*. Heidelberg: Springer Verlag.
- Kampik, A., Grehn, F., 2002. *Augenärztliche Therapie*. Stuttgart: Georg Thieme Verlag.
- Loimayr, C., Schuster, G., Gabriel, C., Hennerbichler-Lugscheider, S., 2012. *Comparison of Cornea Donor Age and post-mortem Time to Endothelial Cell Count*. Poster presented at the European Association of Tissue Banks, Vienna.
- Nagabhushana, S., 2005. *Computer Vision and Image Processing*. New Delhi: New Age International(P) Ltd., Publishers.
- Pachghare, V. K., 2008. *Comprehensive Computer Graphics*. Laxmi Publications
- Rechenberg, I., 1973. *Evolutionsstrategie*. Friedrich Frommann Verlag.
- Schwefel, H.-P., 1994. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Basel: Birkhäuser Verlag.
- Winkler, S., 2009. *Evolutionary System Identification: Modern Concepts and Practical Applications*. Linz: Trauner Verlag.

## AUTHORS BIOGRAPHIES

**LISA OBRITZBERGER** received her Bachelor of Science in Medical and Bioinformatics in 2012 from the University of Applied Sciences Upper Austria (UAS). She is a research associate at the Research & Development GmbH, in Hagenberg. Her research interests include medical image analysis, tissue engineering, machine learning and data mining. 2014 she received her Masters of Science in Biomedical Informatics from UAS.



**SUSANNE SCHALLER** received her Masters of Science in Biomedical Informatics in 2009 and 2010 from the University of Skovde and Upper Austrian University of Applied Sciences. She is a research and teaching associate at the Research Center Hagenberg of UAS. Her research interests include system biology, medical image analysis, tissue engineering, data mining and machine learning.



**VIKTORIA DORFER** is a senior researcher in the field of bioinformatics at the Research Center Hagenberg, School of Informatics, Communications and Media. After finishing the diploma degree of bioinformatics in 2007 she was a team member of various projects and is currently pursuing her PhD studies in bioinformatics.



**CLAUDIA LOIMAYR** works at the Red Cross Blood Transfusion Service of Upper Austria in Linz. Her research interests include cornea transplantations and tissue engineering.



**SIMONE HENNERBICHLER** is Head of Production at the Red Cross Blood Transfusion Service of Upper Austria in Linz since 2006. She holds a master's and a PhD degree in Pharmacy from Karl-Franzens-University Graz and is responsible for development and production of blood and tissue products. Previously she was involved in stem cell and tissue regeneration research and teaching.



**STEPHAN M. WINKLER** received his PhD in engineering sciences in 2008 from Johannes Kepler University (JKU) Linz, Austria. His research interests include genetic programming, nonlinear model identification and machine learning. Since 2009, Dr. Winkler is professor at the Department for Medical and Bioinformatics at the University of Applied Sciences (UAS) Upper Austria at Hagenberg Campus; since 2010, Dr. Winkler is head of the Bioinformatics Research Group at UAS, Hagenberg.



# METHODOLOGY FOR FAILURE ANALYSIS IN SERVICE PROCESSES THROUGH SIMULATION

Oroselia Sánchez<sup>(a)</sup>, Idalia Flores<sup>(b)</sup>

<sup>(a)</sup> Universidad Nacional Autónoma de México

<sup>(b)</sup> Universidad Nacional Autónoma de México

<sup>(a)</sup>[oroselfia.sanchez@yahoo.com.mx](mailto:oroselfia.sanchez@yahoo.com.mx), <sup>(b)</sup>[idalia.unam.mx](mailto:idalia.unam.mx)

## ABSTRACT

In this work we study the problem of determining the processes which cause delays in cementing services. Through the use of the Simio platform, a visual simulation model of basic cementing process is developed for different scenarios. Such scenarios are based on the number of orders of cementing services, the capacity of the cementing laboratory and the plant, as well as the number of resources available for the jobs. Multiple scenarios are tested, and their results help us to detect which part of the cementing process could cause delays in the completion of the jobs. Finally, based in these results, it is established the resources that the company needs, as well as the time required to complete a job.

Keywords: simulation, accomplished jobs, cementing process.

## 1. INTRODUCTION

A cementing job (CJ) is necessary to stabilize oil wells, preventing collapses and avoiding contamination of the surroundings. This job is required in every stage of the drilling well and is even necessary to close the well when the oil has been extracted from it.

The cementing service process (CSP) is complex because requires time, personal experience and materials with high costs. Besides, when an unexpected CJ is ordered a technical planning is needed. First, it is important to design a pumping schedule which includes a tuned cement slurry, specific personal and equipment. Every oil well is different. There may be similitude in temperature, gradient, geology between wells, but never is exactly the same. Then it should be designed a tuned job for each oil well.

An ordinary cementing service needs as equipment one batch mixer truck, one high pressure unit and one transporter of cement. On the other hand, the crew needed includes one supervisor, three unit operators, three assistants and one engineer. Sometimes the oil well cementing service requires a large volume of cement slurry, and batch spacers which need high quantity of water and additives for its preparation, and as consequence, a greater number of equipment and

personal is required. In general, the number of equipments for each job has to be planned before.

Ordinarily, the costs of operations are high and a possible failure could cause an elevated economical impact. In fact, it is common to have failures because of delays, human errors, equipment failures, etc. All failures produce a cost penalization and a possible unsatisfied customer, whose effects are difficult to measure.

Sometimes the demand of CJ is high, and exceeds the company's capacity, and when the company accepts more jobs than those they can attend, a failure comes due to the lack of enough human or material resources. It shows the importance of a policy response to different demands in cementing services, avoiding high cost due to failure when a wrong decision is taken. The full knowledge of the capacities of the company in different scenarios is then relevant to formulate this response policy, and this work could help to clarify this knowledge. Moreover, this work could serve to detect which part of the process should be improved when forecasts indicates a possible increase in the demand of CJ.

On the other hand, there exists another kind of service offered in CSP called Pumping Jobs (PJ). This PJ, consist in the rent of a high unit pressure to the customer. This type of job does no need design either the use of the plant but, it impacts on the disposition of the units for CJ because PJ requires much time.

In this paper, Simio simulation is used to analyze different stages considering two kind of demands, urgent and planned, for two kind of jobs, CJ and PJ, in order to determine the process that causes more queuing issues, making us able optimize that result and obtain a policy response.

## 2. CEMENTING PROCESS

The oil well cementing services uses a complex process to deliver a service. However, in this paper we are going to take the basic process of an ordinary CJ and PJ.

Cementing jobs consist firstly in a laboratory design of the cement slurry according to the oil well characteristics. The corresponding instructions of the mix of cement and additives are then sent to the cementing plant. A sample of this cement is inspected



in the laboratory before the cement mix be transported to the oil well, where a high pressure unit (UAP) finish the work pumping the cement into the oil well. On the other hand a PJ provides an UAP to the customer, without the previous stages of design and mixing the cement.

In both kind of CSP, we could have planned and urgent orders. The former ones are regular and planned well in advance. The last ones, occurs without previous notification, but can be estimated the average of them by studying the historical data.

### 3. FAILURE ANALYSIS

Commonly, failures give evidences, whose analysis exposes deficiencies in the system. They could have different origins, such as manufacture, materials selection, and design or operation procedures. When that knowledge is obtained, it is possible to prevent or minimize failures. In any kind of services, it is important to reach the client expectations; however, there are also failures whose occurrence is important to maintain under control because they impact directly on the costs. Nowadays, the majority of the approaches of failure analysis are focused on products and not in services.

Failures in services are more difficult to treat due to the impossibility of make physical tests when a failure occurs. The present work could serve as an alternative method to identify origins of failures in services by making use of simulation analysis.

#### 3.1. Proposed Methodology

The methodology that we propose to analyze failures in services is schematically represented in Fig. 1. Some of the steps are explained below.

*Determine failures:* Determine all kind of failures that could occur in the system.

*Determine causes of failures:* The causes of each failure are determined separately.

*Hierarchy of failures:* In this methodology the analysis is systemic, so it is important to identify those failures that are critical in the system. For determine this hierarchy we can use quantity or quality criteria.

*Simulation of complete process:* This step is necessary for complex systems. The methodology for failure analysis proposed in this paper should accomplish the simulation methodology of reference (Musselman 1992; Sadowski 1989).

*Obtain the set of critical causes:* The objective of this methodology is to determine the set of critical causes. Once identified these critical elements, we can minimize failures by means of operational research procedures.

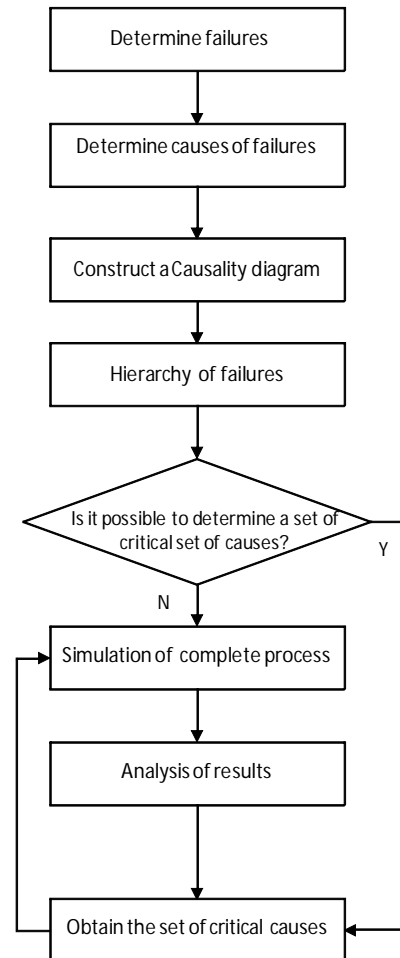


Figure 1: Proposed Methodology for Failure Analysis in Cementing Services Processes.

#### 3.2. Causality Analysis

In this paper, the application of the methodology is applied to the cementing service process (CSP).

The most important failures of CSP are shown in Table 1. The critical criterion is based in the payment of a penalty charge for no productive time. In the case of CSP every hour the customer waits for the job has a certain price. The price is established in the initial contract. Also, CSP has failures which do not generate penalty charge.

Table 1: Classification of failures

| Failure type                | Critical criterion |                 |                     |
|-----------------------------|--------------------|-----------------|---------------------|
|                             | The job is done?   | The job is pay? | Penalty is charged? |
| Unavailability of equipment | Y/N                | Y/N             | Y                   |
| Unavailability of personal  | Y/N                | Y/N             | Y                   |
| Oil Well Tools Failure      | Y                  | Y               | Y                   |
| Cementing Equipment Failure | Y/N                | Y/N             | Y/N                 |
| Personal Failure            | Y                  | Y               | N                   |
| Personal Failure            | Y                  | Y               | N                   |
| Cementing Equipment Failure | Y                  | Y               | N                   |
| Customer Dissatisfaction    | Y                  | Y               | N                   |
| Cement Unconsistant         | Y                  | N               | Y/N                 |
| Others                      | Lost on the road   | Y               | Y                   |
|                             | Later called       | Y               | Y                   |
|                             | Time offset        | Y               | Y                   |
|                             | Forgot additive    | Y               | Y                   |
|                             | Affection owners   | Y               | Y                   |

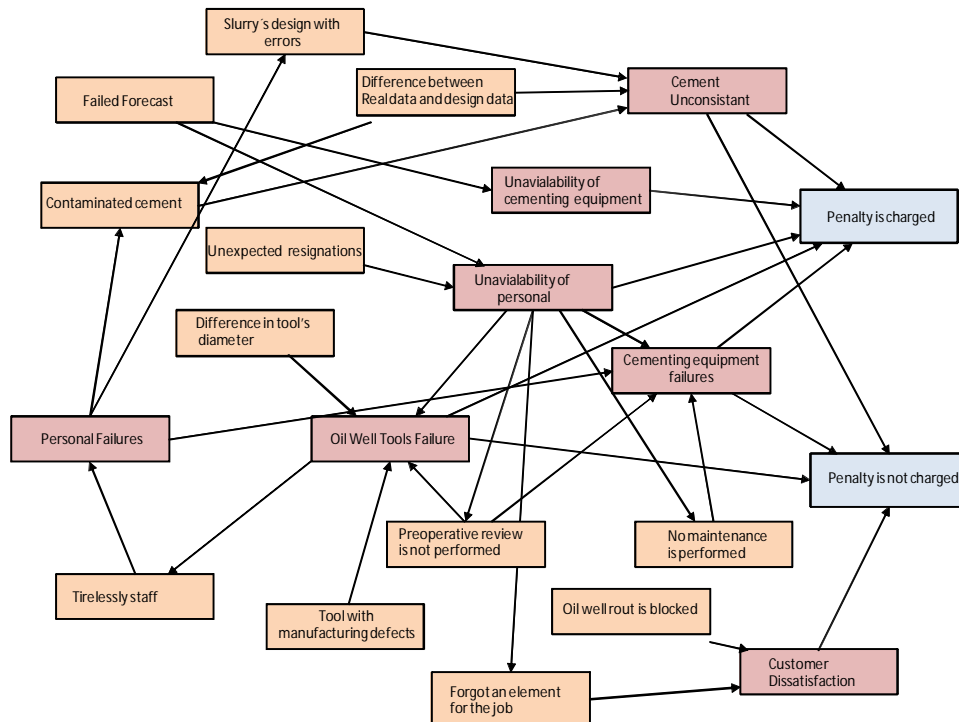


Figure 2: Causality diagram for cementing services failures

Now, we need to determine the causes of every failure type, in separately form. For this analysis we have used the questions suggested by De Castro to get objective information (De Castro, 2004).

A systematic failure analysis is fundamental to prevent or avoid future damages, being important to determine the causes of each failure and the relations between them, which are represented by means of the causality diagram (Guo Li, 2008). In Fig. 2 is illustrated the causality diagram for the CSP, where all the failures are related to the causes. This analysis can provide insights about the problem that we are studying. However, without data of the influence weights in the edges of the network, it is hard to determine the set of critical causes of failures.

#### 4. USING SIMIO

Since the causality diagram is not enough for establish the set of causes, the use of other techniques for get information is needed. Simio is a simulation modeling framework based on intelligent objects (Pedgen, 2011). In this paper, the intelligent objects are built for modeling every step of the CSP described in section 2. Particularly, we use Simio to identify the part of the process which causes queues. It is important to detect possible issues through simulation, because when the problematic part of the process is clearly identified, we can improve the system.

##### 4.1. Simulation of CSP in Simio

The use of simulation can provide a lot of information about the process. We use Simio to simulate the CSP, and analyze how the unavailability of equipment or personal contributes to delays in the completion of the jobs.

The conceptual model of simulation that we are going to use is schematically represented in figure 3 for the processes of cementing jobs (CJ) and pump jobs (PJ). As can be noticed, the CJ requires a previous stage of cement design before the work can be sent to the oil well.

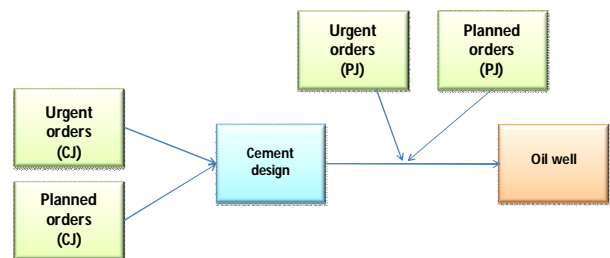


Figure 3: Conceptual simulation model of CSP.

The cement design consists in the series of steps represented in figure 4. The laboratory technician receive the order and prepare the design to be sent to the consistmeters. If some required parameters are not accomplished, the process is repeated by adjusting the design. When the design is ready, it is sent to the plant where employee mix cement and additives in the mix tank in accordance to the laboratory instructions. Then a sample of the mixed cement is sent to the laboratory for quality control. If the quality control is passed, the cement is ready for the oil well, else, the laboratory technician adjust the batch sending instructions to the plant employee and again a sample is send to quality control. When the cement fails the quality control two times, the process is repeated from the beginning by making a new design.

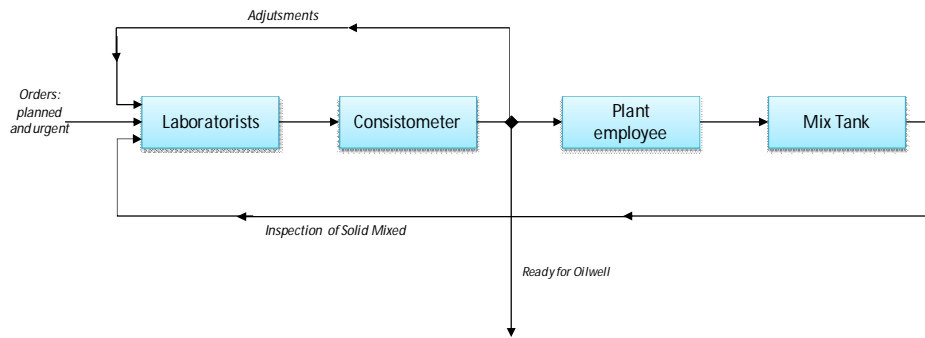


Figure 4: The process of cement design

In Fig. 5 is shown the whole CSP used in the Simio environment. The next parameters were considered in our model:

*Frequency of planned CJ and PJ:* Both considered as constant in time.

*Frequency of urgent CJ and PJ:* It is assumed that they are generated following an exponential probability behavior.

*Capacity of Laboratory technician, Consistometers, Plant Employee, Mix Tank and UAP:* Indicate how many orders can be process at the same time.

*Processing times of Laboratorist, Consistometers, Plant Employee and Mix Tank:* Indicate time necessary to process the order. We take these values as fixed.

*Processing time of UAP:* Time necessary to process an order in the high pressure unit. We consider that this value could be different for CJ and PJ.

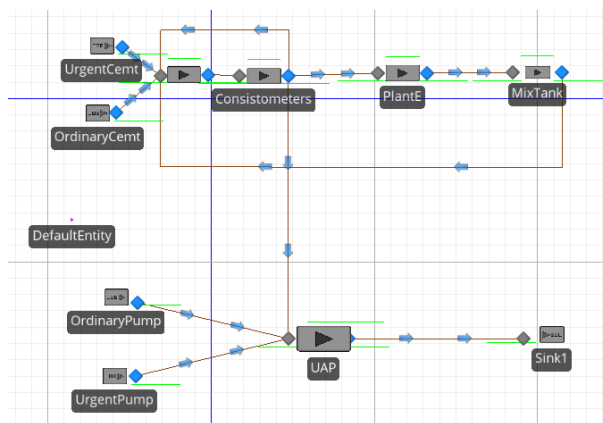


Figure 5: Model of CSP in Simio

*Probability of fail a design:* Assigns a probability for the order to be adjusted in its design.

*Probability of failed inspection:* Assign a probability for the cement mix to be adjusted.

*Probability of fail in adjust batch:* When the sample from the plant fails the quality control, gives the probability that the adjustment of the cement mix be failed.

*Probability of redesign:* Assign a probability of failure the second quality control inspection, and then, start over the process.

Additionally, each step of the system processes first the oldest orders with the purpose of finish them as soon as possible. In the case of the UAP process, CJ have priority over PJ. For this simulation, we will use the parameters given in table 2.

The objective of this simulation is to know:

- Delivery times for CJ and PJ,
- Total Waiting times for each type of orders
- Relationship of equipment number and time of compliance of service.
- Rates of adjustment and redesign

#### 4.2. Verification and Validation of Model

The verification of the simulation model consisted in check its logic behavior according to the set values (Kleijnen, 1995). The most important proofs we did are explained below:

*Verification of model structure:* Visually, the simulation model corresponds to the real system. In this case, the model does not contradict the CSP behavior.

*Parameters verification:* In this point, we review that the proposed parameters values were correctly load into the model.

*Extreme conditions test:* In this point we proof the effectiveness of system decision-making when arrives a lot of orders with different dates, such that the priority values should be executed according to the hierarchy process.

*Behavior proofs:* The first instance simulated, predicted in correct form the results that we expected, based on the probability values set. Besides, we developed a sensitivity analysis varying the values of the parameters in order to get suitable results.

Table 2: Parameters used for the simulation

| Parameter                                    | Value               |
|----------------------------------------------|---------------------|
| Frequency of planned CJ                      | 1 per day           |
| Frequency of planned PJ                      | 1 per 2 days        |
| Frequency of urgent CJ                       | 1 per 2 days        |
| Frequency of urgent PJ                       | 1 per day           |
| Capacity of Laboratorists                    | 1                   |
| Capacity of Consistometers                   | 1                   |
| Capacity of Plant Employee                   | 1                   |
| Capacity of Mix Tank                         | 1                   |
| Capacity of UAP                              | 7                   |
| Processing time of Laboratorist              | 1 hour              |
| Processing time of Consistometer             | 1.5 hours           |
| Processing time of Plant Employee            | 45 minutes          |
| Processing time of Mix Tank                  | 1 hour              |
| Processing time of UAP (CJ)                  | Random(24-48) hours |
| Processing time of UAP (PJ)                  | Random(24-72) hours |
| Probability of fail a design (First Time)    | 50%                 |
| Probability of fail a design (No First Time) | 20%                 |
| Probability of failed inspection             | 10%                 |
| Probability of fail in adjust batch          | 5%                  |
| Probability of redesign                      | 5%                  |

Once the verification is done, we have to validate the simulation model (Kleijnen, 1995). The validation of the model proposed in this paper is based on real-world data. The historical data number of CJ and PJ has a mean of 94 orders requested per month. Our model generates around 90 orders per month. On the other hand, the results of the completion times per order are in validation by an expert.

### 4.3. Results

Through the simulation, we can understand part of the dynamics of the processes. In Figure 6 is shown the cumulative fraction of CJ as function of its completion time for the parameters of Table 2. As can be seen, 50% of CJ become complete in around 50 hours, and above 60 hours an important quantity of the CJ have been completed.

As a first part of the analysis we studied the delivery time for every order when the capacity of UAP is varied. In Table 3 are the total waiting times for deliver 50%, 90% and 100% of the orders arrived in the set period. Notice that the time is reduced when the number of UAP is increased. However this reduction of time is no longer relevant when the number of UAP is greater than seven. The opposite occurs when the number of UAP is five or less, indicating that the queues are out of control, and then, at least there should be six UAP units working.

Table 3: Total Completion Time for CJ and PJ (in hours)

| No. UAP | CJ  |     |      | PJ  |     |      |
|---------|-----|-----|------|-----|-----|------|
|         | 50% | 90% | 100% | 50% | 90% | 100% |
| 5       | 54  | 68  | 96   | 392 | 812 | 1160 |
| 6       | 49  | 62  | 83   | 55  | 90  | 174  |
| 7       | 47  | 57  | 80   | 50  | 70  | 122  |
| 8       | 46  | 56  | 80   | 48  | 68  | 88   |
| 9       | 45  | 56  | 76   | 48  | 67  | 74   |
| 10      | 45  | 55  | 75   | 48  | 66  | 74   |

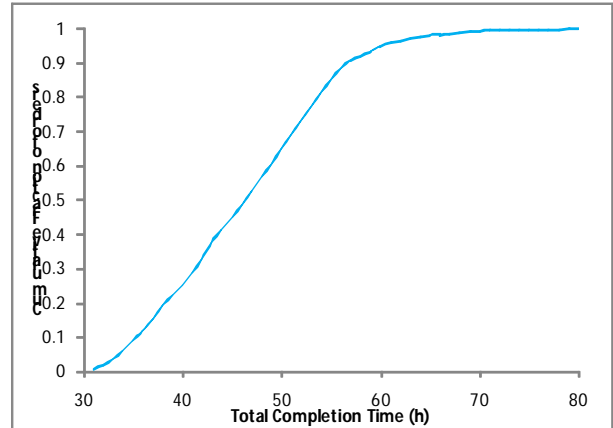


Figure 6: Cumulative fraction of CJ as function of its completion time.

Another critical equipment to be considered as a variable of the failure system is the quantity of Consistometers. For this simulation we set seven UAP and vary the number of consistometer and the results of the completion time are shown in table 4. We can see that the variation between one and two consistometers produce not noticeable difference in the total completion time. Moreover, we do not obtain any benefit by having three consistometers instead of two.

In the simulation results, we got 55.34 % of CJ with some adjustment, and 0.46 % orders were started over and are the cause of most of the delays in completion times.

Table 4: Total Completion Time for CJ (in hours))

| No. Consistometers | CJ  |     |      |
|--------------------|-----|-----|------|
|                    | 50% | 90% | 100% |
| 1                  | 47  | 57  | 80   |
| 2                  | 45  | 56  | 83   |
| 3                  | 45  | 56  | 83   |

## 5. CONCLUSIONS

In this work we have analyzed how the capacity of resources can generate delays in the completion time of cementing and pumping jobs of CSP. For the frequencies jobs showed in table 2, we have found that the number of UAP should be more than seven, and the number of consistometers should not be greater than two. Moreover, the company should promise to complete a CJ in at least 57 hours, and a PJ in at least 70 hours; in this way, the company can assurance that more than 90% of the jobs will be delivered on time, avoiding the penalty charge for no productive time in most cases. Beyond this analysis, it is important to consider the costs and profits involved in CSP, as well as how the promised time of completion impacts on the customer decision of require the company services over

other competitors. These analyses are currently under development.

In summary, we believe that the use of simulation is a very efficient way of obtaining valuable information in order to plan a better response policy of a company.

#### ACKNOWLEDGMENTS

This research was supported by UNAM PAPIIT grant IN116012. The first author specially acknowledge the support from CONACYT scholarship program.

#### REFERENCES

- Alamilla J.L., Sosa E., Sánchez-Magaña C.A., Andrade-Valencia R., Contreras A., 2013, Failure analysis and mechanical performance of an oil pipeline. *Materials and Design*, 50, 766 – 773.
- Brown, J.E., 2009. Identifying cost reduction and performance improvement opportunities through simulation. *Proceedings of the 2009 Winter Simulation Conference*, 2145-2153.
- De Castro P., Fernandes.A.A. 2004. Methodologies for failure analysis: a critical survey. *Materials and Design*, 25, 117-123
- Guo L., Jianmin G., Fumin C., 2008. Construction of Causality Diagram Model for Diagnostics. *IEEE Reliability and Maintainability Symposium*. 28-31 Jan., Las Vegas, NV 97-102.
- Hung, J.Y., Park, N.J., George, K.M., Park, N., 2013. Modeling and Analysis of Repair and Maintenance Processes in Fault Tolerance Systems. *IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, 2-4 Oct. 2013, New York City, NY.261-265.
- Kleijnen J., 1995. Verification and Validation of simulation models. *European Journal of Operational Research*, 82, 145-162.
- Kumbale, M., Hardiman, R., Makarov, Y., 2008. A methodology for simulating power system vulnerability to cascading failures in the steady state. *European Transactions on Electrical Power*, 18, 802-808.
- Musselman, K.J. (1992). Conducting a Successful Simulation Project. *Proceedings of the 1992 Winter Simulation Conference*, 115-121.
- Pasman, H.J., Knegtering, B., Rogers, W.J., 2013. A holistic approach to control process safety risks: Possible ways forward. *Reliability Engineering and System Safety*, 117, 21-29.
- Pegden, C.D., Sturrock, L.L.C., 2011. Introduction to Simio. *Proceedings of the 2011 Winter Simulation Conference*, 11-14 Dec, Phoenix, AZ., 29-38.
- Rossmann HP. Structural failures – liability, learning from failures, technical insurance and legal consequences. Invited lecture 5th special chair AIB-Vin cote 1995, Vrije Universiteit Brussel, 21 February 1995
- Sadowski. R.P. (1989). The Simulation Process: Avoiding the Problems and Pitfalls. *Proceedings of the 1989 Winter Simulation Conference* (Ed.:

MacNair, Musselman, and Heldelberger) Washington, D.C., pp. 72-79

Zhang, K., Wu, R., Ning, J., Zhang, X., Du, H., 2011. Using Simio for Wartime Casualty Treatment Simulation. *IEEE*, 322-325.

#### AUTHORS BIOGRAPHY

**Oroselia Sánchez Sánchez** studied Chemical Engineering at Faculty of Chemistry of Universidad Veracruzana (UV), then a Specialization in Quality Systems at Universidad Nacional Autónoma de México (UNAM). She has worked on Petroleum Industry in Quality, Operation and Management areas. She is currently studying a Master in Operational Research at UNAM. She is also a Professor of the Specialization in Quality Systems at UNAM.

**Idalia Flores de la Mota** She received her Ph.D. in Operations Research at the Faculty of Engineering of the UNAM. She graduated Master with honors and received the Gabino Barreda Medal for the best average of her generation. She has been a referee and a member of various Academic Committees at CONACYT. She has been a referee for journals such as Journal of Applied Research and Technology, the Center of Applied Sciences and Technological Development, UNAM and the Transactions of the Society for Modeling and Simulation International. Her research interests are in simulation and optimization of production and service systems. She is a full time professor at the Postgraduate Program at UNAM.



# HIERARCHICAL FEATURE SELECTION FOR BIOLOGICAL DATA

Witold Jacak<sup>a</sup>, Karin Proell<sup>b</sup>

<sup>a</sup> Department of Software Engineering  
Upper Austrian University of Applied Sciences Hagenberg, Austria  
<sup>b</sup> Department of Medical Informatics and Bioinformatics  
Upper Austrian University of Applied Sciences Hagenberg, Austria

(a) [witold.jacak@fh-hagenberg.at](mailto:witold.jacak@fh-hagenberg.at) (b) [karin.proell@fh-hagenberg.at](mailto:karin.proell@fh-hagenberg.at)

## ABSTRACT

In this paper we present feature selection in biological data by combining unsupervised learning with supervised cross validation. Unsupervised clustering methods are used to perform a clustering of object-data for a chosen subset of input features and given number of clusters. The resulting object clusters are compared with the predefined original object classes and a matching factor (score) is calculated. This score is used as criterion function for heuristic sequential feature selection and a cross selection algorithm.

Keywords: Index Terms— classification, clustering, feature selection, sequential feature selection

## 1. INTRODUCTION

Classification of biological data means to develop a model that will divide biological observations into a set of predetermined classes  $N$ . Typically a biological data set is composed of many variables (features) that represent measures of biological attributes in biological experiments. A common aspect of biological data is its high dimensionality that means data dimension is high, but the sample size is relatively small. This phenomenon is called high dimensionality-small sample problem (Kwak et al. 2002, Maiorana et al. 2008, Zhang et al. 2000). The smaller the sample, the less accurate are the results of classification and the amount of error increases. In traditional statistical classification procedures a probability model must be assumed or estimated in order to calculate the posterior probability upon which the classification decision is performed. One major limitation of the statistical models is that they work well when the underlying assumptions are satisfied. Users must have a good knowledge of both data properties and model capabilities before the models can be successfully applied.

Using all possible features for model creation does not necessarily give the best performance. A model with less features (variables) is faster to construct and easier to interpret, especially in biological data mining where a domain expert should interpret and validate such a result. Using classical supervised clustering and classification methods could lead especially in case of small sample sets, to a faster overfitting of a model during the training phase and to worse prediction performance.

Generally, the feature selection problem deals with choosing those input variables from the measurement space (all input variable) that are most predictive for a given

target and reduce the feature space. The main objective of this process is to retain the optimum number of input variables necessary for the target recognition and to reduce the dimensionality of the measurement space so that an effective and easily computable model can be created for efficient data classification. Appropriate feature (input variable) selection can enhance the effectiveness of an inference model (Kohavi and John 1997).

## 2. FEATURE SELECTION SYSTEM

Feature Selection is a process that identifies a subset of original features. The goal is to reduce the number of features by removing redundant, irrelevant and noisy data in order to improve the predictive accuracy of the model. A typical feature selection process uses four main steps: subset generation, subset evaluation, stopping condition and result validation. Subset generation is a search procedure that produces candidate feature subsets for evaluation based on a certain search strategy. Each candidate subset is evaluated and compared with the previous best one according to a certain objective function. If the new subset turns out to be better it replaces the previous best subset. The process of subset generation and evaluation is repeated until a given stopping condition is satisfied. Then the best subset usually needs to be validated by prior knowledge or different test via real-world datasets (Huan 2005).

*Subset generation:* Search for subsets starts with an empty set and adds features successively or start with a full set and successively remove features or start with both ends and add and remove features simultaneously. For a dataset with  $N$  features there exist  $2^N$  candidate subsets. Such a number of candidate subsets is unfeasible even for moderate numbers of features. Therefore a search strategy is needed in order to direct the feature selection process as it explores the space of all possible combination of features. Therefore different strategies haven applied for candidate subset identification: complete, sequential and random search. In our work we apply traditional sequential forward selection for generating subsets of features. These algorithms use greedy local heuristic search, which incrementally adds and/or deletes features to obtain a subset of relevant features with respect to the response. This search gives up completeness and there is a risk to lose optimal subsets. There are other variations to this greedy hill-climbing approach like sequential backward selection or bi-directional selection. Algorithms with sequential search are easily implemented and produce results faster because the order of the search space is usually  $O(N^2)$ . The problem

with these sequential approaches is that they gravitate toward local minima due to the inability to re-evaluate the usefulness of features that were previously added or discarded (Huan 2005).

*Subset evaluation:* As mentioned above each newly generated feature subset needs to be evaluated by a criterion. This measurement is returned by an objective function which is applied to the feature subset monitoring the “goodness” of a candidate set- a feedback signal used by the search strategy to select new candidates. Objective functions are divided in two groups. Filters evaluates feature subsets by their information content, typically interclass distance, statistical dependence or information-theoretic measures and wrappers use a pattern classifier, which evaluates feature subsets by their predictive accuracy (recognition rate on test data) by statistical resampling or cross-validation (Hua 2005). In our work we make use of the wrapper approach and present a method for objective function specification, called matching factor, combining unsupervised classification of data with prior knowledge about classes of the data.

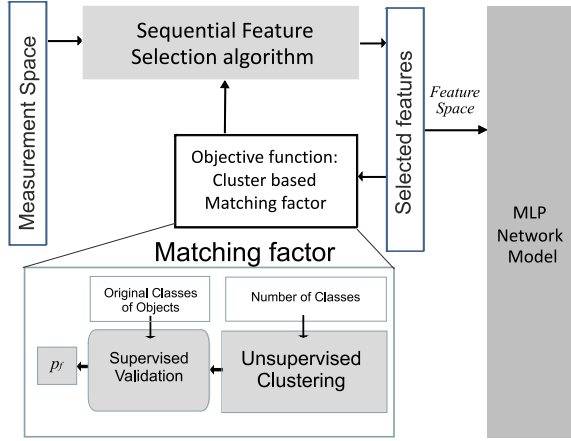


Fig. 1: Feature selection system based on unsupervised learning

*Result validation:* In real-world applications we usually do not have prior knowledge about relevant features to directly measure the results. We have to rely on indirect methods by monitoring the change of predictive accuracy with the change of features. In case of a classification task we can use a classification error rate as a performance indicator for a selected feature subset, simply by conducting the “before-and-after” experiment to compare the error rate of the classifier learned on the full set of features and that learned on the selected subset (Huan 2005). The system we propose is presented in Fig.1.

## 2.1. Objective function

Applying a wrapper approach for objective function we can give a generalized overview (Huan 2005) of the used algorithm for sequential feature selection (see Fig.2).

|                                                                                                            |                               |                                            |
|------------------------------------------------------------------------------------------------------------|-------------------------------|--------------------------------------------|
| input:                                                                                                     | $L(f_0, f_1, \dots, f_{n-1})$ | // training data set with N features       |
|                                                                                                            | $S_0$                         | // a subset from which to start the search |
|                                                                                                            | $\delta$                      | // a stopping criterion                    |
| Output:                                                                                                    | $S_{best}$                    | // an optimal subset                       |
| 01: begin                                                                                                  |                               |                                            |
| 02: initialize: $S_{best} = S_0$ ;                                                                         |                               |                                            |
| 03: $pf_{best} = eval(S_0, D, mf\_calculation)$ ; //evaluate $S_0$ by matching //factor objective function |                               |                                            |
| 04: do begin:                                                                                              |                               |                                            |
| 05: $S = generate(D)$ ; // generate next subset for evaluation                                             |                               |                                            |
| 06: $pf = eval(S, D, mf\_calculation)$ ; //evaluate $S$ by matching //factor objective function            |                               |                                            |
| 07: if ( $pf$ is better than $pf_{best}$ )                                                                 |                               |                                            |
| 08: $pf_{best} = pf$ ;                                                                                     |                               |                                            |
| 09: $S_{best} = S$ ;                                                                                       |                               |                                            |
| 10: end until ( $\delta$ is reached);                                                                      |                               |                                            |
| 11: return $S_{best}$ ;                                                                                    |                               |                                            |
| 12: end                                                                                                    |                               |                                            |

Table 1: Wrapper Algorithm with matching factor calculation as objective function

For each generated subset  $S$ , the sequential forward algorithm evaluates its goodness using the matching factor calculation as an objective function for feature subset  $S$  (see Table 1). Our objective function - called matching factor calculation - applies different clustering methods on each chosen subset of features.

Although we have knowledge of classes of the input data a priori we do not make use of this knowledge initially but have clustering methods (Ye and Liu H 2002) define a partitioning of the input data using simply the well-known number of classes as input parameter to the clustering methods (Thangavel, Shen and Pethalakshmi 2006, Törmä 1996). An average value on the results of the applied clustering methods is calculated which is finally compared to the original assignment of classes of the input data and the matching factor is calculated.

The algorithm for matching factor calculation works as follows and can be seen in full detail in Jacak, Proell and Winkler 2014.

Let  $I = \{1, \dots, N\}$  be a set of indexes of target classes  $L$  and  $J = \{1, \dots, N\}$  be a set of indexes of cluster set  $K$  generated from pure input data of objects with unsupervised learning using for example SOM-SL neural network. Let  $L_i \in L$  be a subset of objects belonging to original class  $i \in I$  and  $K_j \in K$  is a subset of objects belonging to cluster  $j \in J$  created by the clustering algorithms. For each  $i \in I$  and  $j \in J$  we define the Jaccard matching coefficient  $c_{ij}$  as follows

$$c_{ij} = \frac{|L_i \cap K_j|}{|L_i \cup K_j|} \quad (1)$$

We use this coefficient to define the global matching factor  $p_f$  between the generated cluster and the original target classes in an iterative manner. For  $n=1, \dots, N$  we calculate

$$p_{ij}^n = \begin{cases} c_{ij} & \text{if } c_{ij} = \max\{c_{ik} | l \in I_{n-1}, k \in J_{n-1}\} \text{ and} \\ & s_i = \sum(c_{ik} | k \in J_{n-1}) = \min \quad \text{and} \\ & s_j = \sum(c_{ij} | l \in I_{n-1}) = \min \quad (2) \\ 0 & \text{other} \end{cases}$$



where  $I_0=I$ ,  $J_0=J$ , and  $J_n=J_{n-1} - \{j\}$ ,  $I_n=I_{n-1} - \{i\}$ . Calculation stops when  $J_n$  and  $I_n$  are empty sets.

It is easy to observe that the  $N \times N$  sized matrix  $P = [p_{ij}]$  has only  $N$  elements greater zero and represent the best allotment of the generated clusters to the original classes. The global matching factor  $p_f$  can be defined as mean value of non-zero elements of  $P$ :

$$p_f = \text{avg}(p_{ij}) \quad \text{where } p_{ij} > 0 \quad (3)$$

The matching factor describes the score of recognizing the original target classes by unsupervised clustering which is uses subsets of input data without prior knowledge of classes. This factor uses posterior knowledge to calculate the score for validating the goodness on a chosen subset of features which should maximize the matching factor.

## 2.2. Hierarchical feature selection – Cross feature selection

In the above description we applied standard sequential feature selection using the matching factor calculation as objective function. In this section we present an alternate approach for feature selection which can be applied for at least four target classes. We call this approach hierarchical or cross feature selection.

When original target classes  $L = \{L_1, L_2, \dots, L_N\}$  are known then it is possible to construct different subsets of original classes. The subset of original classes

$$\{L_i, L_k, \dots, L_m\} \subset L$$

is grouped together into one new class, called hyper-class

$$LC_i = \cup \{L_i, L_k, \dots, L_m\}$$

This hyper-class contains all objects data from its components. The new non overlapping hyper-classes set  $LC_i$  ( $i=1 \dots K$ ) should cover a set of original classes, i.e.

$$\cup LC_i = \cup L \quad \text{and} \quad \cap LC_i = \emptyset$$

### Fact:

Let  $LC^k$  denote the  $k$ -th experiment ( $k = 1, \dots, M$ ) with chosen hyper-classes, where  $LC^k = \{LC_i^k \mid i=1, \dots, K^k\}$  and  $LC_i^k$  is  $i$ -th hyper class in the  $k$ -th experiment. Let  $F^k \subset \{1, \dots, m\}$  be the set of selected sensitive features for the best recognition of the hyper classes in this experiment, calculated by sequential feature selection algorithm.

When the chosen new hyper-classes sets  $LC^k$  ( $k = 1, \dots, M$ ) meet the condition

$$\cap LC^k = L$$

then the cross selected sensitive features for recognition of the full original target  $L$  could be defined as

$$F = \cup F^k.$$

Such selected features based on cross comparison of subsets of target classes can be compared to feature sets found by the forward/backward sequential selection procedure for original target classes. The union of both candidates' sets of features can be used by exhaustive search to find the optimal final feature space.

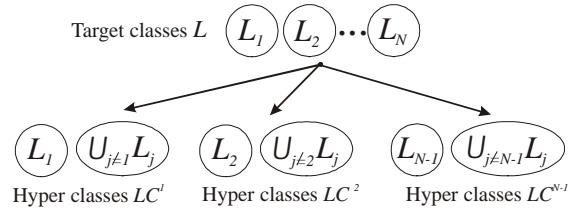


Fig. 2: Construction of hyper-classes for  $N-1$  experiments  
The easiest way to find hyper-classes is the breakdown of the whole collection of classes into two groups. The first group contains only one class and second one a union of all the others (see Fig. 2).

$$LC^i = \{L_i, \cup \{L_j \mid j \neq i\}\}$$

So we have only two hyper classes. Such a construction of hyper-classes conducted for each class separately (for  $i=1 \dots N-1$ ), satisfies the conditions of fact and can be used in cross method of selection.

$$\cap \{LC^i \mid i=1 \dots N-1\} = L$$

If we make such groupings for particular classes, combining parameters for each group represent a new selection for recognition all  $L$  classes.

$$F_{\text{cross selection}} = \cup F^i$$

where  $F^i$  is the set of selected features for recognition of  $LC^i$  hyper-classes.

The hyper-class construction algorithm used in the cross selection method required  $N-1$  times running of process for the sequential feature selection.

## 2.3. Result evaluation

After the feature selection phase a multi-layer perceptron (MLP) network is used to classify the training and validation dataset in the reduced feature space. The classification performance of each feature set is compared with the classes indicated in the original dataset. We used a two-layer perceptron network (MLP) with four neurons in the hidden layer.

## 3. EXPERIMENTS AND RESULTS

In this paper a cross feature selection method independent of the underlying classification model and based on clustering algorithms with posterior validation (matching factor) has been proposed. For our experiments we used a salmon growth experiment using data from fish sampled following 90 days of feeding (Epstein 2012, Gu et al. 2013). A collection of biological data for 288 fish is used. Each fish is described with 89 features max. Features are just labeled to have analyst blind to the actual variables measured. The measurement space – the full feature set - is therefore  $F = \{f_1, \dots, f_{89}\}$ . Missing values in the input samples were replaced by the mean value of the parameters into consideration from the respective class and all data were normalized.

The fish belong to four predefined classes (target classes)  $L = \{A, B, C, D\}$ . Each class corresponds to a certain kind of food. We did the following class labeling depending on the

given values of GMmatrix and GMvalue in the experiment data (see Table 2).

| Class | GMmatrix | GMvalue |
|-------|----------|---------|
| A     | 1        | a       |
| B     | 1        | b       |
| C     | 2        | a       |
| D     | 2        | b       |

Table 2: Class labeling depending on the given values of GMmatrix and GMvalue

Applying sequential forward selection and cross feature selection the following results were provided:

The empirical normal distribution PDF 1 (continuous line) of ratio of matching factor calculated for cross selected features to matching factor for full set of features is presented in Fig 3. This figure presents also the empirical normal distribution PDF 2 (dotted line) of ratio of matching factor calculated for forward sequential selected features to matching factor for full set of features. These ratios can be interpreted as quality of target classes' recognition.

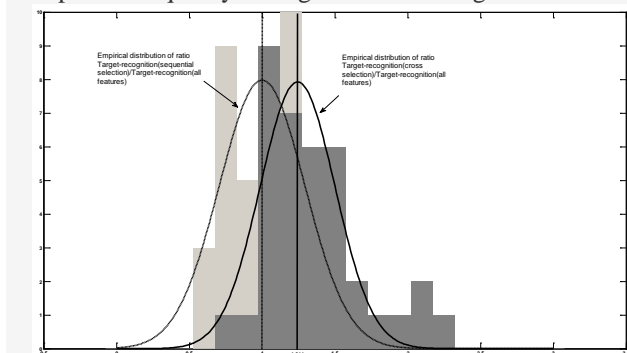


Fig 3: Empirical distribution of target recognition quality of cross selection method (pdf 2-continuous line) and forward sequential selection method (pdf 1-dotted line).

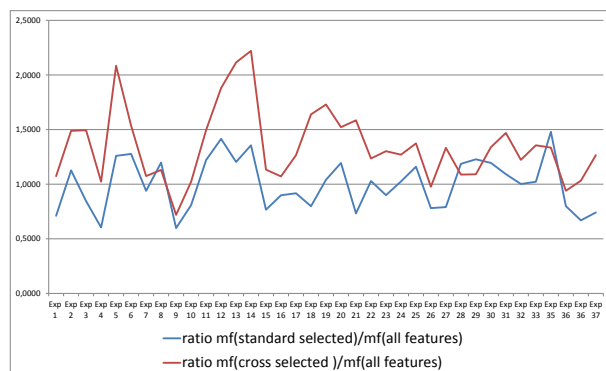


Fig 4: Comparison of matching factor ratio between standard features selection and cross features selection methods

The expectation value of empirical distribution PDF 1 is 1,244 for cross feature selection with matching factor objective function and the expectation value of empirical

distribution PDF 2 is 1,01 for standard sequential feature selection. Cross selection method finds average 2,29 times more features than the standard sequential selection procedure. A direct comparison of recognition quality between cross selection found features and standard forward sequential selection found features shows that the cross selection method has about 20 % better recognition quality (see Fig.4).

#### 4. CONCLUSION

In this paper a feature selection method independent of the underlying classification model and based on clustering algorithms with posterior validation has been proposed. The analysis was performed on biological data. The presented method can be classified as a heuristic method, which obtains an effective feature reduction with almost the same correct classification rate. Heuristics is designed on an unsupervised learning approach. The optimal number of feature space dimensions is therefore difficult to determine. Future works are to extend the analysis to datasets with different number of samples and to further investigate the distance measures to assess the impact on classification performance.

#### REFERENCES

- Epstein M. (ed.), 2012, Proc. Of GMSAFOOD Conference, Vienna
- Gu J., Krogdahl A., Sissener N.H., Kortner T.M., Gelencser E., Hemre G.I., Bakke A.M., 2013, "Effects of oral Bt-maize (MON810) exposure on growth and health parameters in normal and sensitised Atlantic salmon, *Salmo salar L*", British Journal of Nutrition 2013 Apr 28
- Jacak W., Proell, Winkler, K., S., 2014, Neural Networks based Feature Selection in Biological Data Analysis. in: Klempos, R., J. Nikodem, W. Jacak, Z. Chacko, eds. *Advanced Methods and Applications in Computational Intelligence*, Springer, pp.79-93
- Kohavi R., John G. H., 1997, "Wrappers for feature subset selection," Artificial Intelligence, vol. 97, no. 1-2
- Kwak N., 2002, Choi C., "Input Feature Selection for Classification Problems", IEEE Transactions on Neural Networks, Vol. 13, No. 1
- Huan L., 2005, "Toward integrating feature selection algorithms for classification and clustering", Knowledge and Data Engineering, IEEE Transactions on, Vol. 17, Issue 4
- Maiorana F., 2008, "Feature Selection with Kohonen Self Organizing Classification Algorithm", World Academy of Science, Engineering and Technology, Proceedings of WASET, Vol. 45
- Thangavel K., Shen Qiang, Pethalakshmi A., 2006, "Application of Clustering for Feature Selection Based on Rough Set Theory Approach", AIML Journal, Volume (6), Issue (1)
- Törmä M., 1996, *Self-organizing neural networks in feature extraction*", Intern. Arch. of Photogrammetry and Remote Sensing, Vol. XXXI, Part 2

- Ye H., Liu H., 2002, “*A SOM-based method for feature selection*”, Proceedings of the 9th International Conference on Neural Information Processing (ICONIP’02), Vol. 3
- Zhang G. P., 2000, “*Neural Networks for Classification: A Survey*”, IEEE Transactions on Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 30, No. 4

# PROBABILISTIC BEHAVIORAL MODEL FOR THE DETECTION OF CRITICALITIES WHEN USING MORPHINE AND FENTANYL PCA

Henrikas Pranevicius<sup>(a)</sup>, Mindaugas Snipas<sup>(a)</sup>, Tadas Kraujalis<sup>(a)</sup>, Mindaugas Pranevicius<sup>(b)</sup>,  
Osvaldas Pranevicius<sup>(c)</sup>, Vytautas Pilkauskas<sup>(a)</sup>,

<sup>(a)</sup> Department of Applied Informatics Kaunas University of Technology Lithuania

<sup>(b)</sup> Albert Einstein College of Medicine USA

<sup>(c)</sup> Department of Anesthesiology New York Hospital Queens Flushing USA

<sup>(a)</sup>[henrikas.pranevicius@ktu.com](mailto:henrikas.pranevicius@ktu.com), <sup>(b)</sup>[mipran@hotmail.com](mailto:mipran@hotmail.com), <sup>(c)</sup>[opranevicius@aol.com](mailto:opranevicius@aol.com)

## ABSTRACT

When patient controlled analgesia (PCA) was originally introduced, the belief was that frequency of analgesic demand uniquely reflects the level of patient's pain. However frequency of the demand is a random process that has its own distribution with a unique shape and parameters. We used this distribution to simulate the risk of drug concentration exceeding critical threshold. We used quantized state system model to create hybrid aggregate model of PCA. We investigated two randomly selected, real data based, unidentified morphine and fentanyl PCA logs. Based on this data we generated model of the random process that approximated real demand data and created 500 virtual PCA logs. These logs allowed pharmacokinetic simulation of the effect compartment concentration. The proposed methodology allows an estimation of frequency and duration of critical episodes, when target concentration exceeds critical threshold. These estimations might be used to evaluate patient specific risk of postoperative opiate overdose.

Keywords: patient controlled analgesia, hybrid aggregate model, time series, probabilistic distribution.

## 1. INTRODUCTION

Research suggests (Grass, 2005), that different persons show different demand pattern when using PCA (Dahan, Aarts and Smith, 2010). This behavioral pattern varies depending on the variety of simultaneously occurring factors (Boom et al. 2013), including the level of pain, drug concentration in plasma and the effect site, various side effects (e.g., nausea, sedation, respiratory depression), or even the psychological state of the patient, such as anxiety or cognitive impairment. All these simultaneous factors (Woodhouse and Mather, 2000) introduce randomness in the behavioral dose demand pattern and therefore demand sequence could be thought of as a random process what allows utilizing autoregression moving average model ARMA(p,q), commonly used in time series forecasting (Makridakis and Hibon, 1997).

However, for the analysis of random process with arbitrary degree of precision, there must also be available time series of random process of arbitrary length. In our publication here we describe one of the

ways of how to generate and analyze time series of PCA demand of arbitrary length.

## 2. HYBRID AGGREGATE MODEL OF PATIENT CONTROLLED ANALGESIA

### 2.1. Hybrid aggregate model

For simulation of patient controlled analgesia we used hybrid systems simulation method based on PLA formalism (Pranevicius et al. 2011).

PLA is a special case of automaton models. In the application of the PLA approach for system specification, the system is represented as a set of interacting piece-linear aggregates. The PLA is taken as an object defined by a set of states  $Z$ , input signals  $X$ , and output signals  $Y$ . Behaviour of an aggregate is considered in a set of time moments  $t \in T$ . States  $z \in Z$ , input signals  $x \in X$ , and output signals  $y \in Y$  are considered to be time functions. Transition and output operators,  $H$  and  $G$  correspondingly, must be known as well.

The state  $z \in Z$  of the piece-linear aggregate is  $z(t) = (v(t), z_v(t))$ , where  $v(t)$  is a discrete state component taking values on a countable set of values; and  $z_v(t)$  is a continuous component comprising of  $z_1(t), z_2(t), \dots, z_{v_k}(t)$  coordinates.

When there are no inputs, an aggregate state changes as follows:  $v(t) = const$ ,  $\frac{dz_v(t)}{dt} = -a_v$ , where  $a_v = (a_{v_1}, a_{v_2}, \dots, a_{v_k})$  is a constant vector.

For hybrid aggregate model (Pranevicius et al. 2011) continuous coordinate's model is described by the ordinary differential equations (ODE):

To solve ODEs system we'll adopt Quantized State System (QSS) method, which was defined by Ernesto Kofman (Kofman 2004).

Considering ODE system:

$$\frac{dx(t)}{dt} = f[x(t), u(t)],$$

where  $x(t) \in R^n$  is the static vector,  $u(t) \in R^n$  is an input vector, which is a known piecewise constant function.

The QSS method simulates an approximate system, which is called quantized state system:

$$\frac{dx(t)}{dt} = f[q(t), u(t)],$$

where  $q(t)$  is a vector of quantized variables which are quantized versions of the state variables  $x(t)$ . Each component of  $q(t)$  is related to the corresponding component of  $x(t)$  by a hysteretic quantization function. A generic Quantized State System can be represented by the block diagram of Figure 1.

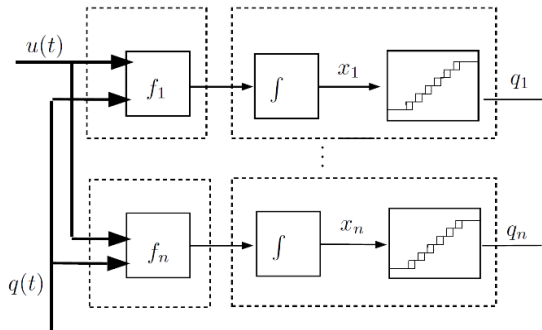


Figure 1: Block diagram representation of a QSS (Kofman 2004)

QSS method was implemented using PLASim simulation library created in our department (Pranevicius, Pilkauskas and Guginis, 2006). The PLASim is an object-oriented library for discrete-event simulation of models created using aggregate formalism. The PLASim's current version written in C# for NET Framework 4.0 and has packages that support random number generation, statistical collection, basic reporting with data visualization and discrete-event simulation. The development of a simulation model is based on sub-classing the SimulationModel class that provides the primary recurring actions within a simulation and event scheduling and handling.

We upgraded module of this library PLASimInternalEvents which implements internal event classes:

InternalWEvent – class of the internal event;  
 ContinuousCoordinate – abstract of the continuous coordinate;

WISum – class of the controlling sum;

ControlSequence – class of the controlling sequence;

InternalEventHandler – handler (delegate) of the internal event.

Controlling sums (ControlSum) initiate internal event of the aggregate -- CreateInternalEvent(w). Timing of the internal event is determined by the parameter w. Determination of the parameter w can be done using object from the ControlSequence class. Generated internal event is placed on the list of internal events internalEventQueue. Internal event for the processing is selected using SimulationModel method. The processing method itself is called up using NextInternalEvent() and InternalEventHandler.

Additional classes were added to construct QSS events:

InternalHEvent – class of QSS events;

InternalHEventList list for QSS events;

Hsum – class of QSS events control sum.

### 2.1.1. Modeling PCA by using hybrid aggregate model

A three compartmental model of drug distribution between the serum and the brain tissue (effect compartment) was used to describe fentanyl and morphine pharmacokinetics/pharmacodynamics, as is shown in Figure 4.

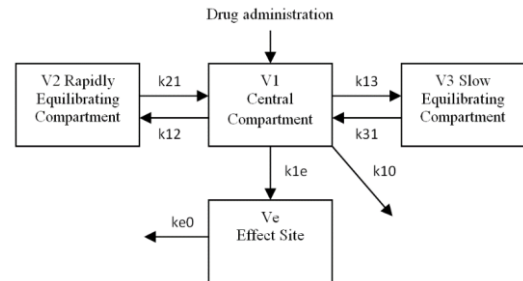


Figure 2: Three compartment pharmacokinetic model

The central compartment (V1) represents a distribution volume and includes rapidly mixing portion of the blood. The peripheral compartments (V2, V3) are composed of tissues and organs, where drug distributes at a slower rate. The effect site is the hypothetical compartment that relates the time course of plasma drug concentration to the time course of drug effect.

Pharmacokinetic model is described by four differential equations:

$$\begin{cases} \frac{dx_1}{dt} = k_{21} \cdot x_2 - k_{12} \cdot x_1 + k_{31} \cdot x_3 - k_{13} \cdot x_1 - k_{10} \cdot x_1 + u(t) \\ \frac{dx_2}{dt} = k_{12} \cdot x_1 - k_{21} \cdot x_2; \\ \frac{dx_3}{dt} = k_{13} \cdot x_1 - k_{31} \cdot x_3; \\ \frac{dx_e}{dt} = k_{e0} \cdot x_1 - k_{e0} \cdot x_e, \end{cases}$$

where  $x_1, x_2, x_3$  and  $x_e$  are the amounts of drug in the central, second, third and effect site compartments, respectively, and  $k_{10}, k_{12}, k_{13}, k_{21}, k_{31}$ , and  $k_{e0}$  are the constants defining the elimination and inter-compartmental transfer rates, and  $u(t)$  the function describing drug delivery.

Aggregate scheme of patient controlled analgesia aggregate model is presented in Figure 5.

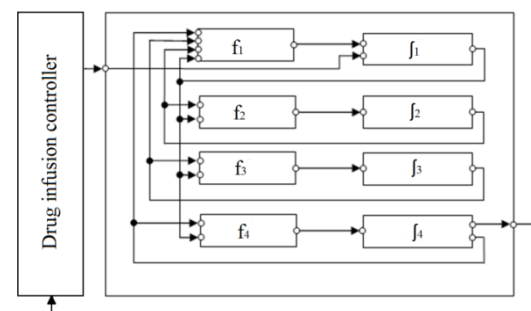


Figure 3: Aggregate scheme of the pharmacokinetic model

### 3. PATIENT'S BEHAVIORAL MODEL

We used two random real data logs from PCA device, one with a prescription for morphine and one for fentanyl. We analyzed the length of time period (in minutes) between the two consecutive drug requirements. For building of a suitable model we used SPSS and MATLAB statistical tools.

#### 3.1. Analysis of morphine PCA demands

Analysis of correlograms shows (see Fig. 4 and Fig. 5), that neither the values of autocorrelation, nor the values of partial autocorrelation function are statistically significant.

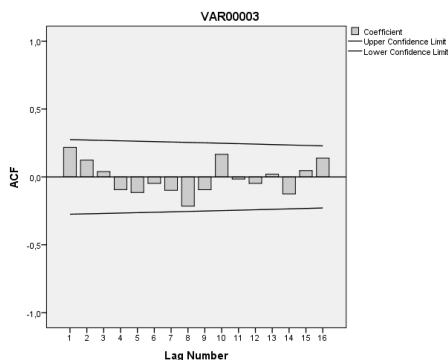


Figure 4: Values of autocorrelation function for morphine PCA log

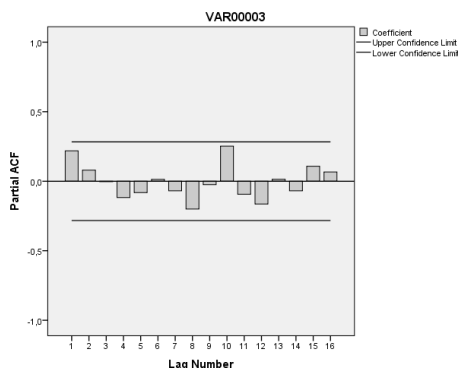


Figure 5: Values of partial autocorrelation function for morphine PCA log

We compared 25 different ARMA(p,q) models, with values of p and q ranging from 0 to 4, and chose the model with lowest value of Bayesian Information Criterion (BIC). BIC, unlike e.g. R squared value, “punishes” models with higher number of parameters, thus it favors more parsimonious models. The values of BIC with different ARMA models are presented in Table 1:

Table 1: BIC values of different ARMA(p,q) models (morphine PCA)

|       | MA(q) |   |   |   |   |
|-------|-------|---|---|---|---|
| AR(p) | 0     | 1 | 2 | 3 | 4 |

|   |       |       |       |       |       |
|---|-------|-------|-------|-------|-------|
| 0 | 7,180 | 7,249 | 7,341 | 7,437 | 7,518 |
| 1 | 7,245 | 7,343 | 7,440 | 7,542 | 7,560 |
| 2 | 7,342 | 7,442 | 7,457 | 7,531 | 7,659 |
| 3 | 7,441 | 7,467 | 7,531 | 7,659 | 7,725 |
| 4 | 7,523 | 7,555 | 7,658 | 7,687 | 7,783 |

Data from Table 1 shows, that ARMA(0,0) has lowest BIC value, which is consistent with the values of autocorrelation and partial autocorrelation functions. This suggests that the following model is optimal:

$$y_t = \mu + kt + \varepsilon_t,$$

where  $\varepsilon_t \sim WN(0, \sigma^2)$ , i.e., identically distributed independent white noise, with mean 0 and variation  $\sigma^2$ . Parameter estimates are presented in Table 2.

Table 2: Parameter estimates for ARMA(0,0) model (morphine PCA)

| Parameters      | Estimate | Significance |
|-----------------|----------|--------------|
| Constant: $\mu$ | 21,029   | 0,034        |
| Numerator: k    | 0,407    | 0,222        |

Parameter estimates of this model suggest, that lag numerator k is not statistically significant, thus optimal model for morphine data is

$$y_t = \mu + kt + \varepsilon_t,$$

where  $\varepsilon_t \sim WN(0, \sigma^2)$ , i.e., identically distributed independent white noise, with mean 0 and variation  $\sigma^2$ .

Parameter estimates (see Table 2) of this model suggest, that lag numerator k does not differ from 0 significantly, thus optimal model for time periods  $y_t$  between two drug requirements during morphine analgesia can be modeled simply as:  $y_t = \varepsilon_t$ , where  $\varepsilon_t \sim WN(0, \sigma^2)$ .

Thus, it seems reasonable to model time periods between two drug requirements as identically distributed independent random numbers. In order to choose the best probability distribution function, we used MATLAB distribution fitting tool.

Research showed, that exponential distribution with parameter  $\lambda = 1/31.4$  (i.e., mean value 31.4) provided the best fit, according to the value of log likelihood, which was -222.34 in this case. Histogram and fitted probability distribution function is presented in Figure 3.



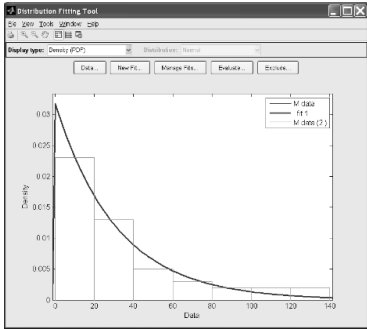


Figure 6: Distribution fitting for data obtained from morphine PCA

Since p value 0.6912 is above 0.05, chi square criterion does not reject the null hypothesis, making exponential distribution a reasonable fit for the data.

```
>> [h,p,s] = chi2gof(M, 'CDF', pd)
h = 0
p = 0.6912
s = chi2stat: 0.7387
edges: [4.9407e-324 12.40 24.80 37.20
124.0000]
O: [17 9 9 15]
E: [16.3128 10.9906 7.4049 15.2917]
```

Exponentially distributed random variables can be generated by using standard functions (e.g., function `expnd()` in MATLAB), or by the inverse function method. If  $r$  is basic random number (i.e., uniformly distributed in interval  $[0;1]$ ), exponentially distributed number  $\varepsilon$  can be generated as  $\varepsilon = -\frac{\ln r}{\lambda}$ .

### 3.2. Analysis of fentanyl PCA demands

Analyzing correlograms of data, obtained from fentanyl PCA, suggests that it can not be assumed to be the white noise. Autocorrelation function damps cyclically, while partial autocorrelation cuts at level 2, which suggests that some kind of ARMA(2,q) process might be suitable.

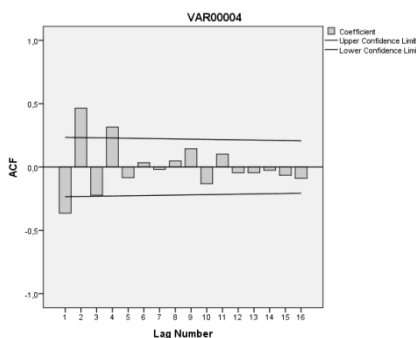


Figure 7: Values of autocorrelation function for fentanyl PCA log

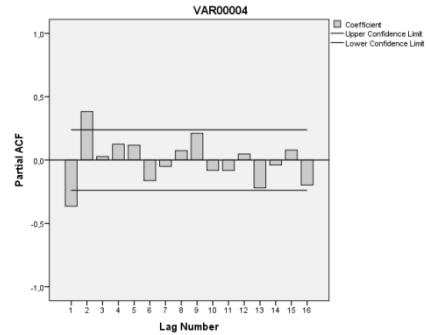


Figure 8: Values of partial autocorrelation function for fentanyl PCA log

Additional research supports this hypothesis, since BIC value is lowest with ARMA(2,0), i.e. AR(2), process suits best (see Table 2.), though it's squared value 0.306 explains smaller part of total variation among data.

Table 3: BIC values of different ARMA(p,q) models (fentanyl PCA)

|       | MA(q) |       |       |       |       |
|-------|-------|-------|-------|-------|-------|
| AR(p) | 0     | 1     | 2     | 3     | 4     |
| 0     | 6,511 | 6,459 | 6,427 | 6,495 | 6,484 |
| 1     | 6,371 | 6,360 | 6,422 | 6,496 | 6,514 |
| 2     | 6,354 | 6,423 | 6,498 | 6,534 | 6,554 |
| 3     | 6,426 | 6,452 | 6,530 | 6,626 | 6,641 |
| 4     | 6,490 | 6,528 | 6,584 | 6,642 | 6,720 |

Residuals of AR(2) model shows no significant autocorrelation, which suggests that residuals are not significantly different from the white noise.

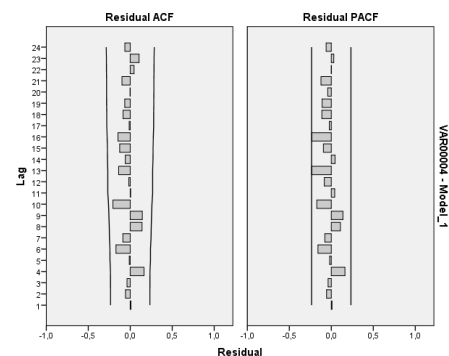


Fig. 9. Values of residuals from fitted ARMA(2,0) model (fentanyl PCA)

Actual five values of patients drug requirements also fitted well into predicted 95 percent confidence intervals, thus ARMA(2,0) might be a reasonable model for the fentanyl analgesia data.

Estimated ARMA(2,0) with linear trend model parameters are presented in Table 4.



Table 4: Parameter estimates for ARMA(2,2) model fitted for data from fentanyl analgesia

| Parameter |       | Estimate | Significance |
|-----------|-------|----------|--------------|
| Constant  |       | 11,570   | 0,26         |
| AR        | Lag 1 | -0,302   | 0,12         |
|           | Lag 2 | 0,203    | 0,11         |
| Numerator | Lag 0 | 0,264    | 0,38         |

All values are statistically significant under standard 0.05 level, so we chose the following model to simulate time period  $y_t$  between two consecutive patient requirements for fentanyl dose by the following process:

$y_t = 11,57 + 0,264t - 0,302y_{t-1} + 0,203y_{t-2} + \varepsilon_t$ ,  
 where  $\varepsilon_t \sim WN(0, \sigma^2)$  i.e., it is identically (though not necessary normally) distributed independent random variables, with mean 0 and variation  $\sigma^2$ .

Our analysis suggests that generalized extreme values distribution, with parameters  $k = -0.021633$ ,  $s = 16.521$  and  $m = -9.426$  provides the best fit for AR(2) model residual, according to the likelihood criterion (the value is equal to -306.101). Histogram and fitted probability distribution function is presented in Fig. 7.

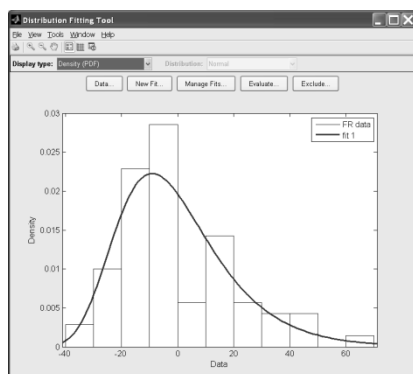


Figure 7: Distribution fitting for residual of ARMA(2;0) model (fentanyl PCA)

Chi square test does not reject our null hypothesis, that residuals are distributed according to a generalized extreme value distribution, since the p value is above the standard 0.05 value.

```
>>pd = fitdist(FR,'GEV')
pd =
generalized extreme value
distribution
 k = -0.021633
 sigma = 16.5207
 mu = -9.42619
>> [h,p,s] = chi2gof(FR,'CDF',pd)
h = 0
p = 0.0586
s =
chi2stat: 5.6748
```

```
edges: [-38.89 -18.42 -8.18 2.05 12.29
22.5260 63.47]
O: [11 15 21 6 6 11]
E: [12.5611 15.1296 14.9153 11.1766
7.1537 9.0636]
```

Moreover, we compared residuals from fitted ARMA(2,0) model with the data generated randomly, that have generalized extreme values distribution with already estimated parameters. Kolmogorov-Smirnov test for both data sets also did not reject the hypothesis that ARMA model residuals have the same distribution as data generated randomly.

```
>> [h,p,s]=kstest2(FR,FRand)
h = 0
p = 0.7246
s = 0.1143
```

The values of generalized extreme values distribution can be generated by the standard functions (e.g., function `gevrand()` in MATLAB), or by the use of an inverse function method.

Generalized extreme values distribution with parameters  $k$ ,  $s$  and  $m$  has the following cumulative distribution function:

$$F(x; k, s, m) = \exp \left\{ - \left[ 1 + k \left( \frac{x - \mu}{\sigma} \right) \right]^{\frac{1}{k}} \right\}.$$

Thus, the value of random variable  $\varepsilon$ , having generalized extreme value distribution, can be generated by the following transformation of standard random number  $r$ :

$$\varepsilon = s \frac{(-\ln r)^{-k-1}}{k} + m.$$

## 4. SIMULATION OF PCA USING REAL PATIENT DATA

### 4.1. Simulation protocol

Simulation of morphine and fentanyl PCA was performed according to aggregate scheme presented in Fig 3. Drug infusion controller simulates patients' behavioral according to models, presented in section 2. Pharmacokinetics of morphine and fentanyl was simulated by three compartment model. We used the following parameters for simulation of morphine pharmacokinetics:

Central compartment volume = 17.8 l,

Time to deliver the bolus dose = 40 sec,

Bolus dose = 1 ml.

Morphine micro rate constants were chosen from (Dahlstrom et al. 1990).

We used the following parameters for the simulation of fentanyl PCA:

Central compartment volume = 6.09 l,

Time to deliver the bolus dose = 40 sec,

Bolus dose = 1 ml.

Fentanyl micro rate constants were chosen from the previous publications (Shafer et al. 1990).

#### 4.2. Evaluation of increased risk

For the evaluation of critical periods during PCA operation, we generated 500 patients demand logs (according to models presented in section 2) and used hybrid simulation technique to estimate drug plasma and effect compartment concentrations, that was done according to the pharmacokinetic multi-compartmental models. Increased risk event was defined as the time when effect compartment concentration exceeded critical (toxic) threshold. From that, cumulative risk could be defined as the time above this threshold and may correlate with the duration and severity of respiratory depression. We chose the following critical concentration thresholds: 0.02 mcg/l for morphine and 0.07 mcg/l for fentanyl; these levels were chosen arbitrarily for the demonstration purposes only. Two parameters were evaluated: the number of times critical concentration threshold was exceeded during 24 hour simulation period and the duration of the periods when concentration exceeds critical threshold.

##### 4.2.1. Evaluation increased risk periods for morphine

500 simulation sessions were performed that modeled morphine concentration at the effect site during 24 hours period; simulations were performed by using personal patients behavioral model (presented in section 2.1) together with pharmacokinetic compartmental model (Dahlstrom et al., 1978). Average number of times that increased risk concentration was reached, was relatively small – 1.78. Histogram (see Fig. 8) resembles geometric distribution, but statistical test rejected the null hypothesis.

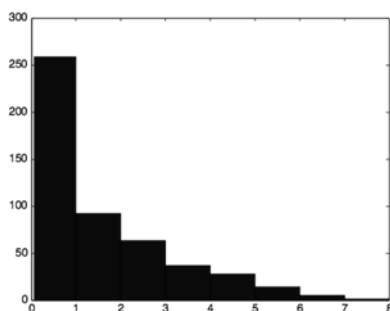


Figure 8: Histogram of the number of increased risk periods during simulated morphine analgesia

The time period spent above critical level is rather lengthy – about 109 minutes on average. More than half (256) of all increased risk periods were longer than 30 minutes. The histogram is presented below in fig.9:

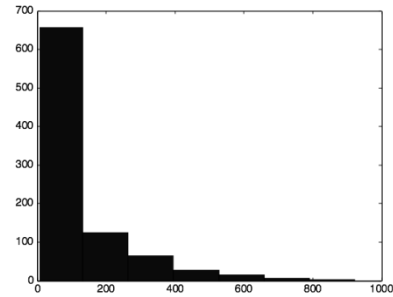


Figure 9: Histogram of the length of increased risk periods during simulated morphine analgesia

Overall proportion of time spent above the critical concentration during morphine PCA simulations was on average about 13.7 percent.

We also compared the estimated effect site morphine concentration from the original demand log with simulated morphine PCA logs (see Fig 10.):

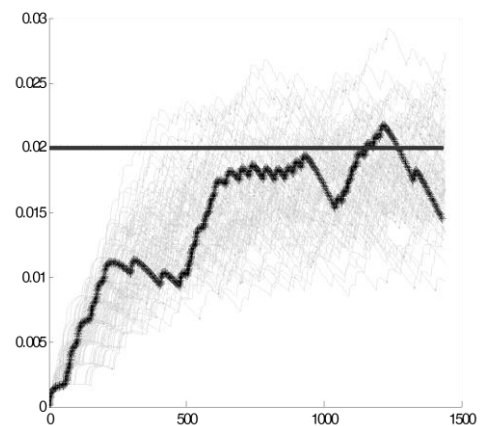


Figure 10: Comparison of estimated patients' drug concentration at the effect site (bold line) using 500 simulated morphine PCA logs. (Critical threshold 0.02)

It appears that the real patients' data fits the pattern of simulations.

##### 4.2.2. Evaluation increased risk periods for fentanyl

500 simulation sessions were performed that modeled fentanyl concentration at effect site during 24 hours period; simulation was done using patients' random behavioral model (presented in section 2.2) together with previously described three compartmental pharmacokinetics model (Shaffer et al., 1990).

Histogram demonstrating the number of times that effect site concentration exceeded critical threshold presented below in fig.11:

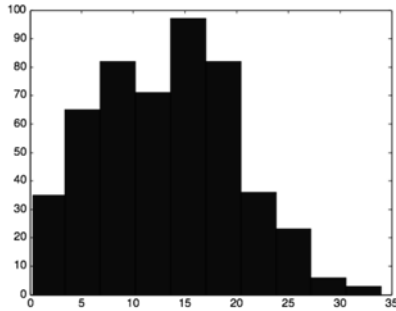


Figure 11: Histogram of the number of increased risk periods during simulated fentanyl analgesia

Average number of times exceeding critical threshold (13.22) is much higher than that of the morphine PCA. Data analysis showed that no standard discrete distribution was suitable to model these data.

However, mean length of the time above the critical threshold is much shorter than that of simulated morphine PCA - only about 6 minutes on average. Histogram is presented below in fig. 12 (no standard distribution was suitable to model the data):

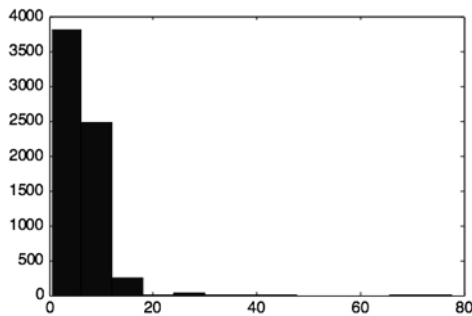


Figure 12: Histogram of the length of periods above critical threshold during fentanyl analgesia simulations

On an average, the time spent above the critical threshold was shorter, about 5.5 percent that of time when comparing morphine analgesia.

We also compared the estimated fentanyl concentration at the effect site obtained from the original demand log with the results obtained from simulated fentanyl PCA (see Fig 13):

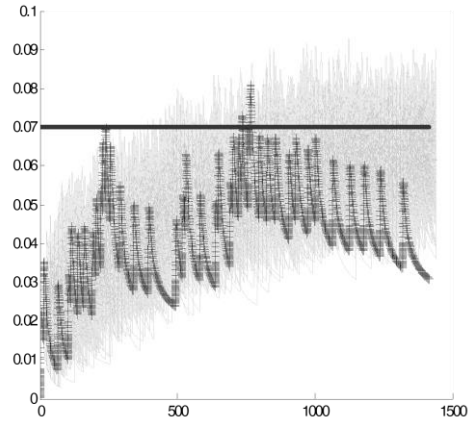


Figure 13: Comparison of the estimated patients' drug concentration at the effect site (bold line) and the results obtained from simulated fentanyl PCA. (Critical threshold 0.07)

As it is shown, patients' data matches simulated drug concentration levels reasonably well.

## 5. DISCUSSION

Although patient's demands for analgesia may be affected by a multiple factors, they can be reasonably approximated by the autoregressive moving average model of a stochastic process. To estimate the parameters of autoregressive moving average model we used the real data logs from PCA device, with duration in excess of 24. It provided ~50 data points in both morphine and fentanyl analgesia, which slightly exceeds a rule-of-thumb minimum sample size of 30 (Box, Jenkins 1994). Estimated model parameter can also be updated once new data (e.g. from PCA log) have arrived.

Even with limited patient's demand data, frequency and duration of the rare events — when effect compartment concentration exceeds critical threshold — can be predicted. Moreover, analysis of PCA logs allowed to uncover the fact, that periods of concentration exceeding critical threshold are more common, but of a shorter duration, when using fentanyl as a drug, not morphine, what corresponds to well known respiratory depression patterns of these two medications (Wong, 2013). Clinical investigation of these findings is warranted in order to establish individual critical concentration threshold for the respiratory suppression.

Our simulations assumed that starting drug concentration equals zero (figs. 10, 13). However, this is usually not the case, as patients commonly receive loading dose in order to achieve therapeutic concentration fast. This frequent clinical scenario was not accounted for in our simulations, although it is important to keep in mind, that the effect of a loading dose is negligible after two to three half-times of the drug.

Moreover, in current analysis we did not investigate whether or not the stochastic demand

process depends on the average concentration in the effect compartment; we intend to do that in the future.

## 6. CONCLUSION

The time between analgesia demands can be viewed as a random variable. Time series of a random process of such variable can be expanded to an arbitrary length, even from the limited real data logs; then these series can be analyzed utilizing autoregression moving average model ARMA(p,q), commonly used in time series forecasting.

Modeling results suggest that periods above critical (toxic) concentration threshold of morphine are less frequent, but of a longer duration as compared to Fentanyl.

## ACKNOWLEDGMENTS

The work described in this paper has been carried out within the framework the Operational Programme for the Development of Human Resources 2007-2013 of Lithuania „Strengthening of capacities of researchers and scientists“ project VP1-3.1-ŠMM-08-K-01-018 „Research and development of Internet technologies and their infrastructure for smart environments of things and services“ (2012- 2015), funded by the European Social Fund (ESF).

## REFERENCES

- Boom, M., Olofsen, E., Neukirchen, M., Fussen, R. et al., 2013. “Fentanyl utility function: a risk-benefit composite of pain relief and breathing responses”. *Anesthesiology*, 119(3), p. 663-674.
- Box, G. and Jenkins G.M., 1994. “Time Series Analysis: Forecasting and Control (Third Ed.)”, Prentice Hall.
- Dahan, A., Aarts, L. and Smith, T., 2010. “Postoperative opioids remain a serious patient safety threat.” *Anesthesiology*, 113(1), 260-261.
- Dahlstrom, B., Paalzow, L.K., Segre, G., Agren, A.A., 1978. “Relation Between Morphine Pharmacokinetics and Analgesia”. *Journal of Pharmacokinetics and Biopharmaceutics*, 6(1), p. 41-53.
- Grass, J.A., 2005. “Patient-controlled analgesia” *Anesthesia & Analgesia*, 101(5S), 44–61.
- Kofman, E., 2004. “Discreet Event Simulation of Hybrid Systems”. *SIAM Journal on Scientific Computing*, No.(25). p.1771–1797.
- Makridakis, S.m and Hibon, M., 1997. “ARMA models and the Box-Jenkins methodology”, *Journal of Forecasting*, 16(3), p. 147-163.
- Pranevicius, H., Simaitis, L., Pranevicius, M. and Pranevicius, O., 2011. “Piece-linear aggregates for formal specification and simulation of hybrid systems: pharmacokinetics patient controlled analgesia” *Electronics and Electrical Engineering* No 4(110).p.81-84.
- Pranevicius, H., Pilkauskas, V. Guginis, G., 2006. “Creating Simulation models Specified by PLA Using UML.” In *Proceedings of the International conference on Operational Research: Simulation and Optimization in Business and Industry*. Technologija, Kaunas. 87-92.
- Shafer, S.L., Varvel J.R., Aziz, N. and Scott, J.C., 1990. “Pharmacokinetics of fentanyl administered by computer controlled infusion pump.” *Anesthesiology* 73, 1091-1 102.
- Wong, M., 2013 “Addressing The Joint Commission's Concern About Opioid-Induced Respiratory Depression” *The Hospitalist*, May 2013.
- Woodhouse, A.; Mather, L.E. 2000. “The minimum effective concentration of opioids: a revisitiation with patient controlled analgesia fentanyl.” *Regional anesthesia and pain medicine*. 25(3), p. 259-267.

## AUTHORS BIOGRAPHY

**HENRIKAS PRANEVICIUS** Professor, Kaunas University of Technology. Business Informatics Department. Habilituated doctor of Technical Sciences at Ryga Electronic and Computer Technik Institute, 1984. Doctor degree in Kaunas Politechnical institute at 1970. Area of research activity: the use formal methods for performance and behavior analysis of complex systems including telecommunication, business, logistic and bioinformatics systems.

# MODELING OF MECHATRONIC SYSTEMS IN MATLAB (SIMULINK / SIMMECHANICS)

I.I. Kalapyshina <sup>(a)</sup>, A.D. Perechesova <sup>(b)</sup>, K.A. Nuzhdin <sup>(c)</sup>, V.M. Musalimov <sup>(d)</sup>, G.B. Zamoruev <sup>(e)</sup>

<sup>(a)</sup> ITMO University, National Research University SPbSPU

<sup>(b)</sup> ITMO University, SPbF IZMIRAN

<sup>(c)</sup> ITMO University, FAM - robotics

<sup>(d)</sup> ITMO University

<sup>(e)</sup> ITMO University

<sup>(a)</sup> [weir521@yandex.ru](mailto:weir521@yandex.ru), <sup>(b)</sup> [perechesova@gmail.com](mailto:perechesova@gmail.com),

<sup>(c)</sup> [nuzhdink@yandex.ru](mailto:nuzhdink@yandex.ru), <sup>(d)</sup> [musvm@yandex.ru](mailto:musvm@yandex.ru), <sup>(e)</sup> [georgyz09@gmail.com](mailto:georgyz09@gmail.com)

## ABSTRACT

Article reviews the training manual “Modeling of mechatronic systems in MATLAB (SIMULINK / SIMMECHANICS)”, which demonstrates the approach to the modeling of mechanical systems using the SimMechanics/Simulink/MATLAB environment. The article describes the structure of training manual and the main stages learning of design and analysis of mechanisms that are found on the basic knowledge of students from courses of university. Thus, the described in the textbook examples of the analysis of mechanisms using SimMechanics library helps to perpetuate the previously learned knowledge. The paper given special attention to the psychological aspects of students' perception of educational and methodical material, learning is based on the principle "from simple to complex" and follows the ideas of visibility and interaction of sensory perception of reality and intellectual knowledge. Tutorial Examples contained in the article and offered to students for learning based on the principles of modern technical devices, which helps to establish the relationship between theoretical knowledge and practical experience.

## Keywords:

simulation, MATLAB, SimMechanics, analysis of mechanisms, kinematics of mechanisms, dynamics mechanisms, training manual

## 1. INTRODUCTION

One of the most effective systems of software for numerical computations in the present days is MATLAB - matrix laboratory and high level programming language. To increase the functionality of the system, there is a Simulink library - supplement provides block imitating modeling of various systems and devices with the use of modern technology a visual-oriented programming. In Simulink toolbox in addition to general there are specialized packages for solving a specific purpose tasks, for example, SimPowerSystems - for modeling energy systems and devices, Signal Processing - for modeling signal processing systems,

Fuzzy Logic - for the simulation of fuzzy systems and others (Li, et al., 2013).

For modeling systems and devices of the physical nature of greatest interest is the Simscape package. In this package there are blocks that represent physical bodies interacting by means reactions, as opposed to block Simulink, which are mathematical operators and interact with each other by means signals. Simscape includes such sections as SimElectronics - for the simulation of electronic and electromechanical systems, SimHydraulics - for the simulation of hydraulic systems, SimMechanics - for the simulation of mechanical systems.

Implementation of the MATLAB in the learning process of many universities is rapidly and there is a lot of literature on this topic. However, there are special packages are described in sufficient detail to fully understand. Therefore especially important to create educational materials for the study of special package tools covering a certain direction of science and technology.

In recent years the Department of Mechatronics, Saint Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO University) in the educational process and scientific research is widely used package SimMechanics. In packet-based testing, is a method of comparing the results of a particular physical model with its analogue and analytical conclusions on the degree of its approximation. On the basis of experience gained during the modeling of mechanical systems, it was decided to create a tutorial for students and teachers.

## 2. THE STRUCTURE OF THE TRAINING MANUAL

Figure 1 shows the structure of training manual in graphical form. It consists of four training chapters, each chapter has a section dedicated to the theoretical information, detailed solutions examples (3-4 pcs.), as well as tasks for students self-study. There are about 12 completing example and more than 50 self-study tasks.

Theoretical information in the chapters intended for reference purposes as soon as textbook designed for senior baccalaureate courses and magistracy. To help the student there is extensive list of references.

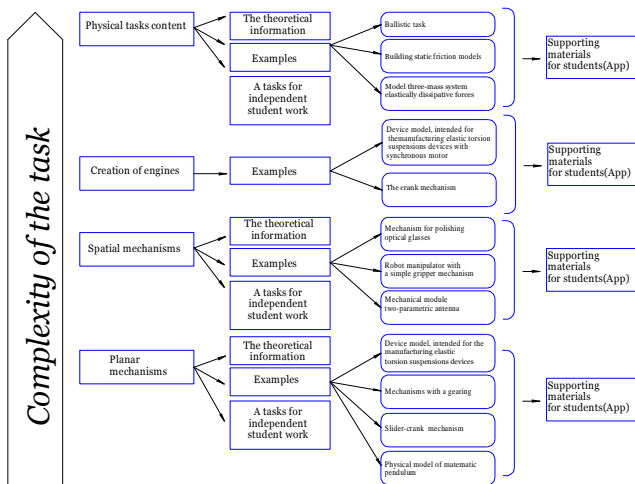


Figure 1. The structure of the training manual "Modeling of mechatronic systems in the MATLAB (Simulink / SimMechanics)»

Task complexity increases gradually from chapter to chapter. However, each set of tasks based on previously studied material, and individual case studies consistently included in tasks that are more complex. It allowing students to retain new knowledge, due to repeated repetition of material. Case Studies gradually disassembled setting, all used blocks, and control programs are given in the annexes. Therefore, the students can successfully carry out the training on their own. Perform tasks for self-study requires maximum use of the knowledge gained. A similar sequence presentation of material allows to control the learning process and to evaluate student's knowledge level adequately.

### 3. MODELING OF MECHANISMS

The first example is for modeling the physical model of a simple pendulum shown in Figure 2 and 3. It is one of the simplest systems for modeling and allows students to get acquainted with the SimMechanics blocks and their interaction with each other.

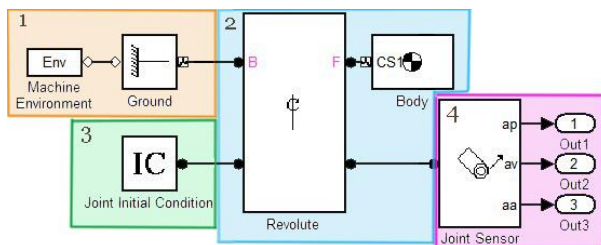


Figure 2. Structural model of a mathematical pendulum in SimMechanics

Modeling and research of pendulums with various modifications is an important task, because the vibrational elements are working bodies in many

measuring devices. For example, in the sources (Kot, et al., 2013; Li, et al., 2013; Dyakonov, 2005; Dyakonov, 2008) by different authors there are models of these pendulums.

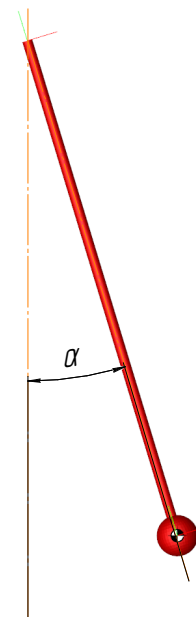


Figure 3. The output - SimMechanics animated model of the mathematical pendulum

Further, the level of complexity of models gradually increased. The following example is a slider-crank mechanism, two-stage gearbox and the weaving devices actuator model designed for the manufacture of elastic torsion suspensions, Fig. 4 and 5 (Musalimov, et al., 2013; Perechesova, 2012; Zamoruev et al., 2012; Kalapyshina, 2012).

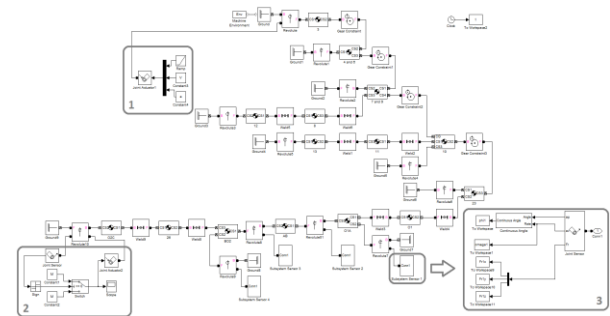


Figure 4. Structural model of the weaving device actuator in SimMechanics

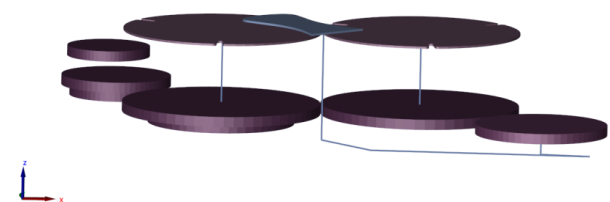


Figure 5. The output SimMechanics - animated model of the weaving device actuator

Model actuator plaiting together the knowledge of students about the work SimMechanics blocks designed to simulate planar mechanisms. It includes all of the subsystems previously discussed, or their equivalent. Value model for learning is that there is a mechanism used in the Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation, Russian Academy of Sciences, St. Petersburg Department (SPbF IZMIRAN).

Sections tasks for independent works devoted to flat mechanisms are based on the mechanisms that invented and investigated by P.L.Chebyshev, and also on the training devices mockups, existing at the Department of Mechatronics ITMO (Fig. 6) (Musalimov, et al., 2013). Such an approach allows the ideas of visualization and helps to establish the relationship the sensory perception understanding of reality and intellectual knowledge. Figure 7 shows an example of a student's independent work, which aims to draw a parallel between the real object and the model in MATLAB.



Figure 6. Study mock of devices, laboratory of the Department of Mechatronics ITMO University;

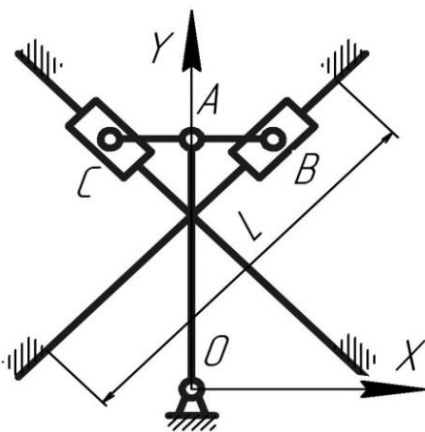


Figure 6. The tasks for independent work of students and Structural model of a student's SimMechanics

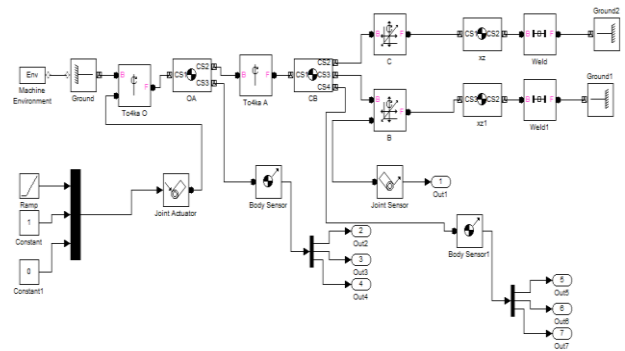
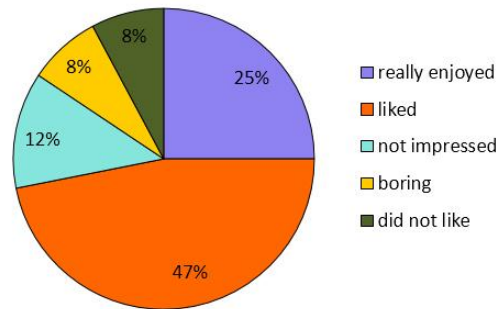


Figure 7. The Structural model of a student's SimMechanics

Opinion survey was conducted among undergraduate students of 4 training groups totaling 64 people. Investigated emotional perceptions and complexity of discipline. Results of the survey are shown in the diagrams (fig.8). It would be concluded a successful testing of course.

#### Emotional perception of the course by students



#### The course complexity of students' opinion

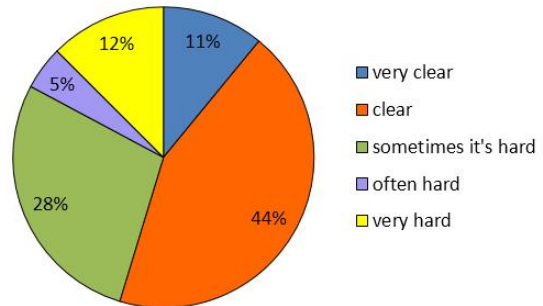


Figure 8. The result of students' opinion survey

Further development of the educational process requires a sequential increase of complexity of the tasks at the same time bringing them closer to the actual mechanisms that are used in modern industry. As a fifth example, students are supposed to analyze and model the mechanical part of a two-parameter antenna (Musalimov, et al., 2013; Perechesova, 2012). As a mechanics object the system under consideration is a representative device with parallel kinematics. His movement occurs in two planes. A sixth example is an actuator for processing of optical elements (Kalapyshina, 2012). The most obvious in this section is an example of modeling the robot with the capture of Fig. 9.



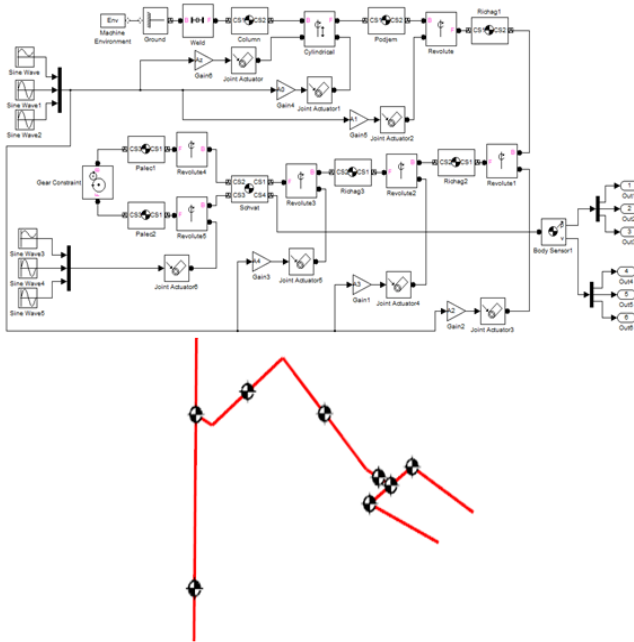


Figure 9. Structural model of the robot with SimMechanics capture and result of - Animated SimMechanics model robot with capture

Robotics is the most actual for the modern industry. Therefore, many research organizations and universities are conduct the research in this area. Examples of modeling of robots-manipulators are described in (Zimmermann, et al., 2009) and in (Xiao, et al., 2009; Rybak, et al., 2013; Ghanbari, et al., 2013; Hasan, et al., 2009) are studies of the mechanisms using the SimMechanics.

In the training manual, also designed section dedicated to solving problems of complicated physical and mathematical laws, the presence of external influences, resistance forces, the presence of higher kinematic pairs (Musalimov, et al., 2013). This section is offered to educate the problem of the three bodies interaction, associated elastic-dissipative forces, static friction model and ballistic task. A good example is SimMechanics model for solving the problem of ballistic Fig. 10 and 11.

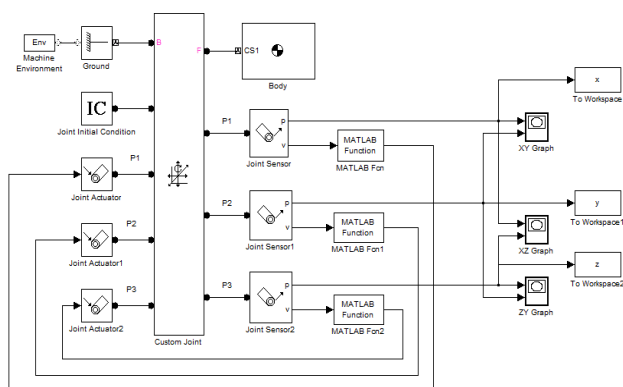


Figure 10. Structural model SimMechanics for ballistic problem

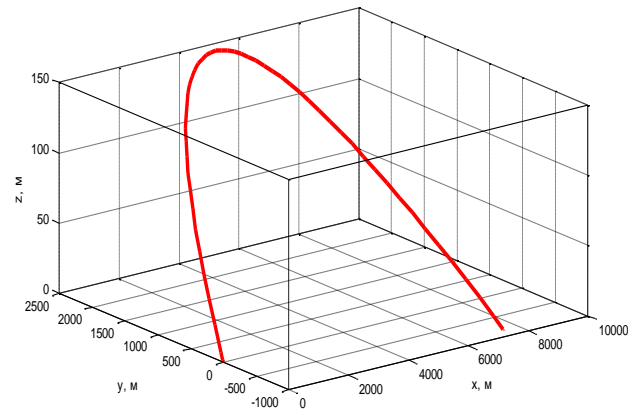


Figure 11. Result - the solving the ballistic problem

The last section of a manual, is dedicated to creating engines. It prepares students for further education of MATLAB SimMechanics library, package Simulink and others.

#### 4. CONCLUSION

Described in the textbook fundamentals of the theory of machines and mechanisms and theoretical mechanics and methods of mathematical modelling of various mechanical systems and physical processes. The manual provides models of mechanisms, mechanical and measuring systems made using the library SimMechanics, package Simulink environment MATLAB. The appendices contain the necessary data for playback students models. To sections of the job AIDS designed for self-study.

The manual is recommended as a textbook for masters of higher educational institutions, enrolled in the direction of masters preparation 200100 "Instrument-making" and the 221000 "mechatronics and robotics", may be useful for students studying courses "Theory of mechanisms and machines", "Designing of mechatronic devices", "Fundamentals of designing and design of devices for use in the master's programs "System simulation in mechatronics", "Modular technology in biomechatronic and robotics", as well as for teachers and researchers dealing with the issues of the course "Theory of mechanisms and machines". The textbook was given the stamp of UMO "Instrument making" resolution 588 from 29.11.2013, For the development of the educational process perspective direction is the creation of electronic educational-methodical complex (UMK), simulation of mechatronic systems in the MATLAB/Simulink/SimMechanics, for distance learning AcademicNT, developed in St. Petersburg national research University of information technologies, mechanics and optics (UNIVERSITY ITMO). This system allows to increase the efficiency and quality of teaching, the unified structure of electronic educational and methodological resources and the use of modern computer technologies.

## REFERENCES

- Dyakonov V.P., 2005. *The MATLAB 6.5 SPI + Simulink 5 u MATLAB 7 + Simulink 6 in mathematics and mathematical modeling*. Moscow: SLON - Press (in Russian).
- Dyakonov V.P., 2008. *The Simulink 5/6/7*. Moscow: DMK – Press (in Russian).
- Ghanbari, A., Solaimani, R., Rahmani, A., Tabatabaie, F., 2013. Design and simulating five-finger robot hand to grasp spherical objects. *Life Science Journal*, Issue SUPPL.3, Vol. 10, pp. 140-145.
- Hasan, R. , Rahideh, A. Shaheed, H., 2013. Modeling and interactional control of the multifingered hand. *ICAC 19th International Conference on Automation and Computing: Future Energy and Automation*, pp 126-131. 2013, London (United Kingdom).
- Kalapyshina, I.I. 2012. Modeling the dynamics of the machine for processing glass. *Scientific and Technical Journal «Priborostroenie»*. no 6 (55). pp. 74-77 (in Russian).
- Kopytenko, Y.A., Sergushin, P.A., Petrishchev, M.S., Levanenko, V.A., Perechesova, A.D., 2010. Device for manufacturing torsion bars with helical anisotropy UISAT-1. *9th International Symposium on Measurement Technology and Intelligent Instruments*, pp. 625-628. 29 June 2009 through 2 July 2009, Saint-Petersburg (Russian Federation).
- Kot, A., 2013. Bi-axial inverted pendulum modelling. *Conference Paper – 14<sup>th</sup> International Carpathian Control Conference*, pp. 171-174. Rytro (Poland)
- Lazarev U.F., 2005. *Modeling of processes and systems in MATLAB*. Saint-Petersburg: BHV (in Russian).
- Li H.X., Lu A.S., 2013. The design of output feedback controller for inverted pendulum system, *Applied Mechanics and Materials. 2nd International Conference on Measurement, Instrumentation and Automation*, pp. 336-338, 2013, Guilin (China).
- Musalimov, V.M., Zamoruev, G.B., Kalapyshina, I.I., Perechesova, A.D., Nuzhdin, K. A., 2013. *The modeling of mechatronic systems by MATLAB (SIMULINK / SIMMECHANICS)*. Saint-Petersburg: ITMO University (in Russian).
- Perechesova, A.D., 2012. *The analysis and synthesis of mechanism for manufacturing Torsion Bars of devices*. Thesis (PhD). ITMO University (in Russian).
- Rybak, L.A., Gaponenko, E.V., Chichvarin, A.V., Strutinsky, V.B., Sidorenko, R.A., 2013. Computer-aided modeling of dynamics of manipulator-tripod with six degree of freedom. *World Applied Sciences Journal*, no 2 (25), pp. 341-346.
- Xiao, X. Li, Y., Tang, H., 2013. Kinematics and interactive simulation system modeling for robot manipulators. *IEEE International Conference on Information and Automation*, pp. 1177-1182. Yinchuan (China).
- Zamoruev, G.B., Kalapyshina, I.I., 2012. Kinematic control of two-parametrical scanning aerie. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, no. 2 (78), pp. 78–83 (in Russian).
- Zimmermann K., Zeidis I., Behn C., 2009. *Mechanics of trrestrial locomotion*. Springer – Verlag Berlin Heidelberg.

## AUTHORS BIOGRAPHY

**V.M. Musalimov.** Professor and Head of Mechatronics Department, ITMO University and Chief research scientist of IPME, RAS.

Participant of International conferences and congresses (Denmark, Poland, Italy, UK, Germany, Portugal, Romania, Russia etc.).

Scientific School Leader of ITMO University “Fundamental and applied problems of precision and reliability of mechanisms and tools”.

Co-leader of regular International Scientific school-conferences “Fundamental and applied problems of precision and reliability of mechanisms and tools”.

Has graduated 26 post-graduates with PhD degree.

Has published more than 200 scientific works, a number of monographs and book references. Has obtained 25 patents.

Gives lectures at the Technical University of Ilmenau regularly (Germany).

Staff lecturer at ITMO University and Institute of Machinery (Saint-Petersburg).

**G.B. Zamoruev.** Graduated in 1962, specialty - Marine machinery. He worked as a design engineer in the turbine building of the ships. Since 1966, he began teaching at the Leningrad Polytechnic Institute at the Department of Theory of mechanisms and machines. In 1972 he defended his thesis. In 1976 he was promoted to the rank of associate professor and to the corresponding post. In 1980 he moved to Leningrad Institute of Fine Mechanics and Optics at the Department of Theory of mechanisms and instruments parts (hereinafter Department of Mechatronics, ITMO University) as an associated professor, where I work now.

At the moment, my overall experience is 52 years old, of which teaching at higher educational institutions - 48 years (the last 34 at ITMO University).

Number of published scientific works (including inventions and manuals) is about fifty.

Research interests:

- geometry and kinematics of devices with higher kinematic pairs;
- kinematics and dynamics of mechanisms (including robotic manipulators);
- simulation of mechanisms, systems and processes;
- methods and computer programs for optimization modeling.

**A.D. Perechesova.** Year of birth 1985. In 2008 I successfully graduated the Mechatronics Department of ITMO University. After University (since 2007) up to present time I am working in SPbF IZMIRAN. From

2008 to 2012 I am studied at the full-time postgraduate of the University ITMO. In 2012 I defended a thesis for scientific degree of PhD (technical sciences), specialty 05.02.18 "The theory of Mechanisms and Machines". The theme of thesis: "The analysis and synthesis of mechanism for manufacturing instruments torsion bars".

Since 2013 I am the research fellow of SPbF IZMIRAN. Since 2013 I am the Docent of the Mechatronics Department of ITMO University.

In 2012 I am won the "Young scientists of ITMO University 2012".

Number of published scientific works (including inventions and manuals) is about thirty.

Research interests include: geophysics, the theory of mechanisms and machines, the optimization theory, the theory of Helically Anisotropic solid.

**I.I. Kalapyshina.** Graduated the Mechatronics department at the University of Information Technologies in 2010. Since this time has been phd-student of the same department. From the 2012 has been working as assistant in Mechatronics department. Number of published scientific works is about fifteen.

Research interests include: the theory of mechanisms and machines, simulation of mechanisms, systems and processes, methods and computer programs for optimization modeling.

**K.A. Nuzhdin.** Graduated the Mechatronics department at the University ITMO in 2010. Since this year has been phd-student of the same department. From the 2011 has been working as assistant in Mechatronics department. Number of published scientific works is about seven.

# ON THE APPROXIMATE SOLUTION FOR A FULLY COUPLED PARABOLIC-HYPERBOLIC PDES SYSTEM APPLIED IN THERMOMECHANICAL BEHAVIOUR OF HYDROGEL MATRIX USED IN DRUG DELIVERY

N. Santatriniaina<sup>(a)</sup>, B. Boutin<sup>(a)</sup>, E. Darrigrand<sup>(a)</sup>, F. Mahé<sup>(a)</sup>, N. Crouseilles<sup>(a)</sup>, D. Pioletti<sup>(b)</sup>, L. Rakotomanana<sup>(a)</sup>

<sup>(a)</sup>University of Rennes 1, Mathematical Research Institute of Rennes, 263 av. du Général Leclerc, 35042 Rennes-France

<sup>(b)</sup>EPFL of Lausanne, Laboratory of Biomechanical Orthopedics, Al. 1137, Ch-1015, Ibo.ch, Switzerland.

<sup>(a)</sup> [Email: nirina.santatriniaina@univ-rennes1.fr](mailto:nirina.santatriniaina@univ-rennes1.fr)

## ABSTRACT

This work deals with numerical methods for a fully coupled system of parabolic-hyperbolic partial differential equations modeling the thermo-mechanics of the hydrogel matrix of drug delivery devices. The model of hydrogel matrix is governed by heat conduction equation, wave equation and a thermodynamically admissible constitutive law. These equations are coupled by source terms. Transient analysis is assumed for the present study. Original numerical approach based on finite element formulation is developed to solve the problem. The model is developed to predict the reversible- irreversible entropy and heat production in the material to control drug delivery devices by following the temperature change. Experimentally, various situations are assumed in order to observe variations of entropy and temperature by using micro-calorimetric test for many frequency rates. Some numerical results in correlation with experimental measurements are presented in this work.

Keywords: finite element formulation, cyclic load, parabolic-hyperbolic equations, drug delivery.

## 1. INTRODUCTION

magneto-elasticity, and thermo-elasticity and for many other phenomena in science engineering, a coupled partial differential equation appears (Mohammed 2009; Jeffrey 2005; Xia 2010). In general, coupled system of parabolic and hyperbolic equations is obtained after deriving conservation equation (physical laws) of the system (Holzapfel 1995). And it is well known that in continuous media theory, the mechanics and the thermodynamics fields are coupled. Most theoretical foundations of such phenomenon were established by Duhamel, Coleman Noll (e.g. Rakotomanana 2009; Claire 2011). In this paper, we are in presence of thermomechanical process in the material. Then, the coupled partial differential equation are obtained by applying conservation law, first and second law of thermodynamics for heat conduction and linear momentum for the wave equation (e.g. Rakotomanana 2009; Betram 1990). The Hydrogel matrix has dissipation properties inducing temperature change under cyclic load (Mohandrea 2013; Philippe 2014). The mechanical loading is accompanied by a production of entropy and heat, which can influence the

temperature and the behavior of the material (Mahaman 2012). Material is mostly viscoelastic. First such phenomenon may be used to optimize drug delivery device induced by the temperature change, as in cartilage joint which is always under cyclic load (Mohandrea 2013; Philippe 2014; chi 2001). Elevation of the temperature may induce defect in the material (prostheses) and may influence the formation of transfer film and contact at the interface between the biological tissues and the prosthesis (Chi 2001). So, we need to establish a predictive model to qualify and quantify the temperature effect in these materials and to determine the maximal temperature admitted for various situations.

Constitutive equation of some viscoelastic material, biological tissues, without temperature effect and undergoing large strain was previously established (Rakotomanana and Pioletti 2009). In order to include the material behavior, we introduce a new constitutive equation in strain gradient continuum taking into account the heat effect. Constitutive equation is first obtained by defining a Helmholtz free energy as in (Rakotomanana and Pioletti 2009) but taking into account an internal variable and temperature effect. We use classical Coleman and Noll approach for thermo-mechanical processes (Holzapfel 1995). To obtain numerical solutions of a thermo-mechanical problem, equations must be solved simultaneously.

Various papers exhibit some smooth solutions of the coupled parabolic-hyperbolic system. The uniqueness, local existence and asymptotic study of the solution were developed in (Hao, Maurizio, Songmu 2007; Yinghui 2013). Many assumptions are used to simplify the problem, some papers solve the problem for one or two dimensions (Shin and Seung 1999; Chi 2001; Kunish 2009), solves the problem by considering that the effect of stress with temperature production is small (in this case the problem is weakly coupled). The last method uses the Eulerian-Lagrangian method in order to obtain numerical solutions for this problem (Mohammed 2009). In this work, we assume that coupling term is not negligible, so the problem is fully coupled and reintroduce the internal variable and its evolution law in order to complete the behavior of the material.

Mechanical transformation is accompanied by production of entropy and heat, which can influence the

temperature and the behavior of the material. For biological tissues, coupling is important. Indeed, some part of energy is converted into heat as dissipation, thus resulting in an irreversible temperature and entropy rise. In such a case, if temperature or entropy gradient is significant, mechanical coupling term cannot be neglected. We establish for the hydrogel the evolution law and then we study the response with a cyclic load.

This paper aim to develop a numerical approach for a fully coupled thermo-mechanical system under cyclically load and dynamical boundary conditions. First, we would like to better understand and quantify the influence of thermo-mechanical coupling on the material behavior and on the dynamic response of the viscoelastic biological tissues substitutes. Quantification is obtained by analyzing its effect on the dissipated energy and on the entropy production within the material for each cycle. We determine the behavior of artificial substitute by studying the correlation of the theoretical results with the experimental measurement. An artificial gel is used to substitute biological tissues, taking into account thermo-mechanical load in the material.

The first part of this paper illustrates the model for the thermo-mechanical coupling in the case of viscoelastic wires subject to a cyclically mechanical load. The numerical approach and the result are presented with the stability study on the evolution of the dissipated energy with the strain rate and on the influence of thermo-mechanical coupling. The second part presents briefly the experimental measurement and the correlation between results on the reversible and irreversible entropy. The thermo-mechanics approaches in this work are based on the Coleman and Noll based on the Clausius-Duhem inequality.

We assume that the Helmholtz's free energy and the equipresence theory by the Truesdell and Toupin to define the constitutive law of the material (Claire; Rakotomanana 2009; Holzapfel 1995). We introduce an internal variable in order to take into account the dependence of free energy, entropy and temperature gradient. These assumptions lead to consistent forms of the second principle of the thermo-mechanics and the Clausius-Duhem inequality. A new constitutive law is established for each internal variable.

## 2. EXPERIMENTAL METHODS

For measuring temperature change in substitute sample under cyclic load, deformation micro-calorimetric test was used. It consists to apply on a cylindrical sample of hydrogel matrix (4 [mm] and 6 [mm] of HEMA-EGDMA hydrogels) a cyclic load. For experimental measurements, various frequencies are tested to observe variations of entropy and temperature by using micro-calorimetric test. We assume that system is adiabatic.

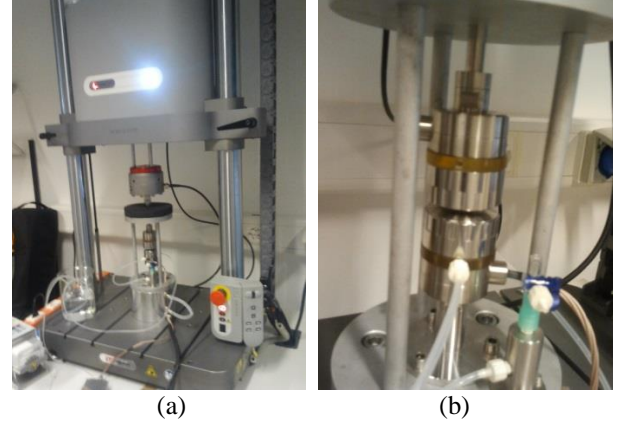


Figure 1: Deformation micro-calorimetric test.

The figure 1 illustrates the micro-calorimetric system: The left side (a) illustrates the global system. The right side (b) illustrates the sample (in middle of the two sensors). This system has a thermal insulation but no in the figure.

## 3. MATHEMATICAL MODEL

Let  $\Omega \subset \mathbb{R}^d$  be a bounded domain in with smooth boundary  $\partial\Omega \subset \mathbb{R}^{d-1}$  (it represents the sample in hydrogel matrix HEMA-EGDMA). For any time, mechanical and heat stress occupies this domain. Time is defined by scalar  $t \in [0, T]$  with  $t \subset \mathbb{R}_+$ .

### 3.1. Constitutive equation

The constitutive equation is based on the thermodynamics potentials. The Helmholtz's free energy is given by

$$\psi(\mathbf{C}, \theta, \xi_i) = \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \quad (1)$$

And the dissipation potential is given by

$$\chi(\mathbf{C}; \dot{\mathbf{C}}, \overline{\nabla\theta}, \overline{\nabla\xi_i}) = \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) \quad (2)$$

The expressions of the invariant  $\mathbb{J}_i$  are given in the appendix. In the equation (1) and (2)  $\mathbf{C}$  denotes the Cauchy-Green strain tensor, given by:

$$\mathbf{C} = \frac{1}{2}(\nabla\mathbf{u} + \nabla^T\mathbf{u} + \nabla^T\mathbf{u}\nabla\mathbf{u}) \quad (3)$$

$$\dot{\mathbf{C}} = \frac{1}{2}(\nabla\dot{\mathbf{u}} + \nabla^T\dot{\mathbf{u}} + \nabla^T\dot{\mathbf{u}}\nabla\mathbf{u} + \nabla^T\mathbf{u}\nabla\dot{\mathbf{u}}) \quad (4)$$

Where  $\mathbf{u}$  denotes the displacement vector.

The second Piola-Kirchhoff stress tensor  $\mathbf{S}$  is given by

$$\mathbf{S} = \mathbf{S}^e + \mathbf{S}^v \quad (5)$$

$$\mathbf{S}^e(\mathbf{C}, \theta, \xi_i) = 2\rho \frac{\partial}{\partial \mathbf{C}} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \quad (6)$$

$$\mathbf{S}^v(\mathbf{C}; \dot{\mathbf{C}}, \overline{\nabla\theta}, \overline{\nabla\xi_i}) = \frac{\partial}{\partial \dot{\mathbf{C}}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) \quad (7)$$

The potential  $\psi(\mathbf{C}, \theta, \xi_i)$  and  $\chi(\mathbf{C}; \dot{\mathbf{C}}, \overline{\nabla\theta}, \overline{\nabla\xi_i})$  are thermodynamically acceptable. They respect the entropy inequality and Clausius-Duhem inequality.

$$\rho \dot{e} + (\mathbf{S}^e + \mathbf{S}^v) : \frac{\dot{\mathbf{C}}}{2} - \frac{\mathbf{Q}}{\theta} \cdot \nabla\theta \geq 0, \forall \mathbf{C}, \dot{\mathbf{C}}$$

Then,

$$\rho [s\dot{\theta} + \dot{\psi}(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i)] + (\mathbf{S}^e + \mathbf{S}^v) : \frac{\dot{\mathbf{C}}}{2} - \frac{\mathbf{Q}}{\theta} \cdot \overline{\nabla\theta} \geq 0, \forall \mathbf{C}, \dot{\mathbf{C}}$$

Where  $s$  denotes the entropy,  
 $s = s(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i)$

And then, the dissipation potential  $\chi(\mathbf{C}; \dot{\mathbf{C}}, \overline{\nabla\theta}, \overline{\nabla\xi_i}) = \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C})$  must be convex, positive and null and verify:

$$\begin{aligned} & \frac{\partial}{\partial \dot{\mathbf{C}}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) : (\dot{\mathbf{C}} - \dot{\mathbf{C}}^*) + \\ & \frac{\partial}{\partial \overline{\nabla\theta}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) : (\overline{\nabla\theta} - \overline{\nabla\theta}^*) + \\ & \frac{\partial}{\partial \overline{\nabla\xi_i}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) : (\overline{\nabla\xi_i} - \overline{\nabla\xi_i}^*) \geq \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) - \\ & \chi(\mathbb{J}_1^*, \mathbb{J}_2^*, \mathbb{J}_3^*, \mathbb{J}_4^*, \mathbb{J}_5^*, \mathbb{J}_6^*, \mathbb{J}_7^*, \overline{\nabla\theta}^*, \overline{\nabla\xi_i}^*; \theta^*, \dot{\mathbf{C}}^*) \forall \mathbf{C}, \dot{\mathbf{C}}, \mathbf{C}^*, \dot{\mathbf{C}}^* \end{aligned}$$

Where  $\chi(\mathbb{J}_1^*, \mathbb{J}_2^*, \mathbb{J}_3^*, \mathbb{J}_4^*, \mathbb{J}_5^*, \mathbb{J}_6^*, \mathbb{J}_7^*, \overline{\nabla\theta}^*, \overline{\nabla\xi_i}^*; \theta^*, \dot{\mathbf{C}}^*)$  denotes the dissipation potential in the dual base.

The state function in  $\Omega \subset (\mathbb{R}^d \times \mathbb{R}_+)$  can be written as,

$$\left\{ \begin{array}{l} \mathbf{S}^e(\mathbf{C}, \theta, \xi_i) = \frac{\partial}{\partial \mathbf{C}} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \\ \mathbf{S}^v(\mathbf{C}; \dot{\mathbf{C}}, \overline{\nabla\theta}, \overline{\nabla\xi_i}) = 2\rho \frac{\partial}{\partial \dot{\mathbf{C}}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) \\ s(\mathbf{C}, \theta, \xi_i) = -\frac{\partial}{\partial \theta} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \\ \frac{\partial}{\partial \overline{\nabla\theta}} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) = 0 \\ \frac{\partial}{\partial \overline{\nabla\xi_i}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) = \frac{\mathbf{Q}}{\theta} \end{array} \right.$$

### 3.2. Governing equation

The governing equation in transient analysis is obtained as a parabolic-hyperbolic coupled system.

$$\left\{ \begin{array}{l} \mathbf{Div} \mathbb{P} + \rho \mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \rho \dot{e} = \mathbf{S} : \frac{\dot{\mathbf{C}}}{2} - \mathbf{Div} \mathbf{Q} + \rho \mathbf{r} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa \theta \nabla \theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{array} \right. \quad (8)$$

Where  $\mathbf{f}$  and  $\mathbf{r}$  denote respectively the force vector and the heat source in  $\Omega \subset (\mathbb{R}^d \times \mathbb{R}_+)$ .

The first Piola-Kirchhoff stress tensor  $\mathbb{P}$  is given by  $\mathbb{P} = \mathbb{F} \mathbf{S}$ . Then, we have  $\mathbb{P} = \mathbb{F} (\mathbf{S}^e + \mathbf{S}^v)$ .

$$\left\{ \begin{array}{l} \mathbf{Div} [\mathbb{F} (\mathbf{S}^e + \mathbf{S}^v)] + \rho \mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \rho \dot{e} = (\mathbf{S}^e + \mathbf{S}^v) : \frac{\dot{\mathbf{C}}}{2} - \mathbf{Div} \mathbf{Q} + \rho \mathbf{r} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa \theta \nabla \theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{array} \right. \quad (9)$$

Where  $\mathbb{F}$  is the strain gradient,  $\mathbb{F} = \nabla \varphi(\mathbf{x}, t)$ ,  $\varphi(\mathbf{x}, t)$  denotes the holonomic transformation of the  $\Omega \subset \mathbb{R}^d$  with the condition  $\mathbf{det} \mathbb{F} > 0$ . We have the relation

$$\begin{aligned} \mathbf{C} &= \mathbb{F}^T \mathbb{F}, \\ \dot{\mathbf{C}} &= \dot{\mathbb{F}}^T \mathbb{F} + \mathbb{F}^T \dot{\mathbb{F}}, \end{aligned}$$

The internal energy  $e$  is given by  $e = s\theta + \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i)$ , and then we have

$$\dot{e} = s\dot{\theta} + \dot{s}\theta + \dot{\psi}(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i),$$

The governing equation can be written as

$$\left\{ \begin{array}{l} \mathbf{Div} [\mathbb{F} (\mathbf{S}^e + \mathbf{S}^v)] + \rho \mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \rho [s\dot{\theta} + \dot{s}\theta + \dot{\psi}(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i)] = (\mathbf{S}^e + \mathbf{S}^v) : \frac{\dot{\mathbf{C}}}{2} \\ - \mathbf{Div} \mathbf{Q} + \rho \mathbf{r} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa \theta \nabla \theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{array} \right. \quad (10)$$

We have,

$$\left\{ \begin{array}{l} \mathbf{Div} [\mathbb{F} (\mathbf{S}^e + \mathbf{S}^v)] + \rho \mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \rho c \frac{\partial \theta}{\partial t} - \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} - \rho \left( \frac{\partial \psi}{\partial \xi_i} - \theta \frac{\partial^2 \psi}{\partial \theta \partial \xi_i} \right) : \dot{\xi}_i - \mathbf{Div} \mathbf{Q} \\ + \rho \mathbf{r} = 0 \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa \theta \nabla \theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{array} \right. \quad (11)$$

The boundary conditions are:

- Mechanical boundary condition
 
$$\left\{ \begin{array}{l} \mathbb{F} (\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = \mathbf{T}_M(\mathbf{n}, \mathbf{M}) \\ \mathbf{u} = 0 \\ \mathbb{F} (\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = 0 \\ \mathbb{F} (\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = 0 \end{array} \right.$$

Where  $\mathbf{n}$  denote the outer unit vector, and  $\mathbf{T}_M(\mathbf{n}, \mathbf{M})$  is the external body force applied in some part of  $\Omega$ .

- Thermal boundary condition

$$\begin{cases} -\kappa\theta\nabla\theta \cdot \mathbf{n} = 0 \\ -\kappa\theta\nabla\theta \cdot \mathbf{n} = 0 \\ \theta = \theta_0 \\ -\kappa\theta\nabla\theta \cdot \mathbf{n} = k_c(\theta - \theta_f) \end{cases}$$

Where  $k_c$  denotes the convection coefficient,  $\theta_f$  denote the reference temperature.

Finally, we have:

$$\begin{cases} \mathbf{Div} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] + \rho\mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^e = 2\rho \frac{\partial}{\partial \mathbf{C}} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^v = \frac{\partial}{\partial \dot{\mathbf{C}}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \rho c \frac{\partial \theta}{\partial t} - \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} - \rho \left( \frac{\partial \psi}{\partial \xi_i} - \theta \frac{\partial^2 \psi}{\partial \theta \partial \xi_i} \right) : \dot{\xi}_i - \mathbf{Div} \mathbf{Q} \\ \quad + \rho \mathbf{r} = 0 \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa\theta\nabla\theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{cases}$$

We consider two cases:

- We consider that the effect of the heat produced by the internal variable  $\xi_i$  is equal to zero,  $\rho \left( \frac{\partial \psi}{\partial \xi_i} - \theta \frac{\partial^2 \psi}{\partial \theta \partial \xi_i} \right) : \dot{\xi}_i = 0$ .

In this case the governing equation is given by:

$$\begin{cases} \mathbf{Div} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] + \rho\mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^e = 2\rho \frac{\partial}{\partial \mathbf{C}} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^v = \frac{\partial}{\partial \dot{\mathbf{C}}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \rho c \frac{\partial \theta}{\partial t} - \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} - \mathbf{Div} \mathbf{Q} + \rho \mathbf{r} = 0 \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa\theta\nabla\theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{cases}$$

- We consider that the effect of the heat produced by the internal variable  $\xi_i$  is equal to zero,  $\rho \left( \frac{\partial \psi}{\partial \xi_i} - \theta \frac{\partial^2 \psi}{\partial \theta \partial \xi_i} \right) : \dot{\xi}_i \neq 0$ .

In this case the governing equation is given by:

$$\begin{cases} \mathbf{Div} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] + \rho\mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^e = 2\rho \frac{\partial}{\partial \mathbf{C}} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^v = \frac{\partial}{\partial \dot{\mathbf{C}}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \rho c \frac{\partial \theta}{\partial t} - \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} - \rho \left( \frac{\partial \psi}{\partial \xi_i} - \theta \frac{\partial^2 \psi}{\partial \theta \partial \xi_i} \right) : \dot{\xi}_i - \mathbf{Div} \mathbf{Q} \\ \quad + \rho \mathbf{r} = 0 \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa\theta\nabla\theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{cases}$$

For the soft biological tissue and for the sample HEMA-EGDMA, we propose the thermodynamics potential in  $\Omega \subset (\mathbb{R}^d \times \mathbb{R}_+)$ .

$$\begin{cases} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) = \frac{\lambda}{2} \exp[\mu(\mathbb{I}_1 - 3)] - \frac{\lambda\mu}{4} (\mathbb{I}_2 - 3) - (\lambda + 2\mu)(\mathbb{I}_3 - 1) - (3\lambda + 2\mu)\kappa(\theta - \theta_0)\mathbb{I}_3 \\ \quad - \frac{c}{2\theta_0}(\theta - \theta_0)^2 \\ \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla\theta}, \overline{\nabla\xi_i}; \theta, \mathbf{C}) = \frac{\eta}{2} \mathbb{J}_2(\mathbb{I}_1 - 3) \\ \quad + \frac{1}{2} \kappa \|\overline{\nabla\theta}\|^2 + \mathfrak{i}(\xi_i) \end{cases}$$

Where  $\mathfrak{i}(\xi_i)$  denotes an indicative function, we assume that the Lamé's constants  $\lambda = \lambda(\xi_i)$  and  $\mu = \mu(\xi_i)$ .

We propose the governing equation of the internal variable  $\xi_i$ : We assume that we have one internal variable, and his evolution law is given by:

$$\begin{cases} \mathfrak{i}(\xi_i) + \frac{\partial \psi}{\partial \xi_i} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathfrak{i}(0) = \mathbf{I} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{cases}$$

#### 4. NUMERICAL METHODS

The numerical method used to solve the problem is based on finite element formulation (FEM). We use the software Comsol Multiphysics to solve the governing equation in  $\Omega \subset (\mathbb{R}^d \times \mathbb{R}_+)$ .

So we introduce the variational principle and the finite element discretization. For the two cases we consider respectively  $\delta \mathbf{u}$ ,  $\delta \theta$  and  $\delta \xi_i$  the virtual displacement, temperature and the internal variable.



- **Case 1:**

$$\left\{ \begin{array}{l} \mathbf{Div} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] + \rho \mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^e = 2\rho \frac{\partial}{\partial \mathbf{C}} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^v = \frac{\partial}{\partial \dot{\mathbf{C}}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla \theta}, \overline{\nabla \xi_i}; \theta, \mathbf{C}) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \left\{ \begin{array}{l} \mathbb{F}(\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = \mathbf{T}_M(\mathbf{n}, \mathbf{M}) \\ \mathbf{u} = 0 \\ \mathbb{F}(\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = 0 \\ \mathbb{F}(\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = 0 \end{array} \right. \\ \rho c \frac{\partial \theta}{\partial t} - \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} - \mathbf{Div} \mathbf{Q} + \rho \mathbf{r} = 0 \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa \theta \nabla \theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \left\{ \begin{array}{l} -\kappa \theta \nabla \theta \cdot \mathbf{n} = 0 \\ -\kappa \theta \nabla \theta \cdot \mathbf{n} = 0 \\ \theta = \theta_0 \\ -\kappa \theta \nabla \theta \cdot \mathbf{n} = k_c(\theta - \theta_f) \end{array} \right. \\ \mathbb{i}(\xi_i) + \frac{\partial \psi}{\partial \xi_i} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbb{i}(0) = \mathbf{I} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{array} \right.$$

Using the Galerkin's Method the weak formulation of t case 1 is given by:

$$\left\{ \begin{array}{l} \int_{\Omega} \mathbf{Div} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] \delta \mathbf{u} \, d\Omega + \int_{\Omega} \rho \mathbf{b} \delta \mathbf{u} \, d\Omega = \rho \int_{\Omega} \frac{\partial^2 \mathbf{u}}{\partial t^2} \delta \mathbf{u} \, d\Omega \\ \int_{\Omega} \rho c \frac{\partial \theta}{\partial t} \delta \theta \, d\Omega - \int_{\Omega} \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} \delta \theta \, d\Omega - \int_{\Omega} \mathbf{Div} \mathbf{Q} \delta \theta \, d\Omega \\ \quad + \int_{\Omega} \rho \mathbf{r} \delta \theta \, d\Omega = 0 \\ \int_{\Omega} \mathbb{i}(\xi_i) \delta \xi_i \, d\Omega + \int_{\Omega} \frac{\partial \psi}{\partial \xi_i} \delta \xi_i \, d\Omega = 0 \end{array} \right.$$

We have,

$$\left\{ \begin{array}{l} - \int_{\Omega} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] : \nabla \delta \mathbf{u} \, d\Omega + \int_{\partial \Omega} \hat{\sigma} \cdot \delta \mathbf{u} \cdot \mathbf{n} \, d(\partial \Omega) \\ \quad + \int_{\Omega} \rho \mathbf{b} \delta \mathbf{u} \, d\Omega = \rho \int_{\Omega} \frac{\partial^2 \mathbf{u}}{\partial t^2} \delta \mathbf{u} \, d\Omega \\ \int_{\Omega} \rho c \frac{\partial \theta}{\partial t} \delta \theta \, d\Omega - \int_{\Omega} \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} \delta \theta \, d\Omega - \int_{\Omega} \mathbf{Q} \cdot \nabla \delta \theta \, d\Omega \\ \quad - \int_{\partial \Omega} \mathbf{Q} \cdot \delta \theta \cdot \mathbf{n} \, d(\partial \Omega) + \int_{\Omega} \rho \mathbf{r} \delta \theta \, d\Omega = 0 \\ \int_{\Omega} \mathbb{i}(\xi_i) \delta \xi_i \, d\Omega + \int_{\Omega} \frac{\partial \psi}{\partial \xi_i} \delta \xi_i \, d\Omega = 0 \end{array} \right.$$

The weak formulation is computed with Comsol multiphysics of the problem and their constitutive law.

- **Case 2:**

$$\left\{ \begin{array}{l} \mathbf{Div} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] + \rho \mathbf{b} = \rho \frac{\partial^2 \mathbf{u}}{\partial t^2} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^e = 2\rho \frac{\partial}{\partial \mathbf{C}} \psi(\mathbb{I}_1(\mathbf{C}), \mathbb{I}_2(\mathbf{C}), \mathbb{I}_3(\mathbf{C}), \theta, \xi_i) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{S}^v = \frac{\partial}{\partial \dot{\mathbf{C}}} \chi(\mathbb{J}_1, \mathbb{J}_2, \mathbb{J}_3, \mathbb{J}_4, \mathbb{J}_5, \mathbb{J}_6, \mathbb{J}_7, \overline{\nabla \theta}, \overline{\nabla \xi_i}; \theta, \mathbf{C}) \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \left\{ \begin{array}{l} \mathbb{F}(\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = \mathbf{T}_M(\mathbf{n}, \mathbf{M}) \\ \mathbf{u} = 0 \\ \mathbb{F}(\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = 0 \\ \mathbb{F}(\mathbf{S}^e + \mathbf{S}^v) \cdot \mathbf{n} = 0 \end{array} \right. \\ \rho c \frac{\partial \theta}{\partial t} - \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} - \rho \left( \frac{\partial \psi}{\partial \xi_i} - \theta \frac{\partial^2 \psi}{\partial \theta \partial \xi_i} \right) : \dot{\xi}_i - \mathbf{Div} \mathbf{Q} \\ \quad + \rho \mathbf{r} = 0 \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbf{Q} = -\kappa \theta \nabla \theta \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \left\{ \begin{array}{l} -\kappa \theta \nabla \theta \cdot \mathbf{n} = 0 \\ -\kappa \theta \nabla \theta \cdot \mathbf{n} = 0 \\ \theta = \theta_0 \\ -\kappa \theta \nabla \theta \cdot \mathbf{n} = k_c(\theta - \theta_f) \end{array} \right. \\ \mathbb{i}(\xi_i) + \frac{\partial \psi}{\partial \xi_i} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \\ \mathbb{i}(0) = \mathbf{I} \text{ in } \Omega \subset (\mathbb{R}^d \times \mathbb{R}_+) \end{array} \right.$$

The weak formulation can be written as:

$$\begin{aligned} & \int_{\Omega} \mathbf{Div} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] \delta \mathbf{u} \, d\Omega + \int_{\Omega} \rho \mathbf{b} \delta \mathbf{u} \, d\Omega = \rho \int_{\Omega} \frac{\partial^2 \mathbf{u}}{\partial t^2} \delta \mathbf{u} \, d\Omega \\ & \int_{\Omega} \rho c \frac{\partial \theta}{\partial t} \delta \theta \, d\Omega - \int_{\Omega} \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} \delta \theta \, d\Omega - \int_{\Omega} \mathbf{Div} \mathbf{Q} \delta \theta \, d\Omega \\ & \quad - \rho \int_{\Omega} \left( \frac{\partial \psi}{\partial \xi_i} - \theta \frac{\partial^2 \psi}{\partial \theta \partial \xi_i} \right) : \dot{\xi}_i \delta \theta \, d\Omega + \int_{\Omega} \rho \mathbf{r} \delta \theta \, d\Omega = 0 \\ & \int_{\Omega} \mathbb{i}(\xi_i) \delta \xi_i \, d\Omega + \int_{\Omega} \frac{\partial \psi}{\partial \xi_i} \delta \xi_i \, d\Omega = 0 \end{aligned}$$

Then,

$$\begin{aligned} & - \int_{\Omega} [\mathbb{F}(\mathbf{S}^e + \mathbf{S}^v)] : \nabla \delta \mathbf{u} \, d\Omega + \int_{\Omega} \rho \mathbf{b} \delta \mathbf{u} \, d\Omega = \rho \int_{\Omega} \frac{\partial^2 \mathbf{u}}{\partial t^2} \delta \mathbf{u} \, d\Omega \\ & \int_{\Omega} \rho c \frac{\partial \theta}{\partial t} \delta \theta \, d\Omega - \int_{\Omega} \rho \theta \frac{\partial^2 \psi}{\partial \theta \partial \mathbf{C}} : \dot{\mathbf{C}} \delta \theta \, d\Omega - \int_{\Omega} \mathbf{Q} \cdot \nabla \delta \theta \, d\Omega \\ & \quad - \int_{\partial \Omega} \mathbf{Q} \cdot \delta \theta \cdot \mathbf{n} \, d(\partial \Omega) = 0 \\ & - \rho \int_{\Omega} \frac{\partial \psi}{\partial \xi_i} : \dot{\xi}_i \delta \theta \, d\Omega - \rho \int_{\Omega} \theta \frac{\partial^2 \psi}{\partial \theta \partial \xi_i} : \dot{\xi}_i \delta \theta \, d\Omega + \int_{\Omega} \rho \mathbf{r} \delta \theta \, d\Omega \\ & \quad + \int_{\Omega} \mathbb{i}(\xi_i) \delta \xi_i \, d\Omega + \int_{\Omega} \frac{\partial \psi}{\partial \xi_i} \delta \xi_i \, d\Omega = 0 \end{aligned}$$

The weak formulation is computed with Comsol multiphysics of the problem and their constitutive law.

## 5. SOME RESULTS AND DISCUSSIONS

We assume two cases of cyclic mechanical loads. For each test, we change the frequency of the load in order to study the material response. We particularly analyze the sensitivity of the material in low frequency. The application of the cyclic load is done in three steps. The

first step is the pre-loading, the second is the loading with the cyclic load and the third step is the unloading. For the first case we use the magnitude and the second case take into account of the magnitude change. Two cyclic mechanical loads were studied. For each test, we change the load frequency. Figures report entropy production during cyclic loading.

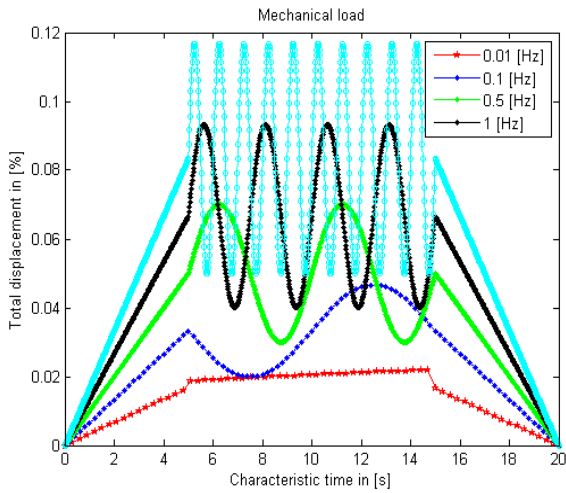


Figure 2: Mechanical load vs. time

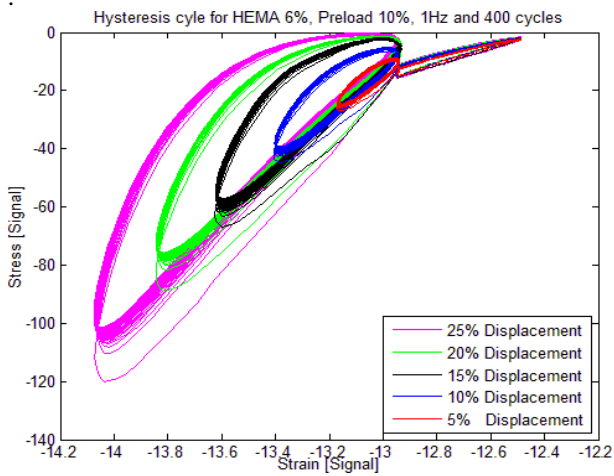


Figure 3: Hysteresis curves of the material

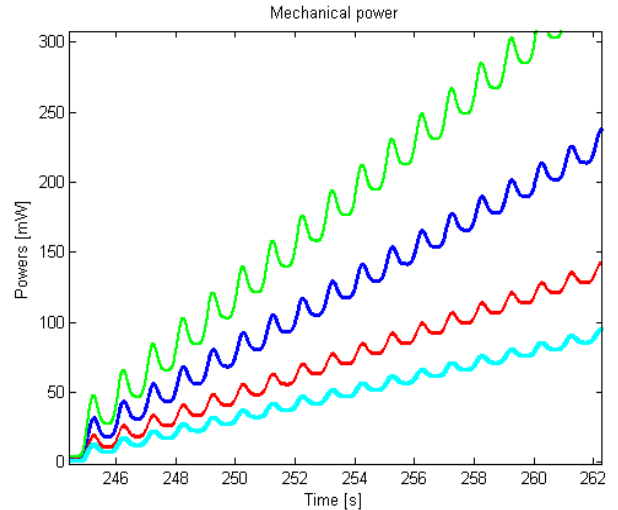


Figure 4: Mechanical Power vs. time

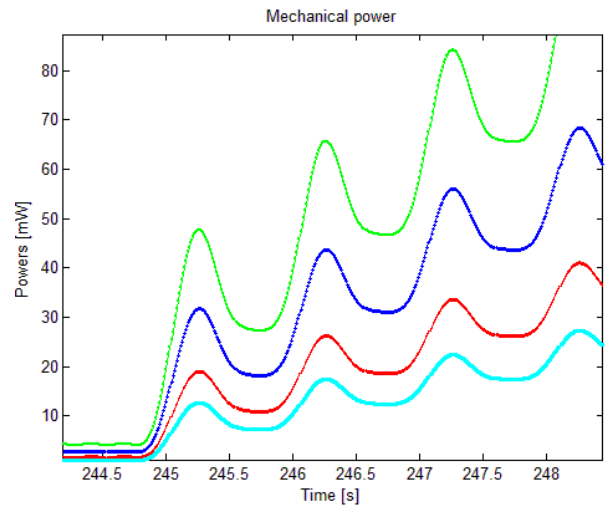


Figure 5: Mechanical Power vs. time, obtained by zooming the figure 4.in order to show the behavior in small time

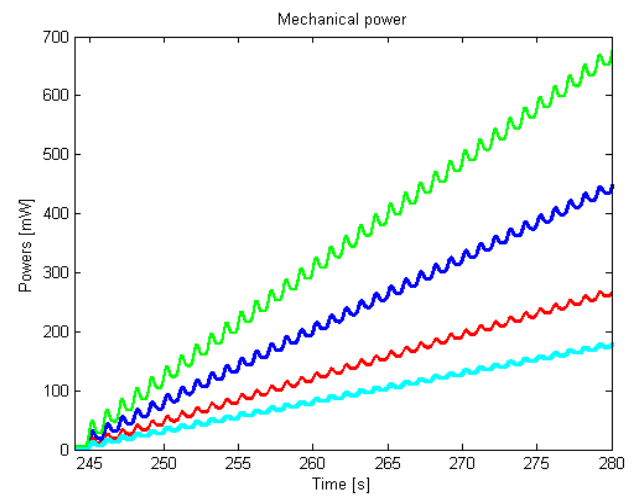


Figure 6: Mechanical Power vs. time

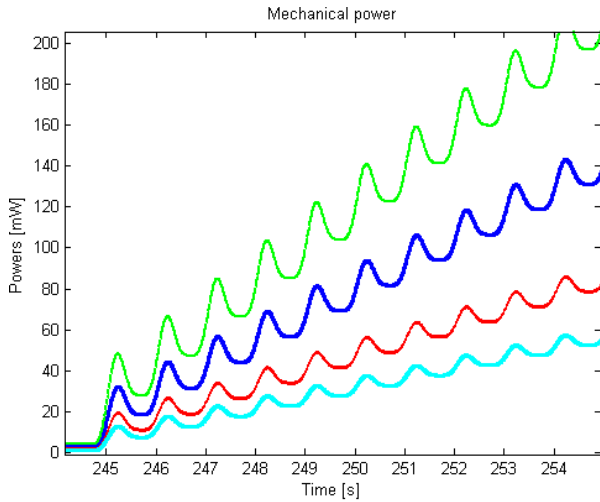


Figure 7: Mechanical Power vs. time, obtained by zooming the figure 6, in order to show the behavior in small time.

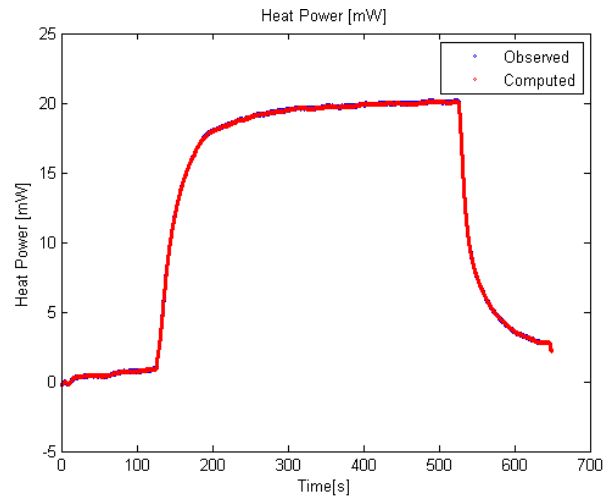


Figure 8: Mechanical Power vs. time

## 6. CONCLUDING REMARKS

Theoretical model was successfully developed in the present work to analyze the thermo-mechanical behavior of hydrogel matrix HEMA-EGDMA. New constitutive equation is proposed in thermo-viscoelasticity by assuming an adiabatic system.

Then, two new thermodynamics potentials and the energy conservation for this material were proposed. The thermo-mechanical of the coupled and transient system was solved by using simplified geometry. Fundamental and general trends can be clearly illustrated with validation with the help of the experimental measurement. New numerical approach to solve the coupled partial differential equation (governing equation) based on finite element method has been developed.

Reciprocal interaction between mechanical and thermal energy has been performed by this approach. This approach has revealed detailed information to qualify the temperature change and characterize the entropy production in the hydrogel matrix under cyclic loads. These results were shown that the heat effects to

the dissipation energy were significant and may exploited to control the drug delivery device driven by the temperature.

## ACKNOWLEDGMENTS

This work is financially supported by Cooperation between the Laboratory of Biomechanics Orthopedics of EPFL Lausanne Switzerland and Mathematical Research Institute of Rennes, University of Rennes, France.

## APPENDIX

### A- Notations

We introduce several notations which will be throughout this work. Let us:

$h_0$  : Henry constant

$\rho$  : Density

$\rho\mathbf{b}$  : Volumic force

$\mathbf{u}$  : Displacement

$\mathbf{r}$  : Heat source

$\mathbf{P}$  : Piola-Kirchhoff's first stress tensor

$\mathbb{F}$  : Strain gradient

$\mathbf{S}$  : Piola-Kirchhoff's second stress tensor

$\mathbf{S}^e$  : Elastic Piola-Kirchhoff's second stress tensor

$\mathbf{S}^v$  : Viscous Piola-Kirchhoff's second stress tensor

$\mathbf{E}$  : Green-Lagrange's strain tensor

$\mathbf{C}$  : Cauchy-Green's strain tensor

$\mathbf{I}$  : Identity matrix

$s$  : Entropy

$\rho$  : Density

$e$  : Internal energy

$\sigma$  : Cauchy's stress tensor

$\theta$  : Temperature

$\theta_f$  : Reference temperature

$\theta_0$  : Prescribed temperature

$k_c$  : Convection coefficient

$\psi$  : Helmholtz free energy

$\chi$  : Dissipation potential

$\xi_i$  : Internal variable

$\kappa$  : Conduction coefficient of the hydrogel

$\mathbf{Q}^{rev}$  : Reversible heat

$\mathbf{Q}^{irr}$  : Irreversible heat

$\mathbf{n}$  : Outer unit normal vector

$\Omega_s$  : Polymer domain

### B- Operators

$\nabla$  : Divergence operator

$\nabla$  : Gradient operator

$\cdot$  : Scalar product

$\times$  : Vector product

$\otimes$  : Tensor product

$\dot{x}$  : Temporal derivative operator of the variable  $x$

$\det(\dots)$  : Determinant

$\text{Tr}(\dots)$  : Trace

### C- Tensor Invariant

$\mathbb{I}_1$  : First invariant of the stress tensor

$\mathbb{I}_2$  : Second invariant stress tensor

$\mathbb{I}_3$  : Third stress tensor

$$\begin{aligned} \mathbb{J}_1 &= \text{Tr}\dot{\mathbf{C}}, \\ \mathbb{J}_2 &= \text{Tr}\dot{\mathbf{C}}^2, \\ \mathbb{J}_3 &= \text{Tr}\dot{\mathbf{C}}^3, \\ \mathbb{J}_4 &= \text{Tr}(\mathbf{C}\dot{\mathbf{C}}), \\ \mathbb{J}_5 &= \text{Tr}(\mathbf{C}^2\dot{\mathbf{C}}), \\ \mathbb{J}_6 &= \text{Tr}(\mathbf{C}\dot{\mathbf{C}}^2), \\ \mathbb{J}_7 &= \text{Tr}(\mathbf{C}^2\dot{\mathbf{C}}^2). \end{aligned}$$

## REFERENCES

- Shin, E. S, S.J.K, 1999, A predictive measure of thermomechanical coupling in elasto-viscoplastic composites. *Composites Sciences and Technology*, 59, pp. 1023-1031.
- Mahaman, M.H.M, 2005, Entropy and temperature gradients thermomechanics: dissipation, heat conduction inequality and heat equation. *Journal Maths Pures Applied*, 59, pp. 1023-1031.
- Claire, C.M, Z.M, W Z, 2011, Thermomechanical coupling in shape memory alloys under cyclic loadings: Experimental analysis and constitutive modeling. *Journal Maths Pures Applied*, 27, pp. 1959-1980.
- Xia, X.C, J-Y.Y 2010, A nonlinear iteration method for solving a two-diemsional nonlinear coupled system of parabolic-hyperbolic equation. *Journal of Computational and Applied Mathematics*, 234, pp. 343-364.
- Yinghui, Y.Z, Z.T 2013, Global existence and asymptotic behavior of smooth solutions to a coupled parabolic-hyperbolic system. *Nonlinear Analysis:Real World Applications*, 14, pp. 465-482.
- Jeffrey, J.R, X.Z, E.Z, 2005, Polynomial decay for a hyperbolic-parabolic coupled system. *Journal Maths Pures Applied*, 84, pp. 407-470.
- Mohammed, M. S, S.J.K, 2009, An Eulerian-Lagrangian method for coupled parabolic-hyperbolic equations. *Applied Numerical Mathematics*, 59, pp. 754-768.
- Chi, C-C.H, J-J.L, C-Y.L, C-H.H, C-K.C 2001, A two-dimensional finite element model for frictional heating analysis of total hip prosthesis. *Materials Sciences and Engeneering*, 17, pp. 11-18.
- Rakotomanana, L.R,D.P, 2000, Non-linear viscoelastic laws for soft biological tissues. *European Journal A/Solids*, 19, pp. 749-759.
- Korgsaard, J.K, 1999, On the representation of symmetric tensor-valued isotropic functions. *International Journal of Engeneering Sciences*, 28(12), pp. 11331-1346.
- Holzappel, G.A.H, G.R, 1995, Fully coupled thermomechanical behavior of viscoelastic solids treated with finite elements. *Pergamon, International Journal of Engeneering Sciences*, 33(7), pp. 1037-1058.
- Hao, H.W, M.G, S.Z, 2007, Convergence to equilibrium for a parabolic-hyperbolic phase-field system with dynamical boundary condition. *Journal of Mathematical Analysis and Applications*, 329, pp. 948-976.
- Kunish, K.K, G.L, J.S, T.F, 2009, Optimal Control of the Coupled Systems of Partial Differential Equations. *International Series of Numerical Mathematics, Verlag*, 158, pp. 1017-1028.
- Philippe, P.A-S, M.N.M, R.S, D.T, D.P, 2014, Intrinsic viscoelasticity increases temperature in knee cartilage under physiological loading. *Journal of theMechanical Behavior of Biomedical Material*, 30, pp. 123-130.
- Mohandreza, M.N.M, V.K, A.V, D.P, 2013, Drug delivery based on hydrogel dissipative properties. *Journal of Biomaterials*.
- Oden, J.T.O, 1972, Finite Element of nonlinear continua. *McGraw-Hill, NewYork*.
- Zenisek, A.Z, 1990, Nonlinear Elliptic and Evolution Problems and Their Finite Element Approximation. *Academic Press London*.
- Betram, A.B, 2005, Elasticity and Plasticity of Large Deformations, An Introduction. *Springer Verlag*.
- Rakotomanana, L.R.R, 2009, Elements de dynamique des solides et structure indéformables. *Presse Polytechniques Univesitaires Romandes*.

## AUTHORS BIOGRAPHY

### N. Santatriniaina



Is a PhD student in Mathematics and applications under supervision of Pr L.Rakotomanana (Full professor at University of Rennes 1) and Pr. D. Pioletti (Full professor at EPFL of Lausanne, Suisse, Laboratory of Biomechanics

Orthopedics) at the Mathematical Research Institute of Rennes, University of Rennes 1, France.

# SIMULATION OPTIMIZATION FOR A VACCINE DISTRIBUTION STRATEGY AGAINST THE SPREAD OF A(H1N1) EPIDEMIC.

**Javier Lara de Paz<sup>(a)</sup>, Idalia Flores<sup>(b)</sup>**

<sup>(a)</sup>Engineering School

National Autonomous University of Mexico

<sup>(a)</sup>[javier\\_lara@unam.mx](mailto:javier_lara@unam.mx)

<sup>(b)</sup>Engineering School

National Autonomous University of Mexico

[idalia@unam.mx](mailto:idalia@unam.mx)

## ABSTRACT

An optimal strategy in order to contain the epidemiological risk of the A(H1N1) disease is analyzed. We scope two different epidemic propagation models, the epidemiological SIR model proposed by Kermack and McKendrick (Kermack and McKendrick 1927) and a small-world network model, assuming that the second one shows a better performance since it considers the interactions between agents and also presents a more realistically dynamic (Moore and Newman 2000) than with the SIR model. To verify our assumption, establishing the proper disease parameters, a simulation is done for different epidemiological scenarios with both models. Furthermore, the results of simulation runs are employed for an optimal containment model based on a vaccination or self-isolation strategy. Additionally, other simulation were performed in order to obtain an optimal policy using our previous results from the spreading models simulation, getting a better insight of the diseases behavior during an outbreak.

Keywords: complex network, epidemiology, spreading models, small world networks

## 1. INTRODUCTION

In 2009 the human flu A(H1N1) epidemics expanded rapidly all over the five continents, with a predominance of America by the current connectivity conditions, becoming a pandemic in just a few weeks. Public Health Institutions showed a lot of concern because the lack of knowledge about this new subtype of influenza virus, which showed a high spreading factor and transmissibility in a short elapse of time. Governments, the whole society and the World Health Organization (WHO) considered this matter as a high priority issue. This motivated to search strategies needed to design adequate containment policies at the proper moment of the outbreak in order to save, not only lives, but expenses as well. But before carrying out any analysis of cost-effectiveness, it is desirable to understand the dynamic of the epidemic propagation. It is necessary to understand the disease behavior as a mathematical model of the epidemic. We implemented spatial agent-based and real time situation simulations to provide pandemic risk assessments and also we

proposed optimal intervention and prevention strategies as well as to estimate costs due to the planned policy. The most common ways to control an outbreak are vaccination and isolation, each one represents a corresponding cost and benefit. An optimal intervention should consider these factors in the scheme of the epidemic containment policy, resulting on a “cost-effectiveness” analysis. The A(H1N1) disease is characterized by pre-symptomatic, infectious and recovery stages that makes detection and control cases more difficult (Kuperman et al. 2001). Therefore we explore some outgoing features of spreading dynamics utilizing the epidemiological SIR model in order to evaluate their parameters of contagious conditions, applying on a dynamically changing small-world-like network (Kuperman et al. 2001). Interactions within a population is studied by means of different kind of social networks, mathematically based on the structure of the population, which in our case presents three subpopulations that interact in proportion to their sizes. With these zero dimensional models, it has been possible to study the epidemic features, the asymptotic solution for the density of infected people, the effect on stochastic fluctuations on the modulation of an epidemic situation as well as the thresholds values. Another classical approach to the epidemic spread describes spatially extended subpopulations, such as elements on a lattice. We consider a population rarely falls into either of these categories, being neither well mixed nor lattices. Watts and Strogatz (Watts and Strogatz, 1998) introduced small world network analysis in order to study the dynamic of many social processes such as disease spreading, formation of public opinion, distribution of wealth, transmission of cultural traits, etc. In the case of epidemiological models, it has been shown that small world networks present a much faster epidemic propagation than other studied epidemic models, such as reaction-diffusion models, or discrete models based on regular lattices of a social network (Kuperman et al., 2001). In this paper we use the epidemiological SIR-Kermack-McKendrick model (KMK) to acquire the epidemic features which are well studied, later, with these parameters we simulate the spread by means of a small-world network which, as mentioned before, presents a much faster epidemic propagation. Therefore we will be able to compare our

results obtained with the KMK model. Understanding these patterns will help direct the allocation of scarce resources, thus facilitating appropriate community preparation; furthermore it will help to design an optimal containment strategy, which is the scope of this work. With this study we are able to examine the relative importance of vaccination and self-isolation, two common measures for controlling the spread of infectious diseases (Yarmand et al. 2010). Since our epidemic propagation models, either the epidemiological SIR model and the small world network model, require realistic parameters to estimate appropriate behaviors of the threshold and propagation of the disease, we focused on a specific mostly homogenous population from the Mexico's City metropolitan zone in the range of ages from 20 – 64 years, which represents numerically the biggest social group, moreover, it seems to be the most susceptible population among the other age groups. Besides, having more realistic data of the targeted population supports the validation of our results. Additionally, simulation results validation is also sustained on the corresponding literature (Córdova et al. 2010).

Once we have obtained results from both epidemic propagation models (epidemiological SIR model and small world network model), by means of simulating with different parameters, we focused on an optimal cost-effectiveness intervention. (Yarmand et al., 2010), proposed a linear model considering two control measures: vaccination and self-isolation, having as decision variables the proportion of the targeted population under vaccination intervention and the fraction of the population being self-isolated. We considered this optimal model as adequate to scope the effectiveness of either of both models inasmuch as it presents a very useful simplicity for comparing our results. Accordingly, thus we find that the objective function is the linear summation of the relative costs of vaccination and self-isolation by considering the relative marginal costs, associated with each of the decision variables. Constraints are related to the number of individuals under treatment and to the percentage of total infective individuals (attack rate). Hence the present work seeks a strategy for mitigating the severity of the A(H1N1) influenza pandemic based on the threshold number  $R_0$ .

## 2. LITERATURE REVIEW

In this section we present an overview of the literature pertaining to contributions related to the disease spreading through a social network, especially on a small work network. Also a review is done on references about epidemiological containment strategies.

D. Watts and Strogatz (Watts, Strogatz 1998) designed in 1998 a model of dynamic networks for collective phenomena called "Small world", which differences the homogeneous agents as it had been treated before in heterogeneous agents, whose interactions are random, producing an intermediate

between regular networks and completely disordered ones, being better its approximation for social networks, neuronal networks, electricity transmission networks, and even for diseases' spread networks. Their work was related with the model realized by L. Sattenspiel and C. Simon in 1987 (Sattenspiel and Simon, 1998), which defined mathematically the spread and prevalence of an infection inside a heterogeneous population, considering the possible interactions within the population. Their work is the main reference for the construction of the diseases spread. M. Kuperman and G. Abramson (Kuperman and Abramson, 2001) used in the year 2001 this approach to networks in order to build the spread of an epidemic, in like manner, Matt J. Keeling and Ken T. D. Eames (Keeling and Eames, 2001) proposed epidemiologic models related with networks. Meanwhile, Christopher Moore and M. E. J. Newman (Moore and Newman, 2000) presented a model that approached the spread of an epidemic using a model of percolation and networks like the "Small world". Alexei Vázquez (Vázquez, 2006) came up with a work that supported the idea of analyzing the topology of an infection spread based on the analysis of a network with homogeneous nodes with a correlation with the measures of the K network, concluding that the rate of contagion between an infected node and a susceptible node is proportional to the average measures of the K network, therefore suggesting that the spread of an epidemic can be researched with the model of the networks "Small world". Eduardo Cuestas, Mario Vilaró, and Pablo Serra (Cuestas et al., 2011) proposed a model of predictability of the spatial and temporal spread of the epidemic of the influenza H1N1 in Argentina with the method of percolation. The researchers Mauricio Canals and Andrea Canals (Canals and Canals, 2010) also designed a model based on a model of percolation for the epidemic of the influenza H1N1, and the obtained data would be compared geographically by the WHO (World Health Organization), indicating with this the validity of this kind of approach. Dionne M. Aleman, Theodor G. Wibisono, and Brian Schwartz (Aleman et al., 2011) made a simulation of the spread of a disease during a "pandemic" outbreak based on non-homogeneous agents. In regard to the works that have been realized in Mexico, the most notorious one is the one made by Córdova Villalobos, et al, (Córdova et al., 2010), in which a compilation of 42 experts of many fields in health sciences is done. These studies are about epidemics from several points of view, allowing the comprehension and planning against the influenza A (H1N1). At the National University of México (UNAM) there is a project that studies collective phenomena called FENOMECA, and in this project G. Cruz-Pacheco, L. Esteva, A. A. Minzonil, P. Panayotaros, and N. F. Smyth (Cruz-Pacheco 2005) proposed a mathematical model for the spread of an epidemic.

Thus once that the spread of an outbreak has been studied, many projects have been based on these



models and proposed strategies to contain it. Kleczkowski A., Olés K., Gudowska-Nowak E. and Giññigan C.A. (Kleczkowski et al., 2011), designed a strategy based on a small world network that would reduce the costs of controlling the spread. Meantime, Pinar Keskinocak, Ali EkiciJulie, and L. Swann (Keskinocak et al., 2013) designed a plan for food distribution during an influenza spread. Sean Carr and Stephen Roberts (Carr and Roberts, 2010) developed a simulation to locate clinics and hospitals, and the distribution of resources during the spread of the epidemic of the influenza H1N1. Yarmand, Ivy, Roberts, Bengtson and Bengtson (Yarmand et al., 2010) studied and analyzed the cost vs. the effectiveness between vaccination and isolation related to the epidemic of the H1N1. Another research about the distribution and optimization of vaccines during an outbreak of H1N1, was made by Yarmand, Julie S. Ivy, Brian Denton, Alun L. and Lloyd (Yarmand et al. 2014), this research considered the optimal location of the vaccination in two phases in different geographical regions that are under uncertainty.

### 3. EPIDEMIC SPREADING MODELS

#### 3.1. SIR kinetic model

Traditionally the process of disease spreading relied on differential equations whose solutions describe the time propagation of the disease within uniformly mixed populations which means that all individuals in the population are equally likely, hence any infected person is equally likely to infect any other person. The spreading process itself is modeled using rate equations, describing population flows between epidemiological classes of individuals, such as susceptible (S), exposed (E), infective (I), and recovered (R). The simplest of these types of models is the widely-utilized Susceptible-Infected-Removed (SIR) model, in which susceptible individuals (S) may become infected (I) and continue to infect others until finally removed (R) from the population due to recovery, death, or self-containment. This compartmental model, however, doesn't consider heterogeneity nor the spatially or the contact patterns between individuals effects, but only that an infected agent transmits the disease to a susceptible agent with a time rate  $t$ . The total population is given by  $S(t)+I(t)+R(t) = N(t)$ . This model assumes a closed population, meaning there are no births, deaths or travel into or out of the population. Between S and I stages, the transition rate is  $\beta I$ , where  $\beta$  is the contact rate, which takes into account the probability of getting the disease in a contact between a susceptible and an infectious subject. Stages I and R interact with the transition rate of recovery or death  $\gamma$ . If the duration of the infection is denoted  $\tau$ , then  $\gamma = 1/\tau$ , since an individual experiences one recovery in  $\tau$  units of time. The SIR system described above can be expressed by the following set of ordinary differential equations:

$$\frac{dS}{dt} = -\beta IS$$

$$\frac{dI}{dt} = \beta IS - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

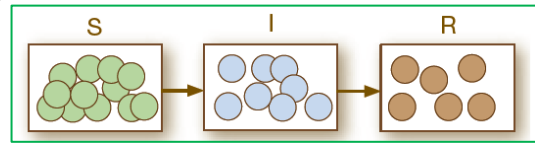


fig.1 SIR compartmental model

A key term in the study of epidemics is the basic reproduction number  $R_0$  (also called basic reproduction ratio). This ratio is derived as the expected number of new secondary infections from a single infection in a population where all subjects are susceptible. This quantity describes the epidemic threshold such that when  $R_0 > 1$ , the population is vulnerable to a large scale epidemic, although not guaranteed to experience one. Conversely when  $R_0 < 1$ , the population only experiences small local outbreaks. Hence  $R_0$  serves as an order parameter for the phase transition. In the SIR model the  $R_0$  number is calculated by:

$$R_0 = \beta \frac{N}{\gamma} = N * \frac{\ln\left(\frac{S(0)}{S(t)}\right)}{(N - S(t))}$$

For setting up the parameters for the simulation,  $R_0$  was set to  $R_0=1.4$ .

#### 3.2 SIR model on dynamic Small-World network

As mentioned in the previous sections a population rarely shows a well- mixed behavior, nonetheless a population shows spatiality between individuals, heterogeneity. Therefore an outlook on social contacts arises for describing the spread of a disease. Instead of relying on this kind of mean-field models we used an epidemic propagation model based on the effects of contacts patterns between individuals, described by contact networks, where the vertices correspond to individuals and the edges to contacts between them (Chen, 2005), resulting on much insight in the context of spreading processes obtaining a more realistic network topologies such as small-world networks. For developing the model we capture some outlining features of the A(H1N1) epidemic spreading dynamics by utilizing the SIR mechanism on a dynamically changing small-world contact network (Saramäki and Kaski, 2005). For a formulation of our spreading model, we consider a social network, such as various networks, displays the small world property, which means that long-range contacts between individuals result in short average distances along the edges of the network. In epidemiology these long-range contacts can be considered either infrequent contacts or random encounters taking place in an underlying regular short-range network structure, which in turn can be interpreted as groups of people having regular or frequent contacts. Therefore, we define our network as a regular one-dimensional ring-shaped lattice of  $N$



vertices with fixed coordination number  $2z = 4$ , with additional temporary long-range links, a clustering coefficient  $C(k)$  between  $0 < C(k) < 1$ , and changing their configuration at random with a rewiring probability  $P$ . The rewiring mechanism is equivalent to link exchanging method [13], at each time step, we randomly select a pair of edges A–B and C–D. These two edges are then rewired to be A–D and B–C. To prevent multiple edges connecting the same pair of nodes, if A–D or B–C already exists in the network, this step is aborted and a new pair of edges is randomly selected. (Figure 2).

For the spread dynamic we use the SIR mechanism, such as any node is labeled as susceptible S, or as infected I or as recovered R. Initially, the number of susceptible individuals is  $N - I$ , where I corresponds to the initial outbreak size. The dynamics of the model is such that at every discrete time step of the network with length  $\Delta t$ , it is randomly rewired and each infected individual in the network infects its new nearest neighbors, if susceptible, so that each infection occurs with an attack rate transmission (ART) determined by the own features of the A(H1N1). We established this ART = 5% and 10%. Once a vertex becomes infected, after some virus check frequency it becomes recovered R with a random gain resistance chance. At this stage the individual recovers and can no longer be infected or infect others (the individual is resistant to the virus). The process ends until any individual is no longer infected.

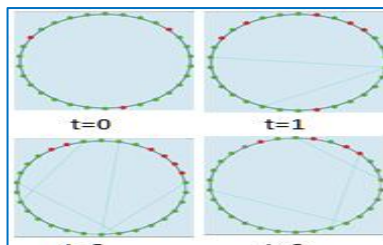


Figure 2: Schematic of epidemic spreading on a dynamic small-world contact network with coordination number  $2z = 4$ . At times  $t = 0, 2, 3, 4$  three vertex are infected (solid red circle). Then the infection spreads to neighboring vertices as well as randomly chosen far-off vertices with a rewiring probability  $P$ . Recovered vertices are shown as solid grey circles.

### 3.3. Optimization model

Additionally to the propagation model we needed to prepare an optimization model in order to obtain the optimal strategy based on a cost-effectiveness scope (Ferguson et al. 2005). As mentioned above, a linear model was used being decision variables vaccination fraction and self-isolation fraction among infected population denoted by  $p_1$  and  $p_2$  respectively. Constrains are the number of individuals under treatment, percentage of total infective individuals, and the threshold number  $R_0$ , those data were previously estimated by means of the epidemic spread model simulation. And the objective function is related to

relative costs of the considered interventions,  $c_1$  for vaccination and  $c_2$  for self-isolation, obtaining an optimization model as the following:

$$\min Z = c_1 p_1 + c_2 p_2$$

**Subject to:**

*Max number of people under treatment (related to surge capacity)  $\leq V1$*

*% of infected people among the population  $\leq V2$*

*Susceptible  $\leq R$*

Since  $c_1$  and  $c_2$  are relative costs, then we have

$$c_1 + c_2 = 1$$

$$0 \leq V1 \leq \text{total population}$$

$$0 \leq V2 \leq 1$$

$$R = (1 - \% \text{ susceptible population}) * R_0$$

Where  $V1$ ,  $V2$ ,  $c_1$  and  $c_2$ , are determined by the modeler based on our results of the prior epidemic models simulation. We also assumed costs  $c_1$  and  $c_2$  being both 0.5 (Yarmand, 2010).

## 4. SIMULATION

As mentioned, we prepared two simulation models for the spread of the A(N1H1) epidemics, one using the KMK epidemiological SIR model (Kuperman and Abramson 2005; Kermack and McKendrick, 1927) and the other uses the SIR model mechanical on the small-world network dynamic, as in the Watts and Strogatz model (Moore and Newmann 2000; Kuperman and Abramson 2005; Saramäki J, Kaski 2005; Xiao et al. 2003). Once we established the parameters, NetLogo 5.0.4 software was used to perform the simulation for different scenarios of propagations disease. This software was employed, mainly, because of its agent-based framework, which was of great relevance for the scope of this research, also the versatility of this software was proven, and given that it contains several libraries in which it is possible to develop different kind of models with remarkable simplicity without loss of generality. Besides, it has not been enough employed in the epidemic spread simulation field. Thereby we were able to perform our simulation, to obtain the needed scenarios for a later searching of the optimal epidemical mitigation strategy. The KMK virus spread model included in the model libraries was used for the search of features for the A(H1N1) epidemic. As mentioned, the NetLogo agent-based framework was useful to develop our small-world network virus spread model.

For the optimization model Lingo software was employed to optimize the simulation results.

### 4.2. Model parameters

A simulation was made considering the A(H1N1) disease parameters based on the WHO assumptions for our targeted Mexico city's population and on the gathered data during the 2009 outbreak. Definitions and values of the corresponding parameters needed for the KMK disease spread model dynamic as well as for the small-world network model are depicted in Table 1

Table 1: Definitions and values of parameters for KMK model

| Parameters                      | Definition                                                                                         | Estimated value                                                                              |
|---------------------------------|----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| $1/\gamma$                      | Infectious period                                                                                  | 1.5 days                                                                                     |
| $1/\kappa$                      | Latency period                                                                                     | 1.9 days                                                                                     |
| Average recovery time           | Total time since the contagious moment                                                             | 3.4 days (84 hours)                                                                          |
| Recovery chance                 | Inverse of mortality rate                                                                          | 98%                                                                                          |
| Attack Rate<br>Transmission ART | Is the infection chance                                                                            | 7.5%                                                                                         |
| Transmission risk               | Corresponds to the infection chance. This parameter affects directly the transmission rate $\beta$ | is adjusted during the simulation process to acquire the corresponding $R_0$ threshold value |

### 4.3. Simulation runs

Once we established the parameters, NetLogo 5.0.4 software was used to perform the simulation for different scenarios of propagations disease. We used the NetLogo epiDEM-Basic model which is structured on the KMK model (Yang and Wilensky, 2011). We performed the simulation for an ART= 5%, and a population size of 100 people with the following adjustments: (see Figure 3.). In order to average the  $R_0$ 's outcomes, 30 replications were run.



Figure 3. Adjustments for the KMK model to

With the previously  $R_0$  distribution with KMK model results we setup our small-world network model to fit with these results. (Figure 4)

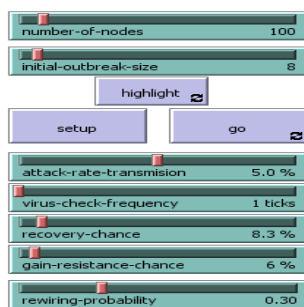


Figure 4. Adjustment controls for the Small-world network model

## 5. RESULTS AND ANALYSIS

The results obtained from our epidemic spread models show a scope of the different scenarios during the A(H1N1) epidemic outbreak using the disease parameters and focused on a specific population. For adjusting the  $R_0$  threshold 30 experiments were performed with the established parameters.

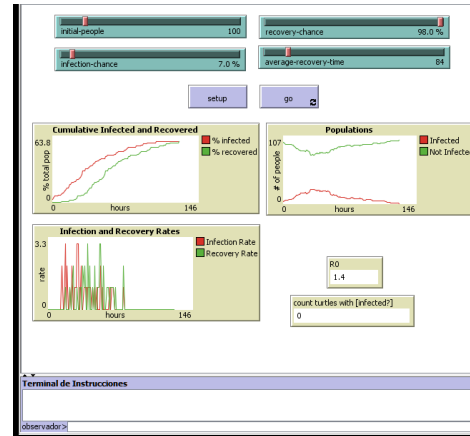


Figure 5. NetLogo Interface simulating KMK SIR model

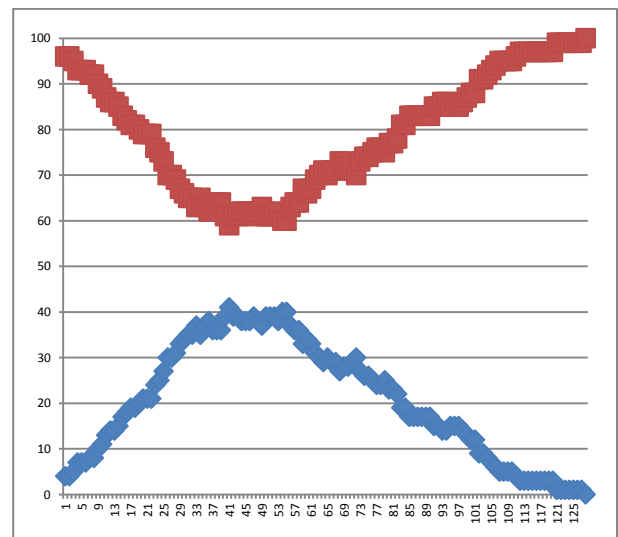


Figure 6. Averaged dynamic of KMK model with ART 5% and  $Rho=1.4$ . Blue series corresponds to the averaged proportion of infected people. Red series corresponds to averaged not infected people. The averaged dynamics last about 127 hours.

Next we performed our SIR adapted to a small-world network model with a population of 100 nodes, an outbreak size of 5 and an ART of 5%.

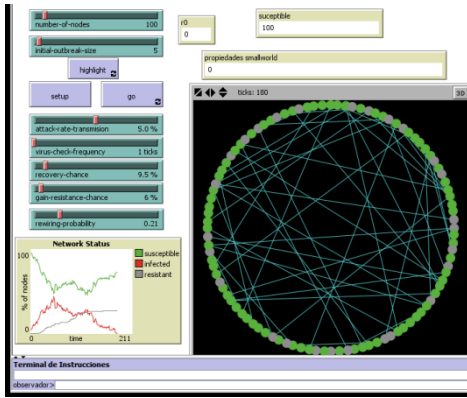


Figure 6. NetLogo Interface simulating small-world network model

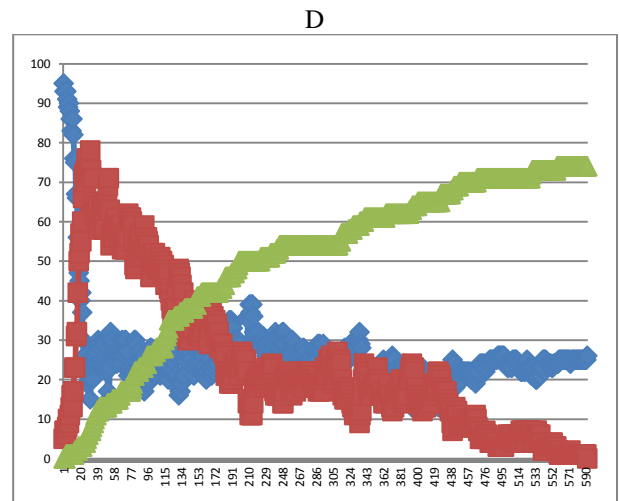
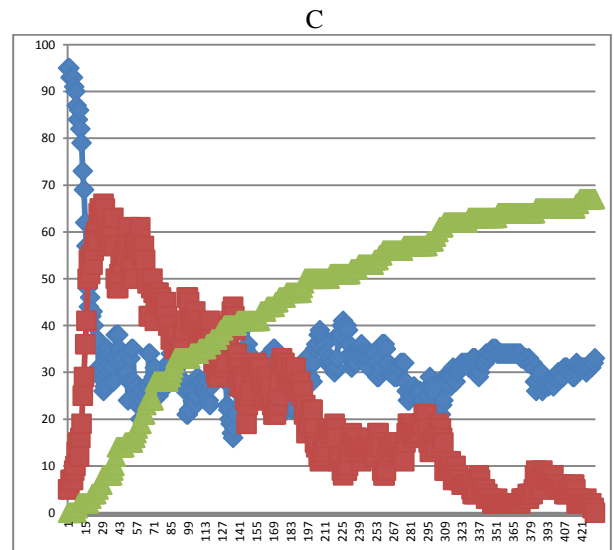
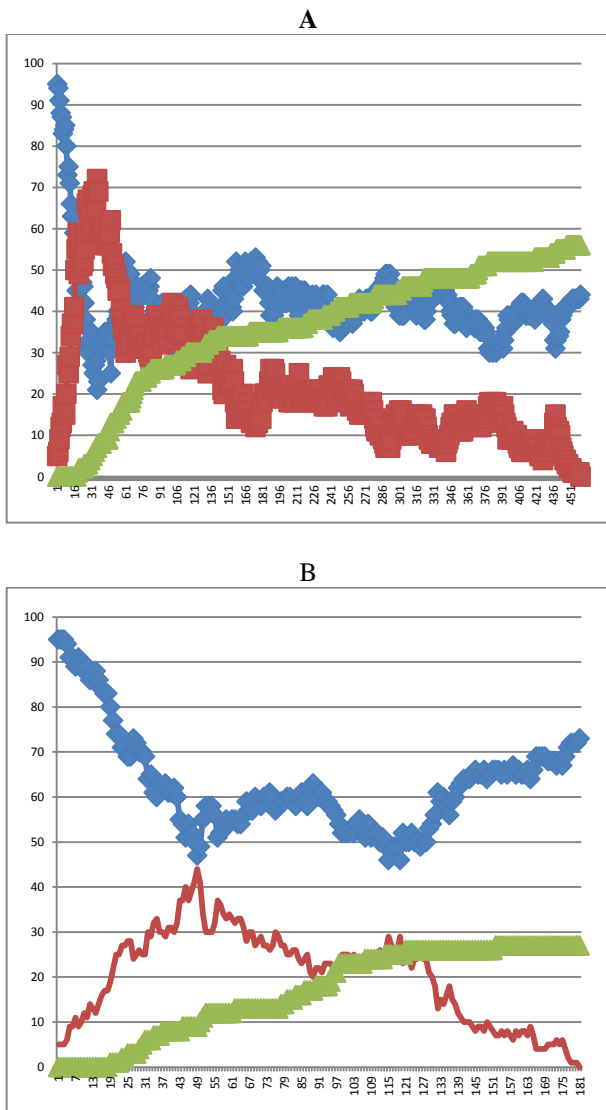


Figure 7. Plots of averaged dynamic. 30 experiments were performed for each rewiring probability. Red series corresponds to the dynamic of infected nodes. Blue series corresponds to susceptible. Plot A corresponds to a rewiring probability  $P=0.5$  with  $Rho=1.4$ . Plot B corresponds to a rewiring probability  $P=0.3$  with  $Rho=1.25$ . Plot C corresponds to a rewiring probability  $P=0.6$  with  $Rho=1.57$ . Plot D corresponds to a rewiring probability  $P=0.7$  with  $Rho=1.75$

After our parameterization of the simulation, several runs were performed in order to obtain the corresponding A(H1N1)  $Rho$  threshold of 1.4. While with the heterogeneous mean-field modeling many parameters needed to be set to acquire the  $Rho$ , also the dynamical behavior of the spread didn't show a regularity in the outcomes, as much as about 30 experiments should be developed in order to average  $Rho=1.4 (\pm 0.2)$  (Córdova et al, 2010).

Conversely, the small-world network model to keep the averaged  $Rho$  at the range of 1.4, just the rewiring probability should be adjusted, obtaining a suitable outcomes at the range of  $P=0.5$  with less than 10 repetitions and with an error of less than 0.07, also

the disease parameters were maintained at its initial values established as mentioned above.

Once we have a good method to estimate the threshold number  $Rho$ , it is possible to design a containment strategy with an optimization model considering that  $Rho$  depends on the susceptible population. Fraser (Fraser, 2007) recommends that to develop a vaccine distribution strategy, the susceptible population for the A(H1N1) outbreak occurred in 2009 should be of 70%, which corresponds to a  $Rho = 1.4$ .

Some remarkable results emerge from the corresponding plots; When the rewiring probability increases, outbreak duration lasts longer, as well as the threshold  $Rho$  increases. About the behavior of the population stages, one can notice that the susceptible population behaves asymptotically proportional to the rewiring probability.

Comparing with the literature, it has a proper fit with the with real data obtained from the A(H1N1) outbreak occurred from March to September 2009. (Cordova et. al., 2010)

## 6. CONCLUSIONS

Significant difference should be appreciated for both spread models, thus the “small-world” network model shows a better performance and a more accurate estimation than SIR KMK model. Although the small-world model doesn’t consider the disease features, it shows a good performance in the epidemic dynamics. The present work pretends to be a good issue in order to collaborate with a better estimation of the dynamics during a new outbreak of A(H1N1); also to generate more realistic scenarios in order to develop a more accurate proper planning on containment strategies. We suggest for a future research to use the simulation of our model, to build an optimization model with the objective the amount or fraction of the population should be vaccinated and how many should be self-isolated and especially in any time.

## REFERENCES

Moore C. and Newman M. E. J., Phys. Rev. E 61, 5678(2000)

Kuperman G., Abramson, Small-world effect in an epidemiological model, PhysRevLett.86.2909. 2001,

Yarmand, Ivy, Roberts, Bengtson and Bengtson, , COST-EFFECTIVENESS ANALYSIS OF VACCINATION AND SELF-ISOLATION IN CASE OF H1N1, *Proceedings of the 2010 Winter Simulation Conference*. 2010

Yarmand, Julie S. Ivy, Brian Denton, Alun L. Lloyd. “Optimal two-phase vaccine allocation to geographically different regions under uncertainty”, European Journal of Operational Research, Volume 233, Issue 1, 16 February 2014, Pages 208-219.

Cordova, Valdespino and Ponce de Leon, La Epidemia de Influenza A/H1N1 en México, *Ed. Médica Panamericana*,. ISBN 978-607-7743-12-5, 2010.

F. Ball, D. Mollison, and G. Scalia-Tomba, 1997, *Ann. Appl. Prob.* 7, 46

Saramäki J, Kaski K, Modeling development of epidemics with dynamic small-world networks, *Journal of Theoretical Biology* 234 (3), 413-421, 69, 2005.

Ferguson NM, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437:209–214. 2005.

Ferguson NM, et al. (2006) Strategies for mitigating an influenza pandemic. *Nature* 442:448–452.

D. Watts and S. Strogatz. "Collective dynamics of 'small world' networks", *Nature* 393, 440-442 (1998)

KermackWO, McKendrick AG. Contributions to mathematical theory of epidemics-I. 1927. *Bull Math Biol* 1991;53(1-2):33-55.

Murray, J. D., *Mathematical Biology, 3rd Edition*, Springer, New York, 2001.

Erdős, P.; Rényi, A. On random graphs. *Publ. Math. Debrecen* 6, 1959, 290-297.

Chen Jiahao, Epidemics on small-world networks. Unpublished term paper for PHYS, 2005.

Santos F C, Rodrigues J F and Pacheco J M 2005 *Phys. Rev. E* 72 056128.

Fraser, C., C. A. Donnelly, S. Cauchemez, and et al. 2009. Pandemic potential of a strain of influenza A(H1N1): early findings. *Science* 324:1557-1561.

Zanette D. H., Critical behavior of propagation on small-world networks. *Phys. Rev. E* 64, 050901(R) – Published 24 October 2001.

Jones G. T., A NetLogo model for the Study of the Evolution of Cooperation in Social Networks, North American Association for Computational Social and Organizational Science, Atlanta, GA (2007).

Kleczkowski A, Oleś K, Gudowska-Nowak E, Gilligan CA. “Searching for the most cost-effective strategy for controlling epidemics spreading on regular and small-world networks”, *J R Soc Interface*. 2012 Jan 7;9(66):158-69. doi: 10.1098/rsif.2011.0216. Epub 2011 Jun 8.

Cruz-Pacheco, Esteva L., Minzoni A.A., Panayotaros P., Smyth N.F., “Modelación matemática de la epidemia”, proyecto FENOMECE, National Autonomous University of Mexico (UNAM), 2005.

Linyuan Lü, Duan-Bing Chen, Tao Zhou, The small world yields the most effective information spreading, *New Journal of Physics* 13 (2011) 123005.

Sattenspiel, L. & Simon, C. P. “The spread and persistence of infectious diseases in structured populations”. *Math. Biosci.* 90, 341–366 (1988).

Xiao Fan Wang, Guanrong Chen, *Complex Networks: Small-World, Scale-Free and Beyond*, IEEE circuits and systems magazine, First Quarter 2003.

Vazquez A., Spreading dynamics on small-world networks with connectivity fluctuations and correlations, *Physical Review E* 74, 056101 2006.

Cuestas E, Vilaró M, Serra P. “Predictibilidad de la propagación espacial y temporal de la epidemia de influenza A/H1N1 en la Argentina por el método de percolación”, *Revista Argentina de Microbiología* (2011) 43: 186-190.

Cruz-Pacheco, L. Esteva, A.A. Minzoni<sup>1</sup>, P. Panayotaros, N.F. Smyth, "Modelación matemática de la epidemia", proyecto FENOMECC, UNAM

Aleman D., Wibisono T. G., Schwartz B., "A Nonhomogeneous Agent-Based Simulation Approach to Modeling the Spread of Disease in a Pandemic Outbreak", *Interfaces*. Volume 41 Issue 3, May-June 2011

Canals M., Canals A., "Percolación de la epidemia de influenza H1N1 en el mundo: Utilidad de los modelos predictivos basados en la conectividad espacial", *Rev. Médica de Chile*; 2010, 138: 573-580.

Carr S., Roberts S., "Planning for infectious disease outbreaks: a geographic disease spread, clinic location, and resource allocation simulation", *Proceedings of the 2010 Winter Simulation Conference*.

Keskinocak P., EkiciJulie A., Swann L., "Modeling Influenza Pandemic and Planning Food Distribution", *Manufacturing & Service Operations Management*, Articles in Advance, Received: June 16, 2009 Accepted: June 18, 2013, Published Online: October 31, 2013.

Yang, C. and Wilensky, U. (2011). NetLogo epiDEM Basic model. <http://ccl.northwestern.edu/netlogo/models/epiDEMBasic>. Center for Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL.

Wilensky, U. (1999). NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL.

## ACKNOWLEDGMENTS

This research was supported by UNAM PAPIIT grant IN116012. The first author specially acknowledges the support from CONACYT scholarship program.

## AUTHORS BIOGRAPHY

**Javier Lara de Paz** studied Physics at the Faculty of Science of National Autonomous University of Mexico (UNAM), then a Specialization in Applied Statistics the UNAM. He has been professor of Science at the Technological Institute of Gustavo. A. Madero in Mexico City. Currently he is studying a Master in Operational Research at UNAM. She is also a Professor of the Specialization in Quality Systems at UNAM.

**Idalia Flores de la Mota** She received her Ph.D. in Operations Research at the Faculty of Engineering of the UNAM. She graduated Master with honors and received the Gabino Barreda Medal for the best average of her generation. She has been a referee and a member of various Academic Committees at CONACYT. She has been a referee for journals such as *Journal of Applied Research and Technology*, the *Center of Applied Sciences and Technological Development*, *UNAM* and the *Transactions of the Society for*

*Modeling and Simulation International*. Her research interests are in simulation and optimization of production and service systems. She is a full time professor at the Postgraduate Program at UNAM.

# KINEMATICS AND DYNAMICS ANALYSIS OF DIFFERENT MECHANISMS USING THE SIMMECHANICS / SIMULINK / MATLAB

A.D. Perechesova <sup>(a)</sup>, I.I. Kalapyshina <sup>(b)</sup>, K.A. Nuzhdin <sup>(c)</sup>

<sup>(a)</sup> ITMO University, SPbF IZMIRAN

<sup>(b)</sup> ITMO University, National Research University SPbSPU

<sup>(c)</sup> ITMO University, FAM - robotics

<sup>(a)</sup> [perechesova@gmail.com](mailto:perechesova@gmail.com), <sup>(b)</sup> [weir521@yandex.ru](mailto:weir521@yandex.ru), <sup>(c)</sup> [nuzhdink@yandex.ru](mailto:nuzhdink@yandex.ru)

## ABSTRACT

The article shows the possibilities of modeling of spatial motions of a solid machines and a mechanisms using the library SimMechanics/Simulink/MATLAB. A two ways of kinematics analysis of the two-parametric antenna are shown: using the library SimMechanics and using the principles of matrix equations. Also work presents the comparison of the results. The article presents the SimMechanics model of tribometric system and the SimMechanics model of executing mechanism of the device for manufacturing torsion bars with Helical Anisotropy. The envelope of the cross-section of torsion bars is an ellipse with longer axis of 0.046 mm and shorter axis of 0.033 mm, the total length - 100 mm, the density of braiding -7 knots/mm.

Keywords: SimMechanics, MATLAB, modeling, two-parametric antenna, tribometric system, device for manufacturing torsion bars with Helical Anisotropy

## 1. INTRODUCTION

It is more effective to use modeling with the help of the SimMechanics library of the Simulink/MATLAB software for studying kinematics and dynamics of different mechanisms (Musalimov et al., 2013). Differential equations are presented in SimMechanics model diagram. Mechanical system is presented as associated block diagram. Blocks stand for models of mechanical joints, their position can change during cycle. SimMechanics translates our block to equivalent mathematical model. SimMechanics works with mechanical forces, but not with the signals.

This saves you the time and effort of developing the mathematical model yourself.

You can record simulation animations in Microsoft Audio Video Interleave® (AVI) format using the SimMechanics visualization. You can display signals generated during simulation and measure different parameters by the Scope blocks and the Sensor blocks.

## 2. OPERATION PRINCIPLES OF THE SIMMECHANICS BASIC BLOCKS

The vectorial method can be used to describe blocks of the library SimMechanics Joints part.

The Revolute block (fig. 1, a) represents a single rotational degrees of freedom about a specified axis between two bodies. A matrix of direction cosine like (1) can be used to describe block.

$$M_{ba} = \begin{pmatrix} \cos(\overset{\wedge}{x_b, x_a}) & \cos(\overset{\wedge}{x_b, y_a}) & \cos(\overset{\wedge}{x_b, z_a}) \\ \cos(\overset{\wedge}{y_b, x_a}) & \cos(\overset{\wedge}{y_b, y_a}) & \cos(\overset{\wedge}{y_b, z_a}) \\ \cos(\overset{\wedge}{z_b, x_a}) & \cos(\overset{\wedge}{z_b, y_a}) & \cos(\overset{\wedge}{z_b, z_a}) \end{pmatrix} \quad (1)$$

where  $[x_a, y_a, z_a]$  – initial coordinate system,  $[x_b, y_b, z_b]$  – rotated coordinate system.

The rotational sense is defined by the right-hand rule. The Revolute block on fig. 1 (a) represents a single rotational degrees of freedom about z axis. The matrix of direction cosine of this block determined by (2), witch correspond with fig. 2, a.

$$M_{ba}^z = \begin{pmatrix} \cos(\varphi_{ba}) & -\sin(\varphi_{ba}) & 0 \\ \sin(\varphi_{ba}) & \cos(\varphi_{ba}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

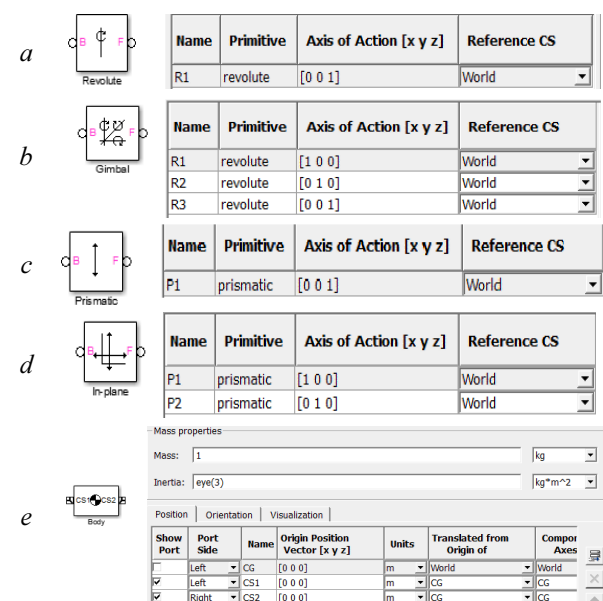


Figure 1. The Joints Blocks

The matrix of direction cosine  $M_{ba}$  must be multiplied by column vector of liner shift  $R_{ba}$  (3) for representation of coordinate systems relative displacement. This displacement depends on joints length  $L$ . The matrix  $R_{ba}$  is given at the Body block (Origin Position Vector) fig. 1 (e).

$$R_{ba} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (3)$$

The Gimbal block (fig.1 b, fig. 2 b) represents a composite joint with three rotational degrees of freedom as three revolute primitives. A matrix of direction cosine like (1) or its equivalent in matrix form (4) can be used to describe the Gimbal block and other blocks of Joints part.

$$M_{ba} = M_{ba}^x \cdot M_{ba}^y \cdot M_{ba}^z \quad (4)$$

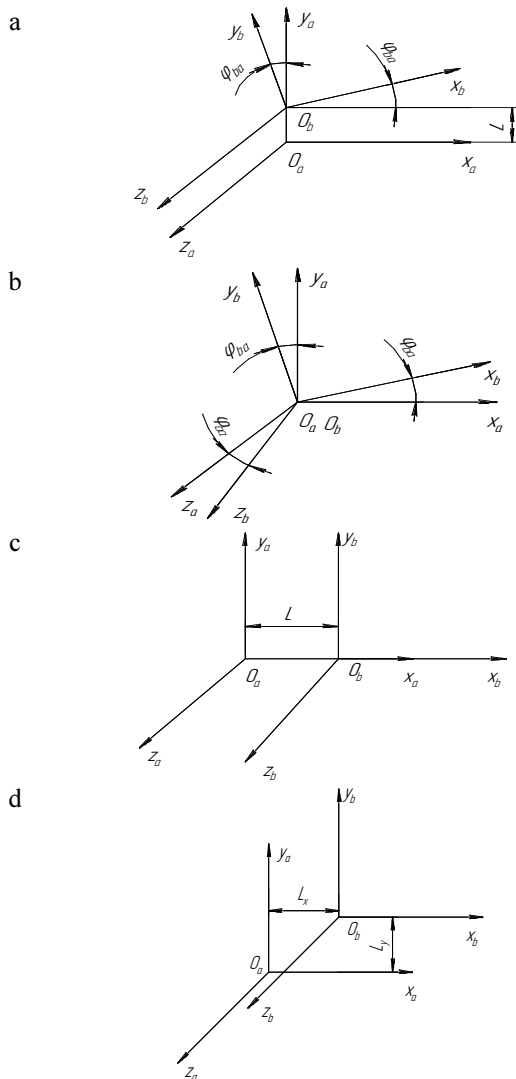


Figure 2. Projection transition of vector  $w$  from  $O_a$  to  $O_b$  coordinate system

A Prismatic block, In-plane block and other describes the kinematic pairs with the translational degree of freedom (fig. 1 c, d; fig. 2 c, d). A motion matrix like (5) can be used to describe blocks. The Prismatic block represents a single translational degrees of freedom along a specified axis between two bodies. The In-Plane block represents a composite joint with two translational degrees of freedom as two prismatic primitives.

$$F_{ba} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

A Body block is an important element when you are creating a mechanism model with the help of the SimMechanics library (fig.1, e). The Body block represents a rigid body with properties you customize. The representation that you specify must include: the body's mass and moment of inertia tensor, the coordinates for the body's center of gravity (CG), one or more Body coordinate systems (CSs). The representation can also include optional body geometry and color information for visualization.

A block Machine Environment set up mechanical environment for machine. A Ground block represents an immobile ground point at rest in the absolute inertial World reference frame [1].

The library SimMechanics designed for studying controlled motion of objects. We know the position and the orientation of the object at a given moment and for this position. This parameters are set in a fixed coordinate system  $O_a$  by coordinates  $x_a, y_a, z_a$  and Euler angles  $\phi_a, \Theta_a, \varphi_a$ . This parameters are using for determine the coupling matrix between  $O_a$  and the  $O_M$  is by. The  $O_M$  is coordinate system concerned with center  $M$  of working body of the mechanism (6).

$$A_M = M_M(x_a, y_a, z_a, \phi_a, \Theta_a, \varphi_a) \quad (6)$$

This relationship can be set through the mechanism local coordinate system.

$$A_M = M_{ab}(q_1)M_{bc}(q_2)\dots M_{(n-1)n}(q_n)M_{nM} = M_M(q_1, q_2, \dots, q_n) \quad (7)$$

When we equate the matrix, we obtain a system of equations for determination the position functions of the mechanism (8).

$$q_j = q_j(x_a, y_a, z_a, \phi_a, \Theta_a, \varphi_a), j = 1, \dots, n \quad (8)$$

We get 9 equations in the general case of the matrix comparison, but independent equations will be  $n$  (its depends of the number of generalized coordinates).

### 3. A TWO WAYS OF KINEMATICS ANALYSIS OF THE TWO-PARAMETRIC ANTENNA

The electromechanical module of the two-parametric antenna is the part of the onboard radar. This



radar working in the UHF waveband range (Zamoruev et al., 2012).

The kinematic configurations of the two-parametric antenna is shown on Fig. 3.

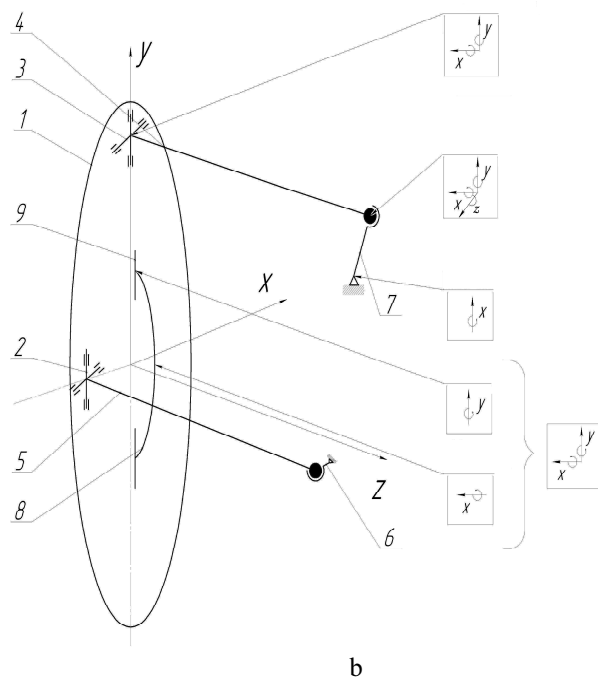
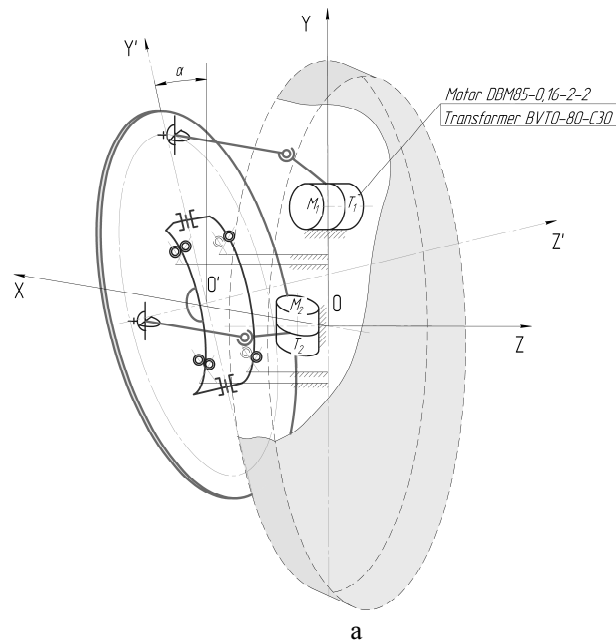


Figure 3. The kinematic configurations of the two-parametric antenna

1 – reflector; 2, 3 – biaxial cardans; 4, 7 and 5, 6 – system of levers are connected in pairs by spherical joint; 4, 5 – traction; 8 – supporting arc; 9 – axle

Reflector rotation about two axes  $OY$  and  $OZ$  is carried out through the hard arc. It is rigidly connected to the reflector. The Reflector has a sliding support that provide mobility of the reflector about the axis  $OY$ . The arc is installed in bearing unit, which consists of four rolling bearings pairs (axis  $OZ$ ).

Rotational motion from motors to reflector is made by con-rods. Each con-rod is a system of levers, which are connected by a joint. The drive lever combined with drive motor and the transformer in a single package (they are on the same shaft). Combining of these elements into a single module allows you to exclude from the system transmissions and shafts. Guided levers connected to the reflector using biaxial cardan.

Movement of the driven member of the reflector 1 (fig.3, b) includes two rotations around the axis  $X$  - axis and the elevation axis  $Y$  - azimuth. Movement is realized by two biaxial kardanov 2, 3 axis ( $Y$  and  $X$ ). The control action of the motor is transmitted to the reflector 1 via a lever system 4, 7, 5, 6 connected in pairs spherical joint has three rotational degrees of freedom. 3 and 4, universal joints provide for a degree of freedom of movement rods 4 and 5. Degrees of freedom for the reflector 1 is provided with respect to the axis  $Y$  axes 9 and the  $X$  axis supporting arc 8 (in fig.3 illustrates in phantom). To simulate the central portion has been converted into a block with two degrees of freedom - located at the geometric center of the reflector 1.

How to mechanics, is considered representative of the instrument devices with parallel kinematics, having two similar (parallel) drive lever. The device is composed of five units (fig. 3, b), gimbal (Grade 4), two - spherical joint (Class 3) and two - flat hinge (Class 5). Back - rotational motion is applied to the vertical and horizontal disposition of cranks. The SimMechanics block diagram of the two-parametric antenna is shown in Fig.4.

The SimMechanics block diagram of the two-parametric antenna is shown on Fig. 4. 1 – Machine Environment, Ground, Universal, Zerkalo3 blocks (corresponds to position 1 in Fig. 3); 2 – Zerkalo3, Universa2, Tjga1 blocks (corresponds to position 1, 3(2), 4(5) in Fig. 3); 3 –Tjga1, Spherical, Shatun1 blocks (corresponds to position 4, 7 (5, 6) in Fig. 3); 4 – Shatun1, Revolute, Ground1 blocks (corresponds to position 7 (6) in Fig. 3); 5 – Joing Actuator, Sine Weve blocks.

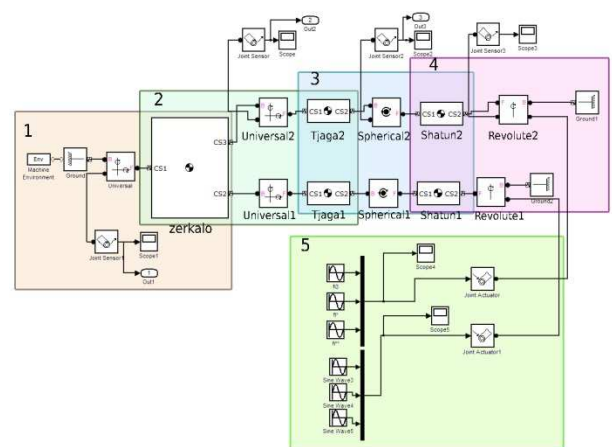
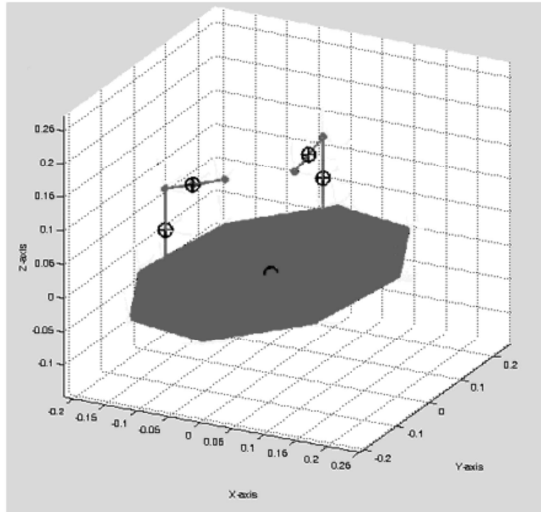
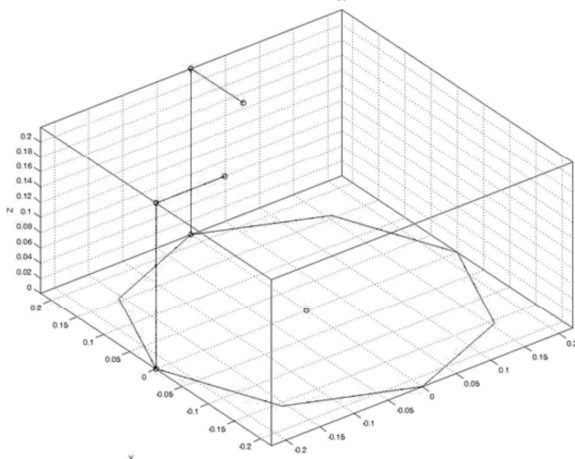


Figure 4. The SimMechanics block diagram of the two-parametric antenna

The SimMechanics animated model of the two-parametric antenna is shown on Fig. 5. Joint Sensor blocks from model (fig. 4) measures the position, velocity, and/or acceleration of a joint primitive in a Joint block. It also measures the reaction force and torque across the Joint.



a



b

Figure 5. The SimMechanics animated model of the two-parametric antenna (a), analytical solution (b)

Was written a system of 12 equations which determine the position of each element of the two-parameter antenna. Fig. 5 (b) shows the zero position of the mechanical system.

Following is a comparison of the analytical model and the SimMechanics model. The main parameter for comparison of the models was adopted the position of the reflector about the axes X and Y, and the input angle of con-rods. Angular positions are set to each rod from  $0^\circ$  to  $40^\circ$ .

Fig. 6 shows results of the comparison of the SimMechanics model and the analytical model: (a) the result of the SimMechanics model; (b) the result of the analytical model; (c) the average values of the position of the SimMechanics and the analytical models.

Figure 6 (a) shows that for small angles deviation of con-rod, there is an overlapping of graphs, which

characterize the deviation of the reflector in the X and Y (two degrees of freedom). The result of the SimMechanics model shows divergence of angles deflection of reflector becomes intensely increasing after  $17^\circ$ . The analytical model (Fig. 6, b) shows the same result. The maximum value of the results variation between the average values of data from the two models equal  $\Delta_{\max} \approx 3'$  (fig.6, c).

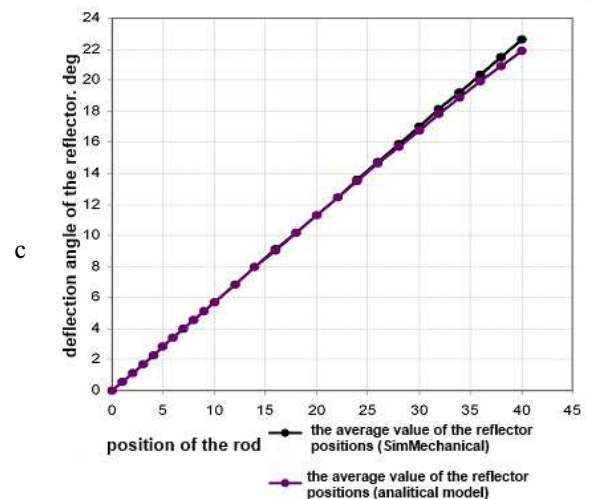
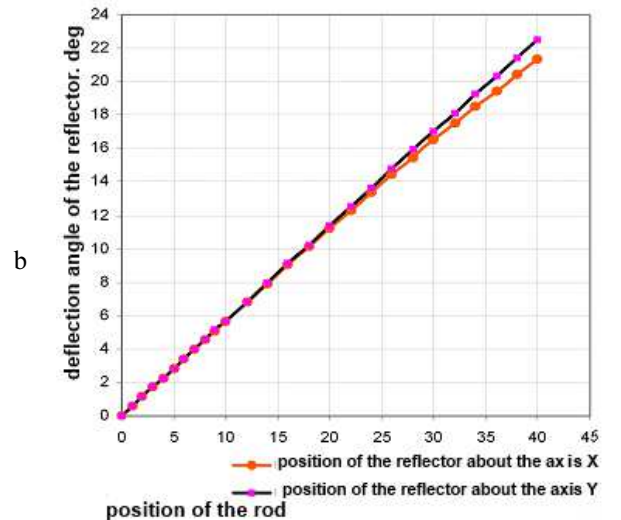
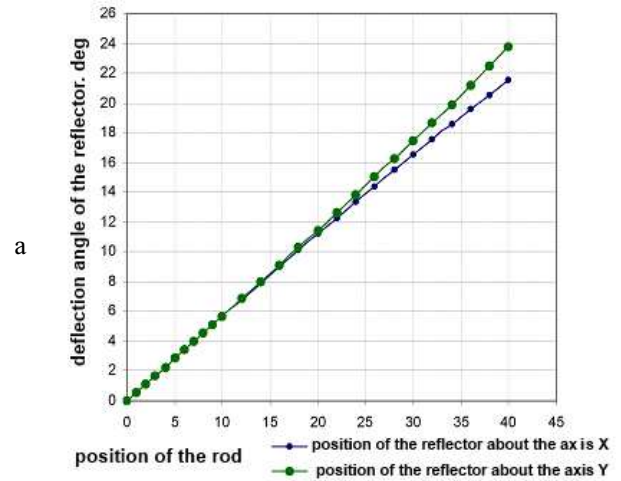


Figure 6. Results of the comparison of the SimMechanics model and the analytical mode

Possible to conclude that the analytical solutions and models SimMechanics are coincide. The maximum value of the results variation  $\Delta_{\max} \approx 31'$  in a range of con-rod positions from  $0^\circ$  to  $40^\circ$ , with intensely divergence after  $25^\circ$  (of rods). It is important, that the position of the con-rod equal  $15^\circ$  is the maximum for the two-parametric antenna specifications.

#### 4. THE SIMMECHANICS MODEL OF THE TRIBOMETRIC SYSTEM

The tribometric system "Tribal-2" was created at the Mechatronics Department of ITMO University. This system designed to investigate and identification of friction processes, which occur because of the reciprocating motion of the test samples relative to each other. The kinematic configuration of the tribometric system is shown in Fig. 7.

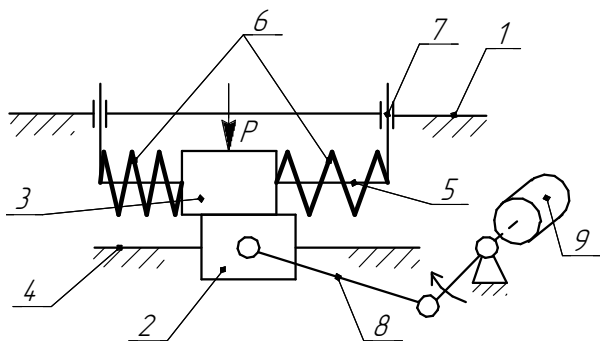


Figure 7. The kinematic configurations of the tribometric system

- 1 - base; 2,3 - upper and lower platforms, accordingly;
- 4 – guide of the lower platform;
- 5 - guide of the upper platform; 6 - spring; 7 - vertical guide; 8 - crankgear; 9 - motor

Tribometric system consists of a base 1, two platforms: the lower 2 and the upper 3. Platforms can reciprocatingly move along the guides 4 and 5. The upper platform can move in the vertical direction by the guide 7. Force of the springs 6 and vertical force (P) acting on the platform. The springs 6 fixed to the guide 5. Motor 9 sets in motion the lower platform by the crank mechanism 8. Upper platform moves by frictional forces. The upper platform seeks to return to its original position by the elastic force of the springs. Thus, platform oscillates. Position of the platforms are determined by linear displacement sensor.

The SimMechanics block diagram of the tribometric system is shown in Fig. 8. The group of blocks 1 is a mechanical part of the system (platforms, springs), which are connected by kinematic pairs. Lower platform controlled by the motor with given parameters (group 3). The external force for the model is accepted as constant, which is given by the mass of the upper platform. Spring and damper unit simulates the elastic-dissipative forces, which depends on the given coefficients of elasticity and damping. The group 2 simulates tribological effects between platforms. The friction force is given by Joint Stiction Actuator block

and depends on relative motion of the platforms (Actuator unit with “clumping” effect). The group of blocks 4 takes the results of platforms motion. More detail, this friction model discussed in (Nuzhdin, 2013).

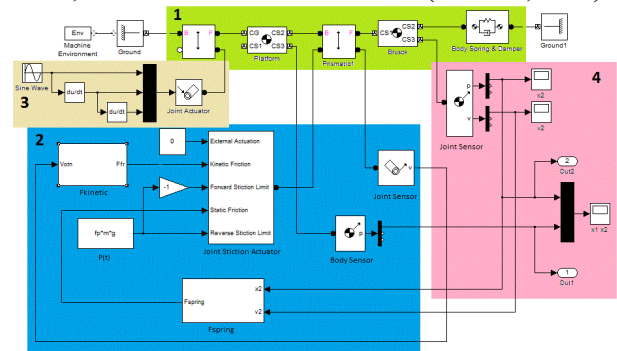


Figure 8. The SimMechanics block diagram of the tribometric system

For the modeling were used relations of the relaxation frictional self-oscillations from (Chichinadze et al., 2003). For modeling the system was used standard blocks of the SimMechanics library of the Simulink/MATLAB software. The rest and sliding friction coefficients, dependence on friction coefficient from velocity of relative motion, elasticity and damping coefficients of the elastic system were taken into account. Results of modeling calculations are shown in Figure 9.

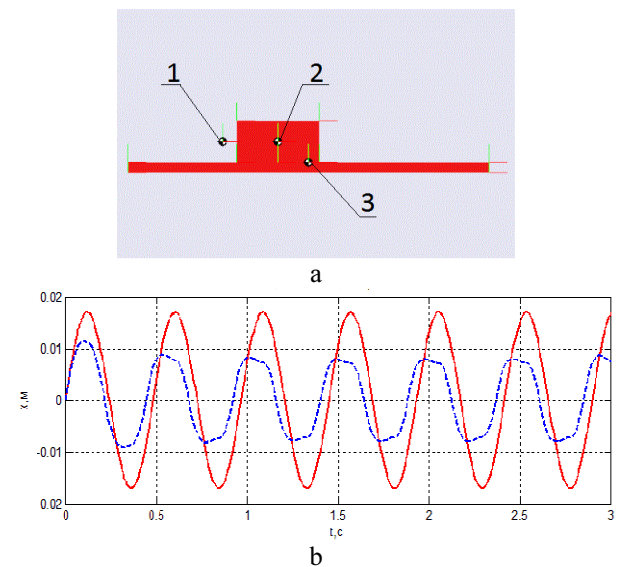


Figure 9. The simulation results:  
 (a) The SimMechanics animated model of the tribometric system: 1 – attachment point of the spring; 2 – the center of the sample on the upper platform; 3 – point of the lower platform; (b) motion of the platforms: red solid line – the lower platform, blue dotted line – the upper platform.

Figure 9 shows the work tribometric system. The results of the friction model at presence of frictional self-oscillations coincide with the analytical solution, which is given in (Nuzhdin, 2011). The considered model of friction can be used to study the dynamic of different mechanical systems, where it is important



consider the effect of friction. Study of dynamic processes of such systems let you make adjustments to their work and choose the most effective parameters, which improves the efficiency of their usage.

### 5. THE SIMMECHANICS MODEL OF EXECUTING MECHANISM OF THE DEVICE FOR MANUFACTURING TORSION BARS WITH HELICAL ANISOTROPY

For the time being, modern materials are using in the manufacturing of devices. The suspension of the magnetosensitive element (MSE) of torsion magnetometer (TM), which is a part of the geophysical complex GI-MTS-1 (SPbF IZMIRAN), obtained by means of device are made of three microfilaments of aramid yarn with a diameter of 0.016 mm (Kopytenko et al., 2010). The main advantage of TM with this MSE is a recording of the magnetic fields and their variations in the frequency range of 0...15 Hz, with the mean square noise level that does not exceed  $1 [pT/\sqrt{Hz}]$ , however device has low climatic factors dependence (temperature, moisture) and has increased resistance to dynamical load during transportation.

Elastic torsion suspension (Fig. 10) – «braid» of three strands (one thread in each strand) used to create the axis of rotation of the magnet indicator. The envelope of the cross-section of torsion bars is an ellipse with longer axis of 0.046 mm and shorter axis of 0.033 mm, the total length - 100 mm, the density of braiding –7 knots/mm.

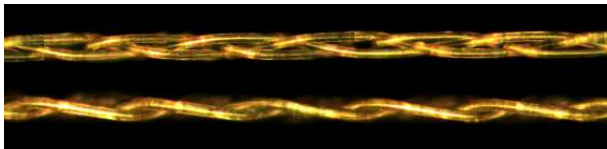


Figure 10. Photos of braided torsion bars: top view and side view

The suspension of the MSE is made of three strings with the help of device for manufacturing torsion bars with Helical Anisotropy (Device). A single filament is used in string, diameter of each doesn't exceed 16 microns.

Physical and kinematic configurations of the Device are shown in Fig. 11 and Fig.12.

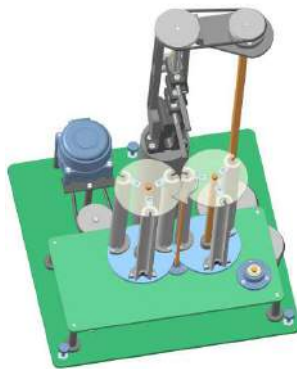


Fig. 11. Physical configuration of the Device

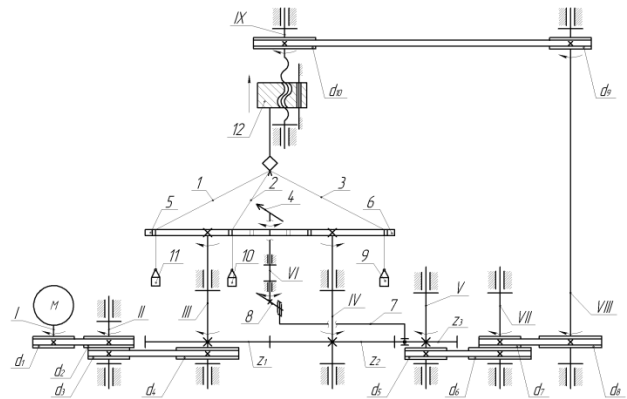


Fig. 12. Kinematic configuration of the Device:

1..3 – braid strings, 4 – cam switch, 5,6 – disks,

7 – con-rod, 8 – yoke, 9..11 – plummets,

12 – screw-nut, I..IX – axles,  $z_1..z_3$  – gear wheels,

$d_1..d_{10}$  – sheaves

The process of braiding is implemented as follows: using the electric motor M through some intermediate transfer mechanism ( $d_1..d_4$ ) revolves two disks 5 and 6, in which slots three strings (1..3) of future braid are passed through. With the help of gear wheels  $z_1$  and  $z_2$  rotation in the mutually-opposite direction of the disks is realized. Synchronously with the rotation of the disks reciprocating rotary motion of arrow-like cam is implemented through the yoke mechanism drive ( $z_3$ , 7, 8) of cam 4. Thus, the cam switch transfers braid string from slots of one disk to free slots of another one. Feed gearing ( $d_1..d_{10}$ , IX, 12) provide a uniform lift of braiding block as strings are being interlaced by braiding mechanism. Constant braid strings (1..3) tension is provided by plummets (9..11).

The sketch of executing mechanism of the device, is shown in Fig.13.

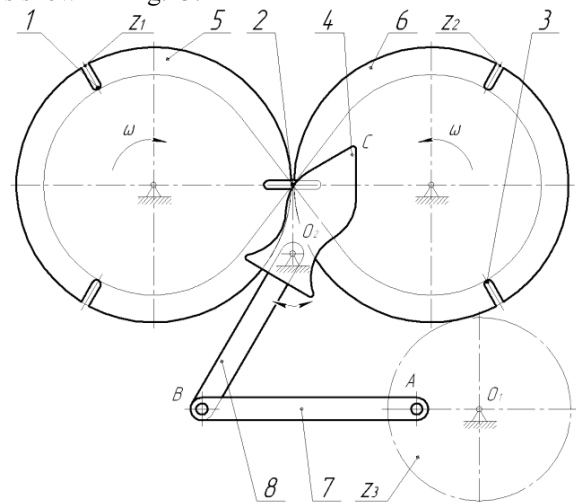


Figure 13. Executing mechanism of the Device:

1..3 – braid strings, 4 – cam switch, 5,6 – disks,

7 – con-rod, 8 – yoke,  $z_1..z_3$  – gear wheels

In work (Perechesova, 2012) parameters of the Device have been optimized. The SimMechanics model of the optimized Device Executing mechanism (fig. 14)

also has been made. Configuration of the Device Executing mechanism is shown in Fig. 14.

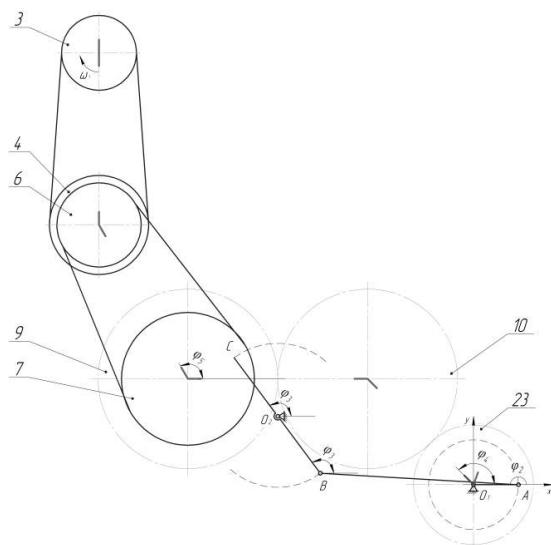


Figure 14. Configuration of the Device Executing mechanism (top view)

Notation of Fig. 14: 3, 4, 6, 7 – sheaves; 9, 10, 23 – gear wheels;  $O_1A$  – crank;  $AB$  – con-rod 24;  $BO_2$  – yoke 25;  $O_2C$  – cam switch 27;  $\varphi_1 - \varphi_5$  – the angles of units rotation. The origin is chosen at the point  $O_1$  for calculating the angles of units rotation.

The mathematical model was used to describe the electromagnetic processes in the synchronous motor.

The SimMechanics animated model of the Device Executing mechanism is shown in Fig. 15.

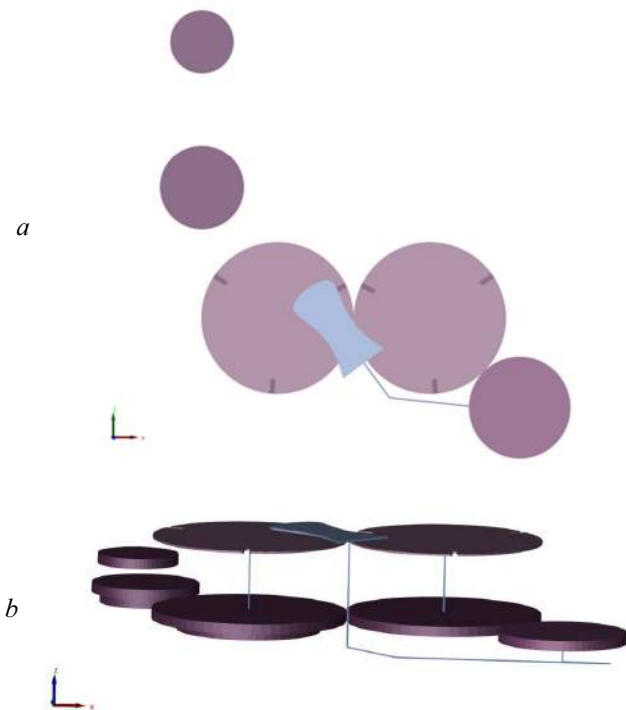


Figure 15. The SimMechanics animated model of the Device Executing mechanism top view (plane  $xy$ ) (a) and an arbitrary view (b)

The SimMechanics block diagram of the Device Executing mechanism is shown in Fig. 16.

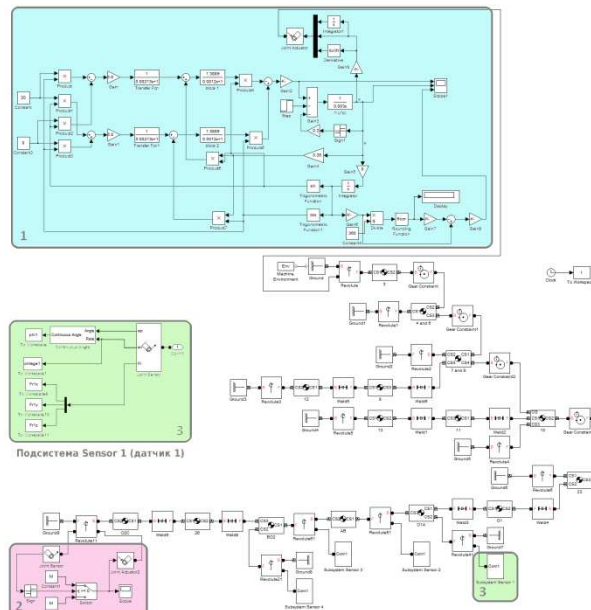


Fig. 16. The SimMechanics block diagram of the Device Executing mechanism with the synchronous motor: 1 - synchronous motor; 2 - load applied to the cam switch  $O_2S$ ; 3 - subsystem Sensor 1; Blocks of subsystem Sensor 1 and blocks of subsystems Sensor 2-4 are similar

Basic laws of changes over time: angle, angular velocity, reaction force of a Joint primitive (axes  $x, y, z$ ) were obtained. During short period (0.11 seconds) there is transient process. Angular velocity from the sensor 1 (crank  $O_1A$ ) equals 1.803 rad/s, which corresponds to the calculated value. The mechanical system works stably within the limits of applied load.

## CONCLUSION

A two ways of kinematics analysis of the two-parametric antenna are shown in paper: using the library SimMechanics and using the principles of matrix equations. Also work presents the comparison of the results. The article presents the SimMechanics model of tribometric system and the SimMechanics model of executing mechanism of the device for manufacturing torsion bars with Helical Anisotropy. Simulation using MATLAB can be successfully applied for designing of mechanical systems.

## REFERENCES

- Chichinadze, A.V., Berliner, E.M., Brown, E.D. etc., 2003. *Friction, wear and lubrication (Tribology and Tribotechnics)*. Moscow: Mashinostroenie (in Russian).
- Kopytenko, Y.A., Sergushin, P.A., Petrishchev, M.S., Levanenko, V.A., Perechesova, A.D., 2010. Device for manufacturing torsion bars with helical anisotropy UISAT-1. *9th International Symposium on Measurement Technology and Intelligent*

- Instruments*, pp. 625-628. 29 June 2009 through 2 July 2009, Saint-Petersburg (Russian Federation).
- Musalimov, V.M., Zamoruev, G.B., Kalapyshina, I.I., Perechesova, A.D., Nuzhdin, K. A., 2013. *The modeling of mechatronic systems by MATLAB (SIMULINK / SIMMECHANICS)*. Saint-Petersburg: ITMO University (in Russian).
- Nuzhdin, K. A., 2011. Development of the tribometric system with feedback. *Vestnik of Lobachevsky state university of Nizhni Novgorod*, no. 4(2). pp. 253-254 (in Russian).
- Nuzhdin, K. A., 2013 Modeling of the relaxation frictional oscillations with the help of SIMMECHANICS. *PME Academy of Sciences. The eleventh session of the International Scientific School "Fundamental and applied problems of reliability and machine diagnosis". The collection of works*, pp. 323-330. October 21-25, Saint-Petersburg (Russian Federation) (in Russian).
- Perechesova, A.D., 2012. *The analysis and synthesis of mechanism for manufacturing Torsion Bars of devices*. Thesis (PhD). ITMO University (in Russian).
- Zamoruev, G.B., Kalapyshina, I.I., 2012. Kinematic control of two-parametrical scanning aeria. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, no. 2 (78), pp. 78–83 (in Russian).

2011 has been working as assistant in Mechatronics department. He has a 7 publications.

#### AUTHORS BIOGRAPHY

**A.D. Perechesova.** Year of birth 1985. In 2008 I successfully graduated the Mechatronics Department of ITMO University. After University (since 2007) up to present time I am working in SPbF IZMIRAN. From 2008 to 2012 I am studied at the full-time postgraduate of the University ITMO. In 2012 I defended a thesis for scientific degree of PhD (technical sciences), specialty 05.02.18 "The theory of Mechanisms and Machines". The theme of thesis: "The analysis and synthesis of mechanism for manufacturing instruments torsion bars".

Since 2013 I am the research fellow of SPbF IZMIRAN. Since 2013 I am the Docent of the Mechatronics Department of ITMO University.

In 2012 I am won the "Young scientists of the ITMO University 2012".

Number of published scientific works (including inventions and manuals) is about thirty.

Research interests include: geophysics, the theory of mechanisms and machines, the optimization theory, the theory of Helically Anisotropic solid.

**I.I. Kalapyshina.** Graduated the Mechatronics department at the University of Information Technologies in 2010. Since this time has been phd-student of the same department. From the 2012 has been working as assistant in Mechatronics department. She has a 14 publications.

**K.A. Nuzhdin.** Graduated the Mechatronics department at the University ITMO in 2010. Since this year has been phd-student of the same department. From the

# A MONO PARAMETER ANALYSIS ON A SIMULATION MODEL TO SUPPORT GALB HEURISTIC OPTIMIZATION ALGORITHMS BASED ON RESOURCE BALANCING

Sergio A. Gallo<sup>(a)</sup>, Giovanni Davoli<sup>(b)</sup>, Andrea Govoni<sup>(c)</sup>, Francesco Longo<sup>(d)</sup>, Riccardo Melloni<sup>(e)</sup>, Teresa Murino<sup>(f)</sup>

<sup>(a, b, c, e)</sup>Department of Engineering “Enzo Ferrari”, DIEF (ex DIMeC), University of Modena and Reggio Emilia, ITALY

<sup>(d)</sup>DIMEG, University of Calabria, Rende, ITALY

<sup>(a, b, c, e)</sup> <sup>(c)</sup>Department of Chemical, Material, and Industrial Production, University of Naples – Engineering Faculty– Federico II Napoli, ITALY

<sup>(a, b, c, e)</sup> [sgallo.giovanni.davoli.andrea.govoni.riccardo.melloni@unimore.it](mailto:sgallo.giovanni.davoli.andrea.govoni.riccardo.melloni@unimore.it), <sup>(d)</sup> [francesco.longo@unical.it](mailto:francesco.longo@unical.it),  
<sup>(f)</sup> [teresa.murino@unina.it](mailto:teresa.murino@unina.it)

## ABSTRACT

This paper is based on previous work that presented an approach and the related model to solve a **GALB** problem, and to analyse a multi model manual assembly line, based on **heuristic algorithm** to optimize the scheduling of tasks to the available stations, respecting a **set of restrictions**, as task/station obligation, and **aiming to optimize a multi objective function** based on time and line balancing costs elements.

The assembly line we considered, process a very large variety of items, that differ for size, features, optional, under the increasing market competition.

In that previous work, quite all strategies about resource scheduling opportunities have been considered.

In the present step, we consider the last system we worked on, with a doubled number of stations, to test the reaction to the values changes of some configuration parameter as Due Date, Tardiness Cost, Tardiness Penal, Man work Direct Labour Cost, or Station Saturation.

Keywords: Simulation Models, Decision Support System, Manual Assembly Lines, GALB Problems.

## 1. INTRODUCTION

The following paper deal with an approach to analyse a multi model manual assembly line, and the following heuristic algorithm.

In particular, at this step we evaluate the reaction of the model as we change values of some configuration parameter.

The scheduling of tasks to the available stations has to respect a set of restrictions, as task/station obligation, and a multi objective function, based on time, balancing utilization rates: Line balancing costs elements can be used as Key Performance Indicator of the balancing performance achieved.

In previous models, Gallo S. A. et al, (2013), in many steps, we developed some model to solve and support the balancing of a manual assembly line, and we start now, from the last development obtained.

So, it could be useful to remind some description.

Tasks assignments to station had to respect efficiency concerns, as the maximization and the balancing of utilization rate, but, moreover, assignment restrictions of specific task to specific station because of

the need of special machines, available just at defined station. This constrains are commons for all items of the mix.

The real original considered system we started from in the original and referable configuration, was an assembly line with six stations, and, with six operators. Lot size can vary largely, and it tends to reduce assembly quantity for single order.

Items advancement on line is done on a accumulating conveyor system, so **line is paced**, but **not synchronized**.

Assembly line process a very large variety of items, defined in 6 families, as the original Assembly Plan, (**AP**), that differ for size, features, optional, lot size. Very low quantity for single order are commonly accepted.

Task times are stochastic, and, based on real observations, and can be correctly approximated by triangular or lognormal distributions.

The duty of respect the defined external Cycle Time, in a first phase, was not considered, and the algorithms define by itself the final Cycle Time, also if the starting Referable Tack Time was calculated as the maximum value between the Ideal Tack Time, and the maximum Task Time value. Tasks that do not respect line cycle time, just cause the line to increase the performed tack time of the single item in the lot, but not generally of the performed Tack Time mean for the lot.

Just after, when the best configuration has to be defined, the External Cycle Time Is considered as we evaluate the AP processing time, and the final data to be compared with a specific Due Date.

Also in the present model no cost of off line completion was considered, but both Tardiness Daily Cost, and a Lateness Penal Cost are introduced.

A precedence diagram supports technological constrains, very similar among items and, not all operation are performed for all items, depending on item features, and optional.

The same tasks of different items, has operating times, that can vary for each item, for the operation declared in the same way. All tasks in the whole annual assembly mix in AP are represented on the precedence diagram, uniform for all items. To assembly a model not all the 34 operations are needed for a specific item, depending on features and optional. Each item has a



defined number of operation which ID number increase as the assembly process goes on.

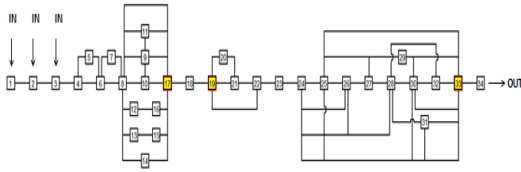


Fig. 1: combinatorial diagram of sequences

At this phase, the model and the system we started from is a U-Shape conveyor, with a doubled number of station compared to the real situation. The idea to double station number, was to enforce balancing opportunities based on the last balancing strategy we applied: the resource/work balance. The constrains position at specific stations has been doubled too, to respect the proportion and homotheticity of the original system.

The performance parameters in the last model were tack time reduction, equal to the production rate maximization, and, on an opposite way, optimize the internal balancing of tasks for stations, or levelling labour level among operators, and they have been considered both in a whole performance or objective function, a cost function.

A virtual model of the assembly line has been built in a simulation environment, to test and measure performances of the heuristic algorithms, where moreover, algorithm code has been implemented in the same software platform to use simulation suite both as a verifying tool, and both as a solving mean, or solution finder, and definitively, as task and resource scheduler.

This suite is Automod®, and the actual number of code lines is quite equal to six thousands.

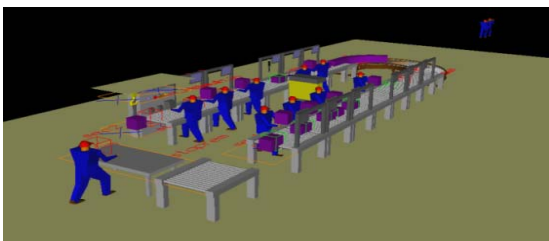


Figure 1: screenshot of the doubled assembly line.

Logics in the heuristic model has been wrote with the aim to be as more general and flexible as possible, to consider the more general problem possible, with easy configurations of data, both processing, both configuration ones.

Simulation models can read AP data from a CSV file, with any useful attribute to be used to characterize the specific Order Line, **OL**, and the configuration: time distribution parameters, item definition, order quantity, order date, etc.. In this way, is easy enough to change configuration.

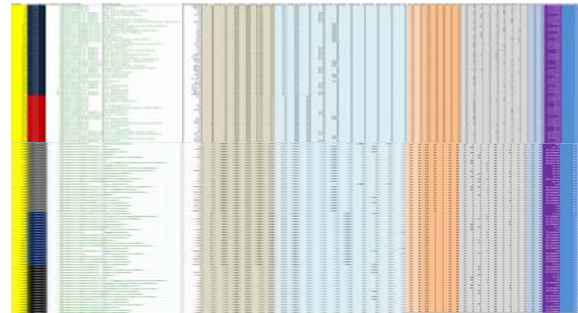


Figure 2: whole mix production plan with parameters values and task times.

We moved from the logic of the last previous model, so, there is a programming code to support all the logics and the heuristic approach to assigning tasks to stations in sequence, attempting to fulfil them, till total station time doesn't overpass calculated the current referable Tack Time. Additional control code check if any constrained task is joined, and in this case, provide to verify the station where to assign that task, if is not the actual one, and to calculate all parameters for intermediate stations (**config\_1**).

Uncharged operators are pointed and memorized, and in **config\_2**, they are assigned to overcharged station/operators, from the higher level to the lower one.

Just the best allocation will survive, and will be considered for following improving strategies.

Moreover, in **config\_3** strategy, if some station is undercharged, based on a defined level, a routine operates a tack time recursive increment till, all the undercharged station become empties, so freed operators could be reassigned to over charged station, as in the previous mentioned case.

In **config\_4**, after operators that were “naturally” free, in other words, that were free before increasing the Reference Tack Time, are tried to be assigned to the most overcharged stations, to help already assigned operators.

In **config\_5**, also operators freed because of the Reference tack Time increase, are tried to be assigned to the current overcharged stations.

In **config\_6**, the best balancing strategy for each line of the Assembly Plan, is choose in the Final configuration. All performance parameters are calculated again.

After the algorithm implementation to the whole assembly mix plan, output performances has been calculated and analyzed, as, particularly, stations utilization coefficients. As many of them showed not very fulfilled values and also the balance was not so good, having to respect constrains in the system, and specificity of and in our data set, not so many other improvement seemed applicable, consequently we considered to compare each station value to others, and to apply an additional balancing strategy based on the resources: heuristic rules to assign to same operators more than one station charge.



assigning the work content to productive units (task assignment, sequence of operations).

Since the times of Henry Ford and the model-T, customer requirements, and consequently, production systems, have changed in a way to increase dramatically customization of their products. The high level of automation of assembly systems and the fixed movement system make the (re)-configuration of an assembly line critical.

In literature, there is a wide variety of algorithms to solve ALBP, any one facing a partial part of the problem, or oriented to a particular system or configuration.

Many of them consider the problem too much statically, just under a one point of view.

But the increasing need to face continuous changes in customer's requirements, as product design, restyling and lot quantity needed, enforced with high customization and reduction of time-to-market, push to test dynamic versions of ALBP solution procedures.

Those modifications imply a very high flexibility level for the line.

ALBP consists of assigning tasks to stations in such a way that (Salveson, 1955):

- each task is assigned to one and only one station;
- the sum of performance task times assigned to each station does not exceed the cycle time;
- the precedence relationships among the tasks are satisfied;
- some performance measures are optimized.

Most procedures consider the types **I and II ALBP**, based on minimization of the number of stations, given a desired cycle time or minimization of the cycle time, given a desired number of stations, respectively.

Because of the simplifying assumptions of this basic problem, this problem was labelled simple assembly line balancing (**SALB**) in the universally accepted review of Baybars (1986). Subsequent works attempted to extend the problem by integrating practice relevant aspects, like U-shaped lines, parallel stations or processing alternatives (Becker and Scholl, 2006), referred to as general assembly line balancing (**GALB**).

Scholl (1995), and Pierreval et al. (2003) proposed a very large and comprehensive reviews of the approaches developed to solve the problem.

Ghosh and Gagnon (1989) defined a taxonomy to classify ALBP solution procedures under two key aspects, mix or variety of items produced on a single line and the nature of performance task times: single model lines or multi/mixed model lines manufacturing more items in batches or simultaneously; deterministic ALBPs, in with performance task times constant, or stochastic ALBPs, with stochastic task times distributed according to a specific distribution function.

ALBP can be solved to optimize both time - and cost, as reported in Amen (2000, 2001) and Erel and

Sarin (1998), which concern the deterministic and stochastic versions of the problem, respectively.

Moodie and Young (1965), Raouf and Tsui (1982), Suresh and Sahu (1994), Suresh et al. (1996) have proposed time-oriented algorithms, improving procedures developed for the single-model deterministic problem, with the aim of minimize stations number and the over time to complete the work off the cycle time.

In any case, relevant incompleteness costs often occur in stochastic assembly lines.

A multi objective cost function often is needed.

Two cases, both described in literature:

- the whole line is stopped till the over work is completed (Silverman and Carter, 1986);
- incomplete products get completed off-line.

Kottas and Lau (1973, 1981) proposed heuristic procedures to minimize both the total labour cost and the expected incompleteness cost. Extensions of the Kottas and Lau's (1973) method were developed by Vrat and Virani (1976), Shtub (1984).

Sarin et al. (1999) proposed, not so general as Kottas and Lau's (1973), a branch and bound heuristic to minimize the total labour cost and the total expected incompleteness cost with good results.

Erel and Sarin (1998) noticed the difficulty of methods in literature to model real conditions, and suggested that newer works should be oriented at useful studies, with impact on real-life assembly lines.

Rekiek (2000) observed that differences among ALBP and real-life statements were the multi-objective nature of the problem, no so considered in literature.

Some studies deal with the re-balancing problem of an existing line, as Sculli (1979, 1984) and, Van Oyen et al. (2001) considered the re-balancing of an existing line, under fluctuations of operator output rates or equipment failures, in short-term problem. The proposed solution to avoid temporary imbalance on the line has been the dynamic work sharing.

Rekiek et al. (2002) demonstrated that the integration between heuristic approaches and multi-attribute decision making techniques is a proven and efficient way for solving assembly lines problems.

The issue of analyse productivity of assembly line through simulation techniques has been faced by De Felice, F. et al (2012) or by Falcone, D. et al (2011), where, through an progressive approach, a simulation model is adopted to produce information about productive capacity, lead time, saturation, Value Added and Non-Value Added activity.

### 3. SYSTEM CONFIGURATION AND PARAMETERS

Now we start to describe the plan and introduce some parameter for the model.

Model parameters (times in hundredth of minute):

$$\text{Station Number} \quad k \in [1, n] \quad (1)$$

$$\text{Task Number} \quad n \in [1, 34] \quad (2)$$

$$N^\circ \text{ of tasks assigned to a single station} \quad i \in [1, h] \quad (3)$$

$$\text{Task Time} \quad T_{Op} \quad (4)$$

$$\begin{aligned} \text{Station Time} \quad T_{Stat} &= \sum_{i=1}^h T_{Op} \\ T_{Stat} &= SUC * RTT \end{aligned} \quad (5)$$

$$\begin{aligned} \text{Operation Unbalancing Coefficient} \\ UC_{Op} &= \frac{T_{Op}}{T_m} \end{aligned} \quad (6)$$

$$\begin{aligned} \text{Station Unbalancing Coefficient} \\ SUC &= \frac{T_{Stat}}{TT_{Line}} \% \\ SUC &= \sum UC_{Op} = \frac{\sum T_{Op}}{TT_{Line}} \end{aligned} \quad (7)$$

$$\text{Line Lead Time} \quad LLT = \sum_{i=1}^k T_{Stat} \quad (8)$$

$$\begin{aligned} \text{Line Tack Time or Cycle Time} \\ LTT &= \text{Max}(\sum_{l=i}^h T_{Op}) = \text{Max}(T_{Stat}) \end{aligned} \quad (9)$$

$$\begin{aligned} \text{Mean/Ideal Tack Time} \\ ITT &= \frac{\sum_{i=1}^n T_{Op}}{k} \end{aligned} \quad (10)$$

$$\begin{aligned} \text{Reference Tack Time} \\ RTT &= \text{MAX} \left( T_{mean} = \frac{\sum T_{Op}}{n}, \text{MAX}(T_{Op_n}) \right) \end{aligned} \quad (11)$$

This last parameter, RTT, is the minimum time, in hundreds of minute, used as a limitation roof in the allocation of tasks, maximum threshold the Station Time cannot overcome. It is the maximum value between the ideal tack time (ITT), and the value of the larger Task Time ( $T_{Op}$ ) for each line of the production plan. This value is increased when we are applying the strategy of emptying undercharged stations. RTT represents 100% of work time that can be assigned to each station.

Other parameter, already declared, is SUC, that is the percentage value calculated as the sum of the Unbalancing Coefficient of any operation assigned to the station. It's value is lower or equal of the RTT one, that is the 100%.

We define, at this step, a new cost function, based on the already used Direct Assembly Cost, DAC, but more over on evaluation of Tardiness.

DAC is the product of the Manpower Direct Cost, multiplied by the station number (or operators when more than one is assigned to a station), multiplied by the volume for the OL, multiplied by RTT, to be representative of both the RTT of the line for each row, and for each model, both the number of resources used:

$$\begin{aligned} \text{Direct\_Assembly\_Cost} \\ DAC &= (RTT \bullet \text{ResNum} \bullet \text{LotQuantity} \bullet \text{ManWorkCost}) \end{aligned} \quad (12)$$

Instead, the  $V\_Whole\_Cost\_xx$  function is defined in this way:

$$\begin{aligned} V\_Whole\_Cost_{<config\_number>} &= DAC_{<config\_number>} + \\ V\_Tardiness_{<config\_number>} &* V\_Daily\_Tardiness\_Cost + \\ V\_penal \end{aligned} \quad (13)$$

Our assembly line is a multi-mixed model, then we face with a **MALBP** (Mixed-Model Assembly Line BP). We will configure our situation as a **MALB-E** problem, given number of K stations, the aim is to maximize the efficiency  $E_{line}$ , i.e. minimizing the Full Cost of assembling the lot.

Our heuristic algorithm is a mix of Work Content and Resource Balancing, that, with the objective function, takes their role and weight, very freely

The simulation code will be used to apply the heuristic logic cascade, to calculate any  $V\_Whole\_Cost$ , distinct for each experiment with specific parameter values. At any step all relevant parameters have been calculated and compared, as in Chutima P. et alter, (2004), Jolai F., et alter, (2008).

#### 4. EXPERIMENTS DESCRIPTION

We decide to define a base configuration with all values for any other parameter set to the middle value.

Values we use are just reasonable, but have to be evaluated, in the future, with a larger extent and ranges, with more values then we did.

Table 1: Values of basic configuration, config 0.

| Level Coeff UnderUtil | Accepted Balancing Level | Efficiency Coef | Due Date | Cost Modulator | Daily Tardiness Cost | Penal     |
|-----------------------|--------------------------|-----------------|----------|----------------|----------------------|-----------|
| 0,200                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 |

All the others values assumed by each parameter can be observed directly on any experiment table, in bold.

We decided to define tree level for each parameter to change with all others values set to the basic configuration.

Values, changed at each experiment, are read on the reading file, "ConfigurationCSV". On this file there are other parameters that we didn't change in this analysis, as the percentage increase value for RTT, the threshold value to divide tasks between two operators when more than one is assigned to a station, and so on.

The strategies to distribute tasks to stations are the same of the previous models.

The approach we adopted has been to define a set of experiments, for all defined set of combination for defined parameters. Then, for each configuration of parameters values, we calculate all of the seven  $V\_Whole\_Cost\_xx$ , one value for any of the seven balancing strategy, config\_1, config\_2, ..., cofig\_7, that are applied and verified one behind the other in each single run of the simulation of the model.

We have to say that the best approach should be to define a multivariate and multi factor approach, ie to evaluate the effect on a simultaneous variation for all parameters values.

Moreover, the values to use to experiment the model behaviour have to be chosen much more accurately.

But, we are already working on a new version of the model that could do this considering a multi factor and multivariate approach and that could analyse outcome in an automated way.

In this work, we want just test, if the heuristic algorithms keep to perform in a good manner, also when some parameter value change.

Any time a new strategy is applied, called config\_xx, all performance parameters, as station/operators utilization coefficients, UCs, Direct Assembly Cost, are calculated, stored and compared to best performing configuration emerged at the previous step for each OL. Just the better, for each item, is the one considered for the final solution.

In last models we operated with defined values for the configuration of the algorithm, so we consider now to evaluate how some variation in the values assumed by some parameter can affect the heuristic performance.

In the following lines, all the experiments configurations.

Table 2: Experiments configuration for “Level Coeff Under” variation: config 1, config 2, config 3.

| Level Coeff UnderUtil | Accepted Balancing Level | Efficiency Coef | Due Date | Cost Modulator | Daily Tardiness Cost | Penal     |          |
|-----------------------|--------------------------|-----------------|----------|----------------|----------------------|-----------|----------|
| 0,200                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 1 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 2 |
| 0,300                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 3 |

Table 3: Experiments configuration for “Accepted Balancing Level” variation: config 4, config 5, config 6.

| Level Coeff UnderUtil | Accepted Balancing Level | Efficiency Coef | Due Date | Cost Modulator | Daily Tardiness Cost | Penal     |          |
|-----------------------|--------------------------|-----------------|----------|----------------|----------------------|-----------|----------|
| 0,250                 | 0,700                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 4 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 5 |
| 0,250                 | 0,900                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 6 |

Table 4: Experiments configuration for “Due Date” variation: config 7, config 8, config 9.

| Level Coeff UnderUtil | Accepted Balancing Level | Efficiency Coef | Due Date | Cost Modulator | Daily Tardiness Cost | Penal     |          |
|-----------------------|--------------------------|-----------------|----------|----------------|----------------------|-----------|----------|
| 0,250                 | 0,800                    | 1,550           | 7,000    | 25,000         | 80000,000            | 80000,000 | config 7 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 8 |
| 0,250                 | 0,800                    | 1,550           | 9,000    | 25,000         | 80000,000            | 80000,000 | config 9 |

Table 5: Experiments configuration for “Cost Modulator” a Labour Cost parameter: config 10, config 11, config 12.

| Level Coeff UnderUtil | Accepted Balancing Level | Efficiency Coef | Due Date | Cost Modulator | Daily Tardiness Cost | Penal     |           |
|-----------------------|--------------------------|-----------------|----------|----------------|----------------------|-----------|-----------|
| 0,250                 | 0,800                    | 1,550           | 8,000    | 10,000         | 80000,000            | 80000,000 | config 10 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 11 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 30,000         | 80000,000            | 80000,000 | config 12 |

Table 6: Experiments configuration for “Level Coeff Under” variation: config 13, config 24, config 15.

| Level Coeff UnderUtil | Accepted Balancing Level | Efficiency Coef | Due Date | Cost Modulator | Daily Tardiness Cost | Penal     |           |
|-----------------------|--------------------------|-----------------|----------|----------------|----------------------|-----------|-----------|
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 40000,000            | 80000,000 | config 13 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000 | config 14 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 120000,000           | 80000,000 | config 15 |

Table 7: Experiments configuration for “Level Coeff Under” variation: config 13, config 24, config 15.

| Level Coeff UnderUtil | Accepted Balancing Level | Efficiency Coef | Due Date | Cost Modulator | Daily Tardiness Cost | Penal      |           |
|-----------------------|--------------------------|-----------------|----------|----------------|----------------------|------------|-----------|
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 40000,000            | 60000,000  | config 16 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 80000,000            | 80000,000  | config 17 |
| 0,250                 | 0,800                    | 1,550           | 8,000    | 25,000         | 120000,000           | 100000,000 | config 18 |

#### 4.1. ANALYSIS AND COMPARISON OF EXPERIMENTAL RESULTS

In the following lines, we will show the results of the experiments we have outlined, with some consideration.

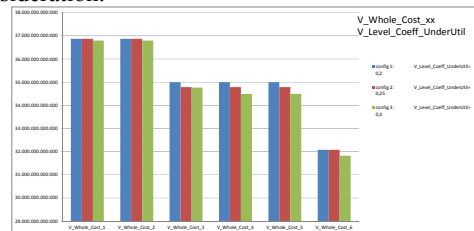


Fig. 4: Whole Cost for config1, config2, config3: Coeff SotUtil variation.

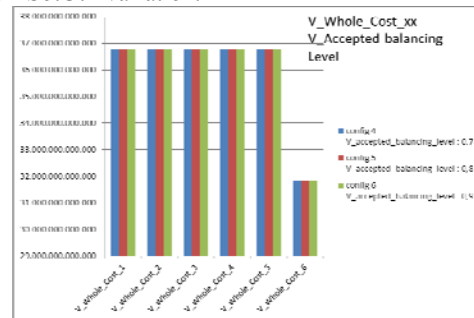


Fig. 5: Whole Cost for config4, config5, config6: V\_Accepted balancing Level variation.

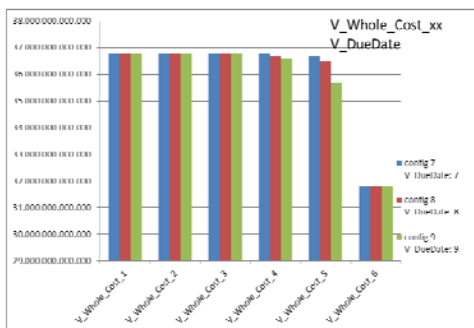


Fig. 6: Whole Cost for config7, config8, config9: Due Date variation.



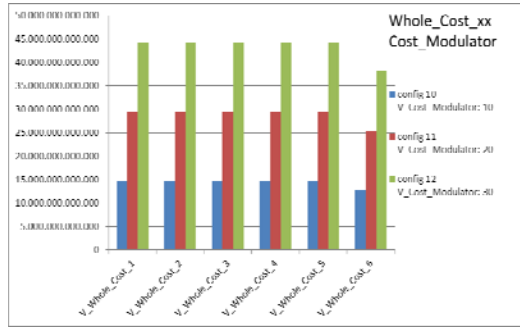


Fig. 7: Whole Cost for config10, config11, config12: Cost Modulator variation.

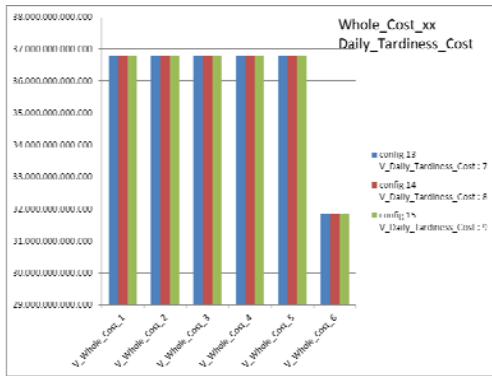


Fig. 8: Whole Cost for config10, config11, config12: Cost Modulator variation.

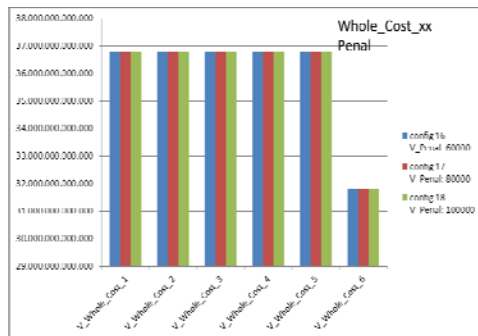


Fig. 9: Whole Cost for config10, config11, config12: Cost Modulator variation.

As we can observe, depending on our set of values, and on the configuration we worked on, some parameter seems to affect more the outline.

In some cases, we can't observe significant differences due to the values we used, this is the case of all the parameter with the exception of the Cost Modulator Parameter that cause a significant difference on the Whole Cost inside each configuration considered, with a reduction for all values just for the aggregation strategy, config\_7.

For all others, we can observe that they do not produce relevant differences when different values are used, but much more they affect the effect of increasing of the performance, of the Whole cost function, as we progress to apply any of the defined balancing

strategies, but with no significant difference based on values changing. This effect is particularly relevant when considering Level Coef Under Utilization and Due Date, while for others some value effect can be observed, and a global effect when we adopt the strategy to aggregate more than one station to a single operator.

We tried also with other value set, and result, that we do not show here, are conform to those presented.

Finally, we can confirm that the algorithm keep to work appropriately also when some parameter value is changed, because never the new balancing strategy shows a worsening performance compared with the precedent applied. At least, keep the same performance.

We have to remember that the balancing opportunities are strongly affected by constrains position and by task time duration that can give big limitation to accomplish the ideal balancing configuration.

## 5. CONCLUSIONS

In this paper a developed a new improvement step in the development of an heuristic logic to solve the issue of an manual assembly line.

We tried to evaluate if the logic and the related code keep to perform also when some parameter value change.

The results show that the logic keep to perform and, that confirm the capability of the proposed algorithm of dealing with the multi objective nature of the re-balancing problem.

Now, in this step we can consider solutions that treat with advantages both in tack time reduction, both on balancing and on saturation improvement, but tha can consider also some aspect related to the respect of eventual Due Date.

We are already developing models that can test the model behavior also in case of multi parameter experiment as a deterministic DOE, or multifactorial Analysis, with data set chose to stress the system and to evaluate model limitation.

Obviously, outlines can change in value as you chance parameters values, or if you do not consider that someone exist, but it is important that the approach keep to be valid: any new balancing strategy can just improve the solution, never worse.

Moreover, because we are oriented to the more general solution, in terms of flexibility, variability of mix, number of resources and stations, number and position of some constrains, we are working on optimization approaches that could include a new data collection and the variation of data of the system randomly with logic, to have a greater validation of the algorithm.

## REFERENCES

Amen, M., 2000. Heuristic methods for cost-oriented assembly line balancing: A survey. *International Journal of Production Economics* 68, 1–14.

- Amen, M., 2001. Heuristic methods for cost-oriented assembly line balancing: A comparison on solution quality and computing time. *International Journal of Production Economics* 69, 255–264.
- Bautista J., Pereira J., 2008. “A Dynamic Programming Based Heuristic for the Assembly Line Balancing Problem”, *International Journal of Production Economics*.
- Baybars, I., 1986. A survey of exact algorithms for the simple assembly line balancing problem. *Management Science* 32, 09–932.
- Becker, C., Scholl, A., 2006. A survey on problems and methods in generalized assembly line balancing. *European Journal of Operational Research* 168, 694–715.
- Chutima, P. Suphapruksapongse, H., 2004. Practical Assembly-Line Balancing in a Monitor Manufacturing Company, *Tharnmasat Int. J. Sc. Tech.*, Vol. 9, No. 2
- De Felice, F., Petrillo, A., 2012. Productivity analysis through simulation technique to optimize an automated assembly line, Proceedings of the IASTED International Conference on Applied Simulation and Modelling, ASM 2012.
- Falcone, D., Silvestri, A., Di Bona, G., Forcina, A., Pacitto, A., 2011. Study and modelling of very flexible lines through simulation, Proceedings of the 2010 Spring Simulation Multiconference - Emerging M and S Applications in Industry and Academia Symposium, EAIA, Pages 11-16. Boston, MA, USA.
- Gallo, S. A., Davoli G., Govoni A., Melloni R., Simulation Models To Support Galb Heuristic Optimization Algorithms Based On Resource Balancing Based On Multi Objective Performance Index, The 25th European Modeling & Simulation Symposium, September, September 25-27, 2013, Athens, Greece.
- Ghosh, S., Gagnon, R.J., 1989. A comprehensive literature review and analysis of the design, balancing and scheduling of assembly systems. *International Journal of Production Research* 27, 637–670.
- Erel, E., Sarin, S.C., 1998. A survey of the assembly line balancing procedures. *Production Planning and Control* 9, 414–434.
- Gökçen, H K. Ağpak, R. 2006. “Balancing of Parallel Assembly Lines”, *International Journal of Production Economics*.
- Kottas, J.F., Lau, H.S., 1973. A cost oriented approach to stochastic line balancing. *AIIE Transactions* 5, 164–171.
- Kottas, J.F., Lau, H.S., 1981. A stochastic line balancing procedure. *International Journal of Production Research* 19, 177–193.
- Jolai, F., Jahangoshai REZAEI M., Vazifeh, A. 2008. Multi-Criteria Decision Making for Assembly Line Balancing, *Springer Science Business Media*.
- Moodie, C.L., Young, H.H., 1965. A heuristic method of assembly line balancing for assumptions of constant or variable work element times. *Journal of Industrial Engineering* 16, 23–29.
- Pierreval, H., Caux, C., Paris, J.L., Viguier, F., 2003. Evolutionary approaches to the design and organization of manufacturing systems. *Computers & Industrial Engineering* 44, 339–364.
- Raouf, A., Tsui, C.L., 1982. A new method for assembly line balancing having stochastic work elements. *Computers & Industrial Engineering* 6, 131–148.
- Rekiek, B., 2000. Design of assembly lines. Memoire presente en vue de l’obtention du grade de docteur en sciences appliquees. *Universite libre de Bruxelles*, Brussels, Belgium.
- Rekiek, B., De Lit, P., Delchambre, A., 2002. Hybrid assembly line design and user’s preferences. *International Journal of Production Research* 40, 1095–1111.
- Salveson, M. E., 1955. The assembly line balancing problem. *Journal of Industrial Engineering* 6, 18–25.
- Sarin, S.C., Erel, E., Dar-El, E.M., 1999. A methodology for solving single-model, stochastic assembly line balancing problem. *OMEGA—The International Journal of Management Science* 27, 525–535.
- Scholl, A., 1995. Balancing and Sequencing of Assembly Lines. *Physica-Verlag, Heidelberg*.
- Scholl, A., Boysen, N., 2009. Designing Parallel Assembly Lines with Split Workplaces: Model and Optimization Procedure. *International Journal of Production Economics*.
- Silverman, F.N., Carter, J.C., 1986. A cost-based methodology for stochastic line balancing with intermittent line stoppages. *Management Science* 32, 455–463.
- Sculli, D., 1979. Dynamic aspects of line balancing. *OMEGA— The International Journal of Management Science* 7, 557–561.
- Sculli, D., 1984. Short term adjustments to production lines. *Computers & Industrial Engineering* 8, 53–63.
- Shtub, A., 1984. The effect of incompleteness cost on the line balancing with multiple manning of work stations. *International Journal of Production Research* 22, 235–245.
- Süer G.A., 1998. Designing Parallel Assembly Lines, *Industrial Engineering Department*, University of Puerto Rico-Mayagüez.
- Suresh, G., Sahu, S., 1994. Stochastic assembly line balancing using simulated annealing. *International Journal of Production Research* 32, 1801–1810.
- Suresh, G., Vinod, V.V., Sahu, S., 1996. A genetic algorithm for assembly line balancing. *Production Planning & Control* 7, 38–46.
- Van Oyen, M.P., Gel, E.S., Hopp, W.J., 2001. Performance opportunity for workforce agility in collaborative and noncollaborative work systems. *IIE Transactions* 33, 761–777.
- Vrat, P., Virani, A., 1976. A cost model for optimal mix of balanced stochastic assembly line and the modular assembly system for a customer oriented production system. *International Journal of Production Research* 14, 445–463.



# BANKSIM: A BANK PAYMENTS SIMULATOR FOR FRAUD DETECTION RESEARCH

Edgar Alonso Lopez-Rojas<sup>(a)</sup> and Stefan Axelsson<sup>(b)</sup>

<sup>(a),(b)</sup>Blekinge Institute of Technology, School of Computing

<sup>(a)</sup>[edgar.lopez@bth.se](mailto:edgar.lopez@bth.se), <sup>(b)</sup>[stefan.axelsson@bth.se](mailto:stefan.axelsson@bth.se)

## ABSTRACT

BankSim is an agent-based simulator of bank payments based on a sample of aggregated transactional data provided by a bank in Spain. The main purpose of BankSim is the generation of synthetic data that can be used for fraud detection research. Statistical and a Social Network Analysis (SNA) of relations between merchants and customers were used to develop and calibrate the model. Our ultimate goal is for BankSim to be usable to model relevant scenarios that combine normal payments and injected known fraud signatures. The data sets generated by BankSim contain no personal information or disclosure of legal and private customer transactions. Therefore, it can be shared by academia, and others, to develop and reason about fraud detection methods. 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 BankSim generates data that usefully approximates the relevant aspects of the real data. We intend to make the simulation and its results available to the research community.

Keywords: Multi-Agent Based Simulation, Bank Payments, Fraud Detection, Credit Card Fraud, Synthetic Data.

## 1. INTRODUCTION

In this paper we present *BankSim*, a **Bank** payment **Simulation**, built on the concept of Multi Agent-Based Simulation (MABS). *BankSim* is based on a sample of aggregated transaction data provided by one bank in Spain with the aim of promoting the development of applications for Big Data. This data contains several thousand records of transactional data covering six months, from November 2012 until April 2013 restricted by zip code location to Madrid and Barcelona. That is, this data is recent enough to reflect current conditions of payments, but aggregated to not pose a risk from a specific customer privacy standpoint.

The defence against fraud is an important topic that has seen some study. In a bank the cost of fraud are of course ultimately transferred to the consumer, and finally impacts the overall economy. Our aim with *BankSim* is to learn the relevant parameters that governs the behaviour of a bank payment system to simulate *normal* behaviour and inject specific fraud scenarios that are interesting to study.

The main contribution, and focus, of this paper is a

method of generating anonymous synthetic data from aggregated transactional data of a bank payment system, that can then be used as part of the necessary data for the development and testing of fraud detection techniques. Even so, the data set generated could also be the basis for research in other fields, such as consumer behaviour, general economic study including social development and forecasting.

Later we plan to address the actual fraud and develop techniques to develop malicious agents to inject fraudulent and anomalous behaviour, and then develop and test different strategies for detecting these instances of fraud. Even though we do not address these issues in this paper, we describe some typical scenarios of credit card fraud that affects bank payments. As this is our ultimate goal, fraud heavily influenced the design of *BankSim*.

The main goal of developing this simulation is that it enables us to share realistic fraud data, without exposing potentially business or personally sensitive information about the actual source. As data relevant for computer security research often is sensitive, for a multitude of reasons, i.e. financial, privacy related, legal, contractual and other, research has historically been hampered by a lack of publicly available relevant data sets. Our aim with this work is to address that situation. However, simulation also have other benefits, it can be much faster and less expensive than trying different scenarios of fraud, detection algorithms, and personnel and security policy approaches in an actual store. The latter also risks incurring e.g. unhappiness amongst the staff, due to trying e.g. an ill advised policy, which leads to even greater expense and unwanted problems.

**Outline:** The rest of this paper is organised as follows: Section 2. introduce the topic of fraud detection for bank payments and present previous and related work. Sections 3. describes the problem, which is the generation of synthetic data of a bank payment system. Section 4. shows a data analysis of the current data. Section 6. presents an implementation of a MABS for our domain and shows the description of some credit card fraud scenarios. We present our results and verification of the simulation in section 7. and finish with a discussion and conclusions, including future work in section 8.

## 2. BACKGROUND AND RELATED WORK

Simulations in the domain of financial markets have traditionally been focused on finding answers to prediction problems such as economic growth, market growth, consumption patterns and so on.

There is currently a lack of research in the area of simulation of bank systems, more specifically for fraud detection.

We have previously analysed the implications of using machine learning techniques for fraud detection using a synthetic data set (Lopez-Rojas and Axelsson, 2012a). We then built a simple simulation of a financial transaction system based on these assumptions, in order to overcome our limitations and lack of real data (Lopez-Rojas and Axelsson, 2012b). However, this work was not based on any underlying data, but rather on assumptions of what such data could contain. We learn the principles of simulation and modelling and successfully applied them to *RetSim* (Lopez-Rojas et al., 2013). *RetSim* is the older brother of *BankSim* and uses data from a retail store to produce a realistic simulation that generates synthetic data.

Here we continued our work and built a realistic simulation based on a real aggregated payment data set that can be used to test diverse fraud detection techniques. All our simulators are part of a financial system chain. They have in common that all are built with the aim of modelling financial activity with the purpose of generating synthetic data sets for fraud detection research. We are continuing to build the needed components to integrate them into a complex financial chain and produce a virtual financial world that covers many domains. This is specifically useful to implement more complex fraud scenarios such as money laundering.

Data mining based methods have previously been used to detect fraud (Phua et al., 2010). This led to the result that machine learning algorithms can identify novel methods of fraud by detecting those transactions that are different (anomalous) in comparison to benign transactions. This problem in machine learning is known as novelty detection. Supervised learning algorithms have previously been used on a synthetic data set to prove the performance of outliers detection (Abe et al., 2006), however this has not been performed on transactional data. There are tools such as IDSG (IDAS Data and Scenario Generator (Lin et al., 2006)) which was developed with 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 their test phase and it has been used to test fraud detection systems.

The most common method today used for preventing illegal financial transactions consists on flagging different clients according to perceived risk and restricting their transactions using thresholds (Bolton and Hand, 2002). Transactions that exceed these thresholds require extra scrutiny whereby the client needs to declare the precedence of the funds. These thresholds are usually set

by legislation without distinction made between different economic sectors or actors. This of course leads to fraudsters adapting their behaviour in order to avoid this kind of control, by e.g. making many smaller transactions that fall just below the threshold. Hence, these and other similar methods have proven insufficient (Magnusson, 2009).

Nowadays with the popularity of social networks, such as *Facebook*, the topic of Social Network Analysis (SNA) has been given special interest in the research community (Alam and Geller, 2012). Social Network Analysis is a topic that is currently being combined with Social Simulation. Both topics support each other for the benefit of representing the interactions and behaviour of agents in the specific context of social networks.

Our approach aims to fill the gap between existing methods and provide researchers with a tool that generates reliable data to experiment with different fraud detection techniques and compare them with other approaches.

## 3. PROBLEM

Fraud and fraud detection is an important problem that has a number of applications in diverse domains. However, in order to investigate, develop, test and improve fraud detection techniques one needs detailed information about the domain and its specific problems.

There is a lack of data sets available for research in fields such as money laundering, financial fraud and illegal payments. Disclosure of personal or private information is only one of the many concerns that those that own relevant data have. This leads to in-house solutions that are not shared with the research community and hence there can be no mutual benefit from free exchange of ideas between the many worlds of the data owners and the research community.

After describing the problem we formulated the main research question that we address on this paper:

**RQ** *How could we model and simulate a bank payment system and generate a realistic and reliable synthetic data set for the purpose of fraud detection?*

## 4. Data Analysis

To better understand the problem we began by performing data analysis of the sample data provided by a bank in Spain. We are interested in finding the necessary and sufficient attributes to enable us to simulate a realistic scenario in which we could reason about and detect interesting cases of fraud.

The bank in Spain, which we will name *Bank Inc.*, provided a web service interface to query aggregated information about bank payments. The web service limited the query to transactions that occurred between November 2012 and April 2013 restricted to transactions that took place in Madrid and Barcelona. The service provided by *Bank Inc.* groups the data by month, week, day of the week and hour. The interface allows three types of queries: *consumption habits*, *customer classification* and *origin and source of transactions*. The basic information provided by the queries are mainly statistical information

about payments such as: number, average, minimum and maximum values. It also provides information regarding zip code location of origin/source, merchant category and customer gender and age. There are 16 merchant categories that differentiate between payments made for example in a restaurant or payments performed while buying cars or other goods.

It was not possible to query information where less than 2 customers made payments. This means there is some missing information about the data provided, but fortunately we know exactly which data is missing, because the response from the web service is different depending on whether the data is missing or restricted.

We initially started by selecting a few zip codes that contain enough information to avoid missing fields. We selected two of the biggest zip codes by number of transactions and amount. We extracted statistical information, presented in table 1. Age Categories are given in table 2 and gender categories are given in table 3. All prices given are in euro.

Due to a lack of space we will focus our presentation of the analysis on one of the biggest zip code by payment volume that we will call Zip Code One (ZC1).

ZC1 is relatively richer in data than the smaller zip codes, it contains 731658 payments during a six month period. This is specially interesting, since we are more likely to find actual cases of fraud.

Table 1: Statistical Analysis Data

| zipcode | gender | age | payments | avgAmountMonth | avgNumCardsMonth |
|---------|--------|-----|----------|----------------|------------------|
| ZC1     | E      | U   | 823      | 31.97          | 90.67            |
| ZC1     | F      | 6   | 12375    | 44.83          | 1002.33          |
| ZC1     | F      | 5   | 39461    | 35.81          | 3297.50          |
| ZC1     | F      | 4   | 72336    | 33.79          | 6514.83          |
| ZC1     | F      | 3   | 94536    | 31.87          | 9337.50          |
| ZC1     | F      | 2   | 128117   | 29.37          | 13457.33         |
| ZC1     | F      | 1   | 41299    | 30.13          | 5002.00          |
| ZC1     | F      | 0   | 1809     | 28.81          | 257.00           |
| ZC1     | M      | 6   | 18030    | 36.93          | 1676.33          |
| ZC1     | M      | 5   | 38097    | 33.29          | 3534.83          |
| ZC1     | M      | 4   | 62314    | 32.39          | 5871.67          |
| ZC1     | M      | 3   | 82222    | 30.38          | 8451.83          |
| ZC1     | M      | 2   | 106404   | 27.42          | 10969.33         |
| ZC1     | M      | 1   | 32031    | 27.70          | 3739.67          |
| ZC1     | M      | 0   | 1516     | 28.37          | 213.83           |
| ZC1     | U      | 6   | 193      | 17.56          | 13.83            |
| ZC1     | U      | 4   | 14       | 23.95          | 3.00             |
| ZC1     | U      | 3   | 54       | 12.03          | 4.00             |
| ZC1     | U      | 2   | 27       | 23.60          | 3.40             |
| ZC2     | E      | U   | 23349    | 5.78           | 482.83           |
| ZC2     | F      | 6   | 13160    | 61.97          | 1373.00          |
| ZC2     | F      | 5   | 27250    | 55.58          | 2766.50          |
| ZC2     | F      | 4   | 50074    | 48.90          | 4508.00          |
| ZC2     | F      | 3   | 63122    | 43.59          | 5746.00          |
| ZC2     | F      | 2   | 91343    | 37.89          | 8026.67          |
| ZC2     | F      | 1   | 37303    | 30.17          | 3152.50          |
| ZC2     | F      | 0   | 1842     | 26.89          | 172.83           |
| ZC2     | M      | 6   | 11176    | 80.01          | 1203.00          |
| ZC2     | M      | 5   | 18854    | 74.22          | 1951.83          |
| ZC2     | M      | 4   | 29474    | 67.89          | 2990.83          |
| ZC2     | M      | 3   | 45850    | 53.18          | 4612.17          |
| ZC2     | M      | 2   | 63568    | 41.72          | 6048.00          |
| ZC2     | M      | 1   | 21538    | 32.88          | 2054.50          |
| ZC2     | M      | 0   | 977      | 28.16          | 92.83            |
| ZC2     | U      | 6   | 67       | 74.08          | 6.33             |
| ZC2     | U      | 5   | 8        | 103.15         | 3.00             |
| ZC2     | U      | 3   | 10       | 24.48          | 4.00             |

## 5. Fraud Scenarios in a Bank Payment System

In this section we describe how three example of fraud that can be implemented in BankSim. These fraud scenarios are based on selected cases from the Grant Thornton report Member and Council (2009). As can be seen in section 6., the different scenarios can be implemented

Table 2: Age Categories

| idAge | Rank    |
|-------|---------|
| 0     | <=18    |
| 1     | 19-25   |
| 2     | 26-35   |
| 3     | 36-45   |
| 4     | 46-55   |
| 5     | 56-65   |
| 6     | >65     |
| U     | Unknown |

Table 3: Gender Categories

| idGender | Description |
|----------|-------------|
| E        | ENTERPRISE  |
| F        | FEMALE      |
| M        | MALE        |
| U        | UNKNOWN     |

Table 4: Categories ZC1

| category             | percentage | avg          | std          |
|----------------------|------------|--------------|--------------|
| Auto                 | 0.0049     | 224.35916667 | 267.52611111 |
| Bars and restaurants | 0.0244     | 31.03238095  | 39.19238095  |
| Books and press      | 0.0014     | 33.34714286  | 45.01428571  |
| Fashion              | 0.0076     | 49.73190476  | 59.08452381  |
| Food                 | 0.0726     | 32.49333333  | 30.87285714  |
| Health               | 0.0179     | 59.39119048  | 113.98619048 |
| Home                 | 0.0021     | 75.48317073  | 121.77292683 |
| Accommodation        | 0.0016     | 97.41071429  | 86.85047619  |
| Hypermarkets         | 0.0178     | 33.06547619  | 35.78166667  |
| Leisure              | 0.0001     | 74.86357143  | 22.01107143  |
| Other services       | 0.009      | 52.9897619   | 76.65309524  |
| Sports and toys      | 0.0043     | 74.8047619   | 75.45452381  |
| Technology           | 0.0031     | 67.28285714  | 108.68452381 |
| Transport            | 0.8176     | 24.56047619  | 20.76928571  |
| Travel               | 0.0004     | 577.46285714 | 518.41885714 |
| Wellness and beauty  | 0.0155     | 44.20809524  | 55.29142857  |

in almost the same way. Furthermore, a fraudster will probably use several different methods of fraud, which means that BankSim needs to be able to model combinations of all fraud scenarios implemented. Although the implementation of these scenarios are out of the scope of this paper, we include a description and explain how to implement them in BankSim.

We will focus on card related frauds. This kind of fraud usually begins when the the important data on the card is compromised: Account name, credit card number, expiration date and verification code. This data can be acquired by a fraudster either by theft of the physical card or by gaining knowledge of the important data associated with the account.

### 5.1. Theft

This scenario includes cases where the customer loses physical possession of her card and a fraudster impersonate the customer purchasing goods or service with the stolen card. In terms of the object model used in BankSim the Theft scenario can be implemented by the following setting: Include in the fraudster the behaviour of sensing customer proximity, then execute the theft and later pur-

chase goods from another merchant with the information from the customer. The volume of fraudulent activity can be modelled changing the specific parameter of number of theft, zip code and frequency. A "red flag" for detection in this case could be a high number of unusual transactions with high value in a short period.

### 5.2. Cloned Card/Skimming

This scenario includes cases where the fraudster creates a clone of the card, letting the user keep the original card but without knowledge of the loss of security. In terms of the object model used in BankSim, the cloned card scenario can be implemented by the following setting: Include in the fraudster the behaviour of sensing customers proximity, then execute the acquisition or cloning of a card and later purchase goods from another merchant with the information from the customer. An alternative way to implement this scenario could be when a merchant is compromised in different ways (e.g. by hacking) and allow a fraudster to steal information from all customers that have been served there on a massive scale. The volume of fraudulent activity can be modelled changing the specific parameter of number of theft and merchant affected, zip code and frequency of use for purchasing. A "red flag" for detection in this case could be similar as previous case, a high number of unusual transactions with high value in a short period. Other methods such as simultaneous payments in different physical locations, or using the card far from previously known locations, could also be flagged.

### 5.3. Internet purchases

This scenario includes cases where the fraudster uses a method called *Carding* to purchase immaterial goods, e.g. music files, redeemable coupons, tickets etc. on the Internet using websites that check the validity of the card instantly. This is to ascertain whether the card data is still valid without having to run the risk of getting caught when using the card while physically present. Similar to cloned cards the customer keeps the original card but without knowledge of the situation. In terms of the object model used in BankSim the cloned card scenario can be implemented by the following setting: Include in the fraudster the behaviour of sense customers proximity, then execute the acquisition of the important information of a card and later on proceed with the method of Carding, to check for validity. A "red flag" for detection in this case could be to have a black list of Carding websites and proceed to cross this information with current user activity to detect any unusual purchases after the Carding was executed.

## 6. MODEL AND METHOD

The design of BankSim was based on the ODD model introduced by Grimm et al. (2006). ODD contains 3 main parts: *Overview*, *Design Concepts* and *Details*.

### 6.1. Overview

#### 6.1.1. Purpose

We aim to produce a simulation that resembles a bank payment system. Our main purpose is to generate a synthetic data set of commercial transactions that can be used for the development and testing of different fraud detection techniques.

If we want to use the real original data for the development of fraud detection methods, it often happens that is difficult to find diverse and enough cases of fraud. However this is not the case of a simulated environment, where fraud can be injected following known patterns of fraud and flagged for easy recognition and evaluation of the performance of the detectors.

#### 6.1.2. Entities, state variables and scales

There are three agents in this simulation: *Merchant*, *Customer* and *Fraudster*.

**Merchant** This agent serves the customer with one category of merchandise specified by the original data. It offers products or services according to the statistics obtained from the specific zip code and time (week, day of the week and/or hour). They are waiting for customers to request products and register the payments.

**Customer** This agent's main objective is to satisfy a need for one of the 16 categories and purchase goods or services from merchants. They possess a payment method which in this case we will be generalised as a credit card.

**Fraudster** The behaviour is determined by the goal of defrauding the customers and/or merchants. The specific behaviour can be extended to fulfil different patterns and can mutate depending on the specific fraud behaviour we are interested in studying. Some of the known fraud behaviour is presented in section 5.

#### 6.1.3. Process overview and scheduling

During a normal step of the simulation, a customer that enters the simulation can decide to purchase an item or service from one of the offered categories. Once the category has been selected, it senses nearby merchants that offer that category and listen to the offers from the merchant. If accepted (with a certain probability of rejection) the transaction takes place and the merchant registers the payment.

The time granularity of the simulation is that each step represents a day of commercial activity, but the original data is so rich that this can be modified to the specific hour of the day. So a normal week has 7 steps and a month will consist of around 30 steps. Notice that in the future we can choose to make the distinction between specific days of the week explicit, since the information from Bank Inc. is good enough to obtain statistics from it. But for now, we are not taking specific day of the week into account to feed the consumption pattern and we treat all days the same.

## 6.2. Design Concepts

The *basic principle* of this model is the concept of a commercial transactions. We can observe an *emergent* social network from the relation between the customers and the merchants. Each of the customers have the *objective* of purchasing articles from the merchants. The merchants *objective* is to serve the customers and commit the payment that result into the generation of a synthetic data set. In our virtual environment the *interaction* between agents is always between merchant and customer. Purchasing articles from another customer or selling articles to another merchant is not included in our model.

Customers can scout for the merchants in any radial direction from their current position in the virtual world and search for a merchant that matches its category selection. If no merchant is found then the transaction can not take place, and the step for this customer ends.

The agents do not perform any specific learning activities. Their behaviour is given by probabilistic Markov models where the probabilities are extracted from the real data set.

## 6.3. Details

### 6.3.1. Initialization

The simulation starts with a number of merchants that match the categories of what a specific zip code offers, an initial number of customers and fraudsters.

### 6.3.2. Input Data

BankSim has different inputs needed in order to run a simulation. The input data concerns the distributions of probabilities for each of the merchants, and the consumer pattern behaviour of the customers specified by gender and age. The items that can be purchased are all grouped into a category using the statistic measures for the payments.

For setting the parameters, we use a parameter file that is loaded as the simulation starts, it contains zip codes that we want to simulate and the malicious parameters. Some parameters can also be set manually in the GUI. The zip codes are queried against the API of the bank and we retrieve information corresponding to the customers: quantity, age and gender distribution. We also query the merchants and obtain sales distributions for each of the merchant categories.

### 6.3.3. Submodels

Figure 1 shows the different use cases of the agents including the misused cases for the fraudsters. This model represent the different actions that an agent can take inside the system.

**Find Merchant** The first step in a simulation for a customer is to find a merchant, each agent decides which category of service they will want to find, so the next step is to sense the environment and find a merchant that provides the category selected. Next search by the customer starts here, i.e. the customers move from merchant to merchant.

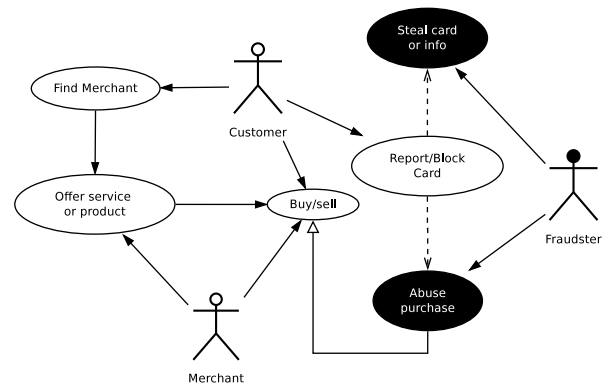


Figure 1: BankSim Use Case Diagram including misuse cases

**Offer service or product** Is performed by the merchant and once a merchant is approached by a customer, it offers a product or service according to the demand specified on the parameters for each category.

**Buy/Sell** Once a customer finds a merchant and after a merchant offers a product, a transaction takes place and it stores the required information for the generation of the synthetic data of transactions.

**Steal card or info** Fraudsters move around the environment of the simulation and find customers to steal the physical card or just the important information of the customer credit card. This information is stored for later use. In this misuse case we aim to emulate the behaviour of a criminal performing a cloning of a card or just stealing the card.

**Abuse purchasing** This misuse case is performed by Fraudsters, they make purchases of goods or services on physical merchants or internet merchants that hides their physical presence.

**Report/Block Card** This use case is performed by Customers, when they realise that abusive behaviour is committed on their accounts, they report the case to the bank and block the card for further abuse.

**Log of transactions** Each time an item or service is purchased from a merchant a transaction is created. A log contains the information about the customer, merchant, amount, location, date and fraud if any.

## 7. RESULTS

BankSim uses the Multi-Agent Based Simulation toolkit MASON which is implemented in Java (Luke, 2005). MASON offers several tools that aid the development of a MABS. We justified our choice mainly for the benefits of supporting multi-platform, parallelisation, good execution speed in comparison with other agent frameworks; which is specially important for computationally

intensive simulations such as BankSim (Railsback et al., 2006). BankSim can be run with a GUI, that helps the user see the states and balance of the customers (purple dots) and easier identify the merchants (green circles) and fraudsters (red dots), as can be seen in the example in figure 2.

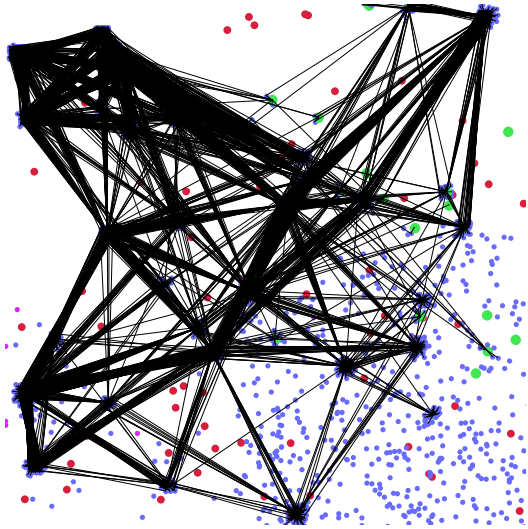


Figure 2: Screenshot of BankSim during a step

The output of BankSim is a CSV file that contains the fields: *Step*, *CustomerId*, *Age*, *Gender*, *zipCodeOrigin*, *merchant*, *zipMerchant*, *category*, *amount* and a special field to flag fraudsters called *fraud*.

### 7.1. Simulated scenarios

We aimed to perform a simulation that would produce a comparable data set to our sample data set which contained payments for over 6 months to match the original data. The simulation was loaded with information from ZC1 (see table 1), which was selected due to the highest amount of payments.

We ran BankSim for 180 steps (approx. six months), several times and calibrated the parameters in order to obtain a distribution that get close enough to be reliable for testing. We collected several log files and selected the most accurate. We injected thieves that aim to steal an average of three cards per step and perform about two fraudulent transactions per day. We produced 594643 records in total. Where 587443 are normal payments and 7200 fraudulent transactions. Since this is a randomised simulation the values are of course not identical to original data.

The result of the simulation for normal transactions is summarised in tables 5, 6 and 7. Remember that the codes for age categories are given in table 2 and gender codes are given in table 3. All prices given are in euro.

### 7.2. Evaluation of the model

We begin the evaluation with the verification and validation of the generated simulation data (Ormerod and

Table 5: Simulated ZC1

| gender | age | payments | avgAmount |
|--------|-----|----------|-----------|
| E      | U   | 1171     | 34.02     |
| F      | 6   | 13795    | 32.13     |
| F      | 5   | 33574    | 31.47     |
| F      | 4   | 57835    | 31.74     |
| F      | 3   | 77333    | 31.97     |
| F      | 2   | 103112   | 32.14     |
| F      | 1   | 32340    | 32.09     |
| F      | 0   | 1818     | 34.75     |
| M      | 6   | 12718    | 31.57     |
| M      | 5   | 28382    | 31.35     |
| M      | 4   | 49780    | 31.99     |
| M      | 3   | 67870    | 31.83     |
| M      | 2   | 81690    | 31.48     |
| M      | 1   | 24924    | 31.84     |
| M      | 0   | 586      | 33.36     |
| U      | 3   | 173      | 32.28     |
| U      | 2   | 164      | 28.83     |
| U      | 1   | 178      | 33.23     |

Table 6: Categories Simulated ZC1

| category             | payments | perc   | avgAmount | std    |
|----------------------|----------|--------|-----------|--------|
| Accommodation        | 1196     | 0.002  | 106.55    | 69.34  |
| Bars and restaurants | 6253     | 0.0105 | 41.15     | 29.55  |
| Books and press      | 885      | 0.0015 | 44.55     | 33.14  |
| Fashion              | 6338     | 0.0107 | 62.35     | 44.36  |
| Food                 | 26254    | 0.0442 | 37.07     | 25.00  |
| Health               | 14437    | 0.0243 | 103.74    | 76.87  |
| Home                 | 1684     | 0.0028 | 113.34    | 83.23  |
| Hypermarkets         | 5818     | 0.0098 | 40.04     | 27.96  |
| Leisure              | 25       | 0      | 73.23     | 20.91  |
| Other services       | 684      | 0.0012 | 75.69     | 54.59  |
| Sports and toys      | 2020     | 0.0034 | 88.50     | 63.13  |
| Technology           | 2212     | 0.0037 | 99.92     | 73.49  |
| Transport            | 505119   | 0.8494 | 26.96     | 17.53  |
| Travel               | 150      | 0.0003 | 669.03    | 494.90 |
| Wellness and beauty  | 14368    | 0.0242 | 57.32     | 41.48  |

Table 7: Fraud Simulated ZC1

| Fraud | payments | per   | total       | per   | avg    | std    |
|-------|----------|-------|-------------|-------|--------|--------|
| 0     | 587443   | 98.78 | 18708432.56 | 83.03 | 31.84  | 31.47  |
| 1     | 7200     | 1.21  | 3822671.17  | 16.96 | 530.92 | 835.52 |

Rosewell, 2009). The verification ensures that the simulation correspond to the described model presented by the chosen scenarios. We described BankSim in section 6. In our model, we have included several characteristics from a real payment system, and successfully generated a distribution of payments that involved the interaction of merchants and customers.

The validation of the model answer the question: *Is the model a realistic model of the real problem we are addressing?* After several runs of the simulation to calibrate it, we are able to answer that question affirmatively. We present a table summarising the generated data in tables 5, 6 and 7.

Table 5 can be compared with table 1, both tables compare the distribution of payments by gender and age. Similar values are found in both tables because we created the agents based on gender and age distribution of the zip



code. However, we did not programme the consumption behaviour of agents based on gender and age. This is because we did not have the statistic standard deviation for the consumption patterns per age and gender, we only have the average. This affects the results, despite that in the overall results we find similar data. But we think the missing information from the real system can be found with further calibration that is at the moment beyond the scope of our work. Figure 3 shows a distribution of gender and age from our simulated data.



Figure 3: ScatterPlot Payments vs Age/Gender

Table 6 is comparable to table 4. We succeed in generating a distribution of categories that resembles the real data. We matched the percentage of categories and simulated similar average and standard deviation to the ones present in the original data. One thing to notice is that the category *auto* did not get any transaction during the simulation, this could be due to the location of the merchant in the environment being random and was perhaps far enough to be hidden from customers that wanted to purchase from this category. A box plot of the simulated categories is shown in figure 4. Since the values of travel are bigger than other categories, we decided to draw the box plot omitting this category in figure 5 to improve the visualization of the simulated data.

The simulated fraud behaviour is presented in table 7. The total amount stolen was around 3.8 million Euros which corresponds to a rather high crime rate of nearly 17% of the total amount of payments. We programmed an aggressive behaviour where few transactions (only 7200 and 1.2% of total) could defraud 17% of the payments with an average of 530 Euros per fraud. For the purpose of fraud detection there is a benefit from the occurrence of enough cases of fraud that can help the investigators to gather the evidence needed to prosecute the criminals. In our case we benefit from the abundance of fraud cases because many detection methods need enough data to train better a classifier that can detect the fraud behaviour.

So in summary, our agent model with its programmed

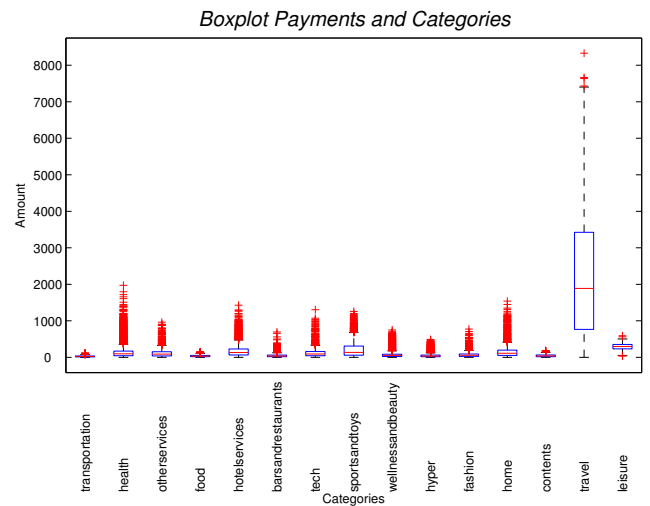


Figure 4: BoxPlot of a BankSim simulation

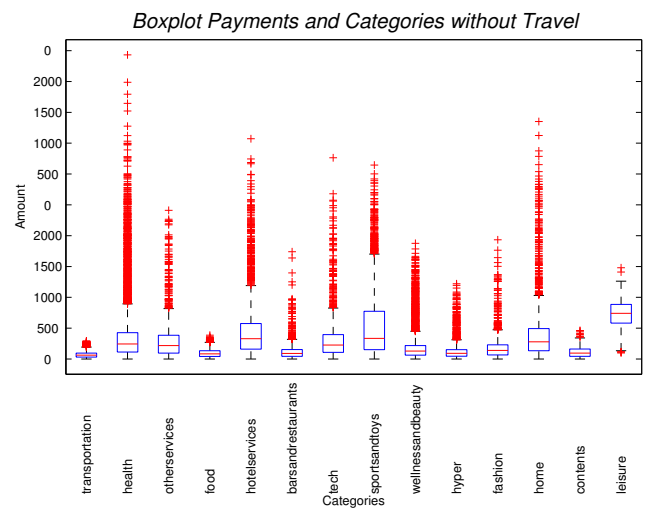


Figure 5: BoxPlot of a BankSim simulation without category Travel

micro behaviour, produces the same type of overall interaction that we can observe in the original data, and furthermore, this interaction give rise to the same macro behaviour for the whole zip code as for a real situation as well.

Since we are running a simulation we argue that the differences are not significant for our purpose, which is to use this distribution to simulate the normal behaviour of payments, and simultaneously combine this with injected anomalies and known patterns of fraud.

## 8. CONCLUSIONS

BankSim is a simulation of bank payments with the objective to generate a synthetic transactional data set that can be used for research into fraud detection. The data sets generated with BankSim can aid academia, financial organisations and governmental agencies to test their fraud detection methods or to compare the performance of dif-



ferent methods under similar conditions using a common public available and standard synthetic data set for the test.

In section 3. we formulated our research question: *How could we model and simulate a bank payment system and generate a realistic and reliable synthetic data set for the purpose of fraud detection?*

In section 6. we presented the model for BankSim, which is based on the ODD methodology. In order to better support our claim and answer our research question we analysed the type of data needed to generate and output as a CVS file (see section 7.) and we evaluated and verified our model in section 7.2.

It is important to know how much information from the real data set is contained in the generated synthetic data. First we do not have access to any specific record of who is purchasing anything and neither the merchant involved in the transaction. We based our simulation purely on the aggregated statistical measures present in the original data that give us an approximate description of how the individual agents behave. This means that Bank Inc. can be sure that the privacy from the customers is preserved when using BankSim.

We argue that BankSim is ready to be used as a generator of synthetic data sets of financial activity of a payments. Data sets generated by BankSim can be used to implement fraud detection scenarios and malicious behaviour scenarios such as a stolen or cloned credit cards or unusual simultaneous activity of purchase in different physical locations. We will make a stable release of BankSim available to the research community together with standard data sets developed for this article and further research.

For the future we plan several improvements of and additions to the current model. BankSim can be calibrated to improve the results presented in section 7. and increase the granularity and the coverage of zip codes that enrich the synthetic data set and make it even more valuable as a realistic data set for fraud detection.

In order to generate records with malicious behaviour we plan to extend BankSim to also generate malicious activity that can come from the merchants, customers, different fraudsters or combinations of these.

Among the additions we consider are: increase the step granularity and add to the simulation more zip codes simultaneously. We intend to make BankSim a complete bank system by adding other bank transactions such as deposit, withdraws and transfers besides the current payments. Unfortunately for this addition there is a lack of real data that we can use for this purpose, but hopefully in the future we will find financial institutions interested in our project that are willing to share this data.

## REFERENCES

Naoki Abe, Bianca Zadrozny, and John Langford. Outlier detection by active learning. *Proceedings of the 12th ACM SIGKDD international conference on Knowl-*

*edge discovery and data mining - KDD 06*, page 504, 2006. doi: 10.1145/1150402.1150459.

SJ Alam and Armando Geller. Networks in agent-based social simulation. *Agent-based models of geographical systems*, pages 77--79, 2012.

R.J. Bolton and D.J. Hand. Statistical fraud detection: A review. *Statistical Science*, 17(3):235--249, 2002.

Volker Grimm, Uta Berger, Finn Bastiansen, Sigrunn Eliassen, Vincent Ginot, Jarl Giske, John Goss-Custard, Tamara Grand, Simone K. Heinz, Geir Huse, Andreas Huth, Jane U. Jepsen, Christian Jørgensen, Wolf M. Mooij, Birgit Müller, Guy Pe'er, Cyril Piou, Steven F. Railsback, Andrew M. Robbins, Martha M. Robbins, Eva Rossmanith, Nadja Rüger, Espen Strand, Sami Souissi, Richard a. Stillman, Rune Vabø, Ute Visser, and Donald L. DeAngelis. A standard protocol for describing individual-based and agent-based models. *Ecological Modelling*, 198(1-2):115--126, September 2006. ISSN 03043800. doi: 10.1016/j.ecolmodel.2006.04.023.

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 and Stefan Axelsson. Money Laundering Detection using Synthetic Data. *The 27th workshop of Swedish Artificial Intelligence Society (SAIS)*, pages 33--40, 2012a.

Edgar Alonso Lopez-Rojas and Stefan Axelsson. Multi Agent Based Simulation (MABS) of Financial Transactions for Anti Money Laundering (AML). *The 17th Nordic Conference on Secure IT Systems*, pages 25--32, 2012b.

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.

Dan Magnusson. The costs of implementing the anti-money laundering regulations in Sweden. *Journal of Money Laundering Control*, 12(2):101--112, 2009. ISSN 1368-5201. doi: 10.1108/13685200910951884.

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](http://www.grantthornton.com/staticfiles/GTCom/files/Industries/Consumer&industrialproducts/Whitepapers/Revivingretail_Strategiesforgrowthin2009.pdf).

Paul Ormerod and Bridget Rosewell. Validation and Verification of Agent-Based Models in the Social Sciences. In Flaminio Squazzoni, editor, *LNCS*, pages 130--140. Springer Berlin / Heidelberg, 2009. ISBN 978-3-642-01108-5.

Clifton Phua, Vincent Lee, Kate Smith, and Ross Gayler. A comprehensive survey of data mining-based fraud

detection research. *Arxiv preprint arXiv:1009.6119*, 2010.

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.

## **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.

### **Dr. Stefan Axelsson**

Stefan Axelsson is a senior lecturer at Blekinge Institute of Technology. He received his M.Sc in computer science and engineering in 1993, and his Ph.D. in computer science in 2005, both from Chalmers University of Technology, in Gothenburg, Sweden. His research interests revolve around computer security, especially the detection of anomalous behaviour in computer networks, financial transactions and ship/cargo movements to name a few. He is also interested in how to combine the application of machine learning and information visualization to better aid the operator in understanding how the system classifies a certain behaviour as anomalous. Stefan has ten years of industry experience, most of it working with systems security issues at Ericsson.

# A MULTI-AGENT MODEL OF TYPICAL COMPETITIVE BIDDING PROCESS IN CONSTRUCTION

Ronald Ekyalimpa<sup>(a)</sup>, Simaan AbouRizk<sup>(b)</sup>

<sup>(a)</sup>University of Alberta, 5-047 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA

<sup>(b)</sup>University of Alberta, 5-080 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA

<sup>(a)</sup>[rekyalimpa@ualberta.ca](mailto:rekyalimpa@ualberta.ca), <sup>(b)</sup>[abourizk@ualberta.ca](mailto:abourizk@ualberta.ca)

## ABSTRACT

At some point in the lifetime of a construction contractor company, it has to engage in a competitive bidding process as a means of acquiring work. The competitiveness of its bid ultimately determines whether or not it is awarded the project. However, prior to this, the company has to make a decision on whether or not to bid on the project. Making this decision is usually challenging, even for experienced practitioners. Moreover, generating a competitive bid is a non-trivial process. These challenges are attributed to the stochasticity and dynamics that surround the bidding process in the construction industry. A Multi-Agent Model is proposed in this paper that emulates the typical bidding process and environment. A design for the model is first presented then details on how it was implemented with AnyLogic simulation system are explained. An experiment is run to demonstrate how the model can be used.

Keywords: competitive bidding, multi-agent model, AnyLogic, construction industry

## 1. INTRODUCTION

Competitiveness in the construction domain is a phenomenon that is broad and very complex. It can be assessed from different perspectives and at different levels: by the owners, the users of the facilities built by the companies, or by the companies themselves. Competition can also be studied at an industry level or at an individual company level. At each of these levels, issues are different, but competition at each level is inter-dependent and equally important for the benefit of individual companies and the industry at large.

Some scholars argue for a top-bottom approach (i.e. solve industry problems first and then address individual companies) in tackling competitiveness, while others advocate a bottom-top approach. The authors propose the adaptation of a holistic concurrent approach that deals with prevailing issues in a systematic fashion at both levels. Although the sequence in which competitiveness problems are addressed at the two different levels is important, it is not the focus of this study and will not be discussed further. An approach that can be adopted to experiment

with competitiveness issues at an individual company level is presented.

Various methods have been proposed and used in the past to manage the performance of a company. Examples include the Balanced Score Card (BSC) (Kaplan and Norton 1996), Key Performance Indicators (KPIs), and EFQM Excellence Model (European Foundation for Quality Management - EFQM 1999). These techniques have a number of pitfalls: 1) they can only tell what might have gone wrong in the past but give no insights into how things might turn out in the future if changes are implemented; 2) some are self-assessing, so they don't compare performance to competitors in the industry or globally; 3) in some techniques, the perspective from which performance is assessed is very limited and certain key measures are left out; and 4) they are not automated, and hence, are cumbersome to use.

The authors propose a simulation-based performance management approach that addresses these problems. This simulation-based approach is comprised of a number of components, one of which is an Agent-Based Model (ABM). The main focus in this paper is to discuss the design and specifications of the ABM for bidding in the construction industry, and its implementation in AnyLogic simulation system. The verification and validation work for this model was successfully done but is not elaborated due to space constraints; highlights of how verification and validation processes were carried out are presented.

The developed model can be used in practice in various ways. First, it can be used to guide individuals in-charge of running operations at a construction company on workload planning issues to achieve a balance between running their companies over capacity and below capacity. The model can also be used to guide practitioners on which type of bidding strategies to adopt to ensure that they match that capacity. The simulation model can be extended to gain insights into the type of projects a construction company should go after in order to match its internal competencies and capacity.

### 1.1 Simulation-Based Approach for Contractor Performance Management

In order to objectively assess an individual company's overall performance, the entire life cycle of the business operations at the company must be considered. This involves examining the process through which the company competes for work and how it executes the work that it has been awarded. The former is referred to as the front-end of the company's business operations, while the latter refers to the tail-end of its business operations. To develop a simulation model that can be used to manage a company's performance, the company's core processes must be abstracted accurately. Figure 1 summarizes a concept model proposed for implementing the simulation model. This model illustrates the components envisaged as part of the developed system. It also details the communication protocols to take place between these components.

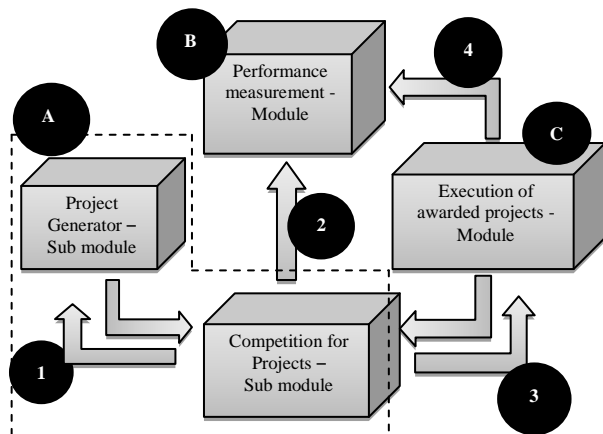


Figure 1: Layout of Simulation-Based Performance System Components/Modules

The letters “A,” “B” and “C” represent the modules that exist in the company simulation system. “A” represents a “Tendering” Module. “B” and “C” represent the “Performance Measurement” and “Operations” Modules, respectively. The Tendering Module models project arrivals and the competition for these projects by companies operating within a virtual construction industry. The Operations Module processes projects awarded to the “Company of Interest.” The Performance Measurement Module collects observations on performance measures from all other modules and generates an overall performance rating for the company at the end of the simulation. The numbers 1-4 represent the communication that takes place between the modules during the simulation: “1” represents bid submissions and companies being notified of the winning bid, “2” represents the collection of statistics/observations on tendering performance as the simulation progresses, “3” represents communication between the Operations Module of the Company of Interest to the user and the Tendering Module. Information transferred includes data on prevailing conditions in the Operations Module at project arrival so that this can aid with the Company of

Interest's bid/no-bid decision, and it represents notification of the Operations Module of projects that have been awarded to the Company of Interest and that need to be processed. The number 4 represents the collection of performance measures (e.g. quality, production efficiency, safety, cost slippage, schedule slippage, etc.) as the simulation advances.

Figure 2 shows details for each of the components and how they interact with each other. This figure also shows the simulation method intended for implementation of each component and the integration of these components into a distributed simulation environment using a synthetic environment, COstruction SYnthetic Environment (COSYE) (AbouRizk and Hague 2009). The Tendering Module is implemented using an Agent-Based approach while the Operations and Performance Modules are implemented using a discrete event simulation (DES) approach.

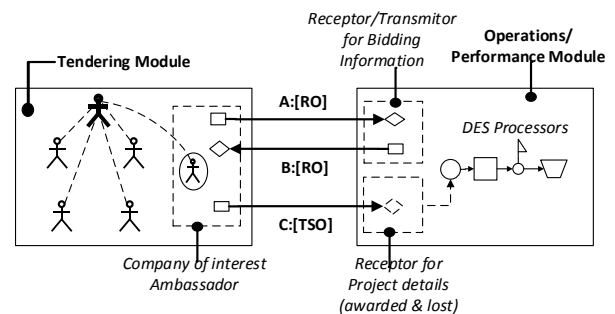


Figure 2: Modeling Paradigms Used to Implement Simulation-Based Performance System

Presenting design and implementation details of the entire company simulation system is not possible in one paper, so only one component of the system will be covered: the system that models the front-end of the business. This front-end system models a typical competitive bidding process involving a finite number of companies operating in a virtual industry.

### 1.2 Competitive Bidding in Construction

Bidding is an offer (often competitive) of a price that one is willing to pay for something or for a demand that something be done (Oxford Press 2014). This process is one through which most work/projects get awarded to contractor companies within the construction domain. It is one of those processes that are least understood by members of the construction domain in both academia and practice, hence the vast amount of research that has been done on the subject. Two threads of research have emerged in the subject of competitive bidding. The first attempted to propose models that emulate the bid/no bid decision that companies undertake on a project-by-project basis. The second thread strived to generate ways of enhancing the chance of a bidder being awarded a project. In principle, this thread delivered research work that can be further sub-categorized into: those that estimated the chance of winning a bidding context and those that proposed ways of estimating the markup to carry in a bid to guarantee success.

### 1.2.1 Bid/No-Bid Decision

When a company encounters a potential project, it will perform a formal or informal evaluation to decide whether or not to pursue that project. This decision is not a trivial one and for this reason has resulted in a number of research studies that have attempted to propose algorithms that can be used to guide this decision making process. Examples of these studies include work done by Egemen and Mohamed (2007; 2008), Bagies and Fortune (2006), Lin and Chen (2004), and El-Mashaleh (2010). Although the approaches proposed in these studies are viable, they pose application limitations, especially due to the number of input factors required to make the decision. Moreover, implementing them in a computer-based simulated environment is likely to be quite challenging as a result of the static nature of some of the variables proposed for use in the decision making process.

Consequently, this study adopted a more robust and easy-to-apply approach for the bid/no-bid decision. It was based on a comparison of project attributes with company tolerance levels for these attributes (i.e. internal company policies) and available production capacity at the time a potential project opportunity is realized.

### 1.2.2 Winning Probability and Markup Generation Algorithms

Research done in the area of competitive bidding along the second thread was most likely started in the 1950s by Lawrence Friedman when he first proposed a probabilistic approach for estimating the chance of a competitor winning a bid given that their bid price is set to a specified value (Friedman 1956). Since that time, other mathematical models for tackling the same problem have emerged. Some of these have been for Friedman's approach, while others have been against it. There are numerous examples of these studies published in journal and conference papers (Benjamin 1972, Dixie 1974, Gates 1967, Morin and Clough 1969, Rosenshine 1972, Fuerst 1976, Wade and Harris 1976, Griffis 1992, Ioannou 1988).

It is worth noting that from a practical applications perspective, most of these studies have major pitfalls. For example, they heavily relied on an assumption that the analyst has a good understanding of their competitor's behavior. In other words, the analyst would need to know how their competitors bid on past competitions and assume that they would stick to the trend. Bid price data is not always accessible and contractors change their behavior from time-to-time depending on the conditions/situation. This makes these models limited from an applications stand-point.

There have been studies done within the construction domain that endeavored to overcome this challenge by attempting to predict the markup to carry in a bid to guarantee success. Examples include work done by Hegazy and Moselhi (1994), Dozzi, AbouRizk, and Schroeder (1996), Marzouk (2002), Li (1996), Fayek (1998), and Cui and Hastak (2006). The

techniques used in these studies included fuzzy logic, artificial neural networks, genetic algorithms, system dynamics, utility theory, etc.

These studies overcame the problems of the lack of past bidding data with the exception of the artificial neural network based algorithms. However, implementing these in a computer-based simulation environment would result in a model that is cumbersome to develop and use as a result of the vast number of input variables required to generate an output. In addition, some factors proposed by the authors are static in nature and in some cases don't represent the reality in bidding practices. As a result, a Monte Carlo simulation-based algorithm was utilized in this study. The approach was first proposed by Winston in 2001 in a book that he wrote on Simulation Modeling Using @Risk (Winston 2001). We modified and extended his model to closely match reality and to suit our development work.

## 2. MODELING APPROACH ADOPTED: ABM

Different methods exist for abstracting and analyzing systems with simulation, but the ultimate choice of method is guided by the nature of the problem at hand, presence of sufficient knowledge and skills in the method and access to a simulation system that supports the method.

The problem in this study is best analyzed using simulation methods because it is stochastic and dynamic in nature. The stochastic nature arises from the fact that output for the system will vary even if the input is the same each time the system is experimented with. The dynamic nature is due to the fact that in each run, future events are affected by past and current events.

In the bidding problem, time between new project arrivals varies stochastically. This stochastic behavior is in part attributed to variations in investment practices amongst owners within a given construction industry. The stochastic behavior is also applicable to the attributes of these projects e.g. cost, duration, complexity, etc. The decision of companies to bid these projects is also stochastic in nature. In fact, this decision has dynamic aspects to it to some extent. This is because a company's decision to bid or not to bid on a project will depend on what happened in the recent past i.e. was it awarded projects or not and how many, and what is happening at the current time i.e. what types of projects are up for bid, what are its internal policies, etc. These facts make the problem suitable for modeling using a simulation-based approach.

Given that the bidding problem comprises a number of distributed constructs such as the community of owners and the construction contractor companies, which operate in an autonomous or semi-autonomous fashion within the same environment, i.e. the construction industry, the problem lends itself to an ABM paradigm. This assertion is made because the construction industry can be mapped to a virtual environment within which agents thrive. The autonomous self-executing constructs (community of

owners and construction contractor companies) can be directly mapped to agents that encapsulate their corresponding behavior. Moreover, the ABM would be developed using simple rules which would result in more complex emergent behavior as agents interact; something that is evident in the course of competition for projects by companies in the construction industry. An ABM paradigm was therefore chosen for use in modeling this problem.

The community of owners is represented by a bid manager agent. This agent is modelled as a singleton. Companies are modelled using three types of agents: “small company agent,” “medium company agent” and “large company agent.”

### 2.1. ABM Design and Specifications

There are numerous tools available for the design and specification of ABMs. It is usually advised that the modeler undertake a comprehensive design process before they embark on developing their ABM, regardless of whether it is a simple or complex one. In this way, the designs can be assessed for accuracy (validity), which in turn guarantees the generation of reliable models. In this study, design aides used include a block diagram and a sequence diagram.

#### 2.1.1 UML Class Diagram for the ABM

A block diagram summarizes the constructs in the real world system that is being abstracted which map onto classes and objects in the model. It also documents the attributes and operations for these constructs. The relationship between these classes is also documented. This concept is adopted from a Universal Markup Language (UML). Figure 3 summarizes the block diagram for the tendering module.

The Figure summarizes the agents (Company and Bid Manager Agent) in the model along with the container in which these agents thrive (Main Agent). A composition relationship exists between the Company Agent and the Main Agent and the Bid Manager Agent and the Main Agent. The diagram also documents a class referred to as a “project entity,” which wraps information about a specific project under tender. Instances of this class are passed around between the Bid Manager Agent and Company Agents in the course of their communication. An association relationship exists between project entity and the Bid Manager Agent and project entity and the Company Agents.

#### 2.1.2 Sequence Diagram for the ABM

Communication between agents is an essential part of ABM. It is advised to design the entire communication protocol between agents for every ABM before development work commences. Sequence diagrams are typically used to design and specify the communication protocols existing between agents in a planned model. Communicating agents could belong to the same type or may be of different types. In this model, it was envisaged that communication would take place between Bid Manager Agent and Company Agents. No

communication was to take place between Company Agents. This was done deliberately to avoid situations identical to collusion which conflict bidding ethics and practices in a practical bidding setting. The sequence diagram designed for this communication is presented in Figure 4.

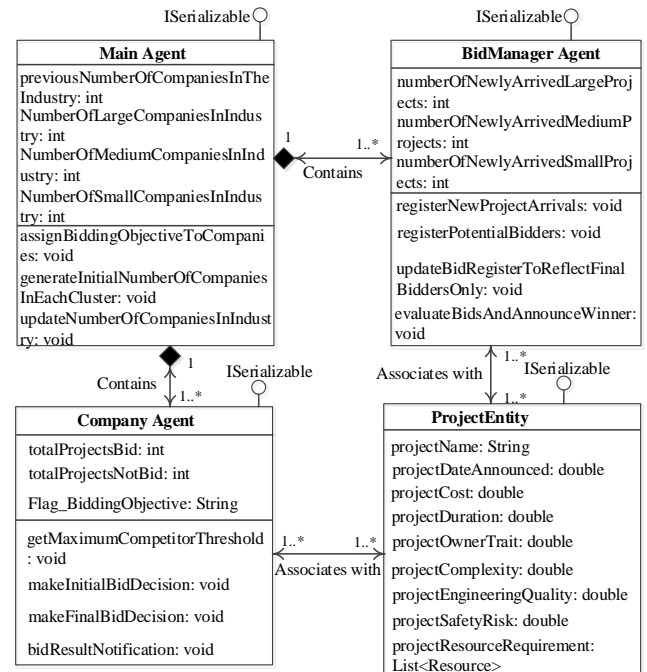


Figure 3: A UML Class Diagram for the Bidding ABM

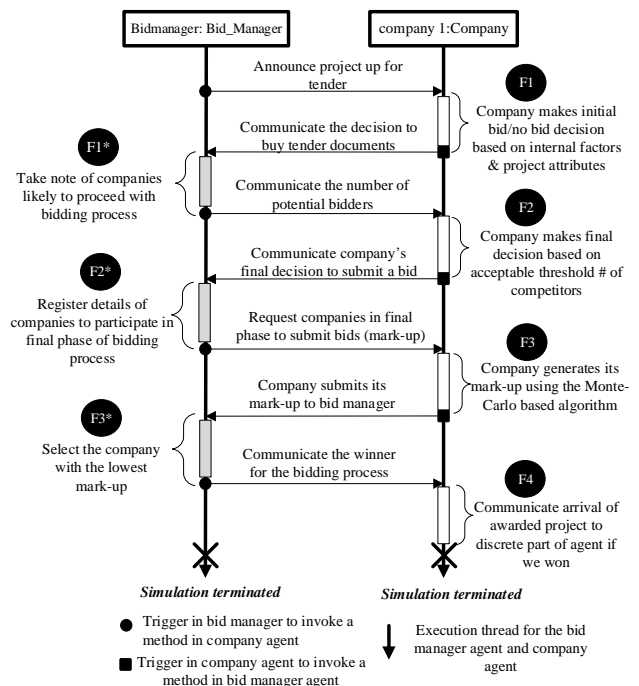


Figure 4: Sequence Diagram Detailing Communication between Bid Manager Agent and Company Agents

For simplicity, the company agents are represented as one block. Only one cycle of communication that is typical of information exchanged between Bid Manager



Agent and Company Agents is presented in this figure. The entire cycle of communication between Bid Manager Agent and Company Agents is triggered by the arrival of a project into the construction industry and terminated when the project is awarded to a winning bidder. There is no apparent advancement in simulation time in the course of communication. Furthermore, the communication between Bid Manager Agent and Company Agents is asynchronous: either agent cannot send two messages in sequence without hearing back from the other agent. The design specifications summarized in this sequence diagram served as a basis for implementing methods within each respective agent. The methods in the Bid Manager facilitate it to announce newly arrived projects to all Company Agents, receive bids from companies that decide to bid, evaluate them and announce the winner. The Company Agent on the other hand has methods that enable it to decide which projects to bid on, generate a bid price for those that it chooses to bid and pass awarded projects to its discrete event process flow model for execution.

## 2.2 Agents within the Competitive Bidding Model

When a multi-agent based approach is used to abstract and analyze a system, constructs within that system that are autonomous, have intelligence and a memory, are represented as agents. In this model, the construction industry (Main Agent) and companies (small, medium and large) are represented as agents. The community of owners is also represented as an agent i.e. the Bid Manager.

### 2.2.1 Bid Manager Agent

The designs for the Bid Manager and all other agents were implemented within the AnyLogic simulation system. The term “Bid Manager” is one that was coined in this study to refer to a construct that is used in the model to represent the community of owners and the role that they play within a typical construction industry. The owners are the party that primarily invests in construction projects. The owners are also in charge of the process required to select a contractor to perform the project—the bidding process.

To model the dynamics of projects arriving in a construction industry, the Bid Manager Agent samples the project arrival events from a statistical distribution. The attributes of these projects also get assigned to a project after they are sampled from statistical distributions. Large, medium and small projects are generated independently. Figure 5 shows the modeling elements (functions) required to achieve the Bid Manager Agent’s behavior.

An event modeling element is embedded within the Bid Manager Agent (see Figure 6) that schedules the arrivals of new large projects by drawing inter-arrival times from a predefined uniform distribution. The parameters in Figure 6 are used to set values with which beta distributions are constructed for generating cost and duration attributes for new projects.

Figure 7 shows the controls embedded within the Bid Manager Agent for defining the resource requirements and setting other project attributes. The controls presented in Figures 6 and 7 are those dedicated to modeling new large projects only. Similar controls exist in this agent for small and medium size projects.

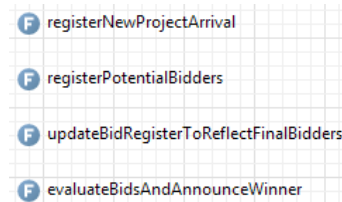


Figure 5: Functions Embedded in the Bid Manager

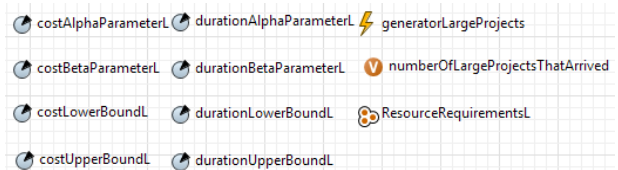


Figure 6: An Event Element and Parameters Used to Model New Large Projects

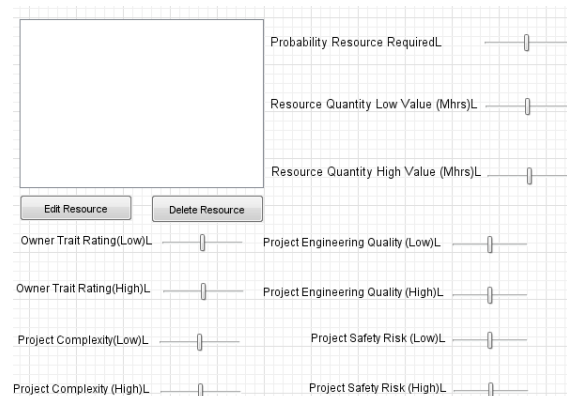


Figure 7: Controls for Modeling Project Resource Requirements and Other Project Attributes

### 2.2.2 Company Agents

Companies are modeled as autonomous intelligent agents that can decide on whether or not to bid on a project, generate a competitive bid if they decide to bid and process projects awarded to them. It was assumed that construction industries are comprised of small, medium and large companies.

Each of these company categories were modeled as an autonomous agent with unique attributes and behaviors. Company Agents are notified by the Bid Manager Agent when a project arrives in the virtual construction industry. The Company Agents then make internal decisions autonomously on whether or not to bid an announced project. If it decides to bid on the project, the Agent makes use of another algorithm to generate a bid price that it submits to the Bid Manager. Once a project is awarded to a company, it is passed on to the discrete event simulation model embedded within the Agent for processing. Figure 8 summarizes the

methods embedded within each Company Agent to exhibit the above behavior. These methods are consistent with those presented in the sequence diagram in Figure 4. These screen shots are taken from the model developed within AnyLogic simulation system.

- F makeInitialBidDecision
- F makeFinalBidDecision
- F submitFinalBid
- F bidResultNotification
- F getCompetitorNumberThreshold

Figure 8: Functions Embedded in Each Company Agent

### 2.2.2.1 The Bid Decision Making Process in Company Agents

The decision on whether or not to bid on a specific project is implemented using a hierarchical approach. The first phase is referred to as an initial bid decision phase and the last is referred to as a final bid decision phase. The initial bid decision is made on a set of criteria that include the trait of the project owner, project complexity rating, project safety risk ratings, engineering quality rating of the project, and availability of sufficient capacity to process the project.

In order for a Company Agent to decide to bid on a project, the attributes of that project, with respect to each of the above mentioned criteria, need to be within the tolerance limits predefined by the company. The final decision regarding whether the Company Agent will proceed to submit a final bid is based on whether the anticipated number of final bidders exceeds the company's set limit for the maximum number of bidders it is willing to compete against.

The screen shot presented in Figure 9 summarizes the parameters in the Company Agent that store the threshold values for its tolerances for each of the different project attributes. The first four parameters represent the definition of a company's production capacity.

- maximumNumberOfProjectsSMP
- maximumProjectQueueCapacitySMP
- maximumNumberOfProjectsLP
- maximumProjectQueueCapacityLP
- ownerTraitThresholdToleratedS
- projectComplexityThresholdToleratedS
- projectEngineeringQualityThresholdToleratedS
- projectSafetyRiskThresholdToleratedS

Figure 9: Parameters Embedded in Each Company Agent

### 2.2.2.2 Processing Projects Awarded to Company Agents

When a project is awarded to a Company Agent, it passes it on to the appropriate discrete event process

flow model for execution. The models are set up to be resource constrained. Projects are modelled as entities (i.e. flow units). Projects granted their requested resources are processed in a cyclic fashion on a day-by-day basis. After it has been processed for its duration, it releases the resource and is destroyed because it is considered to have been fully processed. Figure 10 presents the DES models used for processing small and medium size projects (top layout) and large projects (bottom layout).

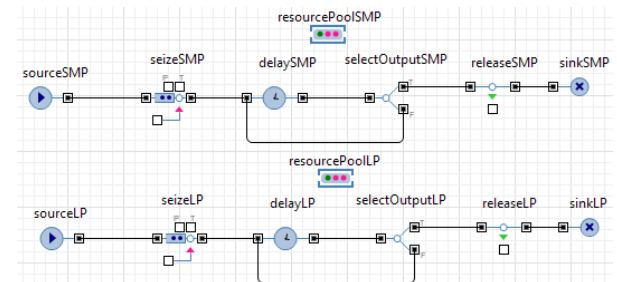


Figure 10: Discrete Event Models Embedded Within Each Company Agent

## 2.4 Construction Industry (Main) Agent

The Main Agent represents the top/global level of the model where all Agents thrive (Figure 11) in the AnyLogic simulation system.

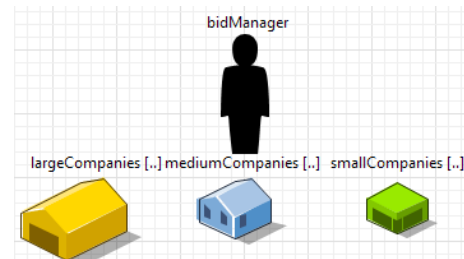


Figure 11: Agents Embedded within the Main Agent i.e. at the Top Level of the Model

Other controls and parameters are set up at this level. These are used, for example, for defining the total number of Company Agents in the industry and within each category. Other controls enable assigning bidding strategies to companies and their tolerance levels for different types of projects.

## 3. MODEL VERIFICATION AND VALIDATION

Verification and validation of the model developed in this paper was done in a fashion similar to that employed in most standard simulation models: trace output was generated in the console of the AnyLogic simulation system.

The trace captures details summarizing the chronology of simulation events processed during the tendering cycle for one project. The details match the logic that was intended to be simulated, hence, the model behavior was verified. For example, bid decisions are made soon after a project has been announced in the market, bid price generation follows that and then the project is awarded to the least bidder.

Validation work was also done which involved sensitivity analysis of the different model parameters. The simulated experiments on sensitivity analysis involved varying the number of projects generated within a specific period, varying the type of projects generated, varying the number of competitors, their bidding strategies and that of the company of interest. The experiments that involved these tests confirmed that the model was valid. Details of this work are not presented within this paper because of space constraints.

#### 4. EXPERIMENTAL SETUP

The objective of running this experiment is to determine a number of issues: 1) the number of projects that arrived in the industry within the simulated period; 2) the distribution of those projects i.e. relative number of small, medium size and large projects; 3) the quality of projects that arrived in the industry with respect to owner trait, complexity, engineering quality and project safety risks; 4) the distribution/allocation of those projects to companies in the industry i.e. proportions that went to small, medium and large companies; and 5) the reason(s) for the distribution/allocation of projects to these companies in the industry.

In order to set up and run the experiment, hypothetical values were chosen for the different model parameters and set prior to simulation. First, values used to model the rate of project arrivals are specified. Then distributions used to generate the attributes for each generated project are summarized. Finally, the tolerances that guide Companies' behavior in making decisions on which projects to bid or not are defined. The values for these parameters are specified as statistical distributions to ensure that there is variation in project instances and decisions made by companies, which is in-line with this phenomenon in a real-life setting.

##### 4.1 Rationale for Project Attribute Selection

When setting up the parameters that will be used to model the dynamics of projects arriving in the industry and the subsequent execution of these projects, attributes were selected so the model closely emulates the true behavior in the industry. From the perspective of work/projects, i.e. keeping all other conditions at a company constant and assuming that they are favorable, the dynamics of projects can still affect the performance of a company in two respects: 1) a company may start to perform badly when it starts suffering from extreme work conditions i.e. too little or too much work; and 2) even when a company strikes the right balance in its workload, its performance may still be affected by the type of work it acquired. If the work is generally bad (i.e. has bad owner traits, is poorly engineered, too complex or has high safety risks), the company may be more susceptible to failure.

##### 4.2 Model Setup and Industry Details

The simulation model was set up to run for 1,000 days. After this time was reached, the model terminated. The

unit of measure of time used in the model was days. A total of 20 Company Agents were created; 50% of these were large, 30% medium size and 20% were small companies. Values used to generate the attributes of newly created projects are summarized in Table 1.

Table 1: Model Inputs – Attributes for New Projects

| Parameter           | Unit of Measure | Small Projects | Med. Projects | Large Projects |
|---------------------|-----------------|----------------|---------------|----------------|
| Inter-arrivals      | Days            | 14~100         | 90~180        | 180~540        |
| Cost                | Million \$      | 10~100         | 100~300       | 250~1000       |
| Duration            | Days            | 300~540        | 450~750       | 600~1200       |
| Owner Trait         | Scale (0-1)     | 0.00~0.50      | 0.40~0.80     | 0.75~1.00      |
| Complexity          | Scale (0-1)     | 0.00~0.40      | 0.25~0.75     | 0.65~1.00      |
| Engineering Quality | Scale (0-1)     | 0.80~1.00      | 0.25~0.85     | 0.00~0.30      |
| Safety Risks        | Scale (0-1)     | 0.00~0.70      | 0.35~0.85     | 0.70~1.00      |

##### 4.4 Company Project Tolerances

Companies have internal policies regarding the type of projects they will bid and work on. These policies are referred to as the company's tolerances to specific types of projects. These tolerances are all inline attributes related to projects, including owner trait, complexity, engineering quality, and safety risks. The ranges of values for each of these criteria are summarized in Table 2. Each of these parameters/criteria is discussed in detail from the perspective of company tolerances for projects.

Table 2: Model Inputs – Company Tolerances for Projects

| Parameter           | Unit of Measure | Company Size |           |           |
|---------------------|-----------------|--------------|-----------|-----------|
|                     |                 | Small        | Medium    | Large     |
| Owner Trait         | Scale (0-1)     | 0.00~1.00    | 0.40~1.00 | 0.80~1.00 |
| Complexity          | Scale (0-1)     | 0.00~0.40    | 0.00~0.85 | 0.00~1.00 |
| Engineering Quality | Scale (0-1)     | 0.15~1.00    | 0.30~1.00 | 0.60~1.00 |
| Safety Risks        | Scale (0-1)     | 0.00~0.50    | 0.00~0.75 | 0.00~0.90 |

Owner trait indicates the degree of unnecessary interruptions to the contractor's work rhythm caused by the owner during project execution. This property is thought to affect the contractor's morale and in turn their productivity. Informal discussions with experienced practitioners in the construction industry revealed that this attribute plays a significant role when

a contractor is deciding whether or not to bid on a project.

The extent to which a project is engineered prior to construction and the quality of the engineering affect the production efficiencies during project execution. This attribute also affects the quality of work done i.e. the amount of rework experienced in a project. The quality of engineering for a project under tender can affect a contractor's decision on whether or not to bid on the project.

The complexity of a project along with the potential safety risks associated with executing a project can also affect a contractor's decision on whether or not to bid on a project. These are defined using scales ranging from 0 to 1. For the complexity attribute, a value of 0 indicates no complexity, while a value of 1.0 indicates an extremely complex project. A value of 0 for project safety risk indicates low likelihood of safety incidents while a value of 1.0 indicates a very high likelihood of safety incidents.

High values for engineering quality and owner trait for any given project are good, while low values are bad. On the other hand, high values for complexity and safety risk are bad, while low values are good.

The semantics discussed above apply to the demand side of projects. On the supply side, issues to do with company tolerances with respect to each of these attributes are dealt with. Rating scales identical to those used for the projects are used for each attribute (from 0 to 1). A low value (0) indicates that the company has an extremely low tolerance for the attribute, while a value of 1.0 indicates a very high tolerance for the negative extreme of the attribute.

Both the demand and supply side values for these parameters influence a company's decision on bidding on a project-by-project basis. Under normal circumstances, a company will decide to proceed to bid on a project based on the engineering quality and owner trait criteria, when the ratings of these attributes for the project are higher than the tolerances set by the company. On the other hand, the company will proceed to bid when the ratings for the project, with respect to complexity and safety risks, are lower than the tolerance levels for the company.

At the start of the simulation, Company Agents are created by the Main Agent. Each Company Agent is assigned tolerance values through a process that involves randomly sampling from a statistical distribution. A uniform distribution is constructed from the ranges provided in Table 1 and used to sample the values for ratings assigned to the company. This means that companies will end up having different tolerance levels for each of the criteria, something that is evident in practice.

## 5. SIMULATION RESULTS

Results generated by the simulation model are displayed in charts to ease their interpretation. The results indicate a number of things: the number of projects generated, the quality of these projects and who these projects

were awarded to, as well as the reasons why they were awarded in that fashion.

### 5.1 Projects Generated and Their Attributes

During the simulated period, a total of 11 projects are generated and passed into the virtual construction industry (see Figure 12).

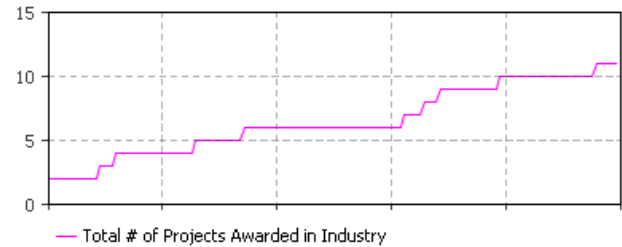


Figure 12: Total Number of New Projects Generated During the Simulated Period

For projects generated, the large size projects had a higher average owner trait, and higher complexity relative to medium size and small projects. The large projects also had the poorest engineering quality and the highest project safety risks compared to medium and small projects. These trends are evident in Figures 13, 14, 15, and 16. These results are consistent with the input data summarized in Table 1 and used in the model.

The trends indicate that small projects have extremely good attributes with the exception of the owner trait. Medium size projects on the other hand possess moderate attributes across the board. All attributes of large projects are bad with the exception of the owner trait. This implies that a company that is inclined to acquire more large projects than small ones would be more susceptible to failure. Values indicated are averages for the projects generated during the simulated period of 1,000 days.

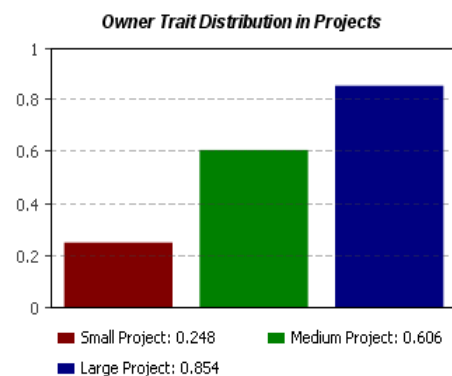


Figure 13: Owner Trait Attributes of Generated Projects

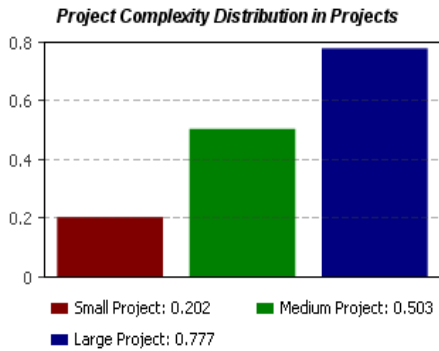


Figure 14: Complexity Attributes of Generated Projects

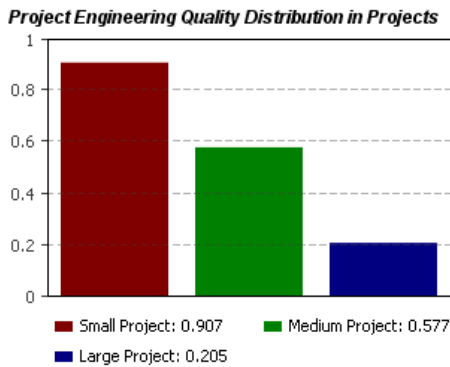


Figure 15: Project Engineering Quality Attributes of Generated Projects

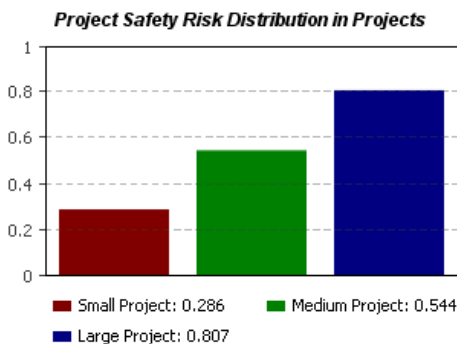


Figure 16: Project Safety Risk Attributes of Generated Projects

## 5.2 Results Summary: Projects Allocated to the Companies

Figure 17 indicates that large companies were awarded a total of 3 projects, medium size companies, 6 projects, and small companies, 2 projects. Despite this small number of projects, the projects awarded to large companies had the highest total values \$1.492 billion. \$754 million and \$124 million worth of projects were awarded to the medium and small companies, respectively. Results displayed in Figure 17 indicates that the bulk of the projects were awarded to medium size companies. Small companies were awarded the least number of projects.

Figure 18 indicates that even though large companies were awarded an intermediate number of projects, the total worth of these projects was the highest (\$1.5 billion). Small companies were awarded

the least number of projects and projects that had the least total value.

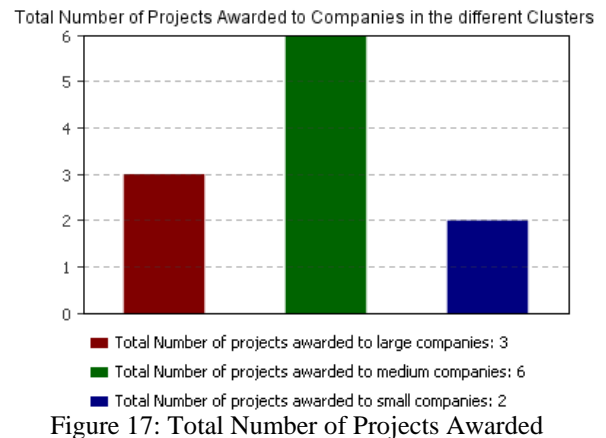


Figure 17: Total Number of Projects Awarded

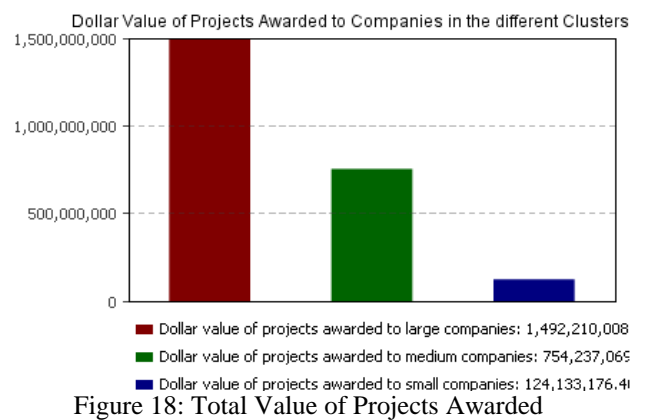


Figure 18: Total Value of Projects Awarded

## 5.3 Reasons for Companies Losing Project Bids

The model tracks the projects that Company Agents either don't bid on or that they bid on and lose. The reasons behind these occurrences are also tracked and reported in the form of pie charts for small companies, medium size companies and large companies, separately (see Figure 19 for an example – small companies). These reasons include insufficient production capacity, competitor number exceeded threshold, bid was not competitive, owner trait very bad, projects too complex, project engineering quality poor, and project safety risks too high.

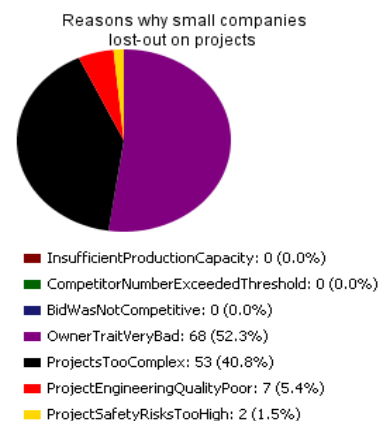


Figure 19: Why Small Companies Lost Projects



To a large extent, all companies opted not to bid on specific projects mainly because the trait of the project owner was bad. The project owner trait in these cases was worse than the tolerances that were predefined by the company. There were no cases in which the companies could not bid on a project due to insufficient production capacity.

## 6. CONCLUSIONS

This paper discussed details of work that was done on the abstraction and implementation of a typical bidding process on a computer using an ABM paradigm in AnyLogic simulation system. The paper also documented the process of abstracting, designing and implementing an ABM that can be useful when analyzing problems in the construction domain.

An experiment set up and run using the model was outlined to show how the model can be put to practical use. The model can be put to various uses. From an owner's perspective, the model can serve as a tool to guide when to invest in projects based on the available company resources in the industry. From a construction contractor's perspective, the model can be used to gain insights into the amount and type of work to go after. These insights can then be used to devise appropriate strategies that ensure that a reasonable work balance is achieved which guarantees good company performance.

In addition, this model can be used as a training tool in both industry and academic institutions to train novices on how to set up internal policies at a company to bid on projects as part of a work acquisition process. The model can also be used as a module within a bigger simulation system to model company performance related issues.

## REFERENCES

- AbouRizk, S., & Hague, S., 2009. An Overview of the COSYE Environment for Construction Simulation. *Winter Simulation Conference*, pp. 2624-2634. Dec. 13-16, Savannah, GA.
- Bagies, A., & Fortune, C., 2006. Bid/No-Bid Decision Modeling for Construction Projects. *22nd Annual ARCOM Conference*, pp. 511-521. Birmingham, UK: UK, Association of Researchers in Construction Management.
- Benjamin, N. B., 1972. Competitive Bidding: The Probability of Winning. *Journal of the Construction Division*, 313-330.
- Cui, Q., & Hastak, M., 2006. Contractor's Bidding Decision Making Through Agent Learning: A System Dynamics Model. *Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, pp. 67-76. Montreal, Quebec, Canada.
- Dixie, J. M., 1974. Bidding Models-The Final Resolution of a Controversy. *Journal of the Construction Division*, 100(CO3), 265-271.
- Dozzi, S. P., AbouRizk, S. M., & Schroeder, S. L., 1996. Utility-Theory Model for Bid Markup Decisions. *Journal of Construction Engineering and Management*, 122(2), 119-124.
- Egemen, M., & Mohamed, A., 2008. SCBMD: A Knowledge-Based System Software for Strategically Correct Bid/No-Bid and Mark-up Size Decisions. *Automation in Construction*, 17, 864-872.
- Egemen, M., & Mohamed, A. N., 2007. A Framework for Contractors to Reach Strategically Correct Bid/No Bid and Mark-up Size Decisions. *Building and Environment*, 42, 1373-1385.
- El-Mashaleh, M. S., 2010. Decision To Bid or Not To Bid:A Data Envelopment Analysis Approach. *Canadian Journal of Civil Engineering*, 37, 37-44.
- European Foundation for Quality Management (EFQM), 1999. *Eight Essentials for Excellence:The Fundamental Concepts and Their Benefits*. Brussels, Belgium.
- Fayek, A. R., 1998. Competitive Bidding Strategy Model and Software System for Bid Preparation. *Journal of Construction Engineering and Management*, 124(1), 1-10.
- Friedman, L., 1956. A Competitive Bidding Strategy. *Operations Research*, 4(1), 104-112.
- Fuerst, M., 1976. Bidding Models: Truths and Comments. *Journal of Construction Division, ASCE*, 102(CO1), 169-177.
- Gates, M., 1967. Bidding Strategies and Probabilities. *Journal of the Construction Division*, 93(CO1), 75-107.
- Griffis, F. H., 1992. Bidding Strategy:Winning Over Key Competitors. *Journal of Construction Engineering and Management, ASCE*, 118(1), 151-165.
- Hegazy, T., & Moselhi, O., 1994. Analogy-Based Solution To Markup Estimation Problem. *Journal of Computing in Civil Engineering, ASCE*, 8(1), 72-87.
- Ioannou, P. G., 1988. Bidding Models-Symmetry and State of Information. *Journal of Construction Engineering and Management*, 114(2), 214-232.
- Kaplan, R. S., & Norton, D. P., 1996. The Balanced Scorecard. *Havard Business Review*, 70, pp. 71-79.
- Li, H. (1996). Neural Network Models for Intelligent Support of Markup Estimation. *International Journal of Engineering Construction and Architectural Management*, 3(1), 69-82.
- Lin, C.-T., & Chen, Y.T., 2004. Bid/No-Bid Decision Making - A Fuzzy Linguistic Approach. *International Journal of Project Management*, 22.
- Marzouk, M. M., 2002. Integrated Monte Carlo/ANN Model for Markup Estimating. *Annual Conference of the Canadian Society for Civil Engineering*. Montreal, Quebec.
- Morin, T. L., & Clough, R. H., 1969. OPBID:Competitive Bidding Strategy Model. *Journal of the Construction Division, ASCE*, 95(CO1), 85-106.
- Oxford Press., 2014. *Oxford Dictionaries*. Oxford, UK: Oxford Press.

- Rosenshine, M., 1972. Bidding Models:Resolution of a Controversy. *Journal of the Construction Division, ASCE, 98*(CO1), 143-148.
- Wade, R. L., & Harris, R. B., 1976. LOMARK:A Bidding Strategy. *Journal of the Construction Division, ASCE, 102*(CO1), 197-211.
- Winston, W. L. (2001). *Simulation Modeling Using @Risk*. Pacific Grove,CA, USA: Thomson Learning (Brooks/Cole).



# MATHEMATICAL MODELING OF THE DIVERSITY IN HUMAN B AND T CELL RECEPTORS USING MACHINE LEARNING

Susanne Schaller<sup>(a)</sup>, Johannes Weinberger<sup>(b)</sup>, Martin Danzer<sup>(b)</sup>, Christian Gabriel<sup>(b)</sup>,  
Rainer Oberbauer<sup>(c)</sup>, Stephan M. Winkler<sup>(a)</sup>

<sup>(a)</sup> Bioinformatics Research Group,  
University of Applied Sciences Upper Austria  
Softwarepark 13, 4232 Hagenberg, Austria

<sup>(b)</sup> Red Cross Transfusion Service of Upper Austria  
Krankenhausstraße 7, 4020 Linz, Austria

<sup>(c)</sup> Elisabethinen Hospital  
Fadingerstraße 1, 4020 Linz, Austria

<sup>(a)</sup> [susanne.schaller@fh-hagenberg.at](mailto:susanne.schaller@fh-hagenberg.at), [stephan.winkler@fh-hagenberg.at](mailto:stephan.winkler@fh-hagenberg.at)

<sup>(b)</sup> [johannes.weinberger@o.ropeskreuz.at](mailto:johannes.weinberger@o.ropeskreuz.at), [martin.danzer@o.ropeskreuz.at](mailto:martin.danzer@o.ropeskreuz.at), [christian.gabriel@o.ropeskreuz.at](mailto:christian.gabriel@o.ropeskreuz.at)

<sup>(c)</sup> [rainer.oberbauer@elisabethinen.or.at](mailto:rainer.oberbauer@elisabethinen.or.at)

## ABSTRACT

We here propose an empirical approach based on the analysis of next-generation sequencing (NGS) data for describing the number of distinct clones of B and T-cell receptors in the human immune system. The status of a human immune system is (amongst other features) defined by the diversity of these receptor cells. It is a well-known issue that NGS data have a higher error rate, and therefore the number of distinct sequences found in sequencing data rises with the number of sequences measured by second generation sequencers. We here present a modeling approach that formulates the number of distinct clones depending on the number of read sequences considering two effects. On the one hand there is a true number of distinct sequences which is asymptotically reached by increasing the number of reads, on the other hand the number of randomly found sequences rises linearly due to read errors. The parameters for this combined model are identified using parameter optimization methods using evolution strategies. This modeling approach is evaluated on the basis of immune status data of several human patients. Additionally, the results are compared to those produced by machine learning methods.

Keywords: B and T cell diversity analysis, model identification, parameter identification, immune system, data mining

## 1. INTRODUCTION: THE HUMAN IMMUNE SYSTEM

The analysis and understanding of the behavior of the human immune system and its key players is one of the most interesting and growing research fields in immunogenetics and medicine. The major cellular components of the adaptive immune response are the B and T cells. These cell types are a subgroup of the white blood cells and are responsible for recognizing foreign

antigens and for the immune response to destruct specific pathogens. B cells have the ability to build antibodies which abrogate foreign cells. Surface membrane-bounded antibodies act as B cell receptors on B cells. Similar to the B cells, the T cells play a central role in cell-mediated immunity and can be separated by the T cell receptor, a protein complex which is essential for the recognition of foreign antigens (Elgert, 2009). The function of T cells is to assist other white blood cells in immunologic processes and manage immunological tolerance and memory function. However, T cells are also able to destroy directly virally infected cells and tumor cell.

Modern DNA sequencing systems are referred to as next-generation sequencing (NGS) (Benichou et al., 2012 and Liu et al., 2012). NGS is a method to determine the nucleotide sequence by high-throughput and parallel processing of material, which leads to decreased run times as well as to massively increased amounts of data produced (Tucker et al., 2009). The B and T cell receptor sequence data are measured using next-generation sequencing technologies. Generated sequences are further pre-processed and provide the basis for modeling the diversity of B and T cells in the human immune system to gain more insights into the cell input and into the error rate using these technologies.

## 2. MATHEMATICAL MODELING OF THE CLONOTYPE DIVERSITY

The research goal of the project presented here is the mathematical modeling of the diversity of B and T cell receptor data. The diversity of B and T cell receptors data provides information about the state of health in humans and supplies detailed insight into the specific existing immune response. This additional know-how helps, e.g., in the diagnostic of autoimmune diseases as

well as in transplantation monitoring, and in the treatment of various infections (Boyd et al., 2009). We perform a modeling approach using data mining and machine learning to identify the relationship to describe the shape and progress of the diversity. So created models shall help to estimate the error rate of the NGS machine and solve additionally biological questions regarding sample preparation and behavior of the NGS technology.

### 2.1 Biological Data

As first step samples are pre-processed and sequenced using high-throughput NGS technology. B and T cell receptor sequences are further processed using the ImMunoGeneTics (IMGT) system (Alamyar, 2013). The output data derived from the IMGT system are used as input for our diversity analyses (see Figure 1). We are interested in the total number of sequences that are sequenced and the distinct number of clonotypes, where a clonotype represents a unique amino acid sequence.

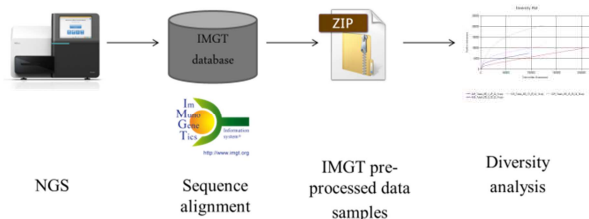


Figure 1: Data sample pre-processing using NGS technology and the IMGT database.

### 2.2 Diversity Analysis

The diversity analysis plays an important role in the analysis of the human immune system. A mathematical analysis of the diversity of immune cells is of high importance due to two reasons: (1) there is a true number of distinct sequences which can be reached by increasing the number of reads (strings of bases), and (2) the number of randomly found sequences rises linearly due to read errors of the sequencers.

Modelling of the diversity for a biological sample can be constructed by randomly choosing  $n$  amino acid sequences from the sample and determining the number of distinct sequences, also known as distinct clonotypes of B and/or T cells found in these  $n$  sequences. This is repeated five times and the number of distinct clones is averaged. The following two examples represent the diversity of B and T cell receptors: Figure 2 represents the diversity of the T cell receptor. The T cell receptor data are gained from blood (4B-TCRB) and kidney (4N-TCRB) samples. Figure 3 represents B cell receptor data.

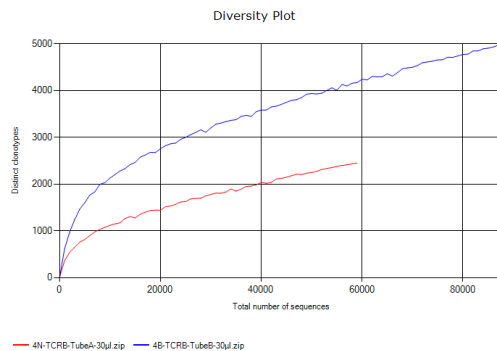


Figure 2: Diversity plot of T cell receptor data (blood and kidney cells).

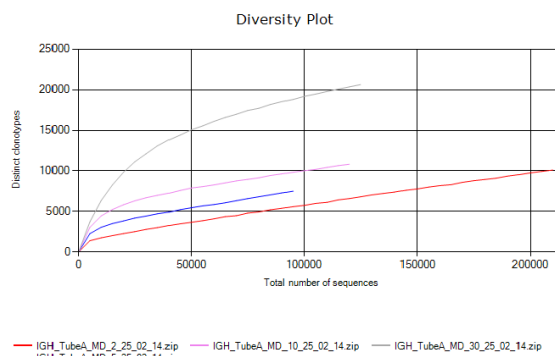


Figure 3: Diversity plot of B cell receptor data (IGH).

## 3. MODEL IDENTIFICATION

A modeling approach has to be defined that explains the number of distinct sequences found in  $n$  sequences; this enables us to distinguish between read errors and “real” distinct sequences.

We are especially interested in white box models. In the following sections we present two modeling approaches for which the correct parameters have to be identified empirically for each sample (Section 3.1). Additionally, we also plan to apply symbolic regression using genetic programming for identifying model structures that estimate the number of distinct clonotypes in a biological sample (Section 3.2).

For validating these modeling approaches and for comparing the models’ quality to the modeling quality achievable using alternative approaches, black box models (Section 3.3) shall also be used.

### 3.1. White Box Modeling Using a Fixed Model Structure and Parameter Optimization

The primary model identification approach used in this work uses different formula structures. The parameters of these models have to be optimized (in this case, using evolution strategies) in order to optimize the fit of the models to the given diversity curves of B and T cell receptor data.

Our modelling possibilities are that the curve initially corresponds to an exponential or logarithmic increase or a multiplicative inverse, while for greater  $n$  the curve rather resembles a linear function. Therefore

we have developed the following model approaches (see Equations 1, 2, 3 and 4) for which the parameters have to be optimized:

$$div(n) = k * n + d \quad (1)$$

$$div(n) = a * (1 - e^{-b*n}) \quad (2)$$

$$div(n) = a * \log(b * n) \quad (3)$$

$$div(n) = a * (1 - e^{-b*n}) + k * n + d \quad (4)$$

For each biological sample, the parameters are optimized using an evolution strategy.

Evolution strategies (ES) have been developed since the 1960s. An ES is an evolutionary algorithm in which genetic operators (mutation, crossover, and selection) are applied on solution candidates until a specific termination criterion is met (Winkler, 2009) and (Borgmann, 2012).

Typically, an ES starts with a population of randomly created individuals; in each generation these individuals are evaluated and new individuals are created using mutation and crossover operators. The next generation's population is formed using successful children. This procedure is repeated over several generations until a termination criterion is met, for example as soon as the quality cannot be improved any more or the maximum number of generations is reached (see Figure 4).

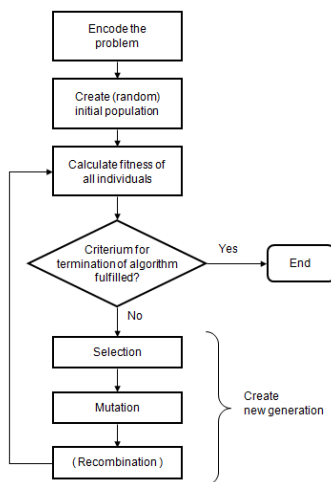


Figure 4: Standard evolution strategy (Winkler, 2009).

One of the goals of this research is to find out, which modeling approach is better in terms of modeling quality, i.e. which approach is more suitable for describing the diversity of B and T cell receptors.

### 3.2 Black Box Modeling

For estimating the quality of the white box models created using the methods summarized in Sections 3.1, the following black box modeling methods are also applied:

- **Random Forests (RF):** RFs are very popular in the field of machine learning as they are very fast and accurate. They create ensembles of decision trees using information gain and variance reduction. Pruning is done using reduced-error pruning methods (Witten, 2005).
- **Genetic Programming:** As an alternative to the approach described in the previous section we also apply a modeling algorithm based on genetic programming (Koza, 1992) using a structure identification framework described in (Winkler, 2008 and Affenzeller et al., 2009) implemented in HeuristicLab (Wagner, 2014).

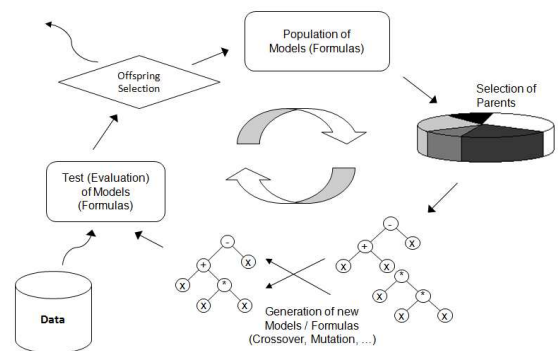


Figure 5: Genetic programming approach.

## 4. RESULTS

### 4.1 White Box Modeling Experimental Results

For testing the approaches described in Section 2.2 and 3, we used three samples containing a different number of sequences and clonotypes (see Table 1).

|              | <b>BB17</b> | <b>SS12</b> | <b>6AB</b> |
|--------------|-------------|-------------|------------|
| # sequences  | 307440      | 284307      | 76273      |
| # clonotypes | 35906       | 25899       | 15580      |

Table 1: Sequence and clonotype information.

Figures 6, 7, and 8 show the diversity analysis of the following samples BB17, SS12, and 6AB as described in Section 2.2. The thick red line represents the original calculated diversity using iteratively 2500 sequences and determines the number of distinct clonotypes. The red thick lines in Figures 6, 7, and 8 show that there first is a strong increase of the number of distinct clonotypes and then the red line starts to continue at a specific number of sequences to increase more or less linearly. Afterwards where the increase seems to be a more linear slope, which represents the number of randomly found sequences rise linearly due to read errors of the sequencers. All the other thin lines show the estimated diversity curve using four different formulas. The parameters such as  $a$ ,  $b$ ,  $k$  and  $d$  are optimized using evolution strategy.

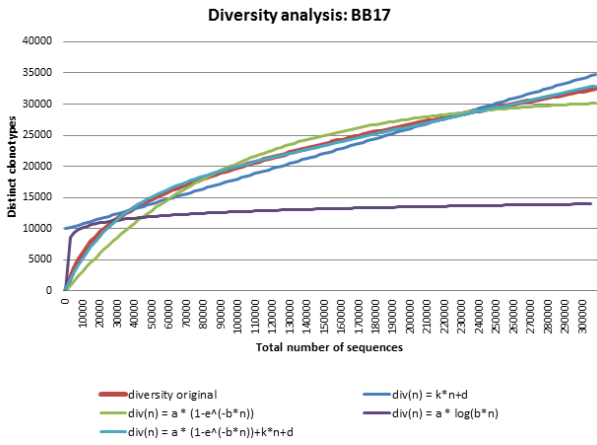


Figure 6: Diversity analysis of sample BB17 using 300 generations,  $\mu=20$ ,  $\lambda=100$  and  $\sigma=0.5$ .

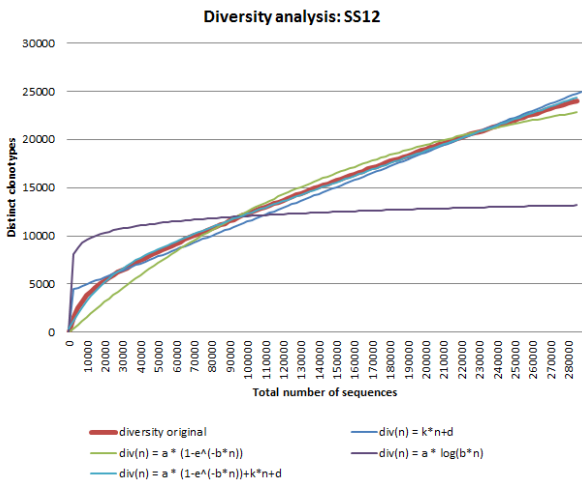


Figure 7: Diversity analysis of sample SS12 using 300 generations,  $\mu=20$ ,  $\lambda=100$  and  $\sigma=0.5$ .

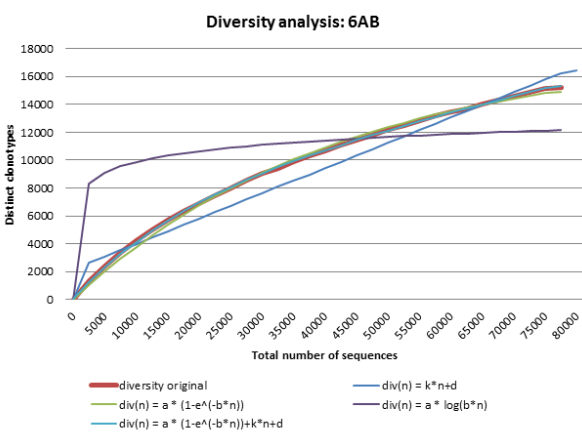


Figure 8: Diversity analysis of sample SS12 using 300 generations,  $\mu=20$ ,  $\lambda=100$  and  $\sigma=0.5$ .

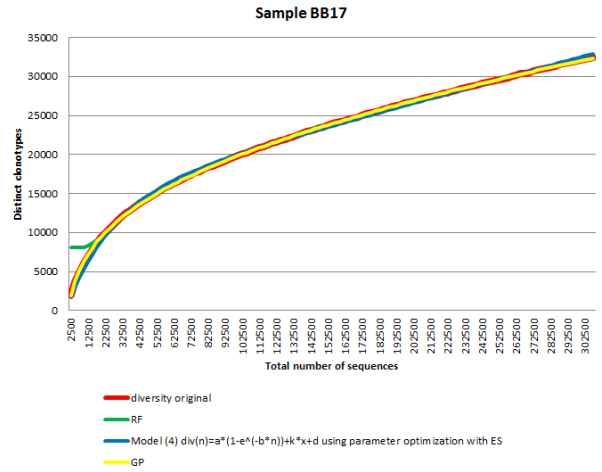


Figure 9: Black box modelling using genetic programming and random forests for sample BB17.

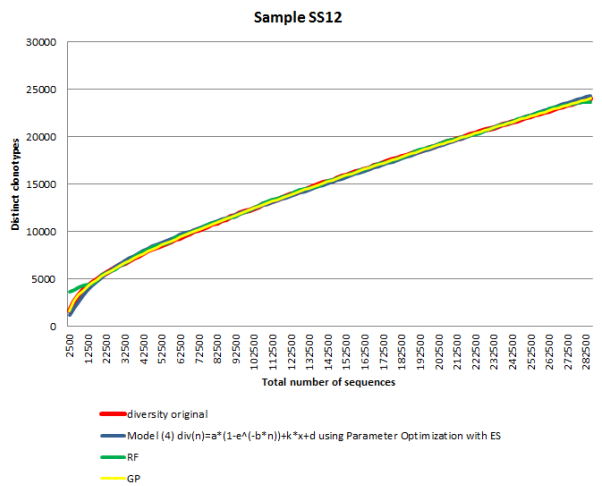


Figure 10: Black box modelling using genetic programming and random forests for sample SS12.

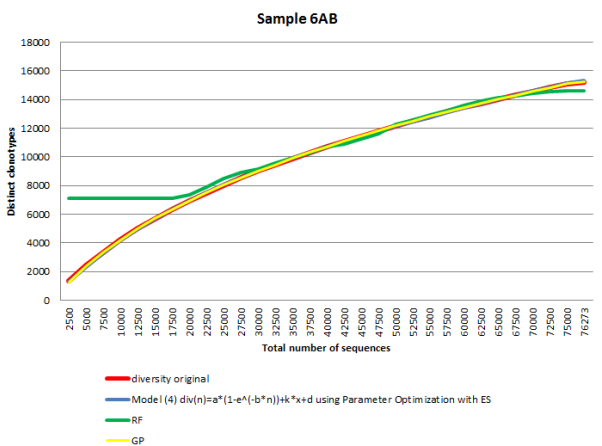


Figure 11: Black box modelling using genetic programming and random forests for sample 6AB.

Here we have used 300 evolution strategy generations and for  $\sigma=0.5$ ,  $\mu=20$  and  $\lambda=100$ . For the parameter optimization with evolution strategy the so-called 1/5 success rule was used.

The diversity analyses of the samples show that the formulae containing an exponential and a linear function fits the original calculated diversity curve best (light blue line in Figures 6, 7, and 8). Figure 8 demonstrates that in a sample containing a lower total number of sequences also the formulae only using an exponential function (light green line in Figures 6, 7 and 8) perform as well as the combination using a linear function.

To estimate the quality of the white box modelling and to validate the modelling approach we have used the following machine learning methods: genetic programming and random forests.

For black box modelling the framework HeuristicLab (<http://dev.heuristiclab.com/>) where all those classification and regression algorithms are integrated and implemented. Implementation details of these methods can be found in (Wagner et al., 2014).

For all machine learning algorithms, a different set of parameters have been tested. Best results can be achieved using for genetic programming the following parameter settings: The mutation rate was set to 15% and an OSGA (offspring selection genetic algorithm) (Affenzeller et al., 2009) with success ratio as well as comparison factor set to 1.0 was used. Crossover and mutation were done using the subtree swapping crossover and the multi symbolic expression tree manipulator, respectively. The function set described in (Winkler, 2009) was used for building composite function expressions (including arithmetic, exponential and logarithm functions).

For random forests the following parameter settings have been set: the number of trees has been set to 30 and randomly seed has been used. Generally, data were split into training and test data and five-fold cross-validation was executed. As is shown in Figure 9, 10 and 11 the thick red line represents the original calculated diversity curve, while green and yellow represent the black box modelling machine learning methods. Genetic programming is able to predict the original calculated diversity curve pretty accurate, while using random forest the estimation of the distinct clonotypes seems to be difficult.

Table 2 and 3 represent statistical analysis of the results using the Pearson's  $R^2$  (Bollen et al., 1981) value for training and testing of the white and black box modelling methods. As it is shown that in training and testing very high Pearson's  $R^2$  value can be achieved, where the value 1 means that the originally calculated diversity curve correlates exactly to the predicted diversity curves using machine learning methods such

as genetic programming and random forests. In Table 2 it is shown that the highest correlation values can be found using a combination of the exponential and linear function. This high correlation can be observed in all three samples.

| Sample/<br>ES | div(n)=k*n+d |          | div(n) = a * ( 1 - e <sup>(-b*n)</sup> ) |          |
|---------------|--------------|----------|------------------------------------------|----------|
|               | training     | test     | training                                 | test     |
| BB17          | 0,993425     | 0,992920 | 0,949096                                 | 0,970529 |
| SS12          | 0,998450     | 0,998406 | 0,991846                                 | 0,990502 |
| 6AB           | 0,992546     | 0,953213 | 0,999550                                 | 0,999569 |

| Sample/<br>ES | div(n)= a * log(b*n) |          | div(n) = a * ( 1 - e <sup>(-b*n)</sup> ) + k*n+d |          |
|---------------|----------------------|----------|--------------------------------------------------|----------|
|               | training             | test     | training                                         | test     |
| BB17          | 0,992889             | 0,985620 | 0,998797                                         | 0,998740 |
| SS12          | 0,980015             | 0,982939 | 0,999275                                         | 0,999172 |
| 6AB           | 0,990095             | 0,992591 | 0,999550                                         | 0,999569 |

Table 2: Person's  $R^2$  values for training and testing of all models using parameter optimization with evolution strategy.

| Sample/<br>ML | GP       |          | RF       |          |
|---------------|----------|----------|----------|----------|
|               | training | test     | training | test     |
| BB17          | 0,999867 | 0,999857 | 0,998261 | 0,981551 |
| SS12          | 0,999983 | 0,999956 | 0,997973 | 0,998479 |
| 6AB           | 0,999988 | 0,999963 | 0,989975 | 0,892114 |

Table 3: Pearson's  $R^2$  values for training and testing of genetic programming and random forests.

## 5. CONCLUSION

In this paper an approach has been described to model the number of distinct clonotypes in dependence of the total number of sequences (diversity) of B and T receptor cells in the human immune system. It has been shown that there is a true number of distinct clonotypes which can be reached by increasing the number of reads using a mathematical model containing an exponential and linear part. Additionally, we have shown that the diversity analysis curve rises linearly due to read errors of NGS machines. The parameters of the mathematical model are optimized using evolution strategy and for evaluation purposes black box modelling methods such as genetic programming and random forests have been used. The statistical analysis of the results shows that there is a high correlation between the machine learning methods and the parameter optimization approach using evolution strategy.

Currently this analysis is integrated in the framework ImmunExplorer. In future work, additional methods for analyzing B and T cells shall be included in



ImmunExplorer to gain more insight in the behaviour of the human immune system.

## ACKNOWLEDGMENTS

The work described in this paper was done within the project “Transplant - Early marker for humoral and cellular rejection of transplanted kidneys” funded by EFRE Regio 13.

## REFERENCES

- Affenzeller, M., Winkler, S., Wagner, S., Beham, A., 2009. Genetic Algorithms and Genetic Programming - Modern Concepts and Practical Applications. *Chapman & Hall/CRC*.
- Alamyar, E., Giudicelli, V., Li, S., Duroux, P., Lefranc, M.P., 2012. IGMT/HighV-QUEST: the IGMT(R) web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome research* 8(1): 26.
- Benichou, J., Ben-Hamo, R., Louzoun, Y., Efroni, S., 2012. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* 135(3):183-91.
- Bollen, K.A., Barb, K.H., 1981. Pearson's R and coarsely categorized measures. *American Sociological Review* – JSTOR.
- Borgmann, D., Weghuber, J., Schaller, S., Jacak, J. and Winkler, S.M., 2012. Identification of Patterns in Microscopy Images of Biological Samples using Evolution Strategies. *Proceedings of the 24th European Modeling and Simulation Symposium EMSS*, pp. 271-276.
- Boyd, S. D., Marshall, E. L., Merker, J. D., Maniar, J. M., Zhang, L. N., Sahaf, B., Fire, A. Z., 2009. Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing. *Science Translational Medicine*, 1(12).
- Chang, C.C., Lin, C.J., 2001. LIBSVM: a library for support vector machines.
- Elgert, K. D., 2009. Immunology: Understanding The Immune System, *John Wiley & Sons*, p. 726.
- Koza, J.R., 1994. Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*.
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., Law, M., 2012. Comparison of Next-Generation Sequencing Systems. *J Biomed Biotechnol*:251364.
- Nelles, O., 2001. Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models.
- Tucker, T., Marra, M., & Friedman, J. M., 2009. Massively parallel sequencing: the next big thing in genetic medicine. *American Journal of Human Genetics*, 85(2), 142–54.

Vapnik, V., 1998. Statistical Learning Theory. *Wiley, New York*.

Wagner, S., Kronberger, G. K., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S. M., Dorfer, V., Affenzeller, M. 2014. Architecture and Design of the HeuristicLab Optimization Environment in Advanced Methods and Applications in Computational Intelligence (Contributions to Book: Part/Chapter/Section 10), (Editors: R. Klempous, J. Nikodem, W. Jacak, Z. Chaczko) - *Springer*, pp. 197-261.

Winkler, S., Affenzeller, M., Wagner, S., 2009. Using enhanced genetic programming techniques for evolving classifiers in the context of medical diagnosis - an empirical study. *Genetic Programming and Evolvable Machines*, pp. 111–140.

Winkler, S., 2009. Evolutionary System Identification - Modern Concepts and Practical Applications. *Schriften der Johannes Kepler Universität Linz, Reihe C: Technik und Naturwissenschaften. Universitätsverlag Rudolf Trauner*.

Witten, I.H., 2005. Data mining: practical machine learning tools and techniques (Second Edition). *Elsevier*.

## AUTHORS BIOGRAPHY



**SUSANNE SCHALLER** received her Masters of Science in Biomedical Informatics in 2009 and 2010 from the University of Skovde and the University of Applied Sciences Upper Austria. She is a research and teaching associate at the Bioinformatics Research Group, in Hagenberg. Her research interests include systems biology, medical image analysis, tissue engineering, data mining, and machine learning.



**JOHANNES WEINBERGER** received his Master in Genetics in 2013 from the University of Salzburg. He is currently PhD student at the University of Salzburg and research associate at the Immunogenetics Department of the Red Cross Blood Transfusion Service of Upper Austria in Linz. His research interests are immunology, next-generation sequencing, BCR/TCR spectratyping, and FACS.



**MARTIN DANZER** received his MSc degree in Genetics from the University of Salzburg. He is actually the Co-Director of the Histocompatibility and Immunogenetics Laboratory in Linz that provides a comprehensive range of diagnostic testing services in the field of transplantation. His research activities focus in particular on immune response gene diversity in the human population and how diversity impacts transplantation.



**RAINER OBERBAUER** received his MD degree in 1990 at the University of Vienna and received an additional MSc in Epidemiology at the Harvard School of Public Health in 2005. Since 2006 he is director of the Department of Internal Medicine III at the Elisabethinen Hospital in Linz. His research areas focus on nephrology, dialysis and transplantation. Since 2010 Mr. Oberbauer has been teaching medicine at the Medical University of Vienna.



**CHRISTIAN GABRIEL** studied medicine in Innsbruck from 1978 to 1983. He worked as resident for general medicine from 1984 to 1989, then joined the Department of Anaesthesiology and Intensive Care at the General Hospital in Linz and served as resident for transfusion medicine in the Red Cross Transfusion Service of Upper Austria. Since 2001 he is the medical director of the Red Cross Transfusion Service of Upper Austria in Linz. His research interest includes intensive care, blood physiology, cell therapies, regenerative medicine, and foremost immunogenetics.



**STEPHAN M. WINKLER** received his PhD in engineering sciences in 2008 from Johannes Kepler University (JKU) Linz, Austria. His research interests include genetic programming, nonlinear model identification and machine learning. Since 2009, Dr. Winkler is professor at the Department for Medical and Bioinformatics at the University of Applied Sciences (UAS) Upper Austria at Hagenberg Campus; since 2010, Dr. Winkler is head of the Bioinformatics Research Group at UAS, Hagenberg.



# A MULTI-TASK ASSIGNMENT METHOD IN CLOUD-BASED SIMULATION

Lei Ren<sup>(a)</sup>, Hejian Ou<sup>(b)</sup>, Jin Cui<sup>(a)</sup>, Bowen Li<sup>(a)</sup>, Baocun Hou<sup>(c)</sup>

<sup>(a)</sup>School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China;

<sup>(b)</sup>The High-Tech Research & Development Center Ministry of Science & Technology, P.R.C, Beijing 100044, China;

<sup>(c)</sup>Beijing Simulation Center, Beijing 100854, China

<sup>(a)</sup>[renlei@buaa.edu.cn](mailto:renlei@buaa.edu.cn), <sup>(b)</sup>[ohj@sina.com](mailto:ohj@sina.com), <sup>(c)</sup>[baocun\\_hou@163.com](mailto:baocun_hou@163.com)

## ABSTRACT

Cloud-based Simulation (Cloud Simulation) can significantly improve the capacity of existing networked modeling and simulation systems. The performance of the Cloud Simulation Platform tightly couples with the partitioning and allocation of simulation modules among different hosts. This paper proposed a Cloud Simulation resources scheduling strategy for initial partitioning based on Genetic Algorithm (GA). The Cloud Simulation resources scheduling model was given by considering the computing resources of involved hosts as well as the predicted computing load of simulation modules. The improved strategy of GA is designed according to the characteristics of the Cloud Simulation multi-task assignment problem. The simulation results demonstrate the feasibility and efficiency of the proposed method.

Keywords: Cloud Computing, Cloud Simulation, Genetic Algorithm, Multi-task assignment

## 1. INTRODUCTION

Cloud-based Simulation is a new network-based and service-oriented simulation model which enables users to acquire simulation resources and capacity on demand through network and Cloud Simulation Platform (Ren, and Zhang 2011; Li, Chai, and Zhang 2012). Cloud Simulation can improve the capacity of existing networked modeling and simulation systems significantly by supporting fine-granularity resource sharing, multi-user, cooperation, fault tolerance and security application mechanism (Li, Chai, and Hou 2009).

The performance of Cloud Simulation Platform has a close relationship with the distribution of simulation modules on involved hosts. A reasonable partitioning can lead to a reduction of simulation run-time. In contrast, a poor performance can emerge due to an inadequate distribution of simulation modules. Therefore, optimal allocation mechanism & algorithm for Cloud Simulation services is one of the key operation technologies for virtual simulation environment.

A lot of strategies for resource scheduling have been proposed in order to improve the performance of certain system.

In the context of computational Grids, researchers employ heuristic algorithms such as Simulated Annealing, Tabu Search and Genetic Algorithm for job scheduling (Abraham, Buyya, and Nath 2000). In addition to optimizing execution time, the cost arising from data transfers between resources as well as execution costs must also be taken into account in Cloud computing environment. Suraj Pandey et al. utilize a particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments (Pandey, Wu, and Guru 2010). In simulation domain, both dynamic and static solutions are presented for partitioning simulation modules of distributed simulation system. Statistic partitioning methods or heuristic assignment mechanisms are widely used in simulation environment. Compared with dynamic partitioning algorithms (Lui and Chan 2002, Low 2002, Deelman and Szymanski 1998), static partitioning approaches are usually employed with regard to specific applications (Peschlow, Honecker, and Martini 2007). However, there is little literature on Cloud Simulation oriented simulation module assignment. Allocation algorithms aimed to improve the efficiency of Cloud Simulation Platform are highly required. In this paper, a multi-task assignment method in Cloud Simulation is proposed based on Genetic Algorithm. A relatively accuracy model and the evaluation function are given. Compared with the experiment result of greedy algorithm, our strategy could reduce the simulation run-time efficiently.

The remainder of the paper is organized as follows. In next section, we exhibit the model of simulation resources assignment problem in Cloud Simulation Platform. In Section 3, we developed a simulation module partitioning mechanism based on Genetic Algorithm. Details of the algorithm are given. Simulation results and discussions are presented in Section 4. Finally, conclusions and suggestions for future work are given in Section 5.

## 2. PROBLEM DEFINITION

Simulation modules assignment mechanism is one of the key issues in Cloud Simulation Platform. Cloud Simulation multi-task assignment problem can be described as a quadruple  $\{M, E, T, C\}$ .  $M = \{M_1, M_2, \dots, M_n\}$  refers to the simulation task which consists of simulation modules  $M_i$  ( $i = 1, 2, \dots, n$ ).  $E = \{N_1, N_2, \dots, N_m\}$  is the collection of involved hosts to execute the simulation task  $M$ .  $E$  consists of computing nodes  $N_i$  ( $i = 1, 2, \dots, m$ ).  $T$  is the time matrix.  $t_{ij}$  represents the time which computing node  $i$  needs to execute simulation module  $j$  ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ). The time matrix is a  $m \times n$  dimensional matrix.

$$T = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ t_{m1} & t_{m2} & \dots & t_{mn} \end{bmatrix} \quad (1)$$

While  $C$  refers to constrains in the simulation modules partitioning problem in Cloud Simulation Platform.

### 2.1. Related Constrains

The Cloud Simulation Multi-task assignment problem is a problem with timing constrains and assignment constrains.

#### 2.1.1. Timing Constrains

If simulation module  $M_i$  is required to be finished before simulation module  $M_j$ , then  $M_i$  is the previous task of simulation module  $M_j$ ,  $M_i$  can be presented as  $Prev(M_j)$ . Similarly,  $M_j$  is the next task of simulation module  $M_i$ ,  $M_j$  can be expressed as  $Next(M_i)$  ( $i = 1, 2, \dots, n; j = 1, 2, \dots, n; i \neq j$ ) (Su, Chen, and Shen 2008). The timing constrain among simulation modules is given by formula (2).

$$\begin{cases} Enforce[\{Prev(M_i), M_i\}, <] \\ Enforce[\{M_i, Next(M_i)\}, <] \\ i = 1, 2, \dots, n \end{cases} \quad (2)$$

#### 2.1.2. Assignment Constrains

As to assignment constrains, decision variable  $x_{ij}$  is 1 when  $M_j$  is assigned to  $N_i$ , otherwise  $x_{ij}$  is 0 ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ).

$$\sum_{i=1}^m x_{ij} = 1, x_{ij} \in \{0, 1\} (j = 1, 2, \dots, n) \quad (3)$$

Formula (3) guarantees that each simulation module of simulation task is assigned to only one computing node of the Cloud Simulation Platform.

### 2.2. Related Constrains

For a certain solution, if simulation module  $j$  is assigned to computing node  $i$ .

$T_{mod}[j]$  is the finishing time of simulation module  $M_j$  ( $j = 0, 1, \dots, n - 1$ ).

$T_{nod}[i]$  is the current time of computing node  $N_i$  ( $i = 0, 1, \dots, m - 1$ ).

After a certain simulation module  $M_j$  is executed on computing node  $N_i$ ,  $T_{nod}[i]$  should be updated to  $T_{nod}^*[i]$  ( $i = 0, 1, \dots, m - 1; j = 0, 1, \dots, n - 1$ ).

$$T_{mod}[j] = \max\{T_{nod}[i], T_{mod}[Prev(M_j)]\} + T[i][j] \quad (4)$$

$$T_{nod}^*[i] = \max\{T_{nod}[i], T_{mod}[Prev(M_j)]\} + T[i][j] \quad (5)$$

Thus, the run-time  $T$  of simulation task can be estimated as

$$T = \max\{T_{nod}[0], T_{nod}[1], \dots, T_{nod}[m - 1]\} \quad (6)$$

## 3. SIMULATION MODULE PARTITIONING STRATEGY

Genetic Algorithm was developed by Holland (Holland 1975) to study the adaptive process of natural systems and to develop artificial systems that mimic the adaptive mechanism of natural systems (Hou, Ansari, and Ren 1994). A standard Genetic Algorithm contains three basic operations: selection, crossover and mutation. In the following part, with the characteristics of the Cloud Simulation multi-task assignment problem considered, a partitioning mechanism based on Genetic Algorithm with improving strategy is presented.

### 3.1. Encoding and Initialization

Assume the population size of Genetic Algorithm is  $N$ . First of all, we initial the assignment array randomly use the following formula.

$$A_k[j] = i \quad (7)$$

In formula (7),  $k = 0, 1, \dots, N - 1$ ;  $j = 0, 1, \dots, n - 1$ ;  $i \in \{0, 1, \dots, m - 1\}$ .

### 3.2. Selection Operation

According to the evaluation function of a solution we discussed in part 3.2, the fitness function can be set as

$$f(A) = \max\{T_{nod}[0], T_{nod}[1], \dots, T_{nod}[m - 1]\} \quad (8)$$

According to the Roulette principle, the probability of a solution  $A_k$  is selected to inherited to next generation is given by

$$P_s(A_k) = \frac{\max\{f(A_0), f(A_1), \dots, f(A_{N-1})\} - f(A_k) + \varepsilon}{\sum_{i=0}^{N-1} (\max\{f(A_0), f(A_1), \dots, f(A_{N-1})\} - f(A_i) + \varepsilon)} \quad (9)$$

In formula (9),  $k = 0, 1, \dots, N - 1$  and  $\varepsilon > 0$ .

This selection mechanism ensures that the population size of Genetic Algorithm is fixed during the whole operation, while solutions with higher fitness value are more likely to survive to next generation.

### 3.3. Crossover Operation

For each iteration, the probability to execute crossover operation is  $P_{c1} = ae^{-\frac{N_c}{b}}$ ,  $a$  and  $b$  are constant. With the increasing of iterations, the probability of perform crossover operation is decreasing.

The probability of a certain solution  $A_k$  to be selected to execute the crossover operation is

$$P_{c2}(A_k) = \frac{(\max\{f(A_0), f(A_1), \dots, f(A_{N-1})\} - f(A_k) + \varphi)^\alpha}{\sum_{i=0}^{N-1} (\max\{f(A_0), f(A_1), \dots, f(A_{N-1})\} - f(A_i) + \varphi)^\alpha} \quad (10)$$

In formula (10),  $\varphi \geq 1$  and  $\alpha > 1$ . Formula (10) can lead to a result that solutions with higher fitness value have a greater probability to perform crossover operation. Crossover operation between two solutions with high fitness value probably leads to a solution with higher fitness value.

Crossover operation utilize two-point crossover mechanism, select  $i$  and  $j$  randomly which satisfies  $0 \leq i \leq j \leq n - 1$ , then exchange the segment between  $A[i]$  and  $A[j]$ .

### 3.4. Mutation Operation

For each iteration, the probability to execute mutation operation is  $P_{m1} = ae^{-\frac{N_c}{b}}$ ,  $a$  and  $b$  are constant. The same as crossover operation, the probability of perform mutation operation is decreasing as well with the increasing of iterations.

The probability of a certain solution  $A_k$  to be selected to execute the mutation operation is

$$P_{m2}(A_k) = \frac{(f(A_k) - \min\{f(A_0), f(A_1), \dots, f(A_{N-1})\} + \delta)^\beta}{\sum_{i=0}^{N-1} (f(A_i) - \min\{f(A_0), f(A_1), \dots, f(A_{N-1})\} + \delta)^\beta} \quad (11)$$

In formula (11),  $\delta \geq 1$  and  $\beta > 1$ . Mutation probability is greater to those solutions with low fitness value. Mutation operation of solutions with low fitness value is more likely to lead to better solutions.

For a certain solution  $A_k$ . Select  $i$  randomly and substitute computing node  $A_k[i]$  with another available computing node.

Based on these rules, the basic steps of the algorithm can be summarized as the pseudo code shown in figure 1.

---

#### Begin

Generate initial population of  $N$  solutions  $A_k(k=0, 1, \dots, N-1)$ .

**While** ( $t < \text{MaxGeneration}$ ) or (stop criterion)

For each  $A_k$  in the population, evaluate its fitness value.

Execute selection operation and update the population.

Select solutions according to  $P_{c2}$ , execute crossover operation and update the population.

Select solutions according to  $P_{m2}$ , execute mutation operation and update the population.

Update the best solution.

---

**End while**

Postprocess results.

**End**

---

Figure 1: Pseudo code for the algorithm

## 4. SIMULATION EXPERIMENTS AND RESULTS

Historical simulation cases on Cloud Simulation Platform can be used to improve the performance of new simulation task. With large amount of simulation cases, the run-time of a simulation module on a certain computing node of the Cloud Simulation Platform is able to be estimated. Thus, the simulation module assignment approach works.

In our experiment, the time matrix  $T_{m \times n}$  is generated randomly. Run-time of different modules of a same computing node is generated randomly. While run-time of a certain module on different computing nodes satisfy a linear relationship. That is, computing node with multi-core and larger memory will lead to a short run-time to a certain simulation module. A Gaussian noise is introduced to the run-time at last.

We conducted the simulation experiment with different parameters for several times. With the following parameters:  $m = 30$ ,  $n = 100$ ,  $\varepsilon = 0.5$ ,  $\varphi = 1$ ,  $\delta = 1$ ,  $a = 0.2$ ,  $b = 100$ ,  $\alpha = 2$ ,  $\beta = 2$ , the relationship between simulation task run-time and the end criterion (max iterations) with certain population size is presented in figure 2. We conducted each simulation experiment for ten times and then calculated the mean value of run-time.

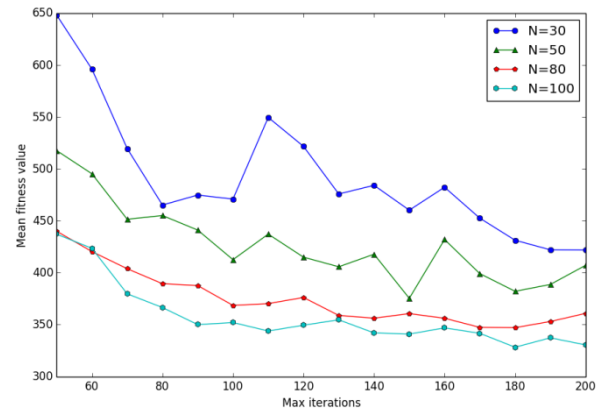


Figure 2: The relationship between simulation task run-time and different end criterion (different max iterations) with certain population size

From figure 2, we could conclude that the simulation run-time has a close relationship with the algorithm population size and end criterion. The simulation task run-time reduced with the increasing of population size of Genetic Algorithm. What's more, the run-time shows downward trend with the max iterations of end criterion increased.

Compared with Greedy Algorithm, the typical experimental results of our partitioning approach are shown in figure 3 (Population size N=50).

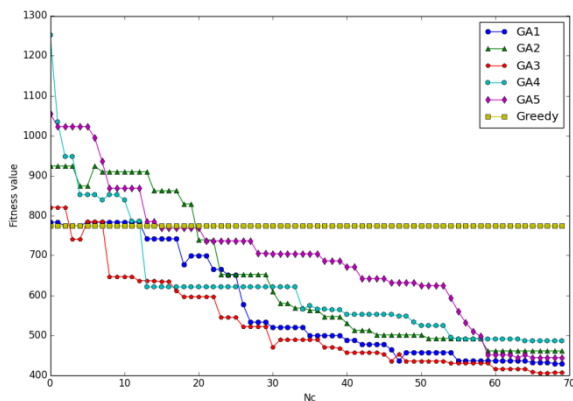


Figure 3: Comparison between Greedy Algorithm and our strategy

We have tested the strategy for five times with an iteration of 70. The mean run-time of the simulation task is 445.09. Meanwhile, the result of Greedy Algorithm is 774.87. Compared with the result of Greedy Algorithm using priority function, our strategy has a better performance with acceptable convergence. Though Greedy Algorithm has an advantage of short execute time, it can be neglected compared with the significant reduction of simulation task run-time by our strategy.

## 5. CONCLUSION AND FUTURE WORKS

In this paper, we revisited the resource scheduling problem in simulation system. With characteristics of the Cloud Simulation Platform taken into account, a multi-task assignment strategy based on Genetic Algorithm is proposed. The probability of crossover operation and mutation operation are modified in order to get better partitioning results. Simulation experiential results demonstrate that the method can efficiently solve the Cloud Simulation multi-task assignment problem. Therefore, our simulation module partitioning mechanism could improve the performance of the Cloud Simulation Platform.

As future research, we plan to focus on improving the simulation performance of Cloud Simulation Platform via utilizing the historical simulation cases. Dynamic partitioning approach is also an important aspect to study in Cloud Simulation Platform.

## ACKNOWLEDGMENTS

The research is supported by the NSFC (National Science Foundation of China) Projects (No.61103096) in China, the National High-Tech Research and Development Plan of China under Grant No. 2013AA041302, and the Fundamental Research Funds for the Central Universities in China.

## REFERENCES

- Abraham, A., Buyya, R., Nath, B., 2000. Nature's heuristics for scheduling jobs on computational grids. *Proceedings of the 8th IEEE International Conference on Advanced Computing and Communications*, pp. 45-52. December 16-19, Cochin (India).
- Deelman, E., Szymanski, B.K., 1998. Dynamic load balancing in parallel discrete event simulation for spatially explicit problems. *Proceedings of the Parallel and Distributed Simulation*, pp. 46-53. December 14-16, TaiWan.
- Su, F., Chen, Y., Shen, L.C., 2008. UAV cooperative multi-task assignment based on ant colony algorithm. *Acta Aeronautica et Astronautica Sinica*, 29(S1): S184-S191.
- Holland, J.H., 1975. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Michigan: University of Michigan Press.
- Hou, E.S.H., Ansari, N., Ren, H., 1994. A genetic algorithm for multiprocessor scheduling. *IEEE Transactions on Parallel and Distributed Systems*, pp. 113-120. December 19-21, TaiWan.
- Ren, L., Zhang, L., 2011. Key issues in cloud simulation platform based on cloud computing. *Proceedings of the 23th European Modeling & Simulation Symposium*, pp. 50-507. September 12-14, Rome( Italy).
- Li, B.H., Chai, X.D., Hou, B.C., 2009. Networked modeling & simulation platform based on concept of cloud computing—cloud simulation platform. *Journal of System Simulation*, 21(17), 5292-5299.
- Li, B.H., Chai, X.D., Zhang, L., 2012. New advances of the research on cloud simulation. *Advanced Methods, Techniques, and Applications in Modeling and Simulation*. Japan: Springer.
- Low, M.Y.H., 2002. Dynamic load-balancing for bsp time warp. *Proceedings of the 35th Annual Simulation Symposium*, pp. 267-274. April 14-18, San Diego (California, USA).
- Lui, J.C.S., Chan, M.F., 2002. An efficient partitioning algorithm for distributed virtual environment systems. *IEEE Transactions on Parallel and Distributed Systems*, pp. 193-211. December 17-20, TaiWan.
- Pandey, S., Wu, L., Guru, S.M., 2010. A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications*, pp. 400-407. April 20-23, Perth (Australia).
- Peschlow, P., Honecker, T., Martini, P., 2007. A flexible dynamic partitioning algorithm for optimistic distributed simulation. *Proceedings of the 21st International Workshop on Principles of Advanced and Distributed Simulation*, pp. 219-228. June 12-15, San Diego (California, USA).

# EMPIRICAL MODELING AND SIMULATION FOR DISCHARGE DYNAMICS ENABLING CATCHMENT-SCALE WATER QUALITY MANAGEMENT

Huma Zia<sup>(a)</sup>, Nick R. Harris<sup>(b)</sup>, Geoff V. Merrett<sup>(c)</sup>

Electronics and Computer Science, University of Southampton, UK

<sup>(a)</sup>[hz2g11@ecs.soton.ac.uk](mailto:hz2g11@ecs.soton.ac.uk), <sup>(b)</sup>[nrh@ecs.soton.ac.uk](mailto:nrh@ecs.soton.ac.uk), <sup>(c)</sup>[gvm@ecs.soton.ac.uk](mailto:gvm@ecs.soton.ac.uk)

## ABSTRACT

Excessive or poorly timed application of irrigation and fertilizers, coupled with inherent inefficiency of nutrient uptake by crops result in nutrient fluxes into the water system. Due to the recent adoption of WSNs in precision agriculture, it is proposed that existing networked agricultural activities can be leveraged into an integrated mechanism by sharing information about discharges and predicting their impact, allowing dynamic decision making for irrigation strategies. Since resource constraints on network nodes (e.g. battery life, computing power etc.) require a simplified predictive model, low-dimensional model parameters are derived from the existing National Resource Conservation Method (NRCS). An M5 decision tree algorithm is then used to develop predictive models for depth ( $Q$ ), response-time ( $t_r$ ) and duration ( $t_d$ ) of the discharge. 10-fold cross-validation of these models demonstrates RRMSE of 10.2%, 30% and 9.6% for  $Q$ ,  $t_r$  and  $t_d$  respectively. Furthermore, performance of these models is validated using multiple linear regression method.

Keywords: discharge prediction, wireless sensor networks, M5 decision trees, simplified model

## 1. INTRODUCTION

Water quality degradation in a catchment is mainly attributed to outdated agricultural practices. Excessive or poorly timed application of pesticides, irrigation water and fertilizer result in nutrient fluxes into the water system with main issues being due to phosphorous (P) and nitrogen (N) losses (EPA 2009). In addition, the inherent inefficiency of nutrient uptake by crops (up to 50% for N and 10% uptake for P) renders nutrient outflows inevitable. This implies that adopting a reutilization mechanism of drainage and nutrients within the farm system can prove to be a valuable strategy to manage these outflows before they end up in rivers (Harper 2012). However, it is challenging to make valid predictions about outflows (what and when to expect) and then make them available across a farm system for their timely reutilization.

Over recent years, wireless sensor networks (WSNs), due to their low cost and real time data availability, have received considerable attention in precision agriculture. It is believed that, despite their limitations, there is huge potential for leveraging existing networked agricultural activities into an

integrated mechanism by sharing information about discharges (Zia, Harris et al. 2013). However, there is no framework to investigate and implement such a mechanism. In this paper, we propose a framework for water quality monitoring control and management (WQMCM) using collaborative WSN's in a catchment to investigate and enable such a mechanism. The basic architecture comprises modules to enable individual networks to evaluate if a correlation exists between neighboring events and the events within their own zone, to predict their impact and then adapt the local monitoring and management strategy. This paper focuses on the development and evaluation of the discharge prediction aspects.

For the prediction of discharges, various physical and mathematical hydrological models have been developed. Although popular in academic research, their dependence on acquiring numerous event and land parameters, the need for calibrating models to individual areas, and the tremendous computational burden involved in running the models makes widespread application complicated and difficult (Basha, Ravela et al. 2008). A computing model running on WSNs requires a simplified underlying physical model, based on fewer and, ideally, real-time field parameters acquired autonomously. In that respect, data-driven techniques based on machine learning, are becoming popular in hydrological modeling (Dawson and Wilby 1998; Wilby, Abrahart et al. 2003; Solomatine and Siek 2006; Galelli and Castelletti 2013), can yield low computational complexity. However, existing models are developed for very large catchments (>1000ha), and hence use years of data as training samples to learn the heterogeneity of such large areas. These models use precipitation and temperature instances, however, do not take into account field conditions which can have visible impact on flow volumes for smaller lands. In this paper, we use a popular NRCS curve number model as a basis for deriving and evaluating simplified model parameters based on real field conditions. In this paper, M5 tree learning algorithm is used to generate the predictive models based on the proposed parameters and fewer training samples. The effect of different feature sets and training sizes, on the prediction performance of the models, are evaluated and discussed.



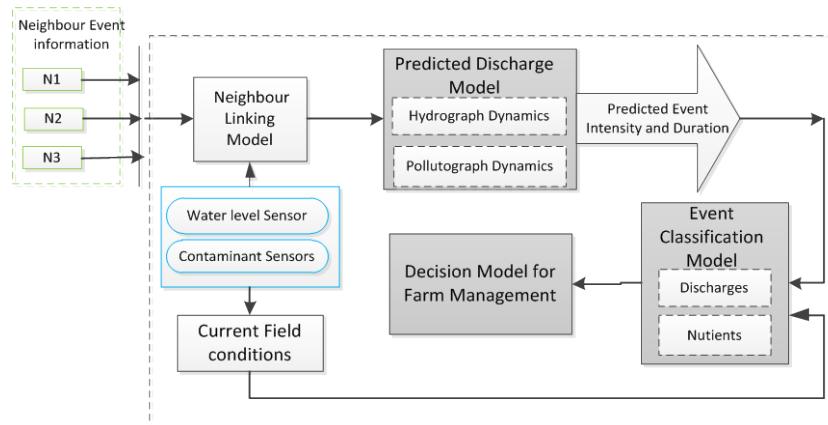


Figure 1: Architecture of the WQMCM framework

## 2. ARCHITECTURE OF WQMCM FRAMEWORK

The fundamental part of the WQMCM framework is that individual networks learn their environment to predict the impact of events elsewhere in the catchment on their own zone of influence and therefore adopt a management strategy. The predicted events can then be classified to allow the adjustment of management strategy accordingly. The overall block diagram of the architecture is illustrated in Figure 1. There are four key modules in this architecture which are briefly discussed below; however, this paper focuses on the discharge predictive model.

### 2.1. Neighbor Linking Model

As a network receives event information shared by its neighbors, it first needs to be able to correlate those neighboring events to events within its own zone of influence. This requires an individual network to learn about its neighbors which are likely to have an impact on it in case of an event. This linking process determines if a linear relationship exists between the discharge sensed by the networks sensors and the event information shared by a particular neighbor. Geographical filtering and linear regression is used to determine these neighbor links.

### 2.2. Predicted Discharge Model

Once a link is established, the next step is to then develop a learning model for predicting discharge dynamics in response to the event information shared by the linked neighbor. The model is termed as 'Predicted Discharge Model' and is developed separately for hydrograph and pollutograph dynamics. This paper focuses on the hydrograph predictive model.

For the prediction of discharges, various physical and mathematical hydrological models have been developed. Although popular in research, their dependence on acquiring numerous input parameters, the need for calibration, and the tremendous computational burden involved in running the models makes wide-spread application complicated and difficult for sensor networks (Basha, Ravela et al. 2008). Furthermore, for implementing the WQMCM

framework, constraints are associated with the practicality of information sharing among neighbors and the transmission costs linked with sharing high-dimensional input parameters for the predictive models.

To develop these models, machine learning algorithms are used on the accumulated training set from the previous stage. The training instances are based on a set of simplified model parameters which are derived from a mathematical hydrological in the later section.

### 2.3. Event Classification Model

Based on the predicted discharge dynamics, the event is then classified as to whether it is reusable or not. It involves the interpretation of discharge loads (of a hydrograph and a pollutograph) into well-defined levels, and then setting thresholds for a local farm based on its field conditions and irrigation or fertilization requirements. An event maybe classified as unusable on the basis of whether it is a high pollutant and a flood event (High risk) or a low pollutant but a low discharge event. On the other hand, an event is classified as usable if either of water or nutrients can be reutilized from the discharges.

### 2.4. Decision Model

The 'Decision Model', then decides either to raise an alert, in case of high pollutant loads, or reutilize the discharges. The model evaluates the economic and environmental benefit of the reuse, in particular, on its field and, in general, on the environment.

The challenge lies in designing a classification and decision model which takes into account local field conditions, predicted event dynamics and its likely benefits or repercussions on the field and then decide for the adapting management practices.

## 3. PREDICTED DISCHARGE MODEL – SIMPLIFICATION OF MODEL PARAMETERS

A runoff or drainage discharge is represented using a hydrograph as shown in Figure 2A. Before the runoff



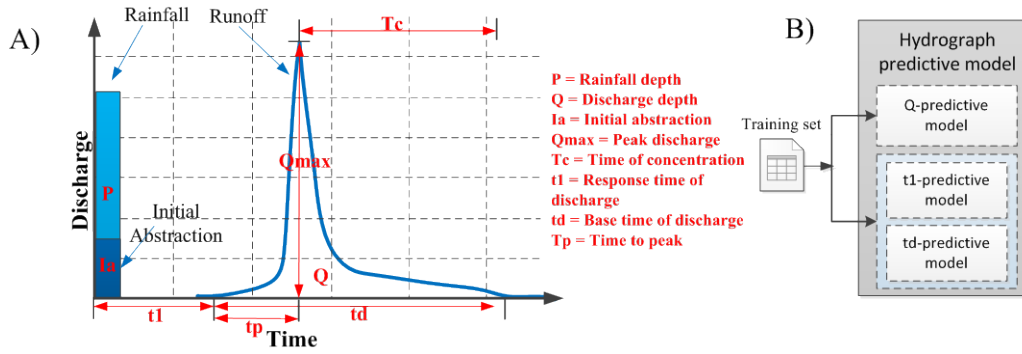


Figure 2: A) A discharge hydrograph, B) Predictive model for hydrograph dynamics

occurs, a certain volume of rainfall (termed as initial abstraction) is either retained in surface depressions and taken up by vegetation, or lost through evaporation, and infiltration. For the hydrograph-predictive model, the parameters of interest are;

1. Depth of discharge ( $Q$ )
2. Response time of discharge ( $t_l$ )
3. Duration of discharge ( $t_d$ )

These parameters provide information which enables effective reutilization of expected discharges. Individual learning models are developed to obtain these parameters termed as  $Q$ -predictive model,  $t_l$ -predictive model and  $t_d$ -predictive model (as shown in Figure 2B).

### 3.1. Mathematical Model for 'Q'

One of the most popular and simpler methods to estimate the volume of direct surface runoff for a given rainfall event, is the Natural Resource Conservation Service (NRCS) Curve Number (CN) method (Hawkins, Hjelmfelt Jr et al. 1985). Using this method,  $Q$  is computed as follows;

$$Q = \frac{\left[ P - 0.2 \left( \frac{1000}{CN} - 10 \right) \right]^2}{P + 0.8 \left( \frac{1000}{CN} - 10 \right)}$$

Where,  $P$  is the rainfall depth and  $CN$  is a coefficient reducing the total precipitation to runoff potential after surface absorption (with values in the range 0-100). The higher the  $CN$  coefficient, the higher is the runoff potential. It is computed considering the type of land use, land treatment, hydrological condition, hydrological soil group, and antecedent soil moisture condition (AMC). The volume of rainfall either retained in surface depressions or lost through evaporation or infiltration, termed as the initial abstraction ( $I_a$ ), is assumed to be 20% of the potential soil moisture retention (Hawkins, Hjelmfelt Jr et al. 1985).

### 3.2. Mathematical Model for ' $t_l$ ' and ' $t_d$ '

As evident from Figure 2A,  $t_d$  is expressed as;

$$t_d = T_c + t_p$$

Where,  $T_c$  is time for runoff to travel from the furthest distance in the watershed to the location where  $Q$  is to be determined, and  $t_p$  is the time to peak discharge. Typically there are three distinct runoff patterns in a watershed such as sheet flow, shallow concentrated flow, and channel flow. Numerical equations based on the underlying physical model are described below.

$$T_c = \frac{0.007(nL)^{0.8}}{(P_2)^{0.5}(S)^{0.4}} + \frac{L}{3600V} + \frac{L}{3600} \left( \frac{n}{1.49(R)^{\frac{2}{3}}(s)^{0.5}} \right)$$

Where,  $L$  is length (ft.) of flow pattern,  $n$  represents land cover,  $P_2$  is 2-year return period 24 hour precipitation (in.) for a region,  $R$  is hydraulic radius (ft.),  $s$  is average ground slope (ft.-vertical/ft.-horizontal),  $T_t$  is travel time (hr.), and  $V$  is average velocity (ft./s) of water.

As per the author's best knowledge, there is no direct mathematical equation to express  $t_p$  in the NRCS method. The other parameter required is  $t_l$ , and once again there is no mathematical expression for this. However, both are extracted from hydrograph plots drawn using the convolution of incremental runoff depth and unit hydrograph flow rate for a specific region. The unit hydrograph is a hypothetical unit response of a watershed (in terms of runoff volume and timing) to a unit input of rainfall. It is specific to a particular watershed, rainfall distribution ( $RD$ ), and rainfall duration ( $P_d$ ) such as 1-hour, 6-hour, or 24-hour (Shaw, Beven et al. 2010).

### 3.3. Limitations in Mathematical Model

The NRCS method, although simpler than the other models, still presents a challenge of acquiring a variety of permanent and transient parameters for every field under observation to determine discharge dynamics (Eq. (1) and Eq. (3)). Under the WQMCM framework paradigm, sharing these parameters among networks is not practical as it incurs high transmission costs resulting in low battery life of the deployed sensors.

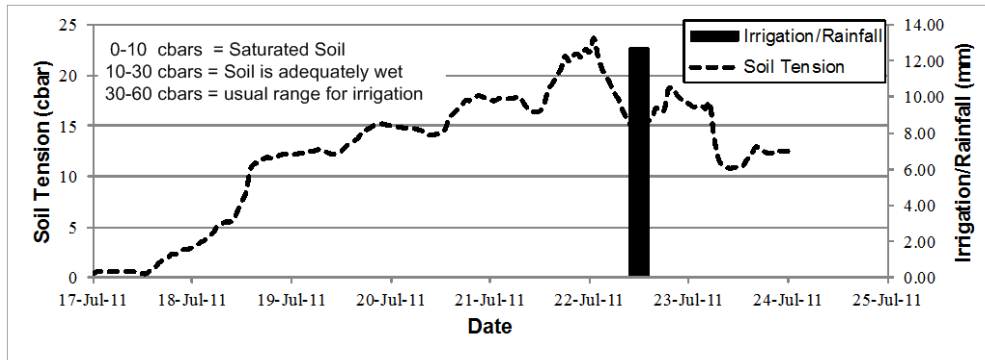


Figure 3: Soil moisture conditions in response to irrigation events in a field

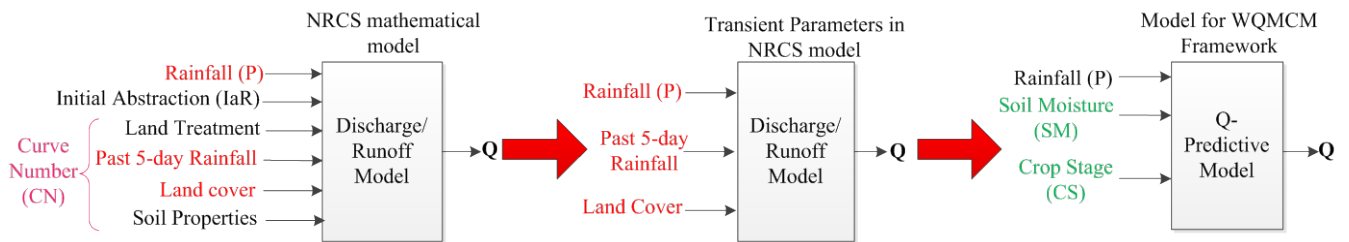


Figure 4: Model simplification for a  $Q$ -predictive model

Moreover, at the time the NRCS method was developed, due to the absence of remote and inexpensive sensing measures, proxy parameters, average values or manual observations were used to represent land conditions. An example is AMC, which is used to determine CN.

This is represented by using the amount of rainfall received in the five days preceding the storm event, which is a subjective judgment, instead of a physical reality (Fennessey and Hawkins 2001). In addition, type and extent of land cover, slope and land treatment etc., is determined by manual observation of the field, which limits autonomous monitoring and renders result prone to error. Furthermore, determining  $t_l$  and  $t_d$  is computationally intensive. This implies that low-dimensional model parameters are required which should take into account real time field conditions in an autonomous manner.

### 3.4. Model Simplification for $Q$ , $t_l$ and $t_d$

During the last decade the area of empirical modeling received an important boost due to developments in the area of WSNs and machine learning. It is anticipated that learning models yield low computational complexity. Here, the authors derive a simplified model based on the NRCS model. This simplification is based on two steps; firstly the transient parameters from the NRCS model parameters are selected for each of the predictive models for  $Q$ ,  $t_l$  and  $t_d$ . This is because learning models are trained only on transient values. After this, the transient parameters are analyzed for likely improvements made possible by using WSNs.

For  $Q$ , model simplification is as shown in Figure 4. The transient parameters in the NRCS model are rainfall depth, past 5-day rainfall and land cover. With increasing adoption of WSNs in agriculture, it is more practical to use this technology to extract real field conditions for prediction. For example, methods such as field imaging and signal attenuation methods have been used to determine the plant biomass autonomously (Vellidis, Savelle et al. 2011). This can be interpreted into the crop stage. Similarly, various applications have used sensors to monitor soil moisture conditions of the field for precision irrigation (Vellidis, Tucker et al. 2008; Zia, Harris et al. 2013). Therefore, it is proposed to use actual soil moisture values instead of the 5-day rainfall index.

In order to validate the limitation of 5-day rainfall index to represent AMC, we have analyzed season long data observed in a precision irrigation application, supplied by the University of Georgia (Vellidis, Tucker et al. 2008). The analysis show that in many cases the soil moisture condition was measured as moderate, although the field did not receive any rainfall or irrigation in the last 5 days. Figure 3 plots a week long data of measured soil tension (represents soil moisture). Using the 5-day rainfall index, on 22<sup>nd</sup> July, dry soil conditions would be estimated, due to the fact that there was no rain in the preceding 5 days. However, the actual soil condition is measured as adequately saturated by the sensors. This leads to incorrect determination of drainage after a rainfall or irrigation. Therefore, rainfall, soil moisture and crop stage are proposed as the simplified model parameters for the prediction of  $Q$  (Figure 5).

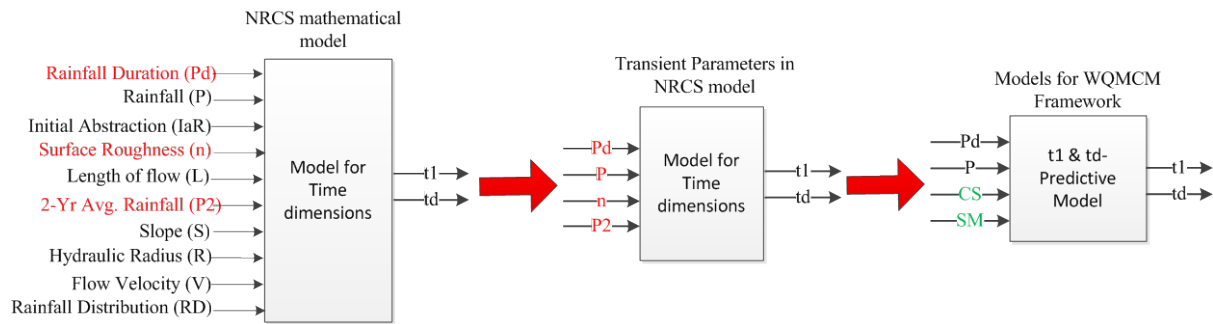


Figure 5: Model simplification for a  $t_1$  and  $t_d$ -predictive models

As already discussed, for  $t_1$  and  $t_d$ , the mathematical model and convolution method requires various parameters and historical data. Firstly the transient parameters are selected which include rainfall duration ( $P_d$ ), rainfall ( $P$ ), surface roughness ( $n$ ) and 2-year average rainfall ( $P_2$ ). This is further corroborated by analyzing an extensive set of simulated data (using NRCS based simulator(Davis), for which a routine in Matlab was written to extract  $t_1$  and  $t_d$ . The data indicated strong correlation of the selected transient parameters with  $t_1$  and  $t_d$ . This is because higher surface roughness inhibits flow rate and increases travel times. It is proposed in this paper that crop stage may well represent the field roughness. Furthermore, instead of relying on historical data for estimating  $P_2$  and  $RD$  for every region, it is proposed to use actual soil moisture conditions. Simulation results can be used to evaluate the effect of this substitution on prediction accuracy of  $t_1$  and  $t_d$ .

#### 4. SIMULATION AND RESULTS FOR HYDROGRAPH PREDICTIVE MODELS

Using machine learning algorithms, the models are trained on the historical data describing the phenomenon in question. Historical data includes known samples that are combinations of inputs and corresponding outputs. The learned model is then used to predict the outputs from the new input values. Examples of the most popular supervised learning methods used in data driven modeling for hydrological predictive systems are: statistical methods, artificial neural networks (ANN), and decision model trees.

An example of a statistical method is multiple linear regression (MLR), which postulates a model (linear function) and then find the parameter values that maximize its fit to the training data (Quinlan 1992). MLR has been used for flood forecasting, in which a model is learnt based on data for parameters such as precipitation, air temperature and river flow (Basha, Ravela et al. 2008). However, the training data is collected from historical data of 7 years. Furthermore, ANNs have been used for rain-runoff modeling (Wilby, Abraham et al. 2003) and stream-flow forecasting (Rasouli, Hsieh et al. 2012). One of the disadvantages of ANNs is that for a decision maker it is very difficult

to analyze the structure of the resulting ANN and to relate it to the outputs. However, there are approaches to numerical prediction that use piece-wise linear approximations which are much easier to interpret. One example is M5 decision model trees (Quinlan 1992), which has been demonstrated as an alternative to ANNs (Solomatine and Dulal 2003).

##### 4.1. Decision model trees for the predictive model

Model trees are an extension of regression trees, and include first order linear models at the leaf nodes, compared to zero-order models in regression trees. Model trees have higher predictive accuracy and are able to make predictions for values outside the training data range, which is not the case with regression trees (Kuzmanovski 2012). The inputs for M5 model trees are mainly selected according to the correlation analysis, which works very well (Solomatine and Dulal 2003). Predictions using M5 in, for example flood forecasting, have been reported to have given an accuracy of 80-95% (Solomatine and Xue 2004). Therefore, we use M5 decision tree algorithm for generating the predictive models in this paper. However, we also use MLR algorithm to compare its prediction performance with the M5 decision tree model.

##### 4.2. Datasets

For training and testing the model, historical data is generated using a simulator based on the NRCS method (Davis), which is developed in Matlab. A combination of various event depths, field conditions and event duration is considered to generate two sets of data – one for  $Q$  predictive model and the other for  $t_1$  and  $t_d$  predictive model. The obtained data set is then modified to substitute  $CN$  with the proposed simplified model parameters of  $CS$  and  $SM$ . To ensure robust evaluation of the model performance, the datasets are randomly sampled, in order to create training and testing subsets, respectively containing one-third and one-fourth of the available data.

##### 4.3. Model evaluation

The prediction accuracy of the learned models is evaluated using multi-assessment criteria (Hwang, Ham et al. 2012). The criteria considered are (i) RMSE (Root

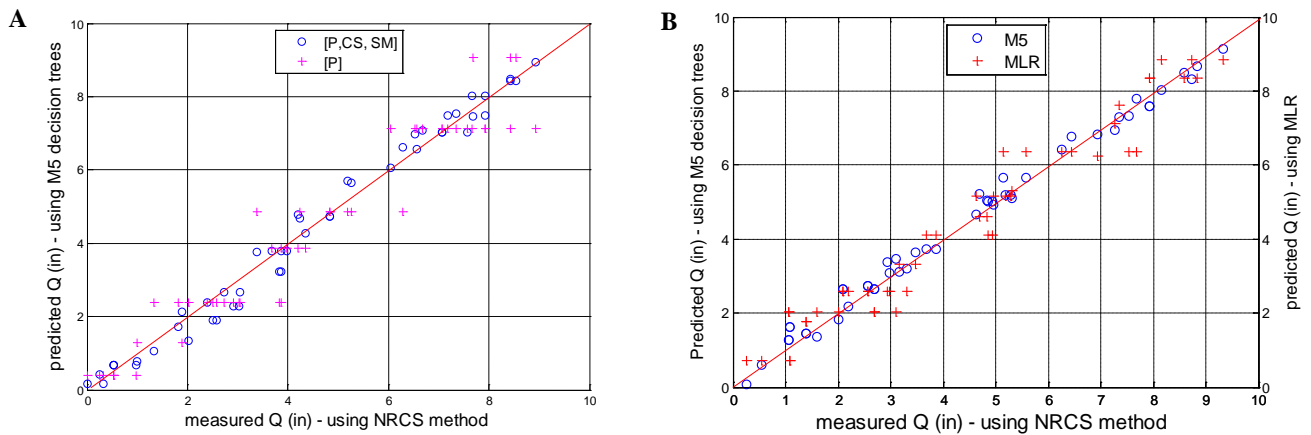


Figure 6: A) Plot of test data for  $Q$ -predictive models developed using M5 decision trees for different model parameters, B) Plot of test data for  $Q$ -predictive models developed using M5 decision trees and MLR algorithm for the proposed parameters.

Mean Square Error), which estimates the concentration of the data around the fitted equation, (ii) Mean Absolute error (MAE), (iii) Relative RMSE (RRMSE), which is the ratio of the variance of the residuals to the variance of the target values themselves and, (iv) R squared value ( $R^2$ ), which shows goodness of fit, i.e., correlation between actual and predicted values. RMSE and MAE are scale dependent measure and have the same unit as the data. A good value for RMSE is stated as half of the standard deviation value for the output data (Singh, Knapp et al. 2005). However, while comparing two forecasting models, a smaller RMSE indicates better forecasting accuracy. This comes out as 1.3 for  $Q$  and  $t_r$ , and 3.2 for  $t_d$ . Values of  $R^2$  and RRMSE can range between 0 and 1, where 1 means perfect forecasting. In this paper, the value of RRMSE is represented as a percentage. The predicted models developed using different model parameters and training set sizes, are evaluated with test data to compare their performance with the NRCS model. For performance evaluation of these models, we use M5 decision tree toolbox developed in MATLAB (Jekabsons 2010)) and the Java-based WEKA machine learning simulator (Hall, Frank et al. 2009).

#### 4.4. Comparative Assessment

The predicted test results of each model developed using the proposed simplified model parameters is compared with the measured test results of the NRCS method. In addition to that, more combinations of feature sets based on reduced parameters from the simplified model parameters are tested to see if further simplification can maintain reasonable prediction accuracy or not. The test results are plotted for each of the predictive models.

In order to compare the prediction performance of the models developed using M5 decision trees with another modeling technique, multiple linear regression model (MLR) is used here. The predicted results are plotted for each of the  $Q$ ,  $t_r$  and  $t_d$  models developed using the M5 decision trees and the MLR model.

#### 4.5. $Q$ -predictive model

For the prediction of  $Q$ , 100 training samples, based on the proposed parameters, are used to generate the M5 decision tree model. The prediction performance for this model with 10-fold cross validation has RMSE value of 0.23 and RRMSE as 10.2%, validate the use of proposed model parameters for  $Q$ . The plot of predicted output for  $Q$ -predictive model, using test data, shows a very good fit ( $R^2=0.98$ ) as compared to the plot of measured output using the NRCS model (as shown in Figure 6 A).

For the sake of comparison, an M5 decision tree learning model is developed using only a single input parameter of  $P$  for the prediction of  $Q$ . The plot for the model shows very poor fit and gives 30% RRMSE. This is because other parameters such as  $CS$  and  $SM$  have a substantial impact on the runoff or drainage expected from a field for a given rainfall. This validates that using  $CS$  and  $SM$ , along with  $P$ , in the learning model can significantly improve its prediction accuracy.

Later, MLR algorithm was used to compare its performance with the M5 decision tree model. For a small training set of 100 samples, both the models give almost comparable performance. As the M5 decision tree model had RMSE as 0.23 and RRMSE as 10.2%, the regression tree model has RMSE of 0.30 and RRMSE as 13.5%. The result for cross validation and test data for the two models is given in Table 1. Figure 6B illustrates the plot of the predicted output for test data using the two models. While trying to test these two models further, we reduced the number of training samples to 50 for developing the models which showed a significant performance difference. M5 decision tree model gives RMSE as 0.317 and  $R^2$  as 0.984, whereas for MLR model, RMSE rises to 0.915 and  $R^2$  drops down to 0.506, which shows a poor fit of the MLR model for smaller training sample size.

#### 4.6. $t_r$ -Predictive and $t_d$ -Predictive model

For  $t_r$ , initially 100 training instances, based on the same model parameters as in the  $Q$ - predictive model

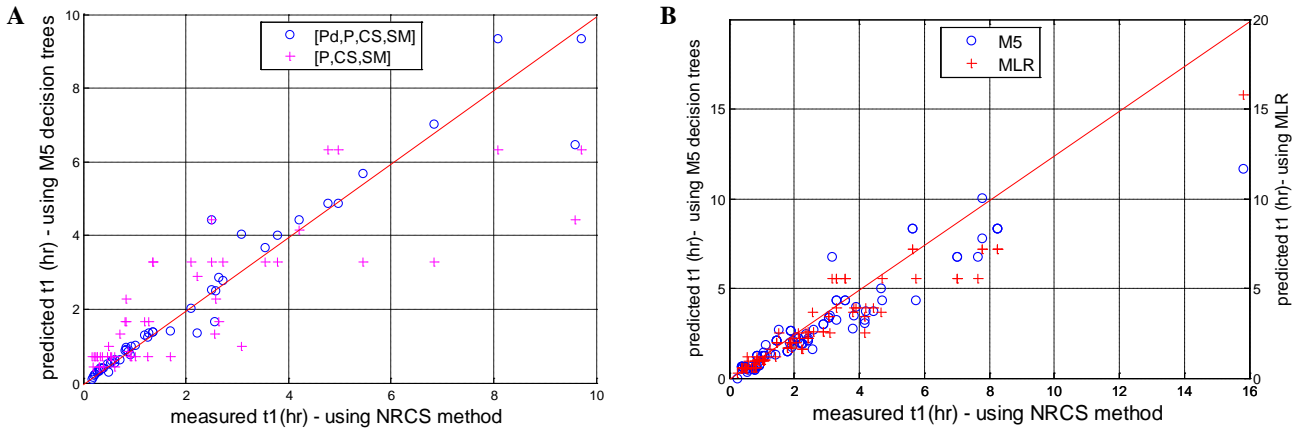


Figure 7: A) Plot of test data for  $t_I$ -predictive models developed using M5 decision trees for different model parameters, B) Plot of test data for  $t_I$ -predictive models developed using M5 decision trees and MLR algorithm for the proposed parameters.

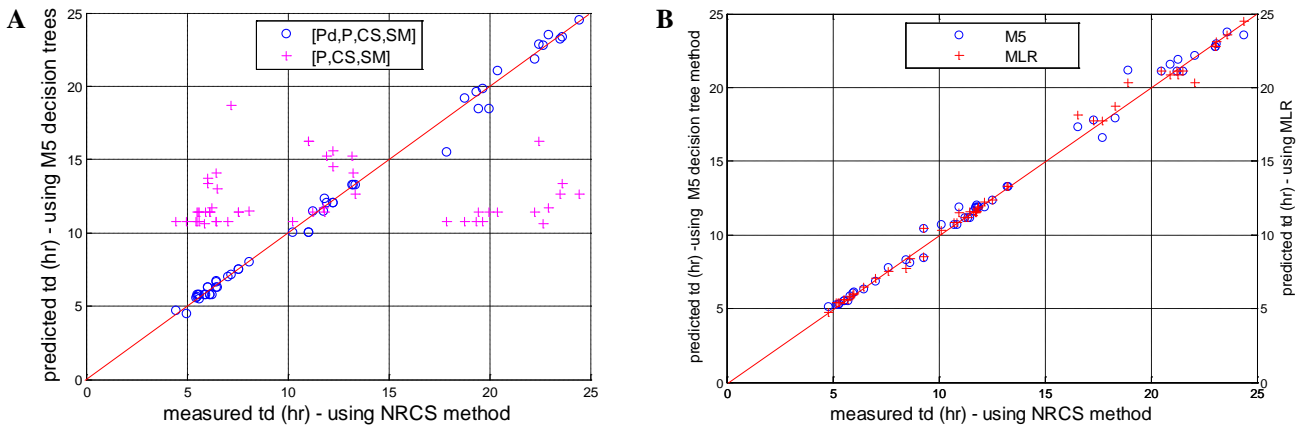


Figure 8: A) Plot of test data for  $t_d$ -predictive models developed using M5 decision trees for different model parameters, B) Plot of test data for  $t_d$ -predictive models developed using M5 decision trees and MLR algorithm for the proposed parameters.

Table 1: Cross validation and test data results for the predictive models developed using M5 decision trees (M5) and multi linear regression method (MLR)

|                         | Model | 10 fold cross-validation |      |      |        | Testing |      |      |        |
|-------------------------|-------|--------------------------|------|------|--------|---------|------|------|--------|
|                         |       | R2                       | MAE  | RMSE | RRMSE  | R2      | MAE  | RMSE | RRMSE  |
| $Q$ -Predictive Model   | M5    | 0.99                     | 0.19 | 0.23 | 10.20% | 0.98    | 0.16 | 0.20 | 8.00%  |
|                         | MLR   | 0.98                     | 0.24 | 0.30 | 13.50% | 0.99    | 0.20 | 0.26 | 10.20% |
| $t_I$ -Predictive Model | M5    | 0.91                     | 0.45 | 0.61 | 30.00% | 0.85    | 0.46 | 0.56 | 38.10% |
|                         | MLR   | 0.74                     | 0.73 | 1.00 | 50.00% | 0.70    | 0.90 | 0.95 | 60.00% |
| $t_d$ -Predictive Model | M5    | 0.99                     | 0.45 | 0.62 | 9.60%  | 0.99    | 0.45 | 0.56 | 7.40%  |
|                         | MLR   | 0.96                     | 0.75 | 1.01 | 15.00% | 0.96    | 0.82 | 1.05 | 13.97% |

( $P$ ,  $CS$ ,  $SM$ ), were used to generate the model for the sake of comparison using M5 decision tree. However, the model performance was very poor with RMSE of 1.47, which is higher than the acceptable value of 1.3, and RRMSE as 74%. This validates that the same model parameters cannot be used for predicting  $t_I$ . Therefore, the proposed parameters ( $P_d$ ,  $P$ ,  $CS$ , and  $SM$ ) were then used to generate the model. This substantially

improved the model performance (RMSE= 0.533, RRMSE=27%). This is further illustrated by plotting the result of test data predicted using the above two models against the measured output of the NRCS model in Figure 7A. R2 for the first model (using  $P$ ,  $CS$  &  $SM$ ) comes as 0.667, whereas for the second model developed using the proposed parameters, it comes as 0.96. The performance of the  $t_I$ -predictive model can be



further improved if the number of training samples is increased to two-third of the total samples, which is around 300. Hence the model trained on 300 samples reduces the RRMSE to 16.8% and the RMSE value to 0.318.

Furthermore, the 5 decision tree model for  $t_f$  is compared with another learning model, the MLR model. The performance evaluation parameters for cross validation and testing data is given in Table 1. There is a substantial difference in performance with 30% and 50% RRMSE respectively for the M5 decision tree model and the MLR model. The plot using test data for the two models, illustrated in Figure 7B, shows poor fit of the MLR model ( $R^2=0.74$ ), specifically for higher values of  $t_f$ . The low performance of MLR model is attributed to its model architecture. In MLR model, the predicted output is simply the mean of the output values associated to the inputs falling in a specific leaf. Whereas, M5 decision model trees show better performance because they have a linear function model in each leaf.

For  $t_d$ , the plot shown in Figure 8A) demonstrates that the results of the model trained on 100 samples of the parameter set consisting of P, CS and, SM, using M5 decision trees algorithm, fits poorly to a 1:1 ratio ( $R^2=0.107$ , RRMSE=98%). However, in comparison to that, the results of the model generated using the proposed parameter set (P<sub>d</sub>, P, CS, & SM), for the same training set, show good performance with  $R^2=0.991$  and RRMSE =8.2%. As illustrated in the model, the prediction of  $t_d$  shows higher correlation with P<sub>d</sub>.

The comparison of the results of M5 decision tree and MLR algorithms for the development of the  $t_d$  – predictive models is given in Table 1. The value of RRMSE and RMSE, for the later model, increases by 50%. RRMSE increases from 9.6% to 15%, and, RMSE changes from 0.62 to 1.01 respectively. However, the models show adequate fit on the plot as illustrated in Figure 8B).

## 5. CONCLUSION

This paper has proposed that individual farm-scale networks can be integrated into a collaborative framework to support catchment-scale water quality monitoring and management to learn and predict the impact of catchment events. This enables reutilization and timely control of nutrient outflows within the farm system. Since a computing model on a sensor network, for the implementation of the collaborative WQMCM framework, requires a simplified underlying physical model therefore, low-dimensional model parameters are derived from the existing NRCS method for the prediction of discharge dynamics. An M5 decision tree algorithm is used to develop predictive models for depth ( $Q$ ), response time ( $t_f$ ) and duration ( $t_d$ ) of the discharge, based on the proposed model parameters. 10-fold cross-validation of these models demonstrates RRSE of 10.2%, 30% and 9.6% for  $Q$ ,  $t_f$  and  $t_d$  respectively. Furthermore, performance of these models is validated using multiple linear regression model.

## ACKNOWLEDGEMENTS

We would like to thank Prof. George Vellidis and Mike Tucker at the University of Georgia for providing us with season long soil moisture data of a field.

## REFERENCES

- Basha, E. A., S. Ravela, et al. (2008). Model-based monitoring for early warning flood detection. Proc. of Embedded network sensor systems Conf., ACM.
- Davis, T. SCS Unit Hydrograph Convolution : Hydrograph Generation and Analysis Tool, Matlab.
- Dawson, C. W. and R. Wilby (1998). "An artificial neural network approach to rainfall-runoff modelling." Hydrological Sciences Journal **43**(1): 47-66.
- EPA (2009). National Water Quality Inventory: Report to Congress ; 2004 Reporting Cycle, US Environmental Protection Agency.
- Fennessey, L. and R. Hawkins (2001). The NRCS Curve Number, a New Look at an Old Tool. Proc. of Pennsylvania Stormwater Management Symp., Villanova Uni.
- Galelli, S. and A. Castelletti (2013). "Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling." Hydrology & Earth System Sciences Discussions **10**(2).
- Hall, M., E. Frank, et al. (2009). "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter **11**(1): 10-18.
- Harper, H. H. (2012). Impacts of Reuse Irrigation on Nutrient Loadings and Transport in Urbanized Drainage Basins, Florida Stormwater Association, Environmental Research & Design, Inc.
- Hawkins, R. H., A. T. Hjelmfelt Jr, et al. (1985). "Runoff probability, storm depth, and curve numbers." Jour. of Irrig. and Drainage Engg. **111**(4).
- Hwang, S. H., D. H. Ham, et al. (2012). "A new measure for assessing the efficiency of hydrological data-driven forecasting models." Hydrological Sciences Journal **57**(7): 1257-1274.
- Jekabsons, G. (2010). "M5PrimeLab: M5' regression tree and model tree toolbox for Matlab." from <http://www.cs.rtu.lv/jekabsons/Files/M5PrimeLab.pdf>.
- Kuzmanovski, V. (2012). Integration of expert knowledge and predictive learning: Modelling water flows in agriculture MS Thesis, Jožef Stefan International Postgraduate School, Ljubljana, Slovenia
- Quinlan, J. R. (1992). Learning with continuous classes. Proceedings of the 5th Australian joint Conference on Artificial Intelligence, Singapore.



- Rasouli, K., W. W. Hsieh, et al. (2012). "Daily streamflow forecasting by machine learning methods with weather and climate inputs." Journal of Hydrology **414-415**(0): 284-293.
- Shaw, E. M., K. J. Beven, et al. (2010). Hydrology in practice, Taylor & Francis US.
- Singh, J., H. V. Knapp, et al. (2005). "Hydrological modeling of the iroquois river watershed using HSPF and SWAT1." Journal of the American Water Resources Asso. **41**(2): 343-360.
- Solomatine, D. and Y. Xue (2004). "M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China." Journal of Hydrologic Engineering **9**(6): 491-501.
- Solomatine, D. P. and K. N. Dulal (2003). "Model trees as an alternative to neural networks in rainfall—runoff modelling." Hydrological Sciences Journal **48**(3): 399-411.
- Solomatine, D. P. and M. B. Siek (2006). "Modular learning models in forecasting natural phenomena." Neural Networks **19**(2): 215-224.
- Vellidis, G., H. Savelle, et al. (2011). "NDVI response of cotton to nitrogen application rates in Georgia." Precision Agriculture: 359.
- Vellidis, G., M. Tucker, et al. (2008). "A real-time wireless smart sensor array for scheduling irrigation." Computers and Electronics in Agriculture **61**(1): 44-50.
- Wilby, R., R. Abrahart, et al. (2003). "Detection of conceptual model rainfall—runoff processes inside an artificial neural network." Hydrological Sciences Journal **48**(2): 163-181.
- Zia, H., N. R. Harris, et al. (2013). "The Impact of Agricultural Activities on Water Quality: A Case for Collaborative Catchment-Scale Management using Integrated Wireless Sensor Networks." Computers and Electronics in Agriculture.

# EVENT-ORIENTED CONTROL FUNCTIONS FOR ENHANCING DEVELOPMENT PROCESS OF WAR-GAME SIMULATORS

Se Jung Kwon<sup>(a)</sup>, Changbeom Choi<sup>(b)</sup>, Tag Gon Kim<sup>(c)</sup>

Dept. of Electrical Engineering, KAIST  
Daejeon, KOREA

<sup>(a)</sup>[sjkwon@smslab.kaist.ac.kr](mailto:sjkwon@smslab.kaist.ac.kr), <sup>(b)</sup>[cbchoi@kaist.ac.kr](mailto:cbchoi@kaist.ac.kr), <sup>(c)</sup>[tkim@ee.kaist.ac.kr](mailto:tkim@ee.kaist.ac.kr)

## ABSTRACT

The development processes of discrete event simulation software may not be straightforward. The processes can be interrupted repeatedly by modified requirements. This paper proposes an approach for making the iterative processes efficient. In order to keep the development cost low, we seek to avoid modifying the simulation model inside as much as possible. Instead, our proposed work enables the same simulators to generate the other execution results adding additional information. This approach can be enabled with the Event-based Simulation concept and the Event-oriented Control Functions that are mapped to concerned events. The simulation engine can handle the input/output level data by accessing events. In this way, users can control their simulation simply by describing the functions mapped to events, instead of modifying simulation models. This paper also includes case studies to support contributions, assuming that a war-game model has been developing.

Keywords: discrete event systems modeling and simulation, event-based simulation control, DEVS formalism, war-game simulator development

## 1. INTRODUCTION

A software development process (or life cycle) means a structure that is imposed on the development of a software product. Many studies and applications in the Software Engineering community have been published for efficient development and maintenance. Similarly, the development processes can be applied for discrete event simulation software, as well.

Among the discrete event simulators, the S/W development processes of the war-game simulators have unusual features in contrast to other simulators. Generally, it is hard for developers to understand the model behavior of military domains. Hence, the stakeholders should suggest the objectives of their ordered simulator and behaviors of real military systems. Stakeholders may be fully able to suggest what to do. However, not only do they propose ambiguous suggestions based on the real systems' behaviors but they also realize the need to change the requirements

based on the results of simulators or other miscellaneous reasons

According to an exploration or alteration of requirements, simulation models must be redeveloped during the development process. The repetitions of model modification and redevelopment cause an increase in the development time and cost. Assuming that the iterative developments and changing requirements are inevitable, this paper attempts to enhance the iterative processes by reducing the cost of developments.

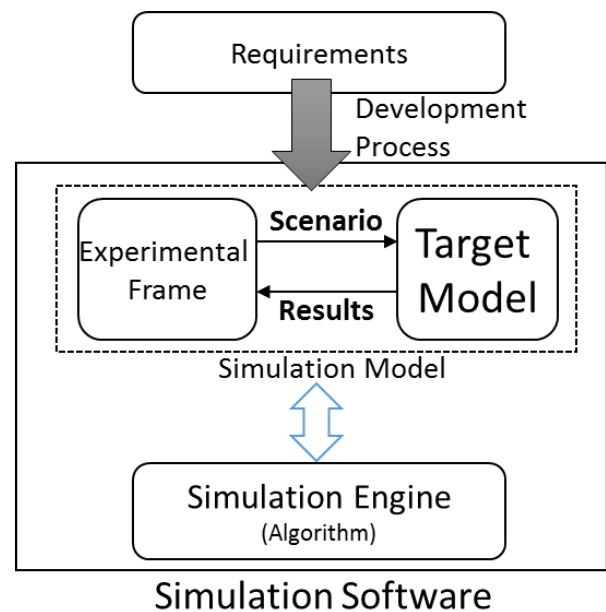


Figure 1: Compositions of Simulation Software

Assuming that the simulation model consists of an Experimental Frame (EF) and a target model, as depicted in Figure 1, an experimental frame generates scenarios for testing or analysis. As the requirements change, developers will try to modify the experimental frame at first and try to make it generate modified scenarios for proper results. However, in most cases, a specific experimental frame cannot deal with new scenarios, even though it would be a better modification approach due to the simplicity. To apply the new requirements right, the inside of target model should be

modified. The modification of the whole models has perfect modifiability, but it causes higher development costs as depicted in Figure 2.

Hence, a key point of this paper is a model development method that has higher modifiability than modification of EF and lower development cost than modification of whole models, like the grey area of the Figure 2. By applying the approach to the existing iterative processes, the cost reduction of the development process can be achieved.

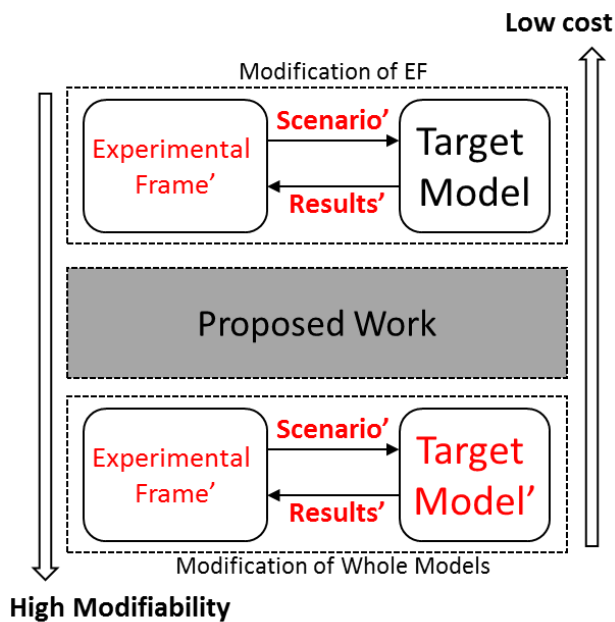


Figure 2: Existing Modification Approaches

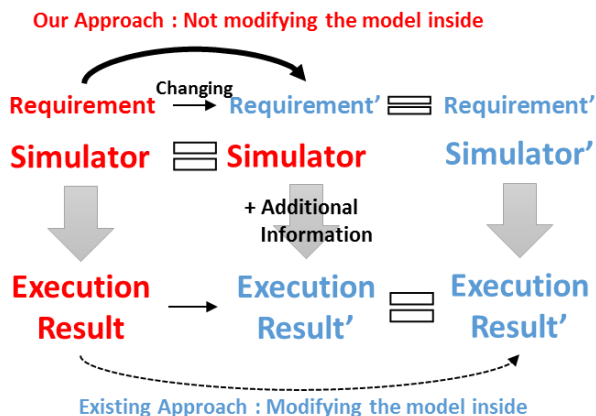


Figure 3: Distinction between Existing Approach and Our Approach as the Requirements Change

Figure 3 describes an approach of this paper, which does not modify the model inside as much as possible. Instead, our approach enables the same simulators to generate the other execution results, according to new requirements. In other words, our approach does not modify syntaxes of models, but it modifies semantics of simulation. This can be possible with described additional information, instead of modification.

For the application of the additional information to the simulation model, this paper adopts an event-based simulation concept, which is a basic execution method for discrete event simulations (Cota and Sargent 1992, Zeigler *et al.* 2000). An event, as a simulation unit of event-based simulation, is listed on an event list and executed in order to affect state transitions of a model. Although the objective models of the simple event-based simulation are event-oriented functions and global states, other object-oriented discrete event models can also be executed with additional interpreter algorithms. Each executed model generates time events and data events for the progression of their simulation. Time events schedule their next execution when the model should be executed, and data events are transmitted to other models with data (messages). The simulation engine cannot access the inside of object-oriented models (e.g. state variables). By accessing the generated events from models, the simulation engine can handle the input/output (I/O) level data. Although object-oriented reusable models can also reduce the development cost (Kim 1996), our approach concerns the I/O level data, not the model inside.

To handle the I/O level data, we define an *Event Control Model* (ECM), which consists of global states and *Event-oriented Control Functions* that are mapped to events. When some models should present different behaviors at the run-time, the same models with additional functions are executed through the event-based simulation engine. When the events are scheduled on the event list at the run-time, the mapped events are passed through the related control functions. For example, the variables of data events can be modulated by the mapped functions, and the behaviors of the models become different without modifying the model inside.

It is true that the model inside has to be modified at one time or another as the requirement is changed repeatedly. Nevertheless, there exist enough empirical evidences that a simulator shows the other behavior with the additional ECM without any modification of the model itself. They can be shown in Chapter 3.2 and Chapter 4.

The contents on which this paper focuses are mainly related to the modulation of variables in data events. By utilizing the events modulation and reducing development costs, these control approaches can make rapid prototyping (Martin 1990) more valuable against the sudden (or planned) changes of scenarios. In particular, the ECM amplifies some development strategies including successive prototyping like the *Sawtooth* model (Rowen 1990).

The target of this paper is a *Discrete Event systems Specification* (DEVS) model (Zeigler *et al.* 2000), which is one of the most frequently utilized system specifications to model discrete event systems in the real world. Since the event-based simulation algorithm is not limited to DEVS and can be applied to other discrete event simulations by designing proper

interpreter algorithms, this paper can also be extended to other discrete event simulations.

This paper is organized as follows: Section 2 presents the event-based simulation concept. Section 3 explains our proposed work, which contains the specifications for the event-oriented control function and its applications, and Section 4 illustrates case studies for the development process of war-game simulators. Finally, Section 5 concludes the paper.

## 2. EVENT-BASED SIMULATION

Event-based simulation is one of the most efficient and basic discrete event simulation strategies because of its simplicity. Event-based simulation works by prescheduling all events in an event list (Zeigler *et al.* 2000). In this view, Discrete Event Systems (DES) can be specified to the event-oriented functions mapped to events (or state transitions), which are as units in contrast with the object-oriented concepts. The event-oriented models consist of functions mapped to events and global variables modified by functions. The simulation engine executes a function mapped to an event by extracting it from the event list, and the executed functions insert newly generated events to the event list for scheduling. Since the event-oriented models, which are not passive models, have scheduling parts for events in contrast to object-oriented models, the event-oriented models are not separated from the execution algorithm. On the contrary, object-oriented models, e.g., DEVS models, are passive models and should be separated from the simulation engine. The following algorithm describes the event-based simulation concept.

```

Tglobal // current simulation time
EventList
// List of sorted event(time, target-function)

Simulation_Run()
while (Event-List is not empty)
 first = top of EventList
 delete the top of EventList
 Tglobal = first.time
 execute first.mapped-function
End while

Schedule_New_Event(time, target-function)
// called by functions mapped to events
create an event with the pair(time, target-function)
insert the event to Event-List

```

Algorithm 1: Simple Algorithm of Event Scheduling

Due to characteristics of object-oriented modeling theory, the models need interpreter algorithms in order to be executed by event-based simulation algorithm as depicted in Figure 4. For example, the execution of DEVS models with event-based simulation can be performed with a pre-process algorithm for flattening the hierarchical structure and a mediation algorithm between different interfaces (Kwon and Kim 2012)

because the DEVS models are structured hierarchically and systematically. Although the interpretation causes a little degradation of simulation performance, executing the object-oriented models by the event-based simulation engine may cause improved speed-up.

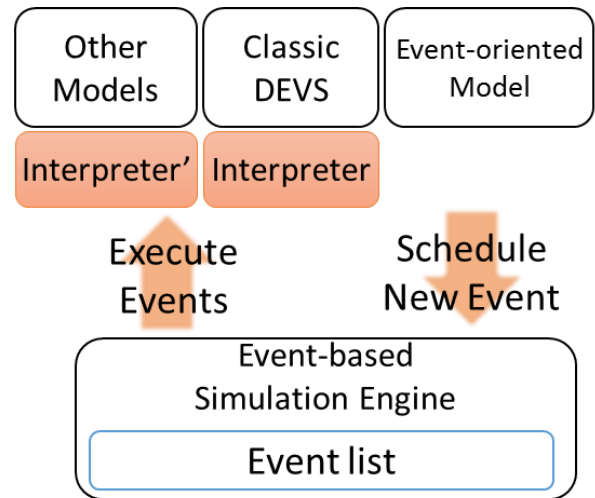


Figure 4: Event-based Simulation

## 3. EVENT-ORIENTED CONTROL FUNCTION AND ITS APPLICATION

This chapter describes the specifications of *Event Control Model* (ECM), including the *Event-oriented Control Function*. Its simulation and its applications are also covered.

### 3.1. Event Control Model with Event-oriented Control Functions

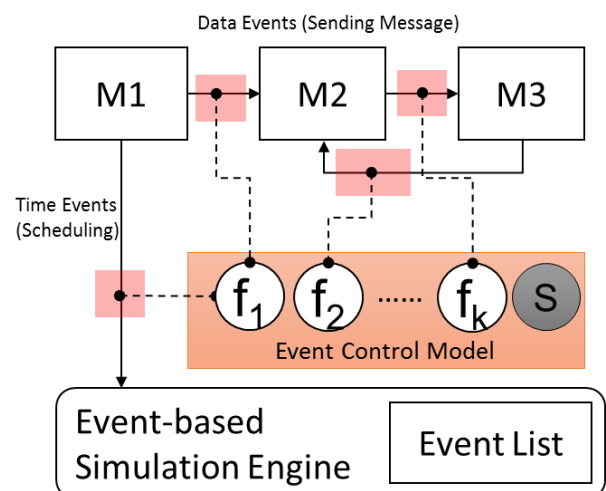


Figure 5: Accessible Points of ECM

The events, occurring state transitions, are listed on the event list of event-based simulation engine as mentioned above. The variables of events inside are public and accessible by the simulation engine as depicted in Figure 5. There are two types of state transitions, time events ( $E_{time}$ ) and data events ( $E_{data}$ ).

When the event occurs, the concerned event-oriented functions try to interfere with the occurred events inside.

We define *Event Control Model* (ECM) for the event-oriented control. The specifications of ECM are the following:

- $Ev = \{E_{time}, E_{data}\}$  is a set of events.
- $E_{time} = \langle T, t_N \rangle$   
 $T$  is a target function (model).  
 $t_N$  is a next scheduled time.
- $E_{data} = \langle T, t_N, \{V_i\} \rangle$   
 $T$  is a target function (model).  
 $t_N$  is a next scheduled time (Usually zero).  
 $\{V_i\}$  are variables.
- $ECM = \langle \{f_i\}, S, SELECT \rangle$   
 $f_i$  is an event-oriented control function.  
 $: Ev \times S \rightarrow \{Ev \cup \emptyset\} \times S$   
 $S$  is set of global states.  
 $SELECT$  is a tie-breaking selection function.  
 $: 2^{\{f_i\}} - \emptyset \rightarrow f_i$

The definitions of two events are a little different. A time event consists of a target model (a source model itself) and a next scheduled time. A data event consists of a target model (a destination model), a next scheduled time and variables. Though the next scheduled time of data events is usually zero, the user can handle the value to affect a time delay to the execution of events for the advance control.

The key point of ECM is the *Event-oriented Control Function*,  $f_i$ . A control function generates a modified event from a generated event of a model, or it can eliminate the event with a certain condition. For the cases that control functions need to store information, the global states are included in the specifications of ECM. The select function exists for resolving the priority problem because two or more functions, mapped to a same event, can be in conflict. The idea of the specifications is borrowed from the event-oriented models and DEVS formalism.

```

Schedule_New_Event(time, target-function(T))
// called by functions mapped to events
create an event(Ev) with the pair(time, T)
for each mapped control functions f_k
 $Ev = ECM.f_k(Ev)$
insert Ev to Event-List

```

Algorithm 2: Modified Algorithm for ECM

The algorithm 2 for executing the ECM is quite simple and has just two lines of added codes compared with the original *Schedule\_New\_Event* function in Algorithm 1. Before the generated events are inserted into the event list, the simulation engine calls the proper functions of the ECM.

Figure 6 shows the simulation environment of proposed work. The target model can be decomposed to

events manually or automatically. Users can select proper events according to the control objectives, and designs the control functions and states. The implemented ECM from the design is executed with the original target model, and it interferes in the event scheduling. The algorithm 2 is embedded in the interpreter algorithm in the event-based simulation engine.

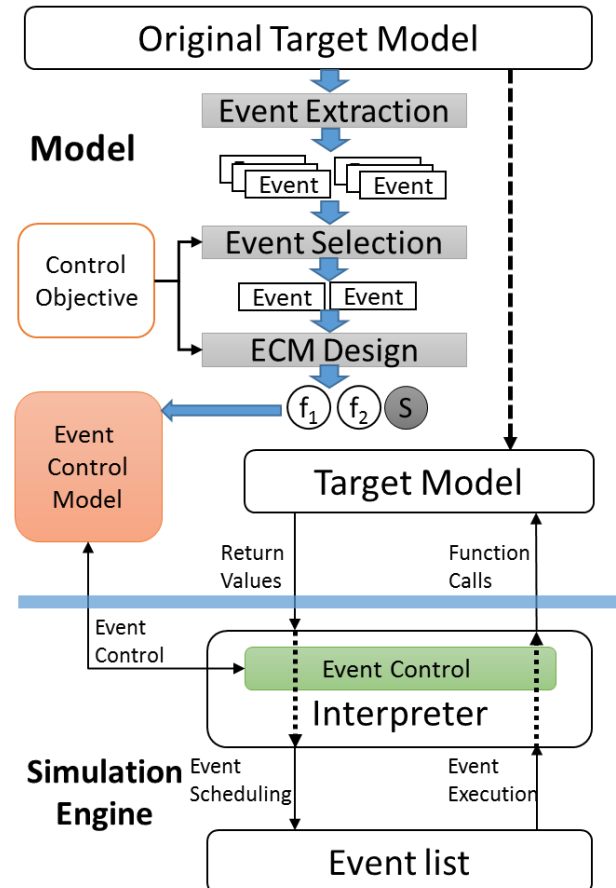


Figure 6: Design Process of the ECM and Proposed Simulation Environment

### 3.2. Applications

A primary usage is modifying variables of messages. Even if the model inside or state variables are unknown or fixed, the variables of I/O messages should be opened to the simulation engine. By describing event-oriented control functions, users can modulate the variables inside events. In the military domain, to simulate the real operations of equipment in detail, the model behavior usually includes the random variables due to the existence of errors of real operations. Though a certain model parameter was classified as non-prime value and became fixed value in the early stage of development, the value can be changed to the randomized value with a simple control function, which generates a modulated event with a random variable from an original event.

The simplest example is found in a Single Server Queuing Server model (known as a GBP model). A control function with a probability function can



modulate the processing time of a job from the Generator (G). This type of usages can be applied to various situations, e.g., fault injections to messages, adding probability functions, and so on. One of the detailed examples is described in Chapter 4.1.

If a simulation model consists of many homogenous events that have to be modulated, the ECM can be applied to not only an event, but also events in a lump. The representative applications of such cases involve the environment variables. The entities, including combat entities of the battlefield, are influenced by environment variables, e.g. natural environment (ex. temperature, humid, wind, etc.) or a certain global effect to the simulation area (ex. communication noises). Instead of an environment model and I/O coupling for environment variables, just a few event functions can simulate the environmental effects. It is simpler than ECM to model environmental characteristics by using global variables. However, if the initial model does not cover the environment variables the first time, the ECM can be much better solutions for adding the variables to the whole or most of the models.

Extending the usages mentioned above, users could reduce the developing costs of new. As the requirements are changing, an unplanned model may need to be newly developed. Perhaps users have no confidence that the model needs to be developed or not. In this case, it can be efficient that they check the results in advance by describing an ECM. It may be no matter that the target model has to save some information into state variables because the proposed ECM has global variables for states. The detailed example is described in Chapter 4.2.

Applying to the development processes, there are mainly two cases using the ECM. One is rapid prototyping for the refinement of ambiguous requirements in the beginning of development. While the prototyping for the software mainly handles requirements for GUI or representation, the prototyping for simulators mainly handles the behaviors of target models. After developers have made the initial model at the early stage, they make several temporal prototypes using ECM quickly and provide them to the stakeholders for acquiring their opinion iteratively. The iterative process can reduce ambiguity of requirement, and the proposed ECM may reduce the cost of iteration.

The other is the unexpected modification of requirements during the development. Similar to the above case, developers can explore whether the modified requirement is feasible or not as preceding researches. The difference is that the ECM in this case may not be discarded and can be succeeded as continuing the development process.

#### 4. CASE STUDY

This chapter will show the applied examples of ECM during development of a war game, which had been developed for the Korean military actually (Seo *et al.* 2011). The brief scenario is illustrated in Figure 7. This

war-game simulation model is developed for analysis and acquisition of underwater warfare. There are four types of combat entities: a submarine, a surface ship, decoys, and a torpedo. The attacking platform is a submarine, and the target platform is a surface ship. The surface ship launches the decoys as counter measures according to stored strategies against the torpedo's possible paths. While the torpedo traces decoys, the surface ship can evade the opponents.

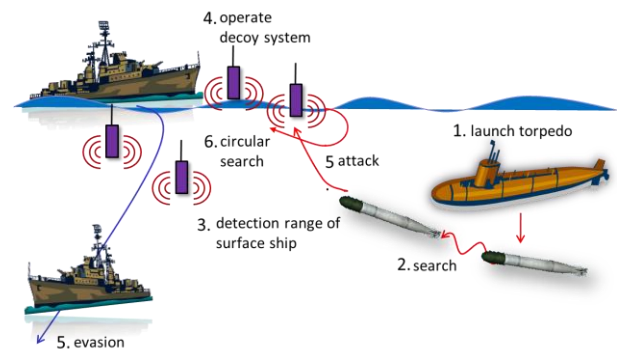


Figure 7: Brief Scenario (Kwon *et al.* 2011, Seo *et al.* 2011)

The objective of this simulation model is to evaluate the counter-measure tactics against the torpedo system. From the results, we can determine how various factors, such as tactics and the performance of underwater weapons, influence the effectiveness of the system. Experimental results can support assessment of anti-torpedo countermeasure effectiveness.

We assume two cases of requirement changing during development process of the simulation model. One is a simple changing of a model parameter. The other is that stakeholders demand development of a new model for an added entity.

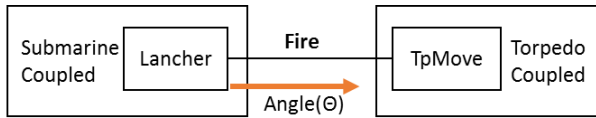
Since the original model was based on DEVS formalism, a little knowledge is needed to understand the example, e.g., the hierarchical structure of DEVS models (*Coupled models* and *Atomic models*). The simulation engine for the war-game model is E-DEVS<sub>Sim++</sub> (Kwon and Kim 2012), which was implemented for executing DEVS<sub>Sim++</sub> (Kim *et al.* 2011) models with event-based simulation. The E-DEVS<sub>Sim++</sub> has been extended for this proposed work. The extended E-DEVS<sub>Sim++</sub> provides mapping API between events and event functions. Users can add event-oriented control functions to mapped events with the original models. The functions should take an object point of an event as function parameters and return whether the event is eliminated or not. Otherwise, there is no limitation of implementation for states of ECM.

#### 4.1. Example 1: Modulating Model Parameters

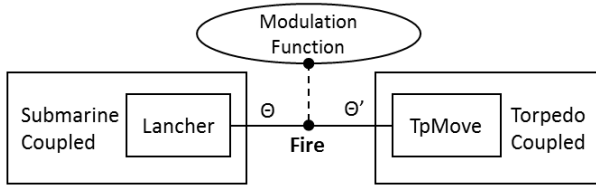
The target parameter cited in this chapter is the angle of fire in an attack command, which is generated by the C2 (Command and Control) model of submarines and transmitted to the torpedo model as a message. We assume that the angle of fire was a fixed value in the



initial model structure, as depicted in Figure 8.(a), and should become randomized with a certain probability function for representing the real error of equipment.



(a) Initially developed model structure



(b) Modifying the parameter with ECM

Figure 8: Modulating the angle of fire

Without modifying the submarine model, adding a control model can handle the angle variable such as the below function. The data event, *Fire*, has a variable for the angle of fire, named as 'Angle'. The control function reads out the value from the *Fire* event and stores the randomized value with the exponential distribution function to the inputted event. The control function is embedded in a C++ class for ECM without any states. It is only the user's job that they register the class to the simulation engine (E-DEVSIm++) without any modification of simulation model.

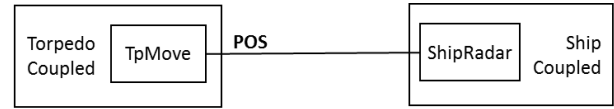
```
bool angleControlFunction(CEvent *ev){
 double v
 = genExponential(ev->GetValue("Angle"));
 ev->SetValue("Angle", v);
}
```

#### 4.2. Example 2: Substitution of a new model development

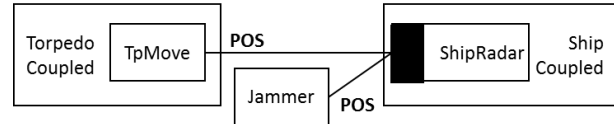
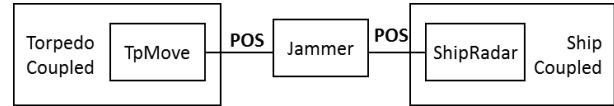
This example is about the extension of the anti-torpedo simulation model by adding a jammer model, which generates air bubbles or noise to prevent the sonar of torpedoes from detecting our forces. Since the original simulation model included only decoy systems as counter measures against a torpedo, there was actually a demand for additional counter measures of the surface-ship model (Kwon *et al.* 2011).

Many instances represent the behavior of the jammer from the initial structure of the simulation model, as depicted in Figure 9.(a) and (b). One of them is that the jammer model is placed for interfering with the position messages between the torpedo model and the radar model of the ship. The other is that radar model gathers the position of the torpedo and the jammer for deciding whether the position information is eliminated or not. The structure of the actual model was developed like the latter. A launched jammer generates its position and sends the message to the radar model.

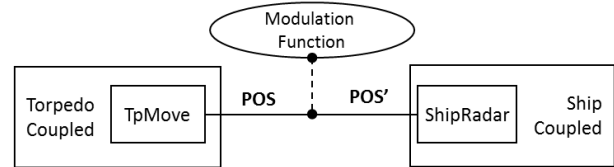
By calculating the distance between two positions, the simulation model decides the success of jamming noise.



(a) Initially developed model structure



(b) Previous solution by adding a new model



(c) Modifying with only ECM

Figure 9: Interfering the position messages

Applying the ECM to this case, the model structure becomes like Figure 9.(c). The event function is attached to the position message, and it decides the elimination of position information instead of the jammer model. Accordingly, users do not have to add a new model or modify the model inside for the extension of simulation model. From the model's pre-development stage, users and stakeholders can know whether their changed requirements are feasible or not. If the model with ECM is enough to be passed for the next step (deployment or analysis), the ECM will be kept. If not, developers may hold the ECM models and redevelop them from the original models. In this time, the model with ECM can be used for testing of the developed model.

```
Location *POS; // x, y, z;
double operatedTime = 0.0;
double lifeTime = 100.0; // Life time of jammer;
int State; // 0: OFF, 1: ON, 2: END
double jamRange; // Operating range

// attached to the fire-order messages of decoys
bool jammerON(CEvent *ev){
 State = 1;
 operatedTime = ev->GetTime();
 POS = (*Location)ev->GetValue("POS");
}

// Attached to the POS messages from the ship
bool jammerControlFunction(CEvent *ev){
```

```

if(State == 1){
 if(ev->GetTime() - operatedTime > lifeTime){
 State = 2; // The life time is over.
 return true;
 }
}
if(GetDist(ev->GetValue("POS"),POS) < jamRange)
 return false; // The event should be eliminated.
return true;
}

```

The actual implemented C++ codes for substituting the jammer model are shown above. To substitute the jammer model, two control functions are needed. The first function substitutes for launching the jammer by attached to the fire-order messages of decoys, which will be operated at the same time with the jammer. When the ship launches the jammer model, the jammer function reads out the operation time and location information by interfering with the data events. At this time, the 'State' is changed from zero to one. The second function is attached to the POS messages from the ship model to the torpedo model. The function will be operating against all the position messages from the ship model, but it does not operate until the 'State' is changed. When the 'State' is changed and is not two, the jammer control function decides whether the location information is eliminated or not by calculating the distance between the jammer and the ship. When the control functions return 'false', the simulation engine knows that the events should be eliminated and throws out the events.

## 5. CONCLUSION

This paper proposes an event-oriented specification for simulation control, *Event Control Model*, enabled by the event-based simulation. Specifically, it can help the development of war-game simulators. The two case studies in Chapter 4 show how that can be possible. The characteristics of war-game simulators have been mentioned as the motivation of this paper. Nonetheless, this paper is not limited to the military domain and it can be extended to the similar domains that have much exclusive knowledge.

Against the ambiguous requirements or the changed requirements of the similar domains, the ECM can interfere with the events that are listed in the simulation engine. The ECM can be applied to various cases, e.g. applying environment variables to a mass of events or a substitution of new models. The applications, which can have event-oriented functions without modifying the model inside, can make various iterative development processes efficient.

It is true that many further studies are needed. We have not proposed a methodology or a full development process using ECM yet. An extended paper about those advanced contents will be published in the near future. The extended methodology should include the criteria for deciding whether the ECM is applicable or not for various cases. The usages of ECM are also extended to

various cases, e.g., events generations, events deletions, logging/proving, and so on. It can be shown with more various case studies that are not limited to the war-game simulators.

## ACKNOWLEDGMENTS

This work was supported by Defense Acquisition Program Administration and Agency for Defense Development under the contract UD140022PD, Korea.

## REFERENCES

- Cota, B.A., Sargent, R.G., 1992. A Modification of the Process Interaction World View. *ACM Transactions on Modeling and Computer Simulation*, 2(2), 109-129.
- Kim, T.G., Ahn, M.S., 1996. Reusable Simulation Models in an Object-Oriented Framework. In: Zobrist, G.W., Leonard, J.V., eds. *Object-Oriented Simulation: Reusability, Adaptability and Maintainability*, USA: IEEE Press.
- Kim, T.G., et al., 2011. DEVSIm++ Toolset for Defense Modeling and Simulation and Interoperation. *The Journal of Defense Modeling and Simulation*, 8(3), 129-142.
- Kwon, S.J., et al., 2011. Effectiveness Analysis of Anti-torpedo Warfare Simulation for Evaluating Mix Strategies of Decoys and Jammers. *Proceedings of AsiaSim '2011*. Seoul (Korea).
- Kwon, S.J., Kim, T.G., 2012. Design and Implementation of Event-based DEVS Execution Environment for Faster Execution of Iterative Simulation. *Proceedings of Spring Simulation Multiconference, Symposium on Theory of Modeling and Simulation (TMS'12)*. March 26-29, Orlando (Florida, USA).
- Martin, J., 1990. *RAD, Rapid Application Development*. USA: MacMillan Publishing Company.
- Rowen, R.B., 1990. Software Project Management under Incomplete and Ambiguous Specifications. *IEEE Transactions on Engineering Management*, 37(1), 10-21.
- Seo, K.M., et al., 2011. Measurement of Effectiveness for an Anti-torpedo Combat System Using a Discrete Event Systems Specification-based Underwater Warfare Simulator. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 8 (3), 157-171.
- Zeigler, B.P., Praehofer, H., Kim, T.G., 2000. *Theory of Modeling and Simulation*. 2<sup>nd</sup> ed. USA: Academic Press.

## AUTHORS BIOGRAPHY

**Se Jung Kwon** received his B.S. in Dept. of Computer Science of KAIST in 2009 and M.S. in Department of Electrical Engineering of KAIST in 2011. He is currently a Ph.D. student in the Department of Electrical Engineering at the KAIST. His research interests include simulation algorithms for DES, DEVS execution environments, and hybrid systems M&S.

# SIMULATION AS A SERVICE IN CONSTRUCTION

Sean Newstead<sup>(a)</sup>, Simaan AbouRizk<sup>(b)</sup>, Stephen Hague<sup>(c)</sup>, Yasser Mohamed<sup>(d)</sup>, Junhao Zou<sup>(e)</sup>

<sup>(a)</sup>University of Alberta, 5-080 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA

<sup>(b)</sup>University of Alberta, 5-080 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA

<sup>(c)</sup>University of Alberta, 5-048 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA

<sup>(d)</sup>University of Alberta, 5-048 Markin CNRL Natural Resources Engineering Facility, Edmonton, Alberta, CANADA

<sup>(e)</sup>City of Edmonton, Edmonton, Alberta, CANADA

<sup>(a)</sup>sean@smaconsulting.ca, <sup>(b)</sup>abourizk@ualberta.ca, <sup>(c)</sup>steve.hague@ualberta.ca, <sup>(d)</sup>yaly@ualberta.ca, <sup>(e)</sup>junhao.zou@edmonton.ca

## ABSTRACT

This paper discusses a novel approach used to provide simulation services within the construction industry. In this application, a private company using a simulation system developed at the University of Alberta applied the concept of special purpose simulation modeling to facilitate the use of simulation tools in decision support and construction management. In particular, the special purpose simulation tool was deployed for utility tunnel construction. Background to the state of the art and the construction problem is first provided, followed by discussion of the simulation strategy used (special purpose modeling), then the service provided to clients using the simulation tools, and a more detailed explanation of the input modeling aspects of the problem is given, as they are found to be critical in providing reliable solutions.

Keywords: special purpose simulation, Symphony, modeling, tunnel construction, decision support system, construction management

## 1. STATE OF THE ART: SIMULATION APPLICATION IN TUNNELING CONSTRUCTION

Computer simulation is sometimes applied in the construction industry to support the decision-making process for different operations. Simulation enables construction practitioners to analyze complex construction processes, evaluate different scenarios, and therefore optimize time and resources for projects. Although simulation has advantages for the construction industry, the challenge is to make simulation accessible to users by presenting it in a simple and more graphical context. In 1973, Halpin introduced CYCLONE which simplified simulation modeling for construction practitioners through the use of graphical representation in modeling (Halpin 1977). CYCLONE models processes based on discrete event simulation. A number of simulation systems have been developed based on CYCLONE, including RESQUE (Chang and Carr 1987) and Stroboscope (Martinez and

Ioannou 1994). These are all general purpose simulation (GPS) tools that can model any process, but a user must have an understanding of simulation techniques to use them effectively. This makes it difficult for industry personnel to use these tools.

Over the years, advancements have been made in construction management simulation tools, with applications to tunneling. Researchers introduced special purpose simulation (SPS) to facilitate modeling of specific types of projects, as it can be developed and customized for various users, and has a more user-friendly interface. For example, Symphony (AbouRizk and Hajjar 1998) is a special purpose simulation tool developed specifically for modeling construction processes. Other advancements include 4D modelling methods and Construction Synthetic Environment (COSYE) (AbouRizk and Hague 2009). Additional innovations were presented in Einstein (2004), and Haas and Einstein (2002) (amongst other publications) where an innovative simulation system for tunnel construction simulation named DAT (Decision Aid for Tunneling) is described. Ioannou (1988) also presented a geologic prediction model for tunneling and risk reduction modelling as well as planning and simulation approaches to augment those predictions.

## 2. SPECIAL PURPOSE SIMULATION MODELING IN SIMPHONY

Simphony is a discrete event simulation system, originally developed by Hajjar and AbouRizk (1999). Simphony supports different modeling constructs to facilitate adoption in various domains; therefore, Simphony at its core was built to facilitate developing modeling templates, which can be developed and customized for various users. The special purpose simulation (SPS) approach enables a practitioner who is knowledgeable in a given domain, but not necessarily in simulation, to easily model a project within that domain using visual modeling tools that have a high degree of resemblance to the actual construction system (AbouRizk and Hajjar, 1998). Examples of special purpose templates (SPS) previously developed and

currently supported in Simphony include a tunneling template, a dewatering template, a program evaluation and review technique (PERT) template, an earthmoving template, a structural steel fabrication template, and a range estimating template. We further illustrate the tunneling template in this paper.

### 3. BACKGROUND TO THE CONSTRUCTION PROBLEM: METHOD

Underground pipe installation typically has two installation methods, trenchless and open cut. The open cut method of installation is suitable (cost effective) to installation depths typically less than 7 meters. Open cut construction requires a large amount of surface area to complete the construction, as a 2:1 slope for a typical open cut angle is typically required. The surface disruption to road traffic and interference with shallow utilities often make trenchless construction more desirable even though it may have a higher unit cost. Trenchless construction is suitable for many depth applications, but is constrained by the type of ground that is present. Trenchless excavation methods vary between hand excavation and machine excavation. For the purpose of this paper, we will focus on the machine excavation application, in particular, the tunnel boring machine (TBM) application, but hand excavation is also used to excavate small sections of the tunnel.

### 4. TBM CONSTRUCTION

The TBM tunnel construction method, as shown in Figure 1, generally starts by laying out the working shaft location, and thus preparing the working site. The working shaft has a predetermined diameter, and is usually excavated with a backhoe and a drilling rig. The backhoe will excavate the first 2 feet and the drilling rig will excavate the remaining depth. Once at the tunnel alignment depth, the working undercut is constructed, which will start with welding rib 0 to the shaft wall. Rib 0 is the outline of the working undercut and shows the alignment of the tunnel. The working undercut is hand excavated and is generally a larger diameter than the actual tunnel diameter. The working undercut is typically 30 meters in length and is lined with a steel rib and wood lagging to support the ground. The excavation of each meter takes place in 2 stages (benching) whereby the top half is excavated first and supported before the bottom half is completed.

Once the full length of the working undercut is completed, the end of the tunnel (the tunnel face) is supported with wood lagging to prevent collapsing. Next, a metal cradle constructed with heavy I-beams is welded together and concreted into the floor for the entire length of the working undercut to the tunnel face. The working undercut and cradle are important as they make up the main staging area to accept and install the TBM. The TBM is brought to site on a large flatbed truck and trailer, and is lifted, lowered down the working shaft, and placed on the cradle in the working undercut typically by a 140-ton crane. The TBM is then pulled forward on the cradle with chains to the tunnel

face. The TBM is then assembled by heavy duty mechanics over roughly 2 weeks. Electricians then come to site to connect a high-voltage power service that is supplied by to the TBM's power transformers underground, using a very large-gauge power cable (mole cable). A final mechanical check and survey for tunnel alignment is completed before excavation takes place. The face boards are removed and the TBM is launched into the ground carefully to make sure tunnel alignment is not compromised. The TBM excavates the dirt at the face and a single train, with usually 4 dirt cars, accepts the excavated material from the conveyer, travels back to the working shaft, and is taken up the working shaft and dumped by a smaller sized crane (usually 75 ton). This section of the TBM tunnel is called the 50-meter start up tunnel, as the TBM will excavate until the 50-meter mark and stop. From here, the remaining conveyer sections are installed and the power transformers are placed on top of a gantry that is dragged behind the TBM by chains. The dirt train and cars all fit underneath the gantry, which accepts the dirt from the extended conveyer unit. The working undercut is now outfitted with a wooden platform on which the train tracks are placed. A switch is installed to allow 2 trains with 5 cars to sit in the working undercut at the same time. Now, as one train exits the tunnel loaded with dirt from the TBM and passes the switch, the other train can enter the tunnel and begin being loaded by the TBM. The other train is unloaded by the crane simultaneously, so as to not hinder production of the TBM. During the time that the train is traveling back to the working undercut, the TBM is installing its concrete segmental liner around the outside of the tunnel diameter to support and finish the tunnel excavation. Once the switch is installed and the wood platform is completed in the working undercut, the main TBM tunnel excavation commences.



Figure 1: Trenchless TBM Tunnel Components and Initial stages of Tunnel Excavation

Prior to the completion of the TBM tunnel, the removal shaft and removal undercut has to be completed to accept the TBM and eventually

disassemble and remove it. These components are completed in the same manner as the working shaft and working undercut, with the exception of the removal undercut being shorter in length than the working undercut, as it is usually only 9 meters in length. Once the TBM is removed from the removal shaft by the 140-ton crane, usually, hand excavated connection tunnels need to be excavated to connect an existing tunnel structure to accept flow in to the new tunnel and then connect again at the opposite end of the tunnel to release flow. These hand tunnels range anywhere from a few meters in length to upwards of 20 meters in length. The connection tunnel is usually of a smaller diameter than the main TBM tunnel and is excavated 1 meter length at a time and lined with metal rib and lagging. Once the existing pipe structure is reached and exposed, the hand tunnel needs to be finished typically with cast-in-place formed concrete. Once the concrete is cured, the entire length of the tunnel is inspected, and any variances are patched with concrete. The working undercut now has the concrete segments banded together in a circle and hand installed for the entire 30-meter length working undercut back to the working shaft location. The bottoms of both shafts are finished with cast-in-place concrete to finish the tunnel and seal the hand connection tunnels to the TBM tunnel. Once this is completed, the existing pipe is cut out (breaking out) at the downstream end first, then at the upstream end to accept flow into the new tunnel for the first time. Manhole barrels are now placed on top of the undercut structures and are stacked up to the ground surface where a manhole cover is placed on top. Fillcrete is poured around the outside of the manhole barrels to seal the gap between the inside of the removal shaft and the manhole. The tunnel is now complete.

## **5. SIMULATION OF TBM TUNNEL CONSTRUCTION**

The tunneling simulation model follows the same process that was explained in the previous section. The model is generally driven by historical data collected for similar tunneling situations. The special purpose simulation model for this problem is illustrated in Figure 1, which shows the major components of the tunnel construction. In special purpose simulation, discrete modeling elements such as work-tasks are added inside these components to accurately reflect how the previously described construction process takes place. The process is therefore flexible to enable a user to make changes to processes taking place for specific projects. These discrete tasks are fitted with durations using standard statistical distributions. The discrete tasks also have labour, equipment, material, and other cost data added to each of them. By providing such information to the tunnel objects, a project estimate can then be produced in standard construction form. Simple processes such as building the shaft liners can generally be modeled with deterministic duration as they do not vary much, while critical path tasks such as the TBM excavation use a fitted distribution, as many factors

affect its value for a particular iteration. A statistical distribution, carefully collected from historical data, provides a reasonable approximation for such tasks.

### **5.1. Tunneling Special Purpose Template Modeling**

The tunneling template is a special purpose template developed in Symphony to simplify planning and analysis of tunnel construction projects (AbouRizk, 2013). The template is comprised of modeling elements, most of which represent the different physical components and resources that exist within a typical tunneling project, for example: a shaft element, tunnel element, crane (site) element and TBM element. The template is made up of modeling elements developed in Visual Studio.NET using Symphony services. The elements resemble the real-life items they represent, making template building easier for users not familiar with simulation. The modeling elements model the process of tunneling operation by capturing resources, scheduling events and releasing resources, collecting statistics or controlling work and non-work times. The template uses a hierarchical approach for design and implementation to match the complex nature of the process.

The templates have two parent-level modeling elements and eleven child-level elements. To model, the user can drag and drop the modeling elements, then align them in a pattern that represents the actual tunnel construction sequence (see Figure 2 for a typical model layout). The modeler can experiment with the template by inputting project information (work method - hand excavation, TBM excavation), project site conditions (depth of tunnel, tunnel diameter, tunnel length, geotechnical conditions along the shafts and tunnel, penetration rates in various ground conditions) and details on resources (number of trains, number of carts). Weather and work shifts (calendar) can also be customized. The individual template modeling elements allow for input of information before simulation and provide results after simulation. Figure 2 shows sample screen shots of the input/output interface. The template simulates the construction of the tunnel based on the information input and output results (i.e. costs, project duration, resource utilization, waiting times, daily advance rates, and volumes of earth excavated and handled). The template also models different dynamics and uncertainties experienced in a typical tunneling project such as equipment breakdown, bad weather interruptions and the details of work shifts – work times, breaks, overtime, weekends and holidays.

## **6. INPUT MODELING IS A CRITICAL PART OF SPECIAL PURPOSE SIMULATION**

Although SPS is quite easy to use and provides good efficiency in building models, one of the significant requirements is accuracy in the input models. For TBM tunneling, given the linear and repetitive nature of the work, errors in the input models could easily produce incorrect results. In providing the service to clients, we

have developed techniques that ensure good input models. These are discussed in this section.

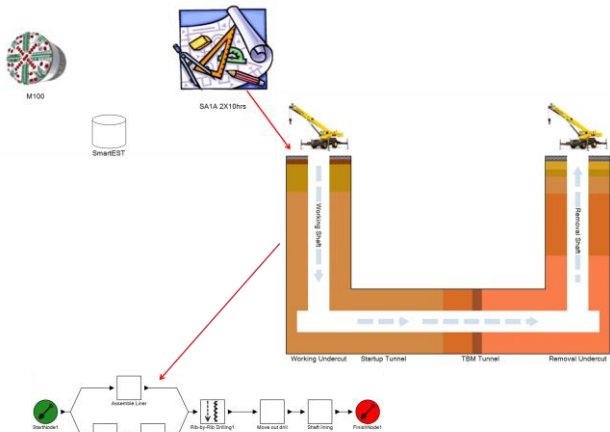


Figure 2: Special Purpose Simulator for Tunneling

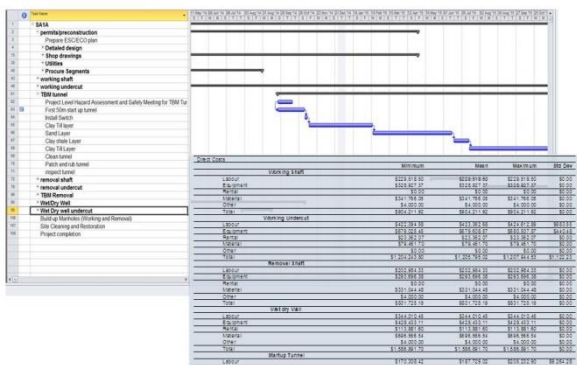


Figure 3: SPS Analysis Driven By Simulation but Product in Standard Form MS Project and SmartEst

### 6.1. Breakdowns/Interruptions

For this application, the critical path tasks were obtained through a Method Productivity Delay Model (Adrain and Boyer 1976) (MPDM) for which a project's delays/breakdowns are categorized and analyzed. The results of the MPDM analysis were fitted to distributions to show the overall delay time, ideal, and method productivities. The ideal productivity is the non-delay productivity that would have been realized if all the delays that occurred were removed. The method productivity is simply the ideal productivity with the delay categories reapplied to it. Each delay category can then be applied in the model as a breakdown element that governs the ideal productivity rate of the excavation.

We have fitted distributions to data representing breakdowns in the process referred to as "interruptions" in the model. We have split the interruptions into two categories: minor interruption and major interruption. A minor interruption is one that lasted between 0.5 and 3.5 hours (less than half a shift) prior to the process resuming its operation, while a major interruption is one that lasted between 4 and 10 hours (half to a full shift). Using statistical analysis, we determined the mean delay time for a minor and major interruption, as well as the

mean time between minor and major interruptions. These results were each fitted to a distribution for use in the simulation model. The fitting process used @RISK to select the data and then determine the best fitting distribution. This was completed by producing a cumulative ascending graph for which the distribution was visually selected using the Bayesian Information Criterion (BIC). The minor breakdowns had a beta distribution fit with shape parameters of 2.0 and 2.0, and range parameters of 1.0 and 4.0, as shown in Figure 4.

The distribution for the time between the occurrences of minor breakdowns is an exponential distribution with a mean of 26.591 hours. The major breakdowns had a beta distribution fit with shape parameters of 2.0 and 5.0, and range parameters of 5.0 and 10.0, as shown in Figure 5.

The distribution for the time between the occurrences of major breakdowns is an exponential distribution with a mean of 41.189 hours. These distributions were placed into Symphony, inside the TBM element, as this is the excavation method and always lies on the critical path.

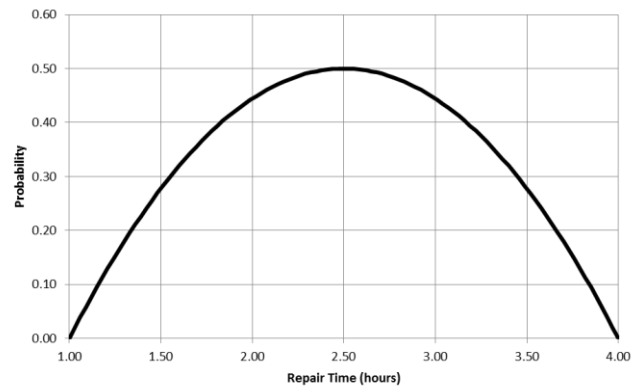


Figure 4: Minor Breakdown Distribution

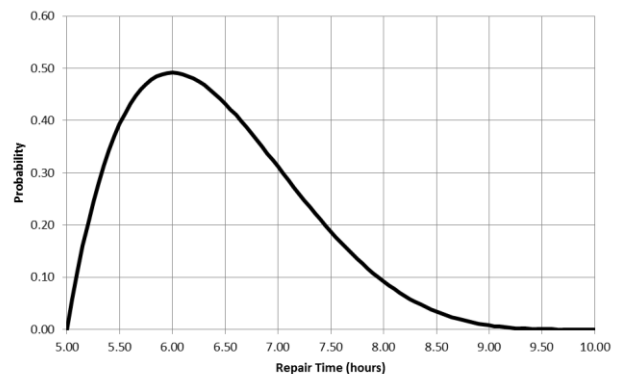


Figure 5: Major Breakdown Distribution

### 6.2. TBM Penetration Rates

Penetration rates for each ground condition have been sampled from the Method Productivity Delay Models (MPDM) studies conducted on various projects. The MPDM studies collect every delay that has occurred on a particular project and categorize it. These categories



are then compared, in hours of delay, to the overall project working hours to obtain a percent delay time. The ideal productivities for similar ground types were used as the penetration rates. The ideal productivity is the productivity that would have been achieved if no interruptions had been realized. Adding this was essential for not double counting for interruptions when simulating.

TBM penetration rates have been governed by surveying intervals to accurately reflect productivity and cost. Surveying has been broken down into 2 categories, laser calibration and moving the laser forward. The surveying duration is beta distributed with shape parameters of 9 and 2, and a range of 60 to 420 (measured in minutes). This distribution will be sampled every 60 meters of excavated common earth. When excavating in material that is predominantly sand, the distribution will be sampled every 15 meters, as there is a higher probability the installed segment liner that the laser is fixed to will settle, causing the laser itself to be misaligned. There is also a higher probability that the TBM creates a void in the ground either above or below the excavation face, which will shift the excavating alignment of the TBM. A curve in the tunnel alignment will drastically alter the surveying intervals. Curved sections will have a laser movement interval every 6 meters, which will constantly take 420 minutes or 7 hours to complete. Soil swell factors also have been accounted for in the special purpose tunneling model. This will affect the rate at which the TBM fills the dirt car trains as bank ground measurement is converted into loose ground measurement. A higher swell material will mean less ground penetration by the TBM and more frequent train travel as the dirt cars fill up faster. The amount of excavated dirt volume and swell factors are displayed for each tunnel ground type. This will also mean that the TBM will sit idle for longer as it is waiting for the train to return.

### 6.3. Unit Rates for Estimates

The SPS provides very easy-to-use features that incorporate estimation into the simulation environment through unit rates. With the use of the SmartEst database, crews are stored with their unit rates. A regular time and overtime rate is provided and will automatically be used according to the type of calendar specified in the scenario. This provides us with the added flexibility to also manually adjust the unit rates or add resources to the table with their corresponding unit rate. By looking at the client's actual charges from a past project, we can update these unit rates within SmartEst to provide accurate year-to-year estimation that is not impacted by inflation. We have revisited and reset proper unit rates to all of the indirect charges associated with the tunneling projects.

Indirect charges for the project appear in the high-level project icon. These are also added through SmartEst, but can be manually entered. By checking the percentage box, this means that the particular indirect

cost will be estimated using the percentage entered in the value column. If the percentage column is not checked, then the value entered in the value column will be a fixed dollar value added to the project.

Common areas of concern for estimating projects can be easily reconfigured in the SPS. These work packages can be given new unit rates that reflect actual performance based on past projects. Overall, the estimate will be a much more accurate representation of how the project will materialize as the simulation models the delays previously mentioned and adjusts the estimate accordingly.

| Name                            | Quantity | Rate | OT Rate |
|---------------------------------|----------|------|---------|
| Equipment operator 3            |          | \$   | \$      |
| Equipment Operator 4            |          | \$   | \$      |
| Equipment Operator 4 (Crane)    |          | \$   | \$      |
| Equipment Operator 5            |          | \$   | \$      |
| Tunnel Foreman 1                |          | \$   | \$      |
| Tunnel Labourer 1               |          | \$   | \$      |
| Tunnel Labourer 2               |          | \$   | \$      |
| Tunnel Labourer 2 (top man)     |          | \$   | \$      |
| Tunnel Differential (TD 1) - 30 |          | \$   | \$      |

Name  
The name of this worker.

Figure 6: SPS Crew Estimation Table

| Name                       | Value | Percent                             |
|----------------------------|-------|-------------------------------------|
| Surveying                  |       | <input checked="" type="checkbox"/> |
| Internal Engineering       |       | <input checked="" type="checkbox"/> |
| Risk and Const Consultants |       | <input checked="" type="checkbox"/> |
| External consultants       |       | <input checked="" type="checkbox"/> |
| In house Drafting          |       | <input checked="" type="checkbox"/> |
| In House Electrical        |       | <input checked="" type="checkbox"/> |
| Temp Water and Sanitary    |       | <input checked="" type="checkbox"/> |
| Temp Power Setup           |       | <input type="checkbox"/>            |
| Overhaul TBM               |       | <input type="checkbox"/>            |
| All Other Indirects        |       | <input checked="" type="checkbox"/> |
| TBM M100                   |       | <input type="checkbox"/>            |

Name  
The name of this cost item.

Figure 7: SPS Indirect Cost Estimation Table

## 7. SIMULATION AS A SERVICE

This paper describes an innovative approach for applying simulation as a service. In this research, simulation is designed specifically for application to a utility tunnel construction project and customized to meet client output needs, using a special purpose simulation approach. The model outputs the information required by the client in a format that is usable for them. This provides the client with decision support information to make their operations more efficient and effective. Additionally, tools such as value engineering and constructability reviews help the client to obtain more accurate estimates and schedules in the project planning stage.

In general terms, simulation tools are effective in modeling tunnel construction, especially because such processes are linear in nature and are composed of repetitive sub-processes. The challenge is to have the decision maker justify the investment in time, resources and costs associated with building a simulation model for smaller tunnel projects (those that are less than \$50 million in cost). For larger projects, the capital investment is significant and the planning time is long, thus providing ample opportunities for deploying simulation in planning the project.

Typically, construction planners rely on two elements once a preliminary design has been produced: the construction costs and the schedule associated with a particular option. Those two elements guide them in selecting the final construction alternative for the project. The simulation in itself is therefore not an end result for a construction planner. It could simply be the means to produce costs and schedule information for decision support. More specifically, schedules are expected to be in a CPM format, and costs in a work-package model, consistent with the models the company uses.

The first author adopted the simulation tools to produce the costs and schedule for clients. First, simulation modeling normally generates more accurate production information, which is the essence of cost and schedule. Second, if special purpose simulation is used, the development time and the consistency by which estimates and schedules can be produced to reflect varying alternative tunnel plans can be significantly reduced, as compared to developing estimates using standard software or schedules using CPM software. Furthermore, since the base model is the same, the estimate and schedule are based on the same foundation, whereas in practice, they are generally developed by different people on different bases.

The consulting firm where the first author works, has applied the above strategy to the modeling of tunnel construction by adopting a special purpose simulation model. The SPS model provides a quick turnaround time, and is a cost effective service to the client, and at the same time, takes advantage of what simulation has to offer in its accuracy of predictions. SPS provides the flexibility to integrate its simulation environment with a basic service that the client may require, such as estimating, scheduling, pre-project planning, or a constructability review. With the SPS, we custom link the simulation of any project with these tools while building the project schedule and estimate in Microsoft Project and SmartEst, respectively. The model is built around the work breakdown structure (WBS) of the project and provides detailed information about each work package. After the WBS is defined, the modeling elements are placed inside their work packages, as shown in Figure 8.

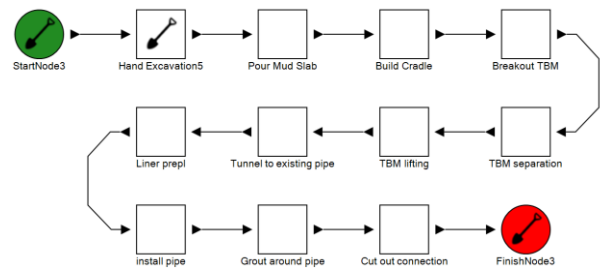


Figure 8: Modeling Elements inside the Removal Shaft Work Package

Each of these modeling elements inside the work packages has its own unique scheduled durations and crew costs. These modeling elements together create their work packages' cost and schedule duration. This is summarized in a summary report that will show the client the start and finish date of each element, the hours needed to complete the element, its resulting daily productivity, and the estimated cost. It should be noted that this summary sheet is particularly unique for this single simulation.

After the client has selected its preferred level of risk, the cost and duration of this work package is added to the overall project. The client can then take these results to a value engineering or constructability session, to provide key decision support to the project team. The project team will then be able to better create different scenarios to construct the project. The scenarios can then be added to the original model to provide further decision support to the project team. Further scenarios can be run to alter the type of shift that the project team needs to hit any certain cost or schedule constraints. The shift change scenarios will be able to show the client any cost/schedule trade-offs that may exist to further provide added value.

The model output provides a schedule (start/finish dates) and cost for each phase of the project as a distribution.

## 8. CASE STUDY APPLICATION

Note: all numbers presented in this case study were scaled and names were removed for confidentiality.

### 8.1. Project Background

The Project X TBM tunnel project is a 630 meter tunnel that is part of the Project A line. The tunnel is to be constructed using the M100 TBM. The tunnel is located along a road, where the working shaft is placed in a parking lot and the removal shaft is located near the pump station at an intersection. A hand excavated connection tunnel will be constructed to connect to the pump station.



Figure 9: Project X Location

## 8.2. Objectives and Approach

The objective of this analysis is to establish production targets, assess feasibility of the project schedule and budget, and establish a base plan for the construction phase of Project X. To achieve those objectives, construction process simulation models were developed, required information was collected and an analysis carried out.

## 8.3. Simulation Model

The TBM tunnel model is composed of 10 major work packages, each with their own properties and parameters that have been derived based on information presented in the geotechnical reports, from historic data, or from expert opinion. These include: working shaft, working undercut, startup tunnel, tunnel excavation, removal shaft, removal undercut and connection tunnel.

## 8.4. Assumptions within the Simulation Model

The simulation models assume the following:

- Double shifts for every task during construction.
- All drawings are completed 1 month in advance to constructing the element.
- The working undercut is two-way hand tunneling.
- The tail tunnel working undercut is half completed under the previous hand tunnel and not part of the TBM tunnel scope.
- Mixed faces are to be encountered between tunneling layers.
- The removal shaft and undercut start during TBM excavation.
- The M100 TBM is charged out at a rate of \$298/meter excavated.
- The unit rate for the precast segments is \$1,008.00/linear meter.

- The connection tunnel and hand installing working shaft segments happen at the same time.
- Building up the manholes, one shaft after another, occurs after opening up the tunnel, assuming all other Project X components are completed prior to finishing the TBM component.
- Most construction risks have been incorporated in the model except for catastrophic events, and contingency.

A number of simulation scenarios were run. All scenarios were run multiple times as is the standard in Monte Carlo simulation techniques.

## 8.5. Base Scenario

The base scenario is composed of a single 8-hour shift that has 2 trains operating with a switch. Under this scenario, the entire project is expected to finish within 2 years, with a total mean cost of approximately \$7,914,347.

The individual work package schedules are as follows. The working shaft would be complete in 20 working days. The working undercut takes 91 working days to complete. The simulation estimated an average of 0.340 meters per shift per tunnel, which includes delays, break downs, etc. This is justified as the undercut sections are usually split in to digging the top section and installing the ribs and lagging, then the bottom section with lagging and spreader.

The TBM in clay takes 33 working days to complete. The average productivity is 0.92973 meters per shift. Layer 1 clay takes 41 working days to complete. The average productivity is 2.4786 meters per shift. Layer 2 clay till takes 40 days to complete. The average productivity is 2.71278 meters per shift. Layer 3 clay takes 14 working days to complete. The average productivity is 2.5000 meters per shift. Layer 4 sandstone takes 89 working days to complete. The average productivity is 2.6337 meters per shift.

The removal shaft takes 20 days to complete excavation. The removal undercut takes 72 working days to complete. The undercut excavation productivity is 0.4259 meters per shift. TBM removal will take 13 days to complete. The connection tunnel takes 47 days to complete. The productivity for the excavation is 0.7994 meters per shift. Building up the man holes for both the working and removal shafts takes 14 working days.

Throughout the tunnel excavation, the TBM was idle for 4.5% of the total work hours. A summary of the TBM tunnel work package productivity by layer is given in Table 1.

Upon presenting the base scenario to the client, the director of construction asked to have the models run so that the construction of the project meets the promised requirements. The following scenarios were produced as summarized in the section below. Two scenarios were run, the first with two 8-hour shifts and the second

with two 10-hour shifts. Both had two trains operating with a switch in the working shaft area.

Table 1: Single 8-Hour Shift TBM Productivities by Soil Layer

| Work Package | Section         | Length (m) | m per shift | Advance per day | Days Required |
|--------------|-----------------|------------|-------------|-----------------|---------------|
| Tunnel       | Start-up (clay) | 36.26      | 0.9297      | 0.9297          | 33            |
|              | Clay            | 119        | 2.4792      | 2.4791          | 41            |
|              | Clay till       | 128        | 2.7128      | 2.7127          | 40            |
|              | Clay shale      | 43         | 2.5000      | 2.4999          | 14            |
|              | Sand stone      | 292        | 2.6337      | 2.6336          | 94            |
|              |                 |            |             |                 | Days 223      |

## 9. CONCLUSION

The SPS makes it possible for simulation to provide a value-adding estimate and schedule solution to the construction industry as a whole. This provides the client with decision support information to make basic client services more efficient and effective. Services such as value engineering and constructability reviews can have multiple project scenarios created for them to allow the client to get more accurate estimates and schedules in the project planning stage. The key to SPS is the data input to the model. Proper background work is needed from past benchmarked projects to act as verification and validation of the model to its intended application.

## REFERENCES

- AbouRizk, S.M. and Hague, S. 2009. An Overview of the COSYE Environment for Construction Simulation. *Proceeding of the 2009 Winter Simulation Conference*. Rossetti, M.D., Hill, R.R., Johansson, B., Dunkin, A., and Ingalls, R.G., eds., IEEE, 2009, 2624-2634.
- AbouRizk, S.M., and D. Hajjar. 1998. A framework for applying simulation in the construction industry. *Canadian Journal of Civil Engineering*, CSCE, 25(3): 604-617.
- AbouRizk, S.M. 2013. Innovations in tunneling construction management applications of simulation. *Proceedings of the EURO:TUN 2013*.
- Adrian, J. and L.T. Boyer. 1976. Modeling method productivity. *ASCE Journal of the Construction Division* 102(1):157-168.
- Chang, D. Y., and Carr, R. I. (1987). "RESQUE: a resource oriented simulation system for multiple resource constrained processes." PMI Seminar/Symposium, Milwaukee, WI, USA, 4-19.
- Einstein, H. H. 2004. The Decision Aids for Tunnelling (DAT) – An Update. *Transportation Research Record* (1892), 199-207.
- Haas, C. and Einstein, H.H. 2002. Updating the Decision Aids for Tunnelling. *Journal of Construction Engineering and Management*, ASCE, 128(1), 40–48.
- Hajjar D. and AbouRizk S. M. 1999. Symphony: An environment for building special purpose construction simulation tools. *Proceedings of Winter Simulation Conference 2*. IEEE: New Brunswick, NJ, 998-1006.
- Halpin, D.W., 1977. CYCLONE: Method for Modeling of Job Site Processes. *Journal of the Construction Division*, ASCE, 103(3) , 489-499.
- Ioannou, P.G. 1988. Geological Exploration and Risk Reduction in Tunnelling. *Journal of Construction Engineering and Management*, ASCE, 114(4), 532-547.
- Martinez, J., and Ioannou, P. G. 1994. General Purpose Simulation with Stroboscope. *Winter Simulation Conference*, ASCE., San Diego, CA, USA, 1159-1166.



# DIGITAL HUMAN MODELS (DHM) TO SIMULATE MEAT PROCESSING TO PREVENT WORK-RELATED MUSCULOSKELETAL DISORDERS (WR-MSDs)

Sergio Amedeo Gallo<sup>(a)</sup>, Riccardo Melloni<sup>(b)</sup>, Simone Mosconi<sup>(c)</sup>, Francesco Longo<sup>(d)</sup>

<sup>(a) (b) (c)</sup> Department of Engineering “Enzo Ferrari” DIEF, University of Modena and Reggio Emilia, Italy

<sup>(d)</sup> Modeling & Simulation Center - Laboratory of Enterprise Solutions DIMEG, University of Calabria, Rende (CS), Italy

<sup>(a)</sup>[sgallo@unimore.it](mailto:sgallo@unimore.it), <sup>(b)</sup>[riccardo.melloni@unimore.it](mailto:riccardo.melloni@unimore.it), <sup>(c)</sup>[simone.mosconi@libero.it](mailto:simone.mosconi@libero.it),  
<sup>(d)</sup>[francesco.longo@unical.it](mailto:francesco.longo@unical.it)

## ABSTRACT

Work-Related Musculoskeletal Disorders (WRMSDs) are the leading cause of Occupational Disease in working populations of the advanced industrialized countries.

The professional pathology of the upper limbs, from the 80's until today, is continuously increased to represent more than 50% of all occupational diseases.

In literature many tools are available to perform the ergonomic assessment of the workplace.

These tools and methods are classified based on the type of analysis that are able to perform and on the results that can provide.

Beside the traditional methods of assessment, based on tabular theoretical methods, that commonly adopt paper check lists that refer to the main methods of analysis (NIOSH, OCRA, Snook & Ciriello, etc..), are available computer tools performing the same analysis, or, much more appropriate when it needs to make a more accurate biomechanical analysis.

These methods provide numerical value indicating the level of risk to which workers are exposed.

Keywords: WRMSD, Digital Human Model, Health and Safety Simulation Models, Ergonomics.

## 1. INTRODUCTION

Traditional methods of assessment, based on tabular and theoretical methods, that commonly adopt paper check lists that refer to the main methods of analysis (NIOSH, OCRA, Snook & Ciriello, etc..), are typically used to assess the risk of the workplace, unfortunately not in a preventive way, but once the operator is already playing activities.

To perform this evaluation, you must go beside the workstation, collect all information necessary for the analysis (times, weights, duration, frequencies, tenure, etc.), sometimes shooting movies or photos.

The analysis of collected data allows to determine the index of exposure of the worker/workplace, but do not support the opportunity to evaluate what are parameters values that cause the score of the exposure level to be off the recommended ranges.

These methods do not indicate how the level of exposure can be correct by the introduction of changes in the organization of the workstation or in the working style of the operator.

To achieve this goal, you need to re define values of parameters in the new configuration of the activity for a new analysis, according to the subjective competencies of the analyst.

This takes a long time and does not allow the relationships between different parameters to emerge.

In addition to the traditional methods, there are some models of biomechanical systems that allow to perform simulation in a quite simple, but quasi-static, way, using humanoids models.

They are very complex tools, able to perform dynamic simulations of work tasks accompanied by ergonomic analysis, under various methods.

These tools can be extraordinarily useful if adopted since the concept or design phase.

The main limitation of the traditional methods of evaluation, based on check lists, is that they are highly subjective whether if you follow thoroughly the procedure: in most cases the evaluator, that is often an external consultant, tends to underestimate parameters values to shorten scores within acceptable ranges.

For this reason, it becomes necessary to introduce alternative means to assess the risk to overload of the upper limbs.

The sneaky issue is that the workload evaluation is done at a specific time, but physical problems emerges further and along the time.

In last years, thanks to the technological progress, it has become possible to study, since the design phase, whole ergonomic aspects of a workstation, using Digital Human Models (DHM).

Example of these tools are the Human CAD, JACK, RAMSIS or DELMIA.

The use of software, since the planning and design phase, helps to understand dynamics that may lead to possible problems, under the medical point of view, and helps to identify, in a timely, solutions to reduce risk, but, and this is our proposal, they, at least, could help to compare alternative solutions, under the efficiency, feasibility and physical point of view.

These software applications are, however, very expensive, both from a point of view of the purchase cost, both as expenditure of very well trained human resources, as researchers or practitioners or analysts.

This makes them unusable by the majority of organizations.

A big issue in using any of available tools, both computerized or not, is the difficulty of define, based on

scientific principles, relationships among the description of the work contents, strength, frequencies, tenure, postures, etc., and the exposure level for a generic, but also defined individual.

In fact, the pre pathological clinic parameters, and their threshold levels, are not definitively known.

So, under this consideration, it is very important to collect, to classify any of related previous studies to base any further study on a larger and validated data base: there are an huge number of papers that deal with this issue, as it is possible observe in the bibliography paragraph, that shows results and considerations on methodologies, instrumentation, physical and biometrical parameters, and on any pre pathological clinical values in the WR-MSDs to support digital simulation and analysis.

## 2. BIBLIOGRAPHY

With the actual available search engines for papers and scientific works and studies, it is much more easier than just few years ago, to find a very scaring number of papers that discuss about WRMSDs.

In many papers the use of DHM and of Simulation Tools is described, as in Dan Lamkull, Lars Hanson and Roland Ortengren (2009), where a comparative study has done between VR models and real word observations. The case study has been conducted on a Volvo System for manual assembling of automobiles.

On another side, Honglun, Shouqian and Yunhe (2007), consider the use of ergonomics simulation systems to perform studies of ergonomics analysis. An ergonomic virtual human model is built, to keep together, in a unified framework, elements of biomechanics, of physiology, of anthropometrical model, of posture and motion model to apply to virtual prototyping and virtual product development.

De Magistris, Micaelli, Evrard, Andriot, Savin, Gaudez and Marsot (2013), define principles to build autonomous dynamic DHMs, to be used to compare both the real task and simulated ones based on operator/manikin's joint angles and applied force in accordance with machinery safety standards. The aim was to examine the error of ergonomics simulations of manual assembly tasks, to correctly predict the real outcomes in the plants, and if outcomes originating from ergonomics simulations could be adopted to increase performance of the real system.

Many works studies special devises, methodologies, data collections tools, to acquire field data to define forces, posture angles or configuration assets, sampling strategies, as in McGorry, Chang and Dempsey (2004), where a special wearable wrist devise is used to collect wrist postures. Trigonometric solution permits determination of wrist angular displacement. Moreover, a people sample was defined and the accuracy of the measure has been showed, as a regression was used to determine the slope and intercept of the relationship between the goniometer and the electromagnetic tracking system for the nine subjects who repeated the evaluation, and a paired t-tests was

used to define signal values and real angles values. In fact, another huge issue is the statistical approach to process data to outlines scientific information.

Riley, Ballard, Cochran and Chang (1983) faced with the influence of the temperature versus the assembly time performed in an assembly process.

In Fogleman, Freivalds and Goldberg (1993) an ergonomic evaluation of knives shapes used in meat cutting tasks, has done. Knives of different shapes, and gloves equipped with sensors were used.

Lewis and Narayan (1993) have studied and designed handles for two commonly hand tools as screwdrivers and chisels. A classification of a people samples, divided in percentiles, and also by sex, has been done. Electromyography analysis were adopted to have an objective measure for strength and forces.

Cimino, Longo and Mirabelli (2009), focus on a methodology for the ergonomic effective design of manufacturing system workstations based on multi-measure approach. An approach based on multiple design parameters, DOE and multiple performance measures is defined to achieve an improved accuracy.

In Boenzi, Digiesi, Mossa, Mummolo and Romano (2013), is descibed an OCRA (ISO 11228-3:2007) approach to evaluate a correct break definition with an additional scheduling for job rotation schedules. Models are applied to automotive industry assembly line, and consist in integer programming models with an objective function.

## 3. DESCRIPTION OF THE ACTIVITIES

### 3.1. Some Preliminary Consideration

In this paper we present the initial outlines and considerations in using DHM and the related software suite, to model and to analyze meat processing activities, that are activities where workers are dramatically exposed to the risks of musculoskeletal disorders (WRMSDs).

We are carrying on this work in partnership with some firms in the food industry field, where we have observed, filmed and analyzed the whole process, and some phase, particularly.

The software suite is Delmia Human, V5 of the Dassault Systèmes, based on the CATIA 3D ambient.

The basic idea, for the truth, not definitively original, is to develop virtual models, validated and verified based on the comparison with classic and traditional methods of analysis, as OCRA, NIOSH; RULA, already done by some consultant and analysts in last years, to be more effective, timely, efficient and cheaper, since the phase of work activity definition.

Moreover, we are acquiring and collecting physiological and anthropometric data, of course, just for the local district of meat processing, to try to define function relationships among them and probabilistic effect and damage curves for workers.

Human DELMIA allows to reconstruct a digital human models (DHM) that can perform the same operations.



At this step, we have focused on the first workstations of the meat sectioning, where pig thighs are trimmed, deboned and prepared to become hams, a very relevant phase for the profitability, for the MSD relevance.

Meat activities are extremely important in the district of Modena, in an amount that are significant at national level, also.

DELMIA software has allowed us to model completely mannequins that replicate all the features of a human subject, up to the joints of the fingers.

Through the software, it has been possible to reproduce actions performed by employees, and then simulate in 3D environment.

We were interested on biomechanical analysis for the efforts on several articular joints, especially.

In this work, we started to test software tool attitudes and possibility, as the evaluation of simulation of different methods both of pig thighs trimming and both ham boning.

We started observing for many times, with many observers, the activity execution, for many phases, for any operators, with a special care to define a methodology to mark up postural frames and related joint values.

The aim is to use all collected data to rebuild in a virtual simulation environment the process and the activities, and, in the future, to achieve a better accuracy when we will trim and arrange models.

We recorded many videos of the same activities, performed by many distinct operators.

We have started to classify all observed operators acting these kind of job, and to define many anthropometric parameters, as well as performance parameters (frequencies, durations, technique, etc.) in the company, to define the different way to do the tasks, to statistically describe the process, to define all relevant parameters and aspects related to the physical exposure, but, so far, we couldn't use this data, because they are not yet complete, and not yet completely elaborate under a statistical point of view.

In fact, till now we have been involved and strongly engaged to explore software behavior and potential to face with our planned targets.

To verify and to define ranges for the response accuracy of the software analysis when different modeling of activities were done, with different mapping for postures, more analysts/researchers started to build models in a blind way, without communicating, to measure and to evaluate errors ranges on the analysis sensitivity.

With respect to the part of the work performed on field, we want underline that we, always, have strived to enforce and to promote collaboration with workers, to achieve a better real observations, and, on another side, we are collaborating with firm Occupational Physicians, and with Occupational Medicine Researchers, to identify the most critical situations from the ergonomic point of view, and to identify most relevant activities to focus the simulation on.

We have started remodeling the "as is" situations, relatively to the layout, locations, equipment, etc. to compare the obtained results among distinct configurations. At this phase, we have not had a big attention to verify models adherence to the real process, as the respect of postures, as well as, we have not yet validate models comparing the simulation assets to the real ones, or outlines and scores and to those reported in the risk assessments documents, but we "just" have tested and explored software potentiality.

In fact, our final target is to overpass all the classical approaches that are used just to evaluate the "as is" observed situation, based on table classification for tasks and posture, that are cited as "state of art methods" for any of the law references in many developed countries, as in Italy is the D.Lgs. 81/2008.

These methods, as OCRA, RULA, Snook & Ciriello, etc. are commonly too much subjective, and not useful and proactive in the project phase of workstations, and of tasks definitions, and that, commonly, get the same overall findings and results.

These methods leaves, as improvement strategy, when scores are too much high to preserve workers health, just the opportunity to introduce work breaks, or to reduce frequencies, and to try to practice jobs rotation.

We also are planning to acquire cutting real efforts applied by operators with a "tricked" knife with load cells inserted between handle and knife blade, or by using dynamometric tables where ham could be laid. The second way, likely, is less accurate and complete to read any of the applied forces.

### 3.2. Tasks Modelling and Software Behavior

In the next figures we show some representations of the software ambient, with snapshots of the layout we considered, and of the activity we considered too.

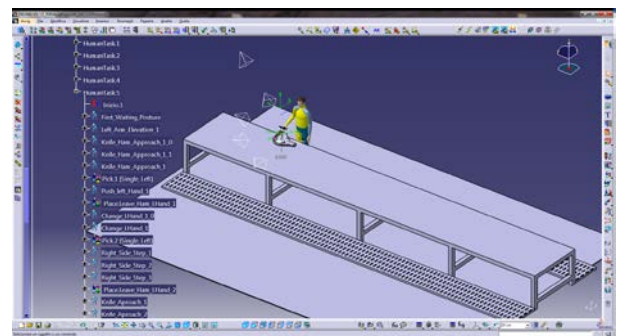


Figure 1: A snapshot of activity layout, of operator/manikin with knife, and of ham

We can observe that there is a table that in the real system is a conveyor, where, on both sides, coming from cold refrigerators storage, thighs, divided in left and right, move ahead, while operators, in a number of eight/ten, execute distinct cutting, and finishing operations.

In the actual situation we have modeled just few operations of the whole number that is required to act this phase.

This is because, the suite we used, that is a best in class suite, both as 3D modeler, both as DHM ambient, to represent and to simulate tasks, layout inter relationship among, work ambient, human issues, activities representation, is very susceptible, touchy, compared to others, that are often simpler but, also, less powerful.

In fact, because it belongs to PLM software, the entire suite can cover quite all of concept, design, manufacturing optimization, layout and work place definition, areas that a product or a product family can require. Moreover, in a very well integrated way, it can supply general CNC code to be used by real manufacturing systems.

Product Lifecycle Management (PLM) is an integrated approach to strategically manage the design, manufacturing, and, also maintenance and end life information for products and services. PLM is supported on computer technologies, but found itself on an integrated approach, and philosophy based on collaborative processes.

PLM access to shared and common information source. It enables the enterprise to extend innovation of product during the entire life cycle, supported on informatics archives and applications reusable several times.

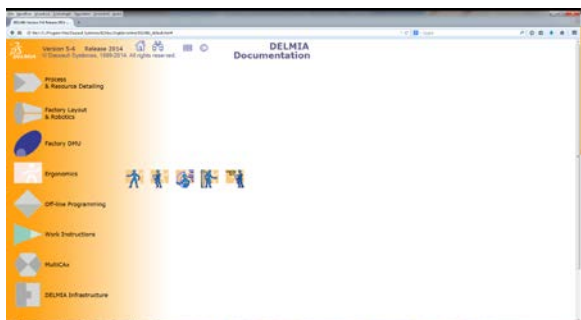


Figure 2: The online guide showing all areas the software can face with: MultiCAX, Factory Layout & Robotics, Digital Mock-Up (DMU), Ergonomics

In the Ergonomics environment, when you model an activity and any of tasks it is composed of, you can rely on many specific “workbenches” specialized to do specific things, that you HAVE to activate appropriately.

When, and it happens often, you are wrong to do this, you can get crazy to understand why.

Three persons have started to model all distinct activities of cutting required by thighs, and, up to now, we completed to define the whole approach, and quite all of the four operations needed have been modeled.

Also some cyclical operation, as the sharpening of the knife has been modeled, at least, for one of the distinct way it can be done.

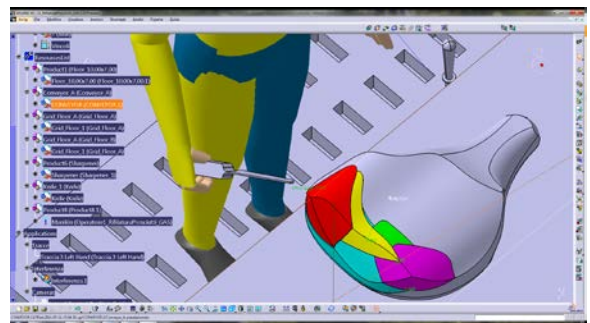


Figure 3: The operator with Analysis colors activated, the modeled 3D Ham, the sharpener and the knife.

In fact, as you can observe in fig. 3, the ham we modeled in the Assembly Design, part of the Mechanical Design workbench, is an assembly of many parts, colored in the figure, that after any cutting and finishing activity, become separated and move on different paths.

The parts that get separated from the ham are six, but someone is processed during the same operation.

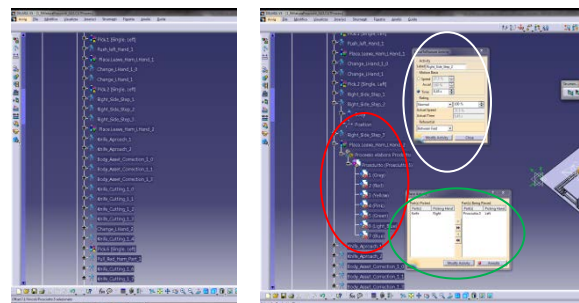


Figure 4: The PPR tree with a list of tasks defined to cut the red part of the ham, and, on right, PPR tree expanded and with configuration windows.

After we modeled first activity, composed of many tasks, we started to test how to perform the promised set of analysis: activating the workbench “Human Activity Analysis” from the “Human Task Simulation” already opened, where, usually, you are modeling activities, the Ergonomic Tools toolbar become available.

From there it is possible select: RULA, Lift-Lower, Push-pull, Biomechanics Single Action, etc. Analysis icons.

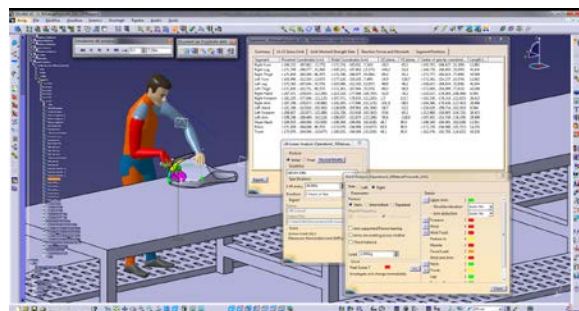


Figure 5: Snapshot of simulation of Red Part Cutting, with some Analysis windows opened.

It is possible produce the analysis you need, starting the activity simulation, with the window of any specific analysis activated: in this way, it's possible observe the updating of simulation and the refreshing of values on windows for any specific analysis.

Is possible to export all data that the software produce during the simulation, with a defined time step.

Moreover, it is possible, at visual level, the activation of coloring on the simulated manikin, when predefined threshold are overpassed. The scale and the type of color can be user defined.

Another thing we were strongly interested to verify, was the opportunity to apply loads or charges to the manikin, over the appropriate weight of parts that are lifted. This interest is because we are, as we told before, planning to acquire all real strength for operators, to be used to feed the model.

We could verify that is possible apply loads to the manikin, define specific values for right and left, specify direction along x, y, z axis.

We started to test the opportunity to fill the field of the load value with formulas, or with table data.

Up to now, on our present experience, it seems possible.

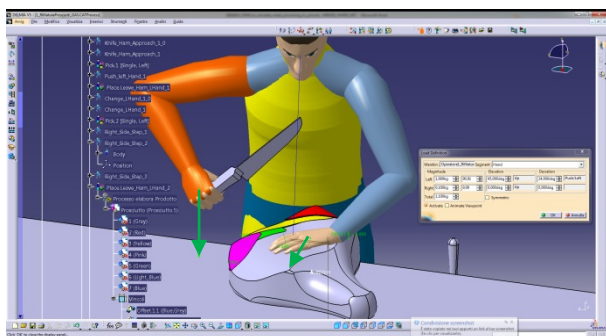


Figure 6: Green arrows represent specific load vectors applied to manikin hands.

Anyway, it is possible change and correct mean values for loads in the analysis configuration, dividing it in more sub parts.

We have verified that in the analysis related to the simulation of the same activity, in any of the suitable analysis methods, the outlines vary depending on the loads values.

We have also compared two different ways to trim hams to define efforts variability and, based on results by simulation, to identify the best solutions, but, no one of the modeled activities has been verified and validated with the real one.

Another consideration to do: any modeled activity goes simulated in a deterministic way, but, when we observe the process, it can vary depending on the operators, depending on the ham characteristics, for the same operator on the specific instance, and so on.

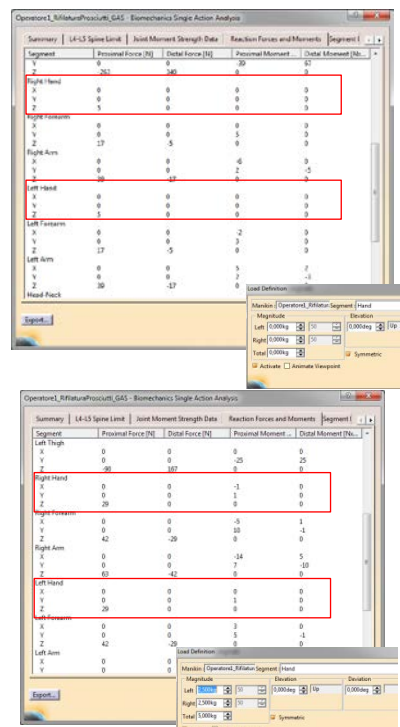


Figure 7: Effects on simulation biomechanical analysis values for two different loads levels.

Then, is very difficult, at the moment, define the "standard" process, and the standard values of the configuration parameters.

The suite we are using do not allow to manage directly this aspect, and a preprocessing phase to define distribution curves for data we are acquiring, needs.

In fact, we are increasing observations to be more accurate in defining the working postures, that we are storing on catalogs, and we are planning to use capture movements tools, also, such as motion capture cams, and systems, available in our Faculty.

Moreover, we are working, in cooperation with Occupational Medicine Researchers, on the identification of the biological and biometric parameters, that can possibly be acquired non-invasively on a sample of workers, in order to identify preclinical data useful to calibrate the digital model.

### 3.3. Conclusions

In this work we tested the attitudes of a PLM suite to support WRMSDs studies, with a special interest in evaluate the preventive opportunity since the definition phase of the integrated system of layout, work place, activity definition.

Up to now, we have experimented the software behavior, its potentialities and possibilities, and we already modeled many of activities observed in the real contest of a meat processing plant.

In this while, anyway, we have observed, analyzed, filmed many instances of process execution; we have started, in cooperation with Occupational Medicine Researchers, to define and to identify all biological and biometric parameters that can be related to produce

clinical and pathological effects on workers, also if under a probabilistic point of view.

Since this is a very ambitious aim, we are interested much more in a definition of comparative opportunity, much more than in to establish absolute effect evaluation, and we are oriented to focus on parameters of both operators, and both of activity configuration, that can possibly be acquired in a non-invasive way, on a defined and available sample of workers, in order to identify preclinical data useful to calibrate the digital model.

We are also developing tools to acquire objective field data, but also subjective ones from workers, from their subjective evaluation of fatigue during the course of activities. But, anyway, we need to proceed on methodological and scientific basis.

Another aim is to feed the model configuration with field data, in a progressively increased automatic way, with the developing of suite software interfaces.

This point is very relevant, especially in the modeling part when the analyst have to define human postures during the work, as when forces have to be defined for the model.

Not often data can be collected with adequate field instrumentations, both for any strength that could be executed by workers, both for biometric values on groups of workers divided in cohorts, that share same particular contest and conditions during a particular time span, and executing the same tasks, with control groups.

Again, in too rare instances, work related factors have been stratified and their effects have been clearly distinguished by infinite others, as is possible with a long term observation, and with the application of adequate statistical test, as the C test, that can follows time series of data.

The effects of the presence or absence of one or more factors should be observed trying to filter effects of other environmental or contingent factors.

For the future we will try to characterize workers for sector with the main anthropometric characteristics relevant to the analysis.

The results shown are the starting point for the work we are doing, at the time, the results are affected in an important way from the method of reproducing working postures.

It is obvious that these are only the first parts of a program that will require much more time to give meaningful results.

This is a first step, and in a short future we will compare our models to the real scenario, and outlines produced by the software to results of ergonomic evaluations of workstations, in the activities of boning hams, carried out with the traditional methods of analysis (NIOSH, OCRA, etc.), in order to identify in a timely with the optimal software workstations to reduce MSDs.

The results are encouraging, even companies involved were very satisfied and encourage us to be hopeful for the future.

## REFERENCES

- Boenzi, F., Digiesi, S., Mossa, G., Mummolo, G., Romano, V.A., 2013. Optimal break and job rotation schedules of high repetitive - Low load manual tasks in assembly lines: An OCRA - Based approach, *7th IFAC Conference on Manufacturing Modelling, Management, and Control, MIM 2013*; Saint Petersburg; Russian Federation; 19 June 2013, Code 99324 1896-1901
- Cimino A., Longo F., Mirabelli G., 2009. A multimeasure-based methodology for the ergonomic effective design of manufacturing system workstations, *International Journal of Industrial Ergonomics*, 39 (2009) 447-455.
- De Magistris G., Micaelli A., Evrard P., Andriot C., Savin J., Gaudez C., Marsot J., 2013. Dynamic control of DHM for ergonomic assessments, *International Journal of Industrial Ergonomics*, 43 (2013) 170-180.
- Fogleman M. T., Freivalds A, Goldberg J. H., 1993. An ergonomic evaluation of knives for two poultry cutting tasks, *International Journal of Industrial Ergonomics*, II (1993) 257-265.
- Honglun H., Shouqian S., Yunhe P., 2007. Research on virtual human in ergonomic simulation, *Computers & Industrial Engineering*, 53 (2007) 350-356.
- Lamkull D., Hanson L., Ortengren R., 2009. A comparative study of digital human modelling simulation results and their outcomes in reality: A case study within manual assembly of automobiles, *International Journal of Industrial Ergonomics*, 39 (2009) 428-441.
- Lewis W. G, Narayan C. V., 1993. Design and sizing of ergonomic handles for hand tools, *Applied Ergonomics*, 1993, 24 (5), 351-356.
- McGorry R. W., Chang C. C., Dempsey P. G., 2004. A technique for estimation of wrist angular displacement in radial/ulnar deviation and flexion/extension, *International Journal of Industrial Ergonomics*, 33 (2004) 21-29.
- Riley M. W., Ballard J. L., Cochran D. J., Chang C. C., 1983. Assembly Task Performance and Ambient Temperature: A Reliability Model. *Reliability Engineering* 4 (1983) 145-158.



# HIERARCHICAL APPROACH TO DEVELOPING A LOGISTIC DISCRETE EVENT SIMULATION MODEL USING AUTOMATED GUIDED VEHICLES

Pavel Raska<sup>(a)</sup>, Zdenek Ulrych<sup>(b)</sup>

<sup>(a)</sup> Department of Industrial Engineering - Faculty of Mechanical Engineering, University of West Bohemia, Univerzitni 22, 306 14 Pilsen

<sup>(b)</sup> Department of Industrial Engineering - Faculty of Mechanical Engineering, University of West Bohemia, Univerzitni 22, 306 14 Pilsen

<sup>(a)</sup>[praska@kp.v.zcu.cz](mailto:praska@kp.v.zcu.cz), <sup>(b)</sup>[ulrychz@kp.v.zcu.cz](mailto:ulrychz@kp.v.zcu.cz)

## ABSTRACT

The paper deals with the hierarchical approach used in developing a logistic discrete event simulation model using AGV (Automated Guided Vehicles) in a real industrial company. The hierarchical approach (using the principles of object oriented programming and inheritance) was used to develop the simulation model. We used template class programming and we also used the principle of inheritance in the simulation model. We used the structured Excel file to set up the input simulation model input parameters and objects in the simulation model. We used a genetic algorithm to find suitable settings for the simulation model. The simulation model was built in Plant Simulation v. 11.

Keywords: Discrete Events Simulation, Automated Guided Vehicles, Hierarchical Approach to Developing a Logistic Discrete Event Simulation Model, Simulation Optimization

## 1. INTRODUCTION

Many industrial companies are addressing the issue of internal logistic optimization. One of the possible modern approaches to this topic is the automatic logistic system using AGVS (Automated Guided Vehicles Systems) (Berman, Schechtmana, and Edana 2009). This system ensures the supply of material between production lines and warehouses. This system can fully replace the human factor associated with transportation and thus eliminate some of the risks. However, changing the mode of transport also brings unpredictable behaviour of this transport. This behaviour can be detected by discrete event simulation.

This paper is focused on description of the program of the implementation of the solved simulation project. This project deals with the newly proposed transport using automated guided vehicles (AGV) (Kesen and Baykoç 2007) in a company producing products in the field of ventilation and climate control. The discrete event simulation model was created in Plant Simulation v. 11 (Siemens).

## 2. SIMULATION

### CHARACTERISTICS

The goal of the project was to verify the proposed method of supplying the production lines using AGV with different numbers of trailers and determining the minimum number of required AGV capable of ensuring the supplying of the production lines. A secondary goal of the project was to verify the behaviour of the proposed transport system. The project was divided into two stages which are mutually linked. Stages included the following main points:

- The first stage was focused on the verification of the transport by AGV to three production lines. This stage was focused on:
  - Mutual familiarization between the sponsor and the project implementer
  - Clarification of the simulation project options to the submitter
  - Defining the suitable input and output parameters and the model statistics
  - Testing AGV transportation for the selected production lines
  - Verification of the proposed solutions to the simulation model
  - Verification of the proposed transportation in the model for the subsequent stage of the project
  - Evaluation of the results obtained from the simulation experiments, report, presentation
- The second stage verified transportation using AGV supplying remaining production lines in the company. This stage was based on the first stage and it is focused primarily on:
  - Defining the remaining suitable input parameters and complementing the output statistics of the simulation model
  - Checking the behaviour of the planned whole internal transport using AGV supplying the rest of the production lines in the company

## PROJECT

- Finding the suitable number of AGV according to specified customer requirements
- The evaluation of the results obtained from the experiments, the final report and the presentation of results.

### 2.1. Description of Simulated System

Large parts are supplied by the trailers and it is not possible to load a big number of these parts to satisfy the needs of the production line for a longer time. Hence more trailers must be used for the transport at once. It is also not possible to control the supply in a way that if the supply falls below a certain level there would be generated a requirement for transport from the warehouse (except for a limited number of some parts). This is caused by the transport time which is longer than the time of consumption of the parts transported to the production line.

The whole system of supplying the production lines is based on a simple principle: the tractor with the trailers continually transports the parts and after the unloading of transported parts goes to the warehouse or to pre-production for new parts which transports them immediately to the production lines. Limited capacity of the buffer (parts storage) at the production line represents a regulation in this case. Each tractor has a defined path using different loading and unloading stations which must be passed. The various types of parts are loaded and unloaded at different stations in the company. The parts can be loaded on the trailer at the loading stations in the warehouse or at the various production departments in the company. Each production line has defined several unloading stations for various parts. Schematic layout of the loading and unloading stations is shown in figure 1

Based on the outputs from the simulation model:

- AGV paths have been changed
- Logical errors in the proposed system were deleted
- The minimum number of required AGV was identified
- There were also identified serious potential risks which could paralyze the AGV transport and thus the production itself

### 2.2. Basic simulation model behaviour

The simulation model was designed in a way that can be managed by the data entered in MS Excel. These data mainly concern: the time characteristics of loading and unloading trailers attached to the AGV (tractor); cycle times of production lines; the number of trailers trailed by the tractor including specification of various types of parts and determining whether the transported part is critical; the maximum size of inventory of parts at the production line; the number of trailers and the transport paths for AGV.

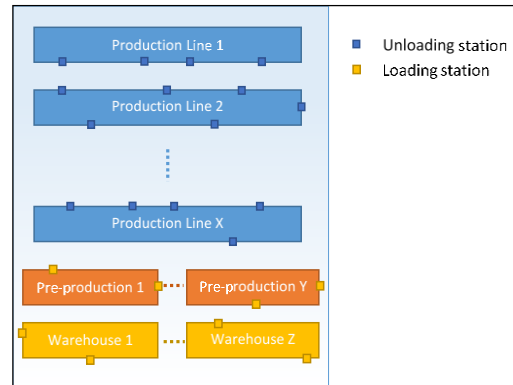


Figure 1: Simple Layout of Loading/Unloading Stations for AGV in the Simulation Model

Production lines are modelled in a simple way using only the cycle time. There are no transport constraints in the simulation model (e.g. transport using the forklifts): the blocking of AGV by the employs entering into the AGV path; discharging the battery in AGV; failure. Modelled constraints are: blocking of AGV arrival at the unloading station because of station occupation by another AGV; blocking one-way transport path because of unsatisfied AGV waiting for the unloading or loading of parts; blocking junction by another AGV; etc.

### 2.3. Transport behaviour

Transport is characterized by the following points:

- AGV takes the same transport path
- AGV transport paths are defined as one-way paths
- Each AGV is responsible for supplying only one production line
- When the AGV unloads all the parts, it runs immediately to load new parts for the production line
- AGV transports the parts on a defined maximum number of connected trailers
- Each trailer transports the same type of part
- AGV cannot overtake another AGV standing in the transport path
- It is possible to load / connect only one trailer to tractor at IN station in one time
- The trailer contains the information about at which IN station it will be loaded or at which OUT station it will be unloaded
- The trailer is unloaded or disconnected at OUT station. If the trailer is disconnected the tractor must connect another empty trailer at the same OUT station
- One or two employees provide unloading / switching at the production line

### 2.4. Production lines behaviour

The exact behaviour of production lines is not modelled in the discrete event simulation model. Production is defined by the cycle time of production line and the bill of material. The behaviour of each production line is defined as follows:



- The production line is defined by the cycle time of production (the time between the completion of two consecutive products)
- There are several unloading / loading stations for each production line
- Key parts are stored in a buffer (parts storage) placed at the production line by the unloading of these parts. This occurs only if there is free capacity of this buffer. When the buffer is full the unloading trailer stops.
- Only one key part of each type is taken from OUT station per one cycle time of production line
- The line stops if any key part is missing

## 2.5. Unloading / switching trailers at production lines

Unloading and switching trailers at the production line is provided by the operator (worker). Each production line has own operator. The operator can move between OUT stations faster than the AGV. The operator is able to unload / switch a single trailer of one AGV at one time. The operator moves between unloading / switching stations. Unloading / switching trailers are modelled as follows:

1. AGV arrives at the production line and stops at first station of unloading / switching.
2. Checking the trailers to be unloaded at the OUT station.
3. If the buffer has a free capacity the trailers set to OUT station are gradually unloaded / switched. If there is no free buffer capacity the parts cannot be unloaded and the AGV waits.
4. Once it is possible to unload parts, AGV waits for the operator until he comes to the OUT station at the production line. When the worker comes to the station the time of unloading (switching) of the entire AGV starts. Loading (switching) time of the entire AGV is defined for each IN / OUT station. Time of unloading / switching AGV is defined by triangular distribution (minimum, medium, maximum).
5. If all trailers are unloaded / switched at the OUT station, AGV moves to the next OUT station at the production line.
6. Next step is to check whether the next OUT station is served by the same operator. If so, the operator moves together with the AGV to the next position OUT station for AGV. If the same operator does not service another OUT station, operator remains at the current position or he goes to his starting point (defined separately for each location)
7. If the operator comes to the OUT station before the AGV, the operator waits until the AGV arrives (operator cannot begin to operate another AGV). In some OUT stations there is defined the maximum time limit for how long the operator can wait for AGV. The operator

can start operating another AGV if AGV does not arrive in the time limit. Proceed to step 3

8. If AGV leaves the last OUT station on the production line, AGV goes to its IN station (loading)
9. After unloading the last trailer of the AGV at last OUT station on the production line it is recorded how long AGV spent on the production line.

## 2.6. Simulation model structure

Principles of object-oriented programming were used by creating a simulation model in Plant Simulation software. We mainly used the following principles:

- Creating custom classes
- Using inheritance

The following classes were created for creating a simulation model:

- Production line template
- Station for loading the trailers template
- Station for unloading the trailers template
- Template of station where each AGV with attached trailers is created

### 2.6.1. Production line template

Production line template representing individual production lines is shown in the next figure 2. The "Production Line" part (orange-tinged part) represents the actual line where up to five inputs exist for up to five different representatives of the critical parts needed to build a product coming from associated unloading stations. "SP" object in this part represents the cycle time of production line. The tables containing data to set the production line are in "Settings of Production Line" section (yellow background). The table named "OutStationList" contains information about which unloading stations the critical parts for the production line come from. Section "Statistics" (green background) contains program code and tables with different statistics about:

- Time required for transport of parts from the loading station to the unloading station at the production line
- Statistics about the unloading part at individual unloading stations
- Statistics about the time needed to unload the entire AGV (all trailers) at the production line
- Statistics about the traffic light located on the production line

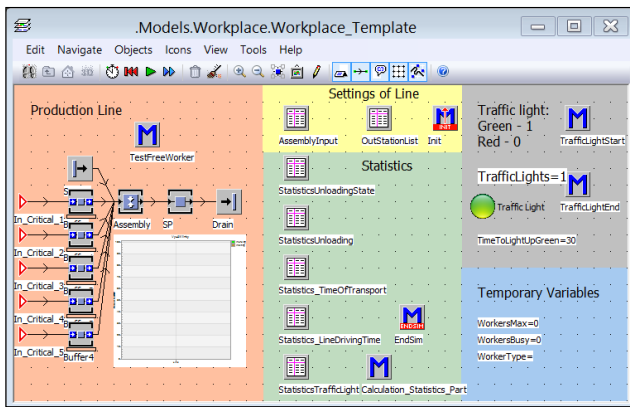


Figure 2: Production Line Template

The section “Temporary Variables” (blue background) contains temporary variables used for the proper operation of each method performed at the production line during the simulation.

The part “Traffic light” (grey background) was used only in the first stage of the project. It was assumed that only one AGV can arrive at the production line and others must wait before the production line. The part controls the switching of the traffic light.

### 2.6.2. Station for loading the trailers template

Another class template is a class representing a loading station for trailers connected to the AGV. The “Settings” section (yellow colour) contains information about the setting of the specific station for loading. This section also contains a table defining parts to be loaded at loading station. The middle yellow part is reserved for the setting of time requirements for the distribution function. This distribution function is used to simulate the loading of a trailer. The bottom yellow part defines whether the operator is needed for the loading. This operator helps to unload trailers on some production lines.

“Loading On the Trailer” part (orange background) provides simulation of loading the trailer. It looks like the simulation process of loading the trailer is quite simple but the opposite is true. We used the code implemented by the programming language SimTalk located in the “Loading” method to model the loading behaviour.

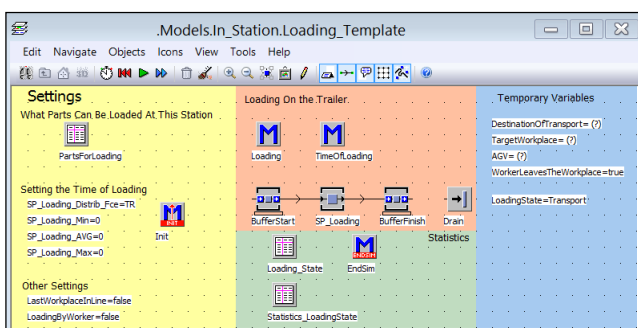


Figure 3: Station for Loading the Trailers Template

The green background section (“Statistics” part) contains tables with the statistics of loading station states stored during the simulation. The section “Temporary

Variables” (blue background) contains temporary variables used for the proper operation of each method performed at the loading station during the simulation.

### 2.6.3. Station for unloading the trailers template

From the logic point of view this class template is the complex model class. This class provides the correct behaviour of the simulation model when the trailers are unloaded. This class must ensure that the trailer can be unloaded when the operator comes to the station. The operator checks the trailer if the part must be unloaded. If so the operator unloads these parts at the unloading station. The operator must check the free place to unload the critical parts before unloading the parts. If there is no free capacity the operator must wait.

If the operator unloads all the trailers this class will ensure that the operator moves together with the AGV to the next unloading station on the production line (only if the operator must unload the parts at the next station). Individual critical parts are merged into a representative critical part of the unloading station and this representative part continues on its production line. This logic is implemented in the “Modelling the Time of Unloading” (orange) section. Almost all the unloading behaviour provides a method called **Unloading**.

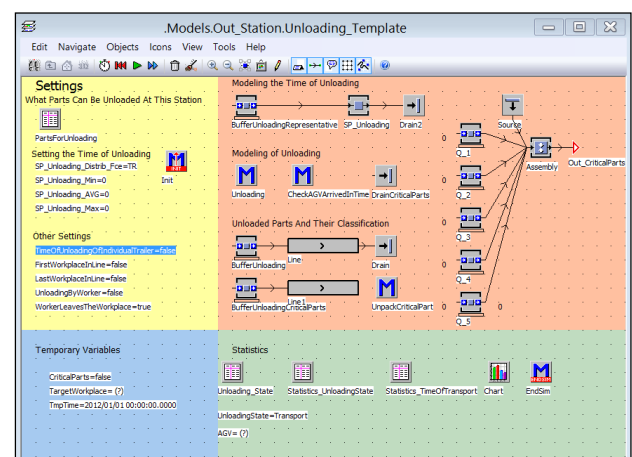


Figure 4: Station for Unloading the Trailers Template

We can find the information about a specific unloading station in the “Settings” section (yellow background). The upper yellow part contains a table that defines the selected unloaded parts at the unloading station. The middle (yellow) part named “Setting the Time of Unloading” is reserved for the detection of time requirements for the distribution function used for unloading a trailer / tractor. At the bottom of the yellow portion is defined:

- Whether the unloading requires operator
- Whether it is the first or last place of unloading on the production line
- Whether the operator leaves the station after finishing unloading or if the operator goes to the starting point provided that the operator does have to unload the same AGV at the following station

The “Statistics” section (green background) contains tables containing the statistics of the status of unloading station states stored during the simulation experiment. The “Temporary Variables” section (blue background) contains temporary variables used for the correct operations of the specific methods ensuring the unloading of the parts at the unloading station during the simulation.

### 2.6.4. Template of station where each AGV with attached trailers is created

The last class template to be created is a class that subsequently ensures the creation of the AGV consisting of a tractor providing supplying the specific production line. The trailers are connected to the tractor. Each trailer conveys predetermined transported parts.

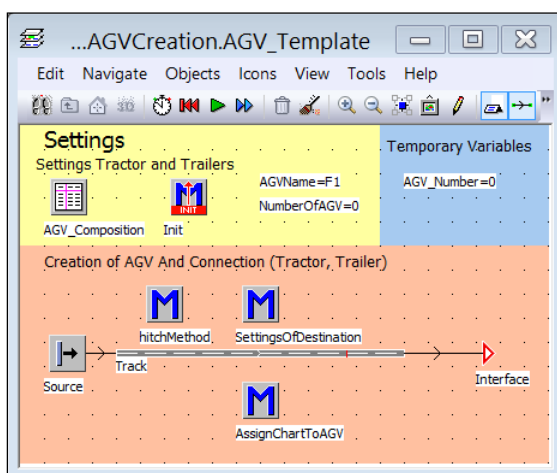


Figure 5: Template of Station Where Each AGV with Attached Trailers Is Created

The “Settings” (yellow) part is again used to define the properties of the AGV (tractor and all the trailers). The “Temporary Variables” (blue) contains temporary variables. The “Creation of AGV and Connection” (orange) part is responsible for the creation of tractors and trailers. Trailers are connected to the tractor and the target station of the AGV is set. The chart showing the statistics is generated for each AGV.

### 2.7. Setting the simulation model behaviour using input parameters

The project goal was to verify the behaviour of the planned method of supplying lines using the AGV. There are tens of loading and unloading stations. More parts for various trailers were always loaded / unloaded at each loading / unloading station and also the number of different AGV transporting different parts was in the tens. It was needed to set many parameters in the simulation model. Hence almost all parameters setting the simulation model behaviour are stored in a structured Excel file. These settings are:

- The time characteristics of loading and unloading trailers
- Cycle times of production lines

- The number of trailers trailed by the tractor transporting parts
- AGV settings:
  - The dimensions of connected trailers
  - The transported parts on the trailer including:
    - The number of loaded parts
    - Identification of critical / non critical part
  - Maximum buffer size (capacity) on the production line for the critical part
- The specification of the overall AGV path around the production hall defined by the names of the loading and unloading stations

Downloading a new set of input data stored in the Excel file into a simulation model is performed using the “Load Data from Excel” button located in the simulation model - see figure 6. All data are loaded into the tables in the simulation model. After that the adaptation of the model (various objects in the model) automatically starts. This adaptation is performed according to previously described simulation model class templates. In other words, it is all about reconfiguration of the yellow parts and possibly some objects in the orange part of the object.

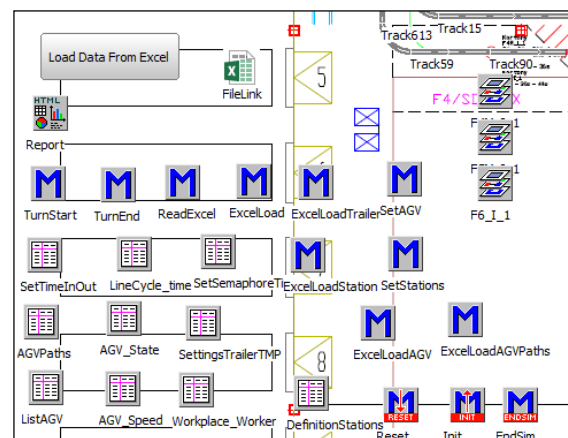


Figure 6: A Part of the Simulation Model Providing Loading of the Input Data into the Simulation Model - Reconfiguration

Operations performed after loading the Excel file into simulation model are following:

- Loading data into hierarchically structured tables of AGV for each AGV type (figure 7):
  - The trailers arrangement – which trailers transport, what and how much can trailers transport, whether it is a critical part for assembly, the size of the buffer containing the parts on the assembly line
  - The default path - which loading and unloading stations AGV must take
  - Retrieving the number of AGV types
- Loading data into structured tables describing the definition of loading and unloading stations:
  - Association of stations with the production line



inherited objects of the class. Objects are logically interrelated. Many settings are defined using the built-in tables. The simulation model was automatically adjusted by the run of the model or by retrieving data from an Excel file. We drastically reduced the number of errors in the model and increased the flexibility of the model needed to find the optimal solution of the implemented simulation project by using this approach. The following figure 8 illustrates a situation where a collapse of the whole transport system occurred (because of blocking the AGV by a different AGV). This is an example of a situation which led to a change in the model.

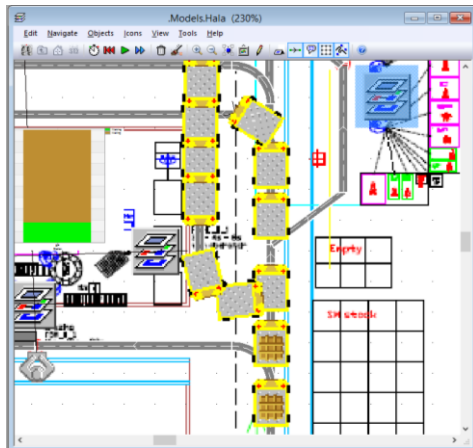


Figure 8: Sample of AGV Collapse in the Simulation Model

### ACKNOWLEDGEMENTS

This paper was created with the subsidy of the project SGS-2012-063 “Integrated production system design as a meta product with use of a multidisciplinary approach and virtual reality” carried out with the support of the Internal Grant Agency of University of West Bohemia.

### REFERENCES

- Berman, S., Schechtmana, E. and Edana, Y. E., 2009. Evaluation of automatic guided vehicle systems. *Robotics and Computer-Integrated Manufacturing*, 25 (3), 522–528.
- Homayouni, S. M., Tang, S. H. and Motlagh, O., 2014. A genetic algorithm for optimization of integrated scheduling of cranes, vehicles, and storage platforms at automated container terminals. *Journal of Computational and Applied Mathematics*, In Press.
- Kesen, S. E. and Baykoç, Ö. F., 2007. Simulation of automated guided vehicle (AGV) systems based on just-in-time (JIT) philosophy in a job-shop environment. *Simulation Modelling Practice and Theory*, 15 (3), 272–284.

### AUTHORS BIOGRAPHY

**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 the 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 research worker in the Regional Technological Institute at the University of West Bohemia in Plzen (Czech Republic). He holds M.Sc., Ph.D. and doc. in Mechanical Engineering at the same university. His research interests are oriented towards the 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.



# A PHYSICS SIMULATION TOOL FOR THE CONTAINER LOADING PROBLEM

António G. Ramos<sup>(a)</sup>, João Jacob<sup>(b)</sup>, Jorge Justo<sup>(c)</sup>, José F. Oliveira<sup>(d)</sup>, Rui Rodrigues<sup>(e)</sup>, A. Miguel Gomes<sup>(f)</sup>

<sup>(a) (b) (d) (e) (f)</sup> INESC-TEC and Faculty of Engineering, University of Porto

<sup>(a) (c)</sup> CIDEM and School of Engineering, Polytechnic of Porto

<sup>(a)</sup>[agr@isep.ipp.pt](mailto:agr@isep.ipp.pt), <sup>(b)</sup>[joajac@fe.up.pt](mailto:joajac@fe.up.pt), <sup>(c)</sup>[jfj@isep.ipp.pt](mailto:jfj@isep.ipp.pt), <sup>(d)</sup>[jfo@fe.up.pt](mailto:jfo@fe.up.pt), <sup>(e)</sup>[rui.rodrigues@fe.up.pt](mailto:rui.rodrigues@fe.up.pt), <sup>(f)</sup>[agomes@fe.up.pt](mailto:agomes@fe.up.pt)

## ABSTRACT

In the Container Loading Problem literature, the cargo dynamic stability constraint has been evaluated by the percentage of boxes with insufficient lateral support. This metric has been used as a proxy for the real-world dynamic stability constraint and has conditioned the algorithms developed for this problem. It has the advantage of not being expensive from a computation perspective. However, guaranteeing that at least three sides of a box are in contact with another box or with the container wall does not necessarily ensure stability during transportation. In this paper we propose a physics simulation tool based on a physics engine that will be used in the evaluation of the dynamic stability constraint. We compare the results of our physics simulation tool with the state-of-the-art simulation engineering software Abaqus Unified FEA, and conclude that our tool is a promising alternative.

Keywords: dynamic stability, physics engine, container loading problem

## 1. INTRODUCTION

The efficient use of transportation resources is of great relevance in the field of logistics, impacting on operational efficiency, customer satisfaction, and transport safety. The Container Loading Problem (CLP) addresses the optimization of the spatial arrangement of cargo inside containers so that the utilization of the space is maximized. The problem belongs to the wider combinatorial optimization class of Cutting and Packing Problems. According to the typology for cutting and packing problems proposed by Wäscher, Haußner, and Schumann (2007), these can be classified according to dimensionality, assortment of large items, assortment of small items, assignment type and shape of small items. In this paper we will focus on three-dimensional rectangular placement problems. The CLP can have two main variants: the maximization of the value of the cargo loaded when the number of containers is not sufficient to accommodate all the cargo, or the minimization of the value of containers when there are sufficient containers to accommodate all the cargo.

In order to be used in real world scenarios a number of constraints found in practice must be considered when addressing the problem. Cargo

stability, weight distribution, cargo positioning or cargo orientation constraints are just some examples (Bortfeldt and Wäscher, 2013).

Stability is considered one of the most important CLP constraints and has received a lot of attention by a large number of authors (Bortfeldt and Wäscher, 2013). Existing approaches to stability can be classified in two main groups, one that only addresses static stability and one that addresses static and dynamic stability. Static stability refers to the ability of each box to maintain the loading position during loading operations, and dynamic stability refers to the ability of each box to maintain the loading position during transportation.

Dynamic stability is usually ensured by placing the boxes with their sides adjacent to other boxes or the container walls. The metric used to evaluate dynamic stability is usually the insufficient lateral support, i.e., the percentage of boxes whose sides are not in contact with other boxes or with the container walls (Bortfeldt and Wäscher, 2013). This approach is used as a proxy of the real-world dynamic stability constraint and has been conditioning the algorithms developed for this problem. However, its effectiveness as a dynamic stability constraint can be easily dismissed. In a wall of boxes, as illustrated in Figure 1, boxes can have 3 sides of lateral support, but in case of acceleration along the  $x$ -axis they would most likely fall.

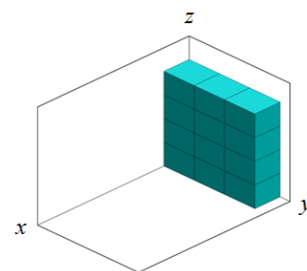


Figure 1: Unstable patterns example

The existing approaches have the benefit of being easy to incorporate in the CLP algorithms without being computationally expensive.

The main objective of this paper is to contribute to narrowing the gap between the real-life dynamic stability constraint and the CLP dynamic stability constraint by developing a physics simulation tool to



emulate the interaction between boxes, and between boxes and the container. This approach can be used to validate new dynamic stability indicators applied to solutions generated by CLP algorithms. The results of the developed tool are validated against state-of-the-art simulation engineering software and analytical results.

The remainder of the paper is organized as follows: Section 2 is devoted to reviewing related work. In section 3 the developed physics simulation tool is presented. Section 4 is dedicated to present the test conditions used to compare the two tools and to report computational results. Finally, Section 5 draws some conclusions from the findings and proposes future work.

## 2. RELATED WORK

A physics engine is a computer software designed to simulate various physical phenomena such as rigid body dynamics, soft body dynamics or fluid dynamics. It manages the forces applied to objects and the interactions between objects by simulating Newtonian physics (Jones, 2011; Seugling and Rolin, 2006). According to Erleben (2002) a physics engine has two main components, collision detection and dynamic simulation. Each one consists of a set of four interacting modules (see Figure 2).

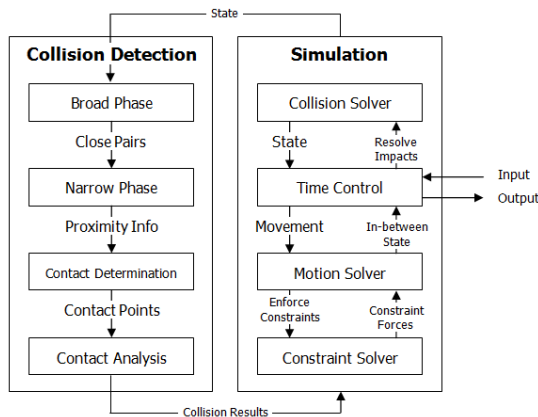


Figure 2: General Purpose Module design (Erleben, 2002)

Their performance is influenced by six essential factors: the simulator paradigm, the integrator, the object representation, the collision detection and contact determination, the material properties and the constraint implementation (Boeing and Bräunl, 2007). These factors are usually developed to address a specific application (Boeing and Bräunl, 2007).

The evaluation or validation of physics engines was addressed by various authors. Seugling and Rolin (2006) and Boeing and Bräunl (2007) evaluate physics engines in a general way, without focus on a particular application, while Hummel et al. (2012) evaluation focused on an interactive application for on-orbit servicing tasks. Pepper et al. (2007) focused on determining and increasing simulation accuracy in urban search and rescue (USAR) robot simulation. The

physics engines evaluated or validated in each paper are presented in Table 1.

Table 1: Physics engines evaluated or validated in literature

|                      | (Seugling and Rolin, 2006) | (Boeing and Bräunl, 2007) | (Pepper et al., 2007) | (Hummel et al., 2012) |
|----------------------|----------------------------|---------------------------|-----------------------|-----------------------|
| Open Dynamics Engine | x                          | x                         |                       | x                     |
| PhysX                | x                          | x                         |                       | x                     |
| Newton Game Dynamics | x                          | x                         |                       | x                     |
| Tokamak              |                            | x                         |                       |                       |
| True Axis            |                            | x                         |                       |                       |
| Bullet Physics       |                            | x                         |                       | x                     |
| JigLib               |                            | x                         |                       |                       |
| Unreal Engine2.0     |                            |                           | x                     |                       |
| Havok Physics        |                            |                           |                       | x                     |

The evaluation or validation of physics engines was carried out by performing and measuring a set of tests. Seugling and Rolin (2006) developed nine tests intended to evaluate three features: energy preservation, constraint handling and collision detection. Boeing and Bräunl (2007) tested the integrator performance, the material properties, the constraint stability, the collision system and the object stacking. Hummel et al. (2012) focused on collision detection, accuracy of collision, constraint stability and collision and friction of complex geometric objects. Pepper et al. (2007) used a set of tests from the National Institute of Standards and Technology (NIST) standard test methods for USAR robots to compare reality and virtual simulation.

From the performed benchmark tests, Seugling and Rolin (2006) reported that Newton Game Dynamics had the best overall results, Boeing and Bräunl (2007) reported that Bullet Physics had a best overall performance and Hummel et al. (2012) consider that Newton Game Dynamics and PhysX can compete with Bullet Physics.

## 3. STABLECARGO SIMULATION TOOL

With the goal of further analysing dynamic stability in the CLP, a tool, designated StableCargo, was developed. It consists of a simulator of the physical behaviour of the cargo in a container when different accelerations are applied to the container, much like those that it sustains in real life situations. This tool is based on the CGFLib (a library for computer graphics based on OpenGL, <http://paginas.fe.up.pt/~ruiodrigo/pub/sw/cgflib/docs/index.html>) and the Bullet physics engine. Both these libraries were chosen as they are cross platform and can run on most hardware. The Bullet physics engine was chosen instead of PhysX

engine due to its support of OpenCL that allows improvement in computational speed in parallel architectures (CPU or GPU based).

The results of this tool are meant to improve the development of spatial optimization algorithms that are responsible for creating container layouts. This way, it is possible to test if a given container layout is stable under a given scenario (i.e., a given set of accelerations) or not. Figure 3 presents the tool workflow.

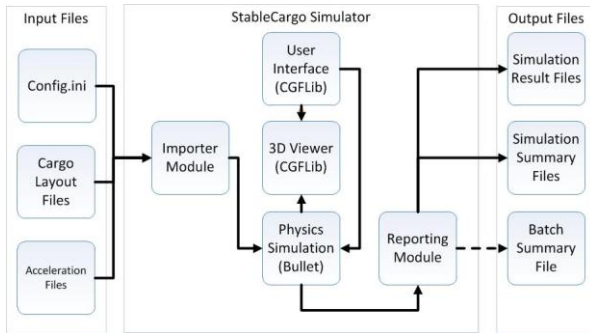


Figure 3: StableCargo's workflow

The overall solution consists of three major components, Input files, Output files and the StableCargo Simulator itself.

### 3.1. The StableCargo Simulator

The StableCargo Simulator was developed in C++. It makes use of the CGFLib, a Computer Graphics library, for 3D rendering and GLUT, for creating a Graphics User Interface to be accessed by the user in order to change simulation parameters in real time.

As Figure 4 depicts, the StableCargo Simulator tool allows for some real-time user interaction. Most notably, it allows the user to select what pair of Layout ("Solution") and Acceleration ("Forces") files are to be used at any given time (after being ran through the Importer Module), as well as applying forces or accelerations in real-time. It also allows to export the current results on demand via the "Export Results"

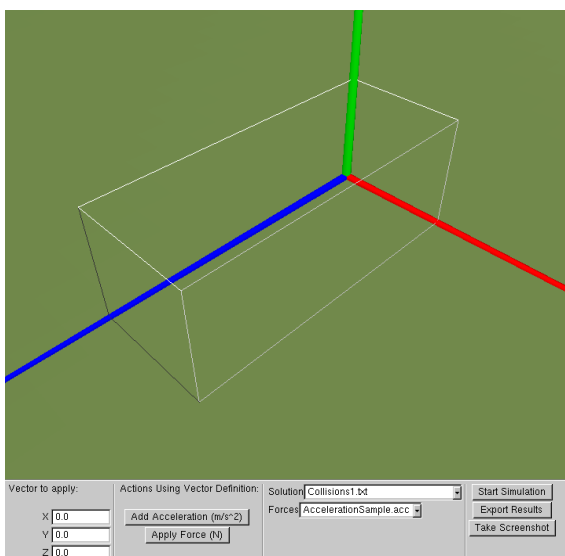


Figure 4: The StableCargo Simulator Tool

button (if the user is interested in only analysing the movement of the container up to a certain moment) and to take a screenshot of the current 3D view of the container ("Take Screenshots" button). The mouse can be used to change the position and rotation of the camera in relation to the container. This way, when the container starts moving, the camera will keep the container framed at the angle defined by the user. The container is drawn using the OpenGL wireframe drawing mode, so as to keep the interior visible at all times. As a physical entity, the container consists of six rigid bodies comprising a compound entity that represents a hollow parallelogram with the dimensions of a standard 20 feet container with a mass of 3700 Kg and friction as specified in the configuration file. In Bullet, friction is declared per physical entity. When two objects are colliding, the friction force is obtained by multiplying the friction coefficients of both objects. This translates to a model compatible with the Newtonian model of a friction coefficient between pairs of objects. Parameterization of the size and mass of a container were considered unnecessary for the scope of this project, as the CPL problem currently under study considers only this type of container. The Reporting Module generates internal statistics concerning each box and each simulation (such as the number of fallen boxes or the maximum kinetic energy of each box), and outputs those statistics to the respective output files and directories.

After selecting "Start Simulation", the application will load the Layout and Acceleration files selected and create each box as a single rigid body (undeformable physical entity) with mass or density as specified by the configuration file (config.ini), with the coordinates of the centre of mass being at the centroid of the respective box. As Figure 5 shows, the visual aspect of the boxes varies, depending on the box type and orientation. Each box type is represented by a unique box image (texture), and the faces are numbered by dots (similar to dice) to show the orientation of the box. This is achieved by blending box textures and markings (dots) using an

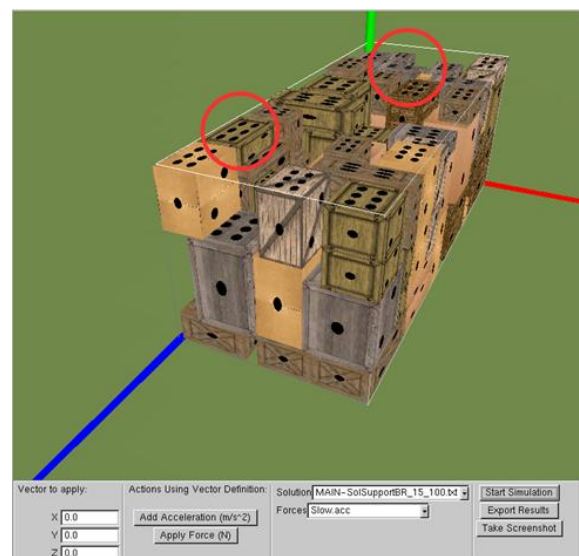


Figure 5: Representation of a loaded Layout File

OpenGL Shading Language shader.

The final placement of the boxes is shown after applying the chosen *Acceleration File* (Figure 6). Of particular interest are the boxes highlighted by the red circles, which have moved and fallen. This is easily visible when directly comparing Figure 5 with 6, but when using the tool, it is possible to see the movement of the boxes in real time, or by checking the “black dots” in the face of each box.

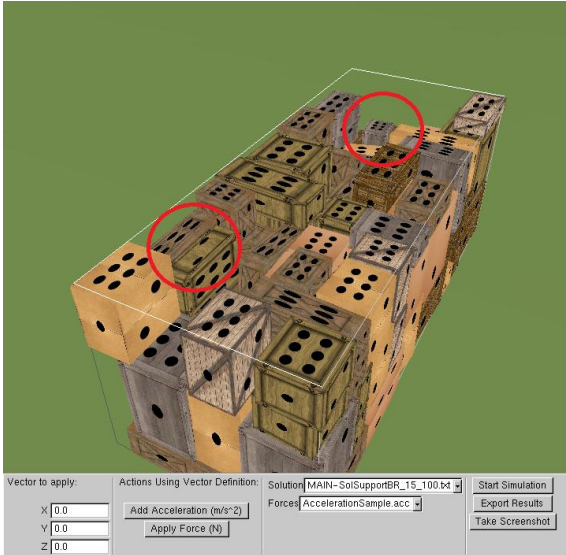


Figure 6: Final Layout after stabilization

### 3.2. Input Files

The input files consist of the *Configuration* (config.ini), *Container Layout* and *Acceleration* files.

The *config.ini* file details several simulation configuration settings that must be set prior to running the tool itself. Parameters consist in:

- [Physics Engine]
  - *DensityValue* - represents the density of the boxes in  $\text{Kg/m}^3$ .
  - *ConstantDensity* – defines whether the *DensityValue* represents the density or the mass of each box.
  - *SimulationStepsPerSecond* – sets how often the simulation is updated.
- [Renderer Engine]
  - *DrawAxis* – defines if the Axes are to be drawn by the renderer.
  - *TexturePack* – holds the path of the texture pack, to be used for skinning the boxes with their respective material.
  - *UseRenderer* – used to enable or disable the graphical visualization. It can be disabled to perform batch simulations, without need of visual feedback or interaction.
- [Export Settings]
  - *ResultsIntervallInSeconds* – Interval of time, in seconds, for the simulator to sample the statistics of each box.
  - *SavePath* – where the results will be saved.

- [Simulation Settings]
  - *StoppingCondition* – defines when the simulation will end. Either by “timeout” (x seconds have passed since the forces have been applied) or by “sleeping” (no box has moved or rotated significantly).
  - *TimeoutInSeconds* – the value in seconds to be used if timeout is the chosen stopping condition.
  - *SleepingThreshold* – the movement threshold to be used in order to consider that the simulation has ended.
  - *BatchSimulation* – specifies if the simulation is a batch simulation, meaning if there will be an attempt to pair all acceleration files and layout files in a folder. This allows for multiple simulations to be done without human intervention.
  - *BatchPath* – the path of the folder containing all acceleration and layout files.
  - *DropThreshold* – how much (in meters) must an object shift its position so that it can be considered to have fallen.
  - *ContainerFriction* – the container friction.
  - *GroundFriction* – the ground friction.
  - *BoxFriction* – the friction of the boxes.

The *Container Layout* file (Figure 7) represents a possible loading scenario of boxes inside a container. It consists of a text file with a one-line header, and multi-line body. Each line of the body represents the position of a box (through the 3D coordinates, in centimetres, of two diagonally opposing corners) and the type of material of that box (identifier of material).

|          |          |     |     |    |     |   |  |
|----------|----------|-----|-----|----|-----|---|--|
| 26281964 | 18642852 |     |     |    |     |   |  |
| 0        | 0        | 0   | 80  | 59 | 106 | 1 |  |
| 80       | 0        | 0   | 160 | 59 | 106 | 1 |  |
| 80       | 0        | 106 | 160 | 59 | 212 | 1 |  |

Figure 7: Example of a Layout file

The *Acceleration* file describes the accelerations the container experiences during a period of time. Each line specifies an initial time (in seconds), duration (in seconds) and 3D acceleration vector in  $\text{m/s}^2$  (Figure 8).

| #    | Initial-Time | Duration | X | Y | Z |
|------|--------------|----------|---|---|---|
| 0    | 1            | 0        | 0 | 0 | 0 |
| 1    | 0.3          | 1        | 0 | 0 | 0 |
| 1.3  | 12           | 2        | 0 | 0 | 0 |
| 13.3 | 0.35         | 2        | 0 | 0 | 0 |

Figure 8: Example of an Acceleration file

### 3.3. Output Files

There are three output files: *Solution*, *Abridged Solution* and *Batch Summary* file.

*Solution* files, are automatically named with the template [SOLUTION]\$LayoutFile\$AccelerationFile. The Solution file consists of raw data for each box with multiple readings per box (as specified by the configuration file), extracted from the Bullet physics engine. It contains the following data:

- *Id* – Identifier of the box. It matches the line of the layout file in which the box was declared.
- *Centre of mass position (X,Y,Z) displacement* – displays the container-relative box displacement, e.g. , the difference between the current 3D position and the initial position of the centre of mass, in relation to the position of the centre of mass of the container, in meters.
- *Total Force (X,Y,Z)* – details the vector of external forces (other than collision, gravity and friction) that might have been applied to the box, in Newton.
- *Angular Velocity (X,Y,Z)* – it is the 3D vector that contains the angular velocity of the box in radians per second.
- *Linear Acceleration (X,Y,Z)* – shows the current linear acceleration the box has in relation to the movement of the container, in  $m/s^2$ .
- *Elapsed time (s)* – represents each sampling interval time.

*Resumed Solution* files, are automatically named as [SOLUTION\_RESUMED]\$LayoutFile\$AccelerationFile and contain information extracted from the Solution File data:

- *Number of fallen boxes* – by comparing the vertical displacement each box suffered during the simulation with the configuration file DropThreshold it is possible to estimate the number of fallen boxes.
- *Centre of Mass Displacement* for each box.
- *Kinetic Energy* of each box, in relation to the container.
- *Total Kinetic Energy* – the sum of the kinetic energies of all the boxes.

*Batch Summary* files, named as [BATCH-RESUMED]\$BatchDirectory are only created if the application is set to run in batch mode. They contain:

- *Layout File Name.*
- *Acceleration File Name.*
- *Number of Fallen Boxes.*
- *Total Kinetic Energy.*

#### 4. BENCHMARK TESTS

The purpose of our physics model is to simulate the movement of a set of boxes inside a shipping container, subject to a set of external forces in typical extreme cases such as vehicle full braking, cornering or lane changing.

As friction is a parameter we considered to be one of the most relevant to evaluate the physics engine performance, a set of benchmark tests involving friction as the main parameter was performed. Numerical results

obtained using our physics simulation tool and a state-of-the-art engineering simulation software (Abaqus Unified FEA) were then compared with the analytical ones, enabling an assessment of the software packages ability in modelling the friction phenomena.

Abaqus FEA is a software suite for finite element analysis and computer-aided engineering. This software suite consists of five core software products:

- Abaqus/CAE, a software application used for pre-processing and visualizing the finite element analysis result.
- Abaqus/Standard, a general-purpose Finite-Element analyzer that employs implicit integration scheme.
- Abaqus/Explicit, a special-purpose Finite-Element analyzer that employs explicit integration scheme to solve highly nonlinear systems with many complex contacts under transient loads.
- Abaqus/CFD, a Computational Fluid Dynamics software application.
- Abaqus/Electromagnetic, a software application which solves advanced computational electromagnetic problems.

This product suite is used in academic work as well as in industrial research projects, namely in the aerospace and automotive industry. In the automotive industry engineering, it can be used to analyse sophisticated nonlinear engineering problems, such as impact/crash events, multibody systems, full vehicle loads and dynamic vibration.

For performing the benchmark tests with Abaqus FEA, rigid elements were selected for the boxes as well as for the container floor. Abaqus/Explicit was selected to perform the analysis, with a fixed time increment value of 0.1 ms. The possibility of contact between all surfaces was considered. Displacement, velocity and acceleration values for the centre of gravity of the box were recorded during all the simulations, which made post-processing the results quite simple.

#### 4.1. Friction equations

Friction can be defined as the phenomenon of resistance of a body on another which delays or prevents relative movement between them. The force that expresses this resistance always acts tangent to the contact surface. Being a force between two bodies, it naturally conforms to the principle of action/reaction. The direction of the reaction force on a body that tends to move in a given direction is always opposite to that direction. In general there are two types of friction: the fluid friction where surfaces are interleaved by a fluid layer (e.g. an oil), and the dry friction, where the two bodies are in direct contact.

In dry friction, if a force  $F$  acts on a block of weight  $W$  that is at rest, it generates reaction forces distributed along the contact surface between the two bodies. These forces have tangential or friction components  $T$ , and normal components  $N$  (see Figure 9).

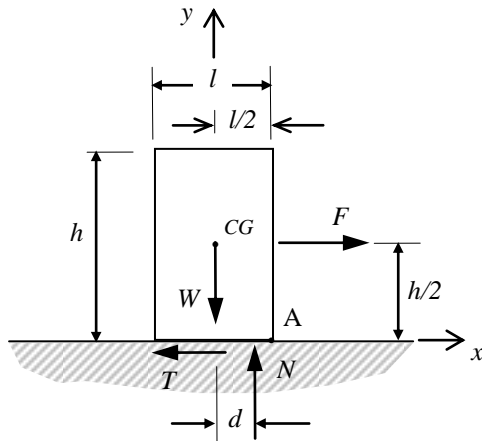


Figure 9: Forces acting on a box

Friction force  $T$  is independent of the surface area of contact, but depends directly on the resultant normal force  $N$ . The coefficient of static friction  $\mu_s$  between two surfaces in contact is determined experimentally. The body is considered to be in the imminence of sliding, if condition (1) is met and in the imminence of rolling about A, if condition (2) is met.

$$\begin{cases} T = \mu_s N \\ d < \frac{l}{2} \end{cases} \quad (1)$$

$$\begin{cases} T < \mu_s N \\ d = \frac{l}{2} \end{cases} \quad (2)$$

To test static friction, two tests are performed. The first has the purpose of testing condition (1), i.e., the imminence of body sliding and the latter has the purpose of testing condition (2), i.e., the imminence of body rolling.

#### 4.2. Sliding Test

In order to test the sliding of a body, one box with dimensions 25 cm × 110 cm × 55 cm was placed in a horizontal plane. A coefficient of static friction between the body and the plane was assigned and the force applied to the box parallel to the plane was incremented until the box started sliding. The acceleration in imminence of (1) in a horizontal plane is equal to (3).

$$a = \mu_s g \quad (3)$$

Figure 10 shows the values of the acceleration in the imminence of sliding for the range of coefficients of static friction 0.1 to 0.8. The analytically calculated value is also represented. Both Abaqus and the

StableCargo tool provided results with a high approximation to analytical values.

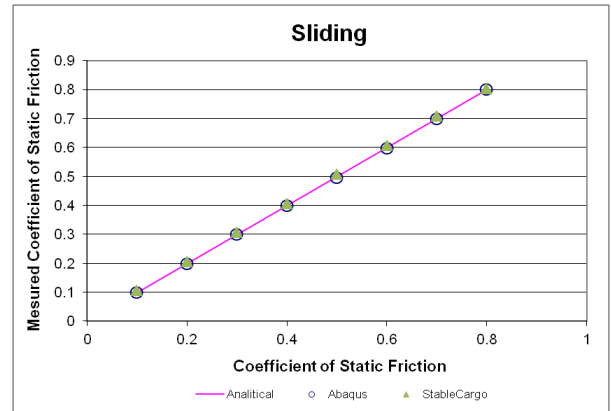


Figure 10: Measured coefficient of static friction

#### 4.3. Body Rolling Test

To analyse the imminence of body rolling, another test was developed. A coefficient of static friction between the body and the plane was assigned and a force, parallel to the plane, was applied to the centre of gravity of the box. This force was incremented until the box started to move. If sliding occurred, then the coefficient of static friction was incremented and the test repeated. When rolling of the body occurred, the coefficient of static friction used was considered the measured coefficient of static friction.

To guarantee that there is no sliding prior to rolling, it can be shown that the height and length ratio of the box must satisfy (4).

$$\frac{l}{h} > \mu_s \quad (4)$$

Figure 11 shows the values of the measured coefficient of static friction in the imminence of rolling for different values of the height and length ratio, ranging from 0.2 to 0.8.

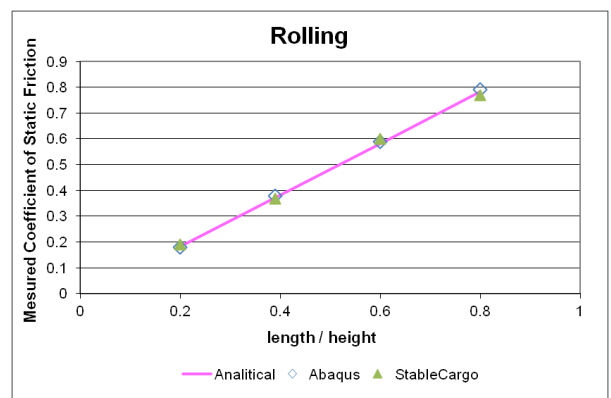


Figure 11: Measured coefficient of static friction for rolling



The analytical values are also represented. Results, obtained with Abaqus and the StableCargo tool, are in very good agreement with the analytical values.

## 5. CONCLUSIONS AND FUTURE WORK

Friction can be considered one of most relevant parameters when analysing a physics engine. To evaluate our physics simulation tool regarding its ability to model friction, a set of benchmark tests was performed, and the results obtained were compared with analytical values and those obtained using the state-of-the-art engineering simulation software Abaqus FEA.

The results obtained are in agreement with both the analytical values and those obtained using Abaqus FEA, offering good prospects for the use of the tool for evaluating dynamic stability within the CLP.

Future work should concentrate on evaluating our physics simulation tool when modelling other events, like collision and rebound that may occur in a shipping container subject to typical extreme situations, such as vehicle full braking, cornering and lane changing. A set of benchmark tests involving these phenomena should be performed and the results compared to those obtained experimentally and with other simulation engineering software, like Abaqus FEA.

## ACKNOWLEDGMENTS

This research was partially supported by ERDF through the Programme COMPETE, by the Portuguese Government through FCT (StableCargo - PTDC/SENTRA/121715/2010) and by the Project BEST CASE - SAESCTN-PIIC&DT/1/2011 which is co-financed by the North Portugal Regional Operational Programme (ON.2 - O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF).

## REFERENCES

- Boeing, A., & Bräunl, T. (2007). Evaluation of real-time physics simulation systems. In *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia - GRAPHITE '07* (Vol. 1, p. 281). New York, New York, USA: ACM Press. doi:10.1145/1321261.1321312
- Bortfeldt, A., & Wäscher, G. (2013). Constraints in container loading – A state-of-the-art review. *European Journal of Operational Research*, 229(1), 1–20. doi:10.1016/j.ejor.2012.12.006
- Erleben, K. (2002). *Module based design for rigid body simulators* (Vol. 1). Retrieved from <http://www.cs.umu.se/kurser/TDBD24/VT05/articler/ErlebenModular2002.pdf>
- Hummel, J., Wolff, R., & Stein, T. (2012). An evaluation of open source physics engines for use in virtual reality assembly simulations. *Advances in Visual ...*, 346–357. doi:10.1007/978-3-642-33191-6\_34
- Jones, M. T. (2011). *Open source physics engines Building believable worlds with open source* (pp. 1–10).
- Pepper, C., Balakirsky, S., & Scrapper, C. (2007). Robot simulation physics validation. In *Proceedings of the 2007 Workshop on Performance Metrics for Intelligent Systems - PerMIS '07* (pp. 97–104). New York, New York, USA: ACM Press. doi:10.1145/1660877.1660890
- Seugling, A., & Rolin, M. (2006). *Evaluation of physics engines and implementation of a physics module in a 3d-authoring tool*. Umea University. Umea University. Retrieved from <http://www8.cs.umu.se/education/examina/Rapporter/SeuglingRolin.pdf>
- Wäscher, G., Haußner, H., & Schumann, H. (2007). An improved typology of cutting and packing problems. *European Journal of Operational Research*, 183(3), 1109–1130. doi:10.1016/j.ejor.2005.12.047

## AUTHORS BIOGRAPHY

António Galvão Ramos graduated in Mechanical Engineering from the Faculty of Engineering, University of Porto, Portugal in 1997 and the M.Sc. degree in Logistics by the Porto Business School, University of Porto, Portugal in 2009. He is an Associate Professor with the Department of Mechanical Engineering (DEM), School of Engineering, Polytechnic of Porto (ISEP), where he has been since 2001. He worked in several multinational companies in Project Management, Operations and Logistics Management. He has been the Director of the Laboratory for Manufacturing Systems (DEM/ISEP) since 2010 and a researcher at the Institute for Systems and Computer Engineering of Porto (INES TEC) and at the Center for Research and Development in Mechanical Engineering (CIDEM/ISEP). He participated/participates in national and international research projects.

João Jacob is an Invited Assistant Professor at the Department of Informatics Engineering of the Faculty of Engineering of the University of Porto and also a PhD student focusing on Serious Games, Mobile Computing, Augmented Reality and Location-Based Games. Additionally he is also a researcher at INESC TEC currently working on a project in the area of physics simulation for the container loading problem.

Jorge Fonseca Justo is a graduate (1992) in Mechanical Engineering, with a M.Sc. degree in Mechanical Engineering, specialization in Structural Engineering, and a Ph.D. degree in Mechanical Engineering from the Faculty of Engineering, University of Porto. He is an Associate Professor with the Department of Mechanical Engineering, School of Engineering, Polytechnic of Porto (ISEP), where he has developed his activity since 1994. He has worked in several industrial projects in the Institute of Mechanical Engineering and Industrial



Management (INEGI) in the areas of structural simulation and machine projects. He is a researcher as well as a member of the managing board at the Centre for Research and Development in Mechanical Engineering (CIDEM). He has guided several master's theses, and has been involved in several research projects, both national and international, involving several companies, which resulted in several publications/communications and a patent.

José Fernando Oliveira is Full Professor at the Department of Industrial Engineering and Management of the Faculty of Engineering of the University of Porto and collaborates with the Business School of the University of Porto. His primary research interests are decision and optimization problems, in particular problems related to the efficient use of raw-materials and other resources (cutting and packing problems) and to decision support systems in industry and services. He regularly publishes the results of his research in the main operations research and management science international scientific journals and keeps a constant activity in consultancy with public and private companies with dozens of successful projects completed. During his more than 25 years long academic career he has mainly taught courses on Statistics, Operations Research and Operations Management and Logistics. He has served as Dean of Studies of The Faculty of Engineering, as main editor of the Portuguese Operational Research scientific journal and is now Vice-President of the Association of European Operational Research Societies. He is member of the General Council of the University of Porto.

Rui Rodrigues graduated in Systems and Informatics Engineering at Minho University in 1998. During his PhD he researched in the area of 3D reconstruction from Images divided between Philips Research, Eindhoven, and Minho University, until he concluded in 2006. He worked in the industry in the field of interactive systems, until he joined FEUP in 2009, where he works currently as Assistant Professor, teaching and researching in the areas of Computer Graphics, Interaction and Gaming. He is also a collaborating researcher at INESC TEC/INESC Porto.

António Miguel Gomes is an Assistant Professor at the Department of Industrial Engineering and Management of the Faculty of Engineering of the University of Porto and researcher at INESC-TEC. His main research areas are Cutting and Packing and Computational Geometry. Currently he is one of the ESICUP (EURO Special Interest Group on Cutting and Packing) coordinators.

# TOWARDS A METHODOLOGY FOR HUMAN BEHAVIOUR ELICITATION: PRELIMINARY RESULTS

João Emílio Almeida<sup>(a,b)</sup>, Rosaldo J. F. Rossetti<sup>(a,b)</sup>, Brígida Mónica Faria<sup>(c)</sup>,  
João Tiago Jacob<sup>(b)</sup>, António Leça Coelho<sup>(d)</sup>

<sup>a</sup>Artificial Intelligence and Computer Science Laboratory (LIACC)

<sup>b</sup>Faculty of Engineering, University of Porto (FEUP)

<sup>c</sup>Polytechnic Institute of Porto (ESTSP-IPP)

<sup>d</sup>Laboratório Nacional de Engenharia Civil (LNEC)

[joao.emilio.almeida@fe.up.pt](mailto:joao.emilio.almeida@fe.up.pt), [rossetti@fe.up.pt](mailto:rossetti@fe.up.pt), [btf@estsp.ipp.pt](mailto:btf@estsp.ipp.pt), [joao.jacob@fe.up.pt](mailto:joao.jacob@fe.up.pt), [alcoelho@lnec.pt](mailto:alcoelho@lnec.pt)

## ABSTRACT

Human behaviour knowledge is an important requirement for implementing realistic evacuation models. Although much work has been done in this field of research there are no universally accepted quantitative methods. In this paper we present a novel methodology for human behaviour elicitation that was coined SPEED (Simulation of Pedestrian and Elicitation of their Emergent Dynamics). An experimental setup to test the concept was envisaged and validated by a group of experts using the Delphi method. A test bed was designed using the Serious Games concept and 22 subjects were selected for a pilot test. Preliminary results are promising, showing that this methodology might be used for the elicitation of human behaviour of subjects when facing an emergency. Moreover, the data acquired is of great importance for the fire safety experts designing new buildings or planning strategies to improve emergency paths. Another possible outcome is to create an artificial population based on human behaviour to populate emergency evacuation simulators.

Keywords: fire safety, human behaviour, behaviour elicitation, decision-making.

## 1. INTRODUCTION

One of the topics within the modelling and simulation field that has an increasing importance among researchers is the study of building evacuation, especially regarding human factors. It was noted by one of the prominent researchers in this area, Prof. Galea, that the development of fully adaptive behaviour models, in which “agents react to the evolving situation rather than simply responding in a pre-programmed manner to a specified scenario,” is one of the grand challenges for the 21<sup>st</sup> century (Galea, 2012).

Evacuation models are used by many architects and engineers when developing new complex buildings, and by emergency managers for improving the safety of existing ones. The behaviour of subjects when facing the need of a rapid exit of a building, due to an emergency situation, whether it is a fire or some other kind of hazards, is one specific field of research that has been gaining much attention and for which no definitive theory is available (Kuligowski, 2011).

Studies refer that many building occupants still lack the proper education and are unaware of the best exit choice strategies when facing a fire or some other emergency. The egress of buildings is a chaotic process depending on many variables, both quantitative and qualitative, that are unknown and hard to determine (Pauls, 1995; SFPE, 2002; Timmermans et al., 2009).

Building safety designers usually define egress paths based on the shortest way to the outside and other safety issues. However, the possibility of a predefined route being blocked due to some unpredictable situation, such as fire, smoke or even partial collapse resulting from earthquake, for instance, is a situation often ignored.

Events such as the fire in the Brazilian discotheque “Kiss”, January 27, 2013, where 242 died, or the 2012’s Halloween party, in Madrid, where an overcrowded concert led to the death of five young women crushed in one of the exit tunnels, are unfortunately more frequent than they should. Such occurrences rely greatly on the lack of information and training of occupants. It should be also noted the need of more information about occupants’ behaviours when facing an emergency and the need to quickly find the best and safest way out of the building (Galea, 2012; Silva et al., 2013a).

Devising sophisticated and advanced evacuation models, using agents resembling human behaviour is a challenge that many researchers are currently pursuing. Carattin and Branningan (2013) state that historical and cultural issues play a great deal of importance in human factors, and for that reason “*behavioural uncertainties are extremely large and include types of uncertainties unknown to physical science models.*” To conduct human behaviour research there are no universally accepted quantitative methods (Ronchi et al., 2013).

Some researchers proposed using computer games to train and also acquire human behaviour (Chitaro and Ranon, 2009; Smith and Trenholme, 2009; Kobes, 2009; Cordeiro et al., 2011a, Almeida et al., 2013). Computer games have a set of features that address these problems. In fact they provide engagement to their players, keeping them focused; additionally they allow players to become experts in the resolution of challenges, and improve their skills. Some work has been done in this domain, using Serious Games (SG).

The availability of game engines such as Unity3D provides a rapid way for prototyping 3D scenarios and of performing experiments to elicit human behaviour in such emergency situations (Rossetti et al., 2013). By using the SG concept, it is possible to record some metrics associated with players' decisions (Almeida et al., 2013).

In this paper we present a novel methodology for the elicitation of the behaviour of pedestrians when exiting a building in case of emergency, designed on the grounds of recent developments in serious games, multi-agent systems, social simulation, and peer-designed agents (PDA). The methodology includes the specification and prototyping of different scenarios simulating various possible situations a building occupant may find in reality. Selecting and implementing scenarios alone represent a very laborious and time-consuming phase of our methodology; for the sake of illustration, in this paper we present only one of such scenarios though.

A preliminary experiment implementing a scenario known as the "exit choice problem" was then set up, with the objective of gaining some insight into how humans behave when trying to leave a room. Two possibilities were considered, namely choosing the left or choosing the right side of a corridor, in different circumstances.

This experimental setup was validated by a panel of international fire safety experts, including researchers, practitioners and fire fighters, using a mix of questionnaire and live interviews based on the Delphi method (Dalkey and Helmer, 1963).

A sample of subjects was selected to play the game that was developed using the Unity 3D game engine. Results were recorded and analysed. During the experience, players' comments were also collected, which will be used to enhance futures scenarios and the game features as this research evolves. Part of the data recorded is expected to be used further to drive the artificial agents trying to recreate player decisions, based on their previous selections and the selected category of behaviour, a process known as peer-designed agents.

The remaining part of this paper is organised as follows. We start by briefly presenting some related concepts that underline and motivate this project. We then put forward the pilot study, describe the setup scenario and characterise the population sample. Results are presented and discussed afterwards. We finally draw conclusions and list some further steps in this research.

## 2. BACKGROUND AND RELATED WORK

### 2.1. Agent-based Modelling and Simulation

Agent-Based Modelling (ABM) and Agent-Based Simulation (ABS) are particularly adequate for representing social relations by using behavioural models exploiting emergent behaviour (Almeida et al., 2011), for which the main modelling metaphor is the so-called agent. An agent is basically an autonomous

entity, capable of perceiving the surrounding environment through a set of specific sensors and of acting upon the same environment, maybe directly affecting its current state, through specific effectors. They feature reasoning mechanisms underlying their decision-making abilities and may exhibit communication channels allowing them to interact between each other (Russell and Norvig, 2009). When multiple agents are put together, they can perform rather socially, building upon a multi-agent system (MAS) which makes ABMS ideal to represent many scenarios of our daily lives, including pedestrians. These concepts briefly introduced are a subset of the Artificial Intelligence (AI) techniques that are used in social simulation (Wooldridge, 2002).

For these reasons, ABMS has been widely used to simulate pedestrian interactions and crowds in a vast range of different scenarios, naturally including evacuations and risk situations (Almeida et al., 2012).

### 2.2. Pedestrian modelling and simulation

Although some attempts of modelling and simulating pedestrian movement have been around for decades, this field of research has recently received a clear boost in attention in a variety of disciplines, not only in the ones traditionally concerned with pedestrians, such as transportation, urban planning and design, but also in applied physics, computer science and artificial intelligence. In the latter case, pedestrian movement is often viewed as an interesting case to show properties of complexity theory and multi-agent models such as aggregate patterns emerging from simple principles applied to microscopic agents (Timmermans, 2009).

The more common representation models use one of the three following formats (Castle 2007):

- Coarse Network (or physics-based models of particle flow) typical of macroscopic models;
- Lattice or Cellular Automata;
- Continuous space.

Pedestrian simulators model real-life situations using collection of various mathematical methods, such as human behaviour models, hazard propagation models and the geometric data of the environment (scenario) in 2D or 3D. Due to heavy computational requirements and long run times, these simulations are performed mostly during the planning stage, helping architects to design safer buildings where human casualties in case of possible hazards are minimized (Jafari et al., 2003).

Pedestrian simulation methods vary with the application domain. The environment is a key aspect to consider when modelling pedestrians. Their behaviour is different indoors and outdoors. Applications of pedestrian simulation range from entertainment to more serious use. Entertainment includes video games where artificial societies are reproduced with diverse purposes, such as the cinema and TV industry so as to create synthetic population in films. In the serious uses, training and education are the most common

applications. Other domains include transportation and urban planning, namely in transportation systems where pedestrian simulation is an important aspect.

Safety is one of the areas where pedestrian simulators have gained and increased development and focus. Aim is obvious and quite straightforward. Architects and engineers will benefit in the designing phase to evaluate the safety conditions of a building in an emergency situation. New advanced and complex buildings, like skyscrapers, with large occupancy and multi-floors, stadiums or other public facilities need tools to validate if they are compliant with fire safety regulations. Evacuation simulators are increasingly used for this purpose (Kuligowski et al., 2010). Some public transportation vehicles such as trains, planes and ships, are nowadays designed bearing in mind the safety issues, using specific developed simulators (Santos and Aguirre, 2004).

### 2.3. The decision-making process in building evacuation

When pedestrians have to leave a building, especially if some sort of emergency requires this action, decisions on which direction to follow must be made. This decision process gets an additional importance when due to safety issues the time to exit is vital. Indeed, if a wrong decision is made, which leads the occupants to a blocked exit or causing them to get stuck with smoke and fire, it may have disastrous consequences.

In complex buildings this problem gains even bigger importance, such as shopping malls, schools, hospitals and hotels (Kobes et al, 2007); this scenario gets worse in case of an emergency such as fire (Xie, 2011).

The behavioural process of way-finding in such emergency situations has been studied by many researchers for the past decades (Pauls, 1995; SFPE, 2002; Kuligowski, 2011). When modelling pedestrians both the collective and the individual issues should be addressed (Hoogendoorn et al., 2004). Timmermans et al. (2009) state that the pedestrians' decision-making process, as well as their movement are of critical importance in the development of pedestrian models that aim to reproduce the reality. Kuligowski (2008, 2011) has brought some insights into the matter regarding human behavioural process during building evacuation. Also, the validation and verification of pedestrian models often relies on data collected by means of direct observations, photographs, videos (Qingge et al., 2007), various sensors (Vasconcelos et al., 2012) as well as stated preference questionnaires (Cordeiro et al., 2011b).

### 2.4. Serious Games and Gamification

The Serious Games (SG) concept emerges with the use of video game frameworks for rapid prototyping applications with purposes other than mere entertainment. Taking advantage of state-of-the-art software technology that includes appealing high-definition graphics and animations, it is possible the

creation of virtual-reality worlds without the need of a battalion of software developers. SGs are nowadays used in a variety of applications including education, training, health, advertising as well as social simulation (Hays, 2005; Frey et al., 2007; Ribeiro et al., 2012a).

The benefits of SGs combined with other training activities include: learners' motivation is higher; completion rates are higher; possibility of accepting new learners; possibility of creating collaborative activities; learn through doing and acquiring experience (Freitas, 2006).

The gamification concept is another new buzzword deriving from the use of game thinking of a game mechanics in a non-game context to engage subjects and make them loyal. Huotari and Hamari (2012) define it as "a process of enhancing a service with affordances for gameful experiences in order to support user's overall value creation". Rossetti et al. (2013) suggest that "gamification is an important instrument towards behaviour persuasion," in the perspective it has the potential to influence behaviour in longer term.

These aspects briefly referred above are a subset of the reasons that are underlined in SG-based applications that make them so common nowadays for many uses, including human behavioural data acquisition.

## 3. TOWARDS SPEED: A METHODOLOGY FOR HUMAN BEHAVIOUR ELICITATION

Due to the high complexity and uncertainty of all environmental and intrinsic variables that affect human behaviour, existing evacuation simulators fail to capture in detail all the dynamics that characterize them. Besides cultural and historical aspects, pedestrians can choose the direction to take; can change at any moment their planned itinerary; and their choices might be affected by any social, economic, or environmental phenomena.

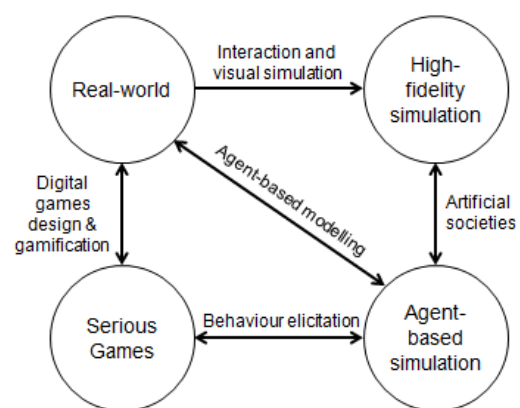


Figure 1 – Methodological perspective combining SG, ABM, ABS, behaviour Elicitation (adapted from Rossetti et al., 2013).

This consists on the basis for the need of a framework to allow consistently acquisition of human behaviour in such hazardous situations that can hardly be recreated due to the inherent risk for the subjects. As stated in previous section, SG and the gamification

concepts seem to be an obvious candidate for the process of behaviour elicitation.

A methodological approach for capturing the complex and uncertain activities of pedestrians in the real-world was devised as a first-class abstraction for behaviour elicitation, using agent-based simulators, virtual-reality simulators and the real world (fig. 1).

The real-world ecosystem is where pedestrians live and interact with themselves and the environment. For practitioners and domain experts simulating extreme situations, testing theories and what-if scenarios, the agent-based modelling and simulation (ABMS) is of paramount importance. The agent-based simulation ecosystem is actually where artificial societies (as a means to represent human behaviours and social interactions) grow and breed. To create such virtual worlds the modelling of agents by means of ABMS needs behaviour knowledge for which the behaviour elicitation is required.

Instead of using high-fidelity simulation that recreates with a high degree of detail the real domain, for which complex and expensive simulators are necessary, such as CAVEs and other sophisticated virtual-reality technology (M.Kobes et al., 2009) we propose the use of simpler SG for behaviour elicitation. Of course we do not mean to replace previous efforts and technologies, but rather complement them with SG techniques to enhance behaviour modelling and analysis (Almeida et al., 2013).

This extended account of the concept of SG is in accordance with the integrative perspective of behaviour elicitation to reveal the decision processes behind the course of action people perform to achieve certain goals and respond to stimuli, or during deliberation. This process is not just a matter of collecting data through logs of different and successive interactions of the player during the game for post processing. Rather, we make use of the intrinsic nature of SG to impel the player to add semantics to every decision and action performed during the game that might better clarify the sequence of cognitive states that resulted in or triggered certain actions. In other words, we ask the player to model his/her own agent, in a process that is known by peer-designed agents (PDA). In brief, we thus integrate serious games into the conceptual framework SPEED (Simulation of Pedestrians and Elicitation of their Emergent Dynamics) by combining behaviour elicitation with the PDAs, allowing players to feature their peer agents with their own idiosyncrasy (Rossetti et al., 2013).

The SPEED framework is thus one instantiation of the aforementioned methodology devised by Rossetti et al. (2013), which aims at the study, simulation and elicitation of the pedestrian interactions and processes namely the reasoning and decision-making aspects related to movement, both indoors and outdoors.

An experimental setup to test the concept was envisaged and validated by a group of experts based on the Delphi method (Dalkey and Helmer, 1963). A test bed was designed and a pilot test devised to implement

the Serious Games concept test bed in which various scenarios for behaviour elicitation are shown to players.

The overall idea is to create a set of scenarios, using the well known concept of “game level”, in which the player is moving from one scenario to another, in a succession that can lead to increasing stages of difficulty. This is one important mechanism explored in the gamification concept, which is currently a fashionable way of engaging subjects and enhance their commitment to the task at hand (Rossetti et al., 2013).

## 4. PILOT TEST

### 4.1. Prototype development

A preliminary simple scenario was devised for behaviour elicitation and as a proof of concept. Players start in a study room, which could be an office in a university building, with two desks, some computers and bookshelves. To implement the prototype we have selected the Unity3D game engine. For the scenario creation we use free-of-charge 3D models available at the Unity Asset Store, a place within the Unity company website (see [www.unity3d.com](http://www.unity3d.com)) where developers share models. Some objects were collected from the Google Warehouse, developed in SketchUp, and afterwards converted in a Unity3D compatible format to be imported into the game framework.

### 4.2. Unity3D and First-Person-Shooter game genre

Unity3D is a powerful WYSIWYG (What You See Is What You Get) game engine that has some interesting features such as: i) free use; ii) powerful graphical interface; iii) capability to import models from other sources; iv) Programming support in JavaScript, C# or Boo.

The First-Person-Shooter (FPS) game genre is very common among video games. FPS is characterised by placing players in a 3D virtual world which is seen through the eyes of an avatar (Silva et al., 2013b). The aim is to give the feeling of immersion in the virtual environment via this first-person view.

The controls for this game use the FPS common standard: a combination of keyboard and mouse to move the player around the environment. To improve usability and having in mind subjects who are not familiar with PC video games and this keyboard plus mouse technique, a joystick was also used as an alternative (Figure 2).



Figure 2 – Thrustmaster joystick T-Flight Stick X PC/PS3 used in our game framework.



The character movement control using the joystick is very straightforward and intuitive: moving back and forward or left and right has the same correspondence in the game. To rotate the avatar, the user has to twist the joystick.

The trigger button was configured as “accept\_button” to confirm the game messages and the button number “2” to present the game main menu. Mouse right button was also another way of leaving game and returning to the main menu.

Main menu was a simple one with a set of buttons, each one giving access to different functionalities of the game: i) training; ii) game (all five levels continuously); iii) one button for each game level (five in total); iv) exit button (to end the game).

### 4.3. Game scenario

The game scenario consists of a typical office room, having two desks with computers, bookshelves, a scoreboard and some other furniture and details to give as much of a realistic look-and-feel immersion as possible. The exit door leads to a corridor with two possible ways; both are similar, having the equidistant way to double exit doors located at the end of the corridor after a turn. The plan is represented in Figure 3.



Figure 3 – Bird view of a building layout: room with corridors leading to two possible emergency exits

Looking at the image, player starts located at the desk that is positioned to the right side, when looking towards the door. Corridor has the same length for both sides. The player chooses either to go left or to go right. When exiting the room, there will have a 90° turn followed by a pathway leading to a double exit-door, with exit signs on the door and on top of it, as it would be in any real scenario. When moving towards the exit, an invisible collider was thoroughly put in place to end the game level and move on to the next scene or return to the main menu.

### 4.4. Experimental setup

The test-bed prototype was run under a Windows 7 commodity laptop with no special specifications. All subjects were tutored in the game controls usage, keyboard and mouse or joystick, and were given a small time to adapt in controlling the FPS character and how

to move around the scenario by loading a tutorial scenario with no goals or restrictions.

During experiments, each player had only one run so as to capture his/her reactions to the gaming experience and controls. To avoid any source of biased data and noise in subjects’ responses, other players were kept apart prior to simulation runs. This way no one possibly had any previous knowledge on the scenario and activities of other players. It is important to stress on this aspect as our expectation was to acquire genuine reactions and register as naturalistic choices as possible.

All subjects were instructed previously about the experiment and signed a written consent.

### 4.5. Game description

The game starts with the player inside the office room. After a small period of time (it was setup for 5 seconds), the fire alarm triggers and the player is urged to exit as quickly as possible, using the nearest exit. In the first level only an exit sign on the top of the office entrance is presented. No other cues are given on which direction to take. Players can equally choose between going left or right (see Fig.4).



Figure 4 – Level 1: no direction sign on the corridor

When the player chooses one direction, moving the character towards the exit, this level finishes as soon as he/she gets near the door (Fig. 5).



Figure 5 – Exit double door at the end of the corridor

The next level starts exactly at the same point, the fire alarm triggers and the player must repeat the process of steering his/her avatar towards the nearest and safest exit. This time, however, there is an emergency sign pointing to the left (Fig.6). The player is expected to follow the exit sign direction but there is no deterrent if he/she chooses the other way.





Figure 6 - Level 2: direction sign pointing left is visible on the corridor wall.

The third level starts similarly the previous one, showing the left turn sign depicted in Fig. 6, and players are expected to follow that direction. However, in this level, there is smoke coming from the end of the corridor (Fig.7). The goal of this level is to present the player a dilemma: should he/she go forward regardless of this obstacle or follow the other direction even though the exit sign is pointing left?



Figure 7 - Level 3: sign pointing left is visible on the corridor as well as smoke

Fourth level is again similar to the previous one. The only difference is that besides the smoke, there is also fire (Fig.8). To prevent players from trying to pass through the fire, this object is set with a collider and is insurmountable. So, all players on this level have no option other than escaping through the right corridor.



Figure 8 - Level 4: sign pointing left is visible on the corridor as well as the huge fire.

The fifth, and last level, presents a different challenge. This time there is neither fire nor smoke, but rather people are running from the left towards the right side of the corridor. The goal is to see the reaction of players when having to select between following the exit direction shown in the sign (left) or following the crowd moving in the opposite direction (see Fig. 9).



Figure 9 - Level 5: People running the opposite direction of the exit sign

#### 4.6. Population Sample

A total of 22 subjects were selected as a sample to test the developed prototype. These testers can be classified according to the parameters presented in Table 1. This data was collected in a questionnaire presented to the players at the end of the game.

Table 1 - Population sample's age and gender

|                    | Values |
|--------------------|--------|
| Male subjects      | 13     |
| Female subjects    | 9      |
| Min age            | 14     |
| Max age            | 55     |
| Mean age           | 35     |
| Age SD             | 14,95  |
| Number of subjects | 22     |

Subjects, prior to playing the game, had to answer a small initial questionnaire. Results are presented in Table 2. The objective was to assess the computer and game expertise as well as fire safety training.

Table 2 – Population sample's skills on computers, gaming and fire safety

|                       | Yes | No |
|-----------------------|-----|----|
| Computer basic skills | 20  | 2  |
| Video game expertise  | 6   | 16 |
| Fire safety training  | 5   | 17 |

#### 4.7. Data Collected

For each scenario, the choice on whether to use the left or the right exit was recorded. Results are presented in Table 3.

Table 3 - Exits used by the players

|                                                               | Left | Right |
|---------------------------------------------------------------|------|-------|
| 1 - no sign on corridor                                       | 8    | 14    |
| 2 - sign pointing left                                        | 21   | 1     |
| 3 - sign pointing left + smoke                                | 12   | 10    |
| 4 - sign pointing left + fire                                 | 0    | 22    |
| 5 - sign pointing left + people running in opposite direction | 6    | 16    |

## 5. PRELIMINARY RESULTS AND ANALYSIS

All subjects found the experiment very interesting and educational. Only five said to have some sort of fire safety training, so the situations presented at the game made them think on which option would be the best and possible consequences in real life. The realistic 3D scenario provided by the game was a positive aspect most players remarked. The possibility of choosing the keyboard plus mouse or the joystick was also noted.

On the negative side, the rotation speed of the joystick was considered to be too slow. This aspect was not changed, although noticed by most players, to avoid biasing the results. It will be improved and turning speed will be further calibrated in future versions of the game. Another aspect was that some players, noticeably the ones more used to playing this genre of games, did the path so quickly that missed the emergency sign pointing left. There was one player that only saw it in the last run. To overcome this issue in future experiments, the starting point of the player will be moved to the middle of the room in front of the door.

To level all players, time of execution was not recorded. In previous experiments (Ribeiro et al., 2012b; Silva et al., 2013c), authors have used the evacuation time as an important metric to evaluate the performance of a player, but we have just realized that gaming experience biases results, so in this setup focus was on understanding and eliciting user choices instead, as well as their reasoning and their underlying way-finding process.

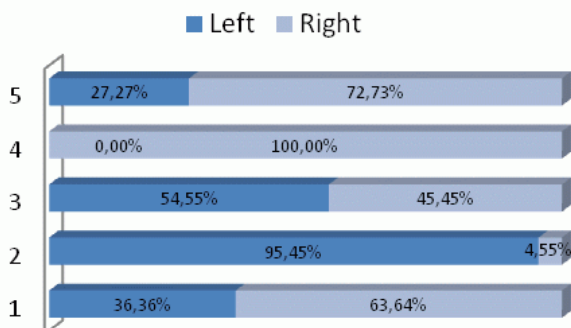


Figure 10 – Results in percentage for each level

The first level goal is to try and find a pattern, if any exists, of occupants when leaving a room turning right or left. Intuition led to the idea that most people prefer right instead of left; and over 72% in this test confirmed that assumption. We suspect the initial location led some players to go forward and left, but further tests are needed to confirm this hypothesis though.

The second level goal is to counteract that reaction (turn right) and force subjects to turn left. Even with the sign pointing that direction, there was one player that missed it and turned right.

The purpose of the third level is to pose a dilemma to the player: if the sign is pointing left but smoke is coming from that direction what should they do? Should they follow the emergency sign or avoid the

smoke? Some players said that the game should have a command to lower the character and pass right through the smoke. More than half of players decided to face the smoke. This happens due to Unity3D limitations; smoke is dark but gets thinner when approaching thus allowing players to surpass it. The fourth level is similar, but this time fire is strong and all players decided to go back and turn to the right exit.

The last level presents another dilemma: people are running the opposite direction as pointed by the exit sign; what should the player do? Follow the runners or ignore them and go on through the left? Despite a great number of players that choose to follow the people running, more than 63%, some said they prefer to go the opposite direction to avoid crowd and respect the sign.

## 6. CONCLUSIONS AND FUTURE WORK

The aim of this experiment was to determine the validity of this methodological approach for human behaviour acquisition using SG in the specific domain of way finding in the evacuation of buildings.

Although the population sample is relatively small, results are promising and point out in the direction of considering that this tool for behavioural acquisition can be considered by social scientists and fire safety engineers. Furthermore when combined with other scenarios, results might be of great importance for researchers. The expert panel consulted using the Delphi method stated the importance of the experiments and the need of real data representative of the behaviours of building occupants when exiting the place.

The aim of the experiments described in this paper is to try and find whether there are patterns in behaviours when making the initial selection of which direction to take when exiting a room: left or right? For the practitioners, both architects and engineers, as well as emergency planners, this tool might be a valuable help in understanding what goes on through occupants' mind when escaping from a place. Will they consider the emergency signs? Will they follow others? Will they return upon smoke or only if they see insurmountable fire?

For fire scientists and researchers, the issue of the exit-selection is of great importance. Some experiments have been carried out, but only in a small scale. Future work will be the creation of a Web version of the game, playable by many people, to gather as much responses as possible, and try to establish a pattern. To this matter, the fact that many of the players confessed not to have followed the emergency signposting, pointing to the nearest and safest exit, must be emphasized.

The very next steps in this research are two-fold: to improve this experiment using players' comments collected during the experiments, and to implement other scenarios for a large sample of subjects to play. Expected results will be thoroughly analysed in order to try and establish some standard behaviour that might emerge from the data. There will be also a questionnaire presenting the same scenarios and asking the subjects

for their decisions. Results of questionnaires and the SG tests will then be compared to see whether answers are similar or not.

## ACKNOWLEDGMENTS

This project has been partially supported by FCT (Fundação para a Ciência e a Tecnologia), the Portuguese Agency for R&D, under grant SFRH/BD/72946/2010. Authors are also grateful to the volunteers that participated in this study.

## REFERENCES

- Almeida, J.E., R. Rossetti, and A.L. Coelho. 2011. "Crowd Simulation Modeling Applied to Emergency and Evacuation Simulations Using Multi-Agent Systems." In *DSIE'11 - 6th Doctoral Symposium on Informatics Engineering*, 93–104. Engineering Faculty of Porto.
- Almeida, J.E., Kokkinogenis, Z., and Rossetti, R. J. F.. 2012. "NetLogo Implementation of an Evacuation Scenario." In *Fourth Workshop on Intelligent Systems and Applications (WISA'2012)*. Madrid, Spain.
- Almeida, J.E., and Rossetti, R. J. F.. 2013. "Mapping 3D Character Location for Tracking Players' Behaviour." In *Fifth Workshop on Intelligent Systems and Applications (WISA 2013)*. Lisbon, Portugal.
- Carattin, E., and Brannigan, V.. 2013. "Science or science fiction? The use of human behavioral models in fire safety regulation." In *Thirteenth International Conference, Interscience Comms*, 553–558. London, UK.
- Castle, C. J. E. 2007. *Guidelines for Assessing Pedestrian Evacuation Software Applications*. Centre for Advanced Spatial Analysis (UCL).
- Cordeiro, E., Coelho, A. L., Rossetti, R. J. F., Almeida, J. E. 2011a. "Human Behavior Under Fire Situations – Portuguese Population." In *2011 Fire and Evacuation Modeling Technical Conference*. Baltimore, Maryland, August 15-16,2011.
- Cordeiro, E., Coelho, A. L., Rossetti, R. J. F., Almeida, J. E. 2011b. "Human Behavior under Fire Situations – A Case-study in the Portuguese Society." In *Advanced Research Workshop: Evacuation and Human Behavior in Emergency Situations*, 63–80. Santander, Spain: GIDAI. Universidad de Cantabria.
- Chittaro, L., and Ranon, R.. 2009. "Serious Games for Training Occupants of a Building in Personal Fire Safety Skills." *2009 Conference in Games and Virtual Worlds for Serious Applications (March)*: 76–83.
- Dalkey, N., and Helmer, O. 1963. "An Experimental Application of the Delphi Method to the Use of Experts." *Management Science* 9 (3).
- Freitas, S. (2006). *Using Games and Simulations for Supporting Learning*. Learning, Media and Technology, 31(4):343– 358.
- Frey, A., Hartig, J., Ketzler, A., Zinkernagel, A., Moosbrugger, H. 2007. The Use of Virtual Environments Based on a Modification of the Computer Game Quake III ArenaR in Psychological Experimenting. *Computers in Human Behavior*, 23(4):2026–2039.
- Galea, E.R. 2012. "Evacuation and Pedestrian Dynamics Guest Editorial – 21st Century Grand Challenges in Evacuation and Pedestrian Dynamics." *Safety Science* 50: 1653–1654.
- Hays, R. 2005. The Effectiveness of Instructional Games: a Literature Review and Discussion. Technical report, Naval Air Warfare Center Training Systems Division Orlando, FL.
- Helbing, D., Molnár, P., Farkas, I.J. Bolay, K. 2001. Self-organizing Pedestrian Movement. *Environment and Planning B: Planning and Design* 28 (3): 361–383. doi:10.1068/b2697.
- Hoogendoorn, S.P., Bovy, P.H.L. 2004. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological* 38 (2) (February): 169-190.
- Huotari, K., and Hamari, J.. 2012. "Defining Gamification: A Service Marketing Perspective." In *16th International Academic MindTrek Conference 2012*. Tampere, Finland.
- Jafari, M., Bakhadyrov, I., and Maher, A. 2003. "Technological Advances in Evacuation Planning and Emergency Management: Current State of the Art". Center for Advanced Infrastructure & Transportation (CAIT) Civil & Environmental Engineering Rutgers, The State University Piscataway, U.S.
- Kobes, M., Helsloot, I., De Vries, B., Oberijé, N., Rosmuller, N. 2007. Fire Response Performance in a Hotel. Behavioural Research. In *Interflam 2007 - 11th International Fire Science and Engineering Conference*, 2:1429–1434.
- Kobes, M., Oberijé, N., and Groenewegen, K. - Ter Morsche. 2009. "Serious Gaming for Behavioural Assessment and Research in Case of Emergency . An Evaluation of Experiments in Virtual Reality ." In *SimTecT 2009 Simulation - Concepts, Capability and Technology*. Adelaide, Australia.
- Kuligowski, E. D. 2008 Modeling Human Behavior during Building Fires. NIST Technical Note 1619.
- Kuligowski, Erica D., Richard Peacock, and Bryan L. Hoskins. 2010. "A Review of Building Evacuation Models, 2nd Edition". Technical Note 1680. Secretary. NIST.
- Kuligowski, E. D. 2011. Predicting Human Behavior During Fires. Fire Technology (November 13).
- Pauls, J. 1995. Movement of People. In *The SFPE Handbook of Fire Protection Engineering*, Second edition, 3–263; 3–285. Society of Fire Protection Engineers, Bethesda, MD.
- Qingge, J., Can G. 2007. Simulating Crowd Evacuation with a Leader-Follower Model. *IJCSES International Journal of Computer Sciences and Engineering Systems*, Vol.1, No.4, October 2007.
- Ribeiro, J., Almeida, J.E., Rossetti, R.J.F., Coelho, A., Coelho, A.L.. 2012a. "Using Serious Games to Train Evacuation Behaviour." In *CISTI 2012 - 7ª Conferencia Ibérica de Sistemas Y Tecnologías de Información*, 771–776. Madrid, España. <http://www.aisti.eu/cisti2012>.
- Ribeiro, J., Almeida, J.E., Rossetti, R.J.F., Coelho, A., Coelho, A.L. 2012b. Towards a Serious Games Evacuation Simulator. In *26th European Conference on Modelling and Simulation ECMS 2012*, ed. Klaus G. Troitzch, Michael Möhring, and Ulf Lotzmann, 697–702. Koblenz, Germany.

- Ronchi, E., Paul a. Reneke, and Peacock, R. D. 2013. "A Method for the Analysis of Behavioural Uncertainty in Evacuation Modelling." *Fire Technology* (2) (July 9).
- Rossetti, R.J.F., Almeida, J.E., Kokkinogenis, Z., Gonçalves, J. 2013. "Playing Transportation Seriously: Applications of Serious Games to Artificial Transportation Systems." *IEEE Intelligent Systems* 28 (4): 107–112.
- Russell, S.J., and Norvig, P. 2010. *Artificial Intelligence: A Modern Approach*. Edited by Editors PRENTICE HALL SERIES IN ARTIFICIAL INTELLIGENCE Stuart Russell and Peter Norvig. Annals of Physics. Third Edit. Vol. 54. MIT Press.
- Santos, G., Aguirre, BE. 2004. "A Critical Review of Emergency Evacuation Simulation Models." In *NIST Workshop on Building Occupant Movement during Fire Emergencies*, (1032) 25-30. Disaster Research Center.
- SFPE. 2002. *SFPE Handbook of Fire Protection Engineering*. Society of Fire Protection Engineers, NFPA, Quincy, MA.
- Silva, J. F. M., Almeida, J. E., Pereira, A., Rossetti, R.J.F., Coelho, A.L. 2013a. Preliminary Experiments with EVA - Serious Games Virtual Fire Drill Simulator. In *27th European Conference on Modelling and Simulation (ECMS 2013)*. Ålesund, Norway.
- Silva, J. F. M., Almeida, J. E., Rossetti, R.J.F., Coelho, A.L. 2013b. "Gamifying Evacuation Drills." In *Third Iberian Workshop on Serious Games and Meaningful Play (SGaMePlay 2013)*. Lisbon, Portugal.
- Silva, J. F. M., Almeida, J. E., Rossetti, R.J.F., Coelho, A.L. 2013c. "A Serious Games for EVAcuation Training." In *IEEE 2nd International Conference on Serious Games and Applications for Health (SeGAH 2013)*. Vilamoura, Algarve, Portugal.
- Smith, S. P., and Trenholme, D. 2009. "Rapid Prototyping a Virtual Fire Drill Environment Using Computer Game Technology." *Fire Safety Journal* 44 (4) (May): 559–569.
- Timmermans, H. 2009. *Pedestrian Behavior: Models, Data Collection and Applications*. Emerald Group Publishing Limited.
- Vasconcelos, G., Petry, M., Almeida, J. E., Pereira, A., Rossetti, R.J.F., Coelho, A.L.. 2012. "Using UWB for Human Trajectory Extraction." In *24th European Modeling & Simulation Symposium - EMSS 2012*. Vienna, Austria.
- Wooldridge, M. 2002. *An Introduction to Multiagent Systems*. John Wiley & Sons, Ltd.
- Xie, H. 2011. *Investigation into the Interaction of People with Signage Systems and Its Implementation within Evacuation Models*. PhD thesis, University of Greenwich.

## AUTHORS BIOGRAPHY

**João Emílio Almeida** holds a BSc in Informatics (1988), and a MSc in Fire Safety Engineering (2008). He is currently a researcher at LIACC, reading for a PhD in Informatics Engineering at the Faculty of Engineering, University of Porto, Portugal. He has co-authored many fire safety projects for complex buildings such as schools, hospitals and shopping centres. His areas of interest include Serious Games, Artificial Intelligence, and Multi-Agent Systems; more specifically he is interested in validation methodologies

for pedestrian and social simulation models. His e-mail is [joao.emilio.almeida@fe.up.pt](mailto:joao.emilio.almeida@fe.up.pt).

**Rosaldo J. F. Rossetti** is a senior research fellow in the Artificial Intelligence and Computer Science Laboratory (LIACC) and an assistant professor with the Department of Informatics Engineering, University of Porto, Portugal. Dr. Rossetti is a member of the Steering Committee of the IEEE Smart Cities initiative and Chair of the Technical Activities sub-committee on Artificial Transportation Systems and Simulation of IEEE ITS Society. His areas of interest include Artificial Intelligence and agent-based modelling and simulation for the analysis and optimisation of complex systems. His e-mail is [rossetti@fe.up.pt](mailto:rossetti@fe.up.pt) and his Web page can be found at <http://www.fe.up.pt.com/~rossetti/>.

**Brígida Mónica Faria** holds a PhD in Informatics from University of Aveiro, is a researcher at LIACC and lecturer at Polytechnic Institute of Porto (ESTSP-IPP). She is also a collaborator researcher at IEETA, University of Aveiro. She can be reached by e-mail at: [btf@estsp.ipp.pt](mailto:btf@estsp.ipp.pt).

**João Tiago Pinheiro Neto Jacob** holds a MSc in Informatics and Computing Engineering (2011) from Faculty of Engineering, University of Porto, Portugal. He is currently a researcher at INESC reading for a PhD in Informatics Engineering at the Faculty of Engineering, University of Porto, Portugal. He specialised in Digital Games Development, Mobile Computing and Simulation. He can be reached by e-mail at: [joao.jacob@fe.up.pt](mailto:joao.jacob@fe.up.pt).

**António Leça Coelho** holds both the Electrotechnical and Civil Engineering degrees, as well as a Master's and PhD in Civil Engineering. He is currently a principal researcher with habilitation at LNEC. His areas of interest include fire safety and risk analysis. He can be reached by e-mail at [alcoelho@lnec.pt](mailto:alcoelho@lnec.pt).



# GPRS NETWORK MONITORING OF SOLAR ENERGY GENERATION AND SUPPLY IN RURAL RESIDENCES

Robson da Cunha Santos<sup>(a,b)</sup>, Gerson Gomes da Cunha<sup>(c)</sup>, Marcos Antônio Cruz Moreira<sup>(a)</sup>, Mônica Castelo Branco<sup>(b)</sup>, David Douglas Nunes Oliveira<sup>(a)</sup>, Matheus Muzitano Reis<sup>(a)</sup>

<sup>(a)</sup>Fluminense Federal Institute, Campus Macaé, Automation and Control Engineering  
164 km Amaral Peixoto Road, Brazil

<sup>(b)</sup>Estácio de Sá University, General Alfredo Bruno Gomes Martins Highway, s/n -Braga - Cabo Frio / RJ, Brazil

<sup>(c)</sup>Federal University of Rio de Janeiro, Alberto Luiz Coimbra Institute Graduate Engineering and Research, Civil Engineering Program/COPPE/UFRJ

<sup>(a)</sup>[profrobsons@yahoo.com.br](mailto:profrobsons@yahoo.com.br), <sup>(b)</sup>[gerson@lamce.ufrj.br](mailto:gerson@lamce.ufrj.br) mail, <sup>(c)</sup>[macruz@iff.edu.br](mailto:macruz@iff.edu.br), <sup>(d)</sup>[xxmoninfante@yahoo.com.br](mailto:xxmoninfante@yahoo.com.br)  
<sup>(e)</sup>[daviddouglas@gmail.com](mailto:daviddouglas@gmail.com), <sup>(f)</sup>[fennattcf@hotmail.com](mailto:fennattcf@hotmail.com)

## ABSTRACT

The paper describes a system for solar power generation, integrated into a "grid-tie" inverter, converting Direct Current "DC", generated by the plates in Alternating Current "AC" on the standards of electrical distribution network. The energy generated is quantified by smart metering module that transmits the information in specified periods of time to a microcontroller via GSM modem. The modem provides the measured data on the internet, using networks and cellular antennas. The monitoring, fault detection and maintenance are performed by a supervisory station. Employed board types, best inverter selection and studies about control equipment and devices have been described. The article covers and explores the global trend of implementing smart distribution electrical energy networks and the incentive to use solar renewable energy. There is the possibility of the excess energy produced by the system be purchased by the local power utility.

Keywords: Smart Grid, Microcontroller, Supervisory, Solar Energy, GSM.

## 1. INTRODUCTION

In Brazil, concessionaires of electricity distribution provide their customers a fully manual system that does not offer the customer control of its consumption, nor is there an opportunity of interacting with the current system. Cited by many researchers and checked in CRESESB (Reference Center for Solar Energy and Wind Sérgio Salvo de Brito) site, one observes that solar power has not yet reached a level of economic competitiveness with conventional energy sources, due to a number of factors, including lack of incentives (tax reduction) and low efficiency of solar panels marketed today (Pinho, 2014). Despite these factors, one cannot consider using these panels merely as an alternative energy source, but also as a technically feasible option.

For it is possible to have a return on investment in the long term by reducing pollution through the steep fall in emissions carbon and eliminating the need of flooding deforestation of large areas such as the hydro entails.

The research is justified for developing an automated system, which enables the consumer no longer being just a passive agent of the generation and transmission of electric power business, bringing the possibility of him/her being in charge of its central micro generator installed at his residence.

Installing a variable frequency drive connected to the power grid causes the consumer to make available all mains power generated during the day. This energy is born by the concessionaire for distribution to other consumers and overnight the consumer - that generated energy during the day - shall take back the energy he generated during the day. The energy meters installed in rural homes works amounting consumption or regressing consumption if the unit is power generating.

At the end of the month the negative (energy generated greater than consumed) or positive (energy consumed greater than that generated) balance occurs. To current Brazilian legislation, in case of negative balance, nothing will be paid by the consumer, but it is now under study for some years adjusting the law so that the consumer can receive for its energy generated more than consumed.

The installation of the new Grid Tie inverter technology eliminates the need for installing batteries, which have a high cost and a low life (about five years). In older versions of solar panels installing the battery bank accounted for nearly fifty percent of the total investment. Regarding modern data acquisition systems, portability and mobility are essential for most applications. In this sense, it has become essential to add internet connection to the new and modern data acquisition systems. In view of this new perspective, this paper proposes an acquisition data system based on cutting-edge and innovative technology inside current

marketplace. This includes data capture using microcontroller and data transmission via cellular, specifically using the GPRS / Internet technology. These data are checked from a supervisory system and this enables making decisions for the case of low efficiency of the system and to check for fraud or intervention required for maintenance.

## 2. DESIGN PARTS

Power generation through solar panels is somewhat already well consolidated around the world. However, there is still much room for modernization of systems, as more modern ones are created, and use becomes increasingly common due to the continuous increase in system efficiency.

### 2.1. Solar panels

The solar panel can be defined as a device designed to convert solar radiation into energy. For relying solely on the sun, the most abundant energy source on the planet, it is the cleanest method known of power generation (Sukhatme, 2008). The Photovoltaic Solar Panels are associations of photovoltaic cells that convert sunlight into electricity. Most of these solar cells are made of silicon or gallium arsenide, and present dark color, for greater capture of light energy. The silicon solar cells produce less energy compared to gallium, but their cost is proportionately lower.



Figure 1: Solar Panels (Pinho,2004)

In the previous figure are shown two examples of solar panels, the difference between them is visible once the left figure, it is polycrystalline panel and the right figure is in a monocrystalline panel, the difference is in the manufacturing process (Pinho, 2014).

### 2.2. Inversor Grid Tie

Before the development of Grid Tie Inverters (grid connected) there were only standalone inverters (Off-Grid Inverter) that are not connected to the power grid and were designed to create an AC voltage from DC power stored in batteries. The later ones are used in autonomous systems such as detached houses, boats, and UPS systems (emergency power supply). These devices are not simultaneously connected to the mains power supply grid. As seen in the following figure, one notices that the load has no connection to the power network; its energy demand is supplied exclusively by the battery.

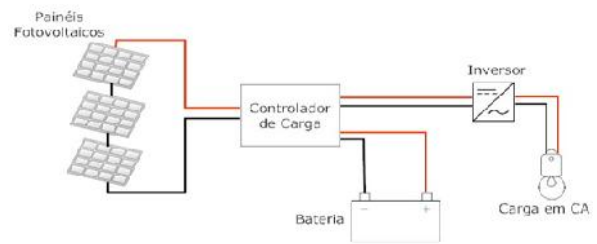


Figure 2: Stand-alone photovoltaic system (the author)

The grid-injection inverters (Grid Tie Inverter) create an AC voltage directly from the Photovoltaic panels; without using batteries. They inject energy into the grid, thus allowing trading of electricity. The main difference between a conventional inverter (Off-Grid inverter) and Grid-Tie inverter is that the latter is able to interconnect with the network utility, due to its ability to synchronize its frequency and its output voltage to the network in which to connect. They are also able to break away from the utility grid when it does not supply power. The panels can be connected directly to the inverter. There is no spending on batteries and charge controllers.

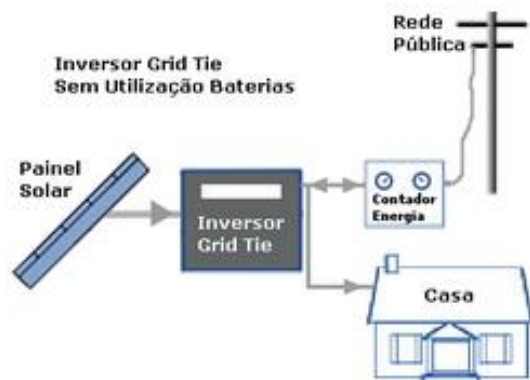


Figure 3: Grid Tie Inverter Connection (the author)

### 2.3. Smart Meter

The meter from manufacturer Elster A102C model displayed in the following figure provides an economical solution for domestic applications. The meter provides a communication port IRDA (Infra Red Data Aquisition), optical output for reading different information on the meter operation. In the liquid crystal display one may view the total energy measured in kWh.

This can be configured for single-phase measurement of direct or direct/reverse energy both for residential and small commercial / industrial users. Furthermore, it provides the measure of active and reactive power. The meter has an output that can provide signals for power controllers, demand controllers or remote reading systems.

The meter from manufacturer Elster A102C model displayed in the following figure provides an economical solution for domestic applications.





Figure 4: Elster A102C meter (Elster,2012)

#### 2.4. Plataforma Arduino

The Arduino project was created in Italy in 2005 with the goal of providing a low cost, easy using platform for electronics prototyping, to be used by anyone interested in creating interactive desings with objects and their environment. (Banzi,2011).

The Arduino electronics board contains multiple inputs and outputs, analog and digital, and serial interface via USB connection to communicate with the computer. The smart card is an element of the AVR microcontroller family that allows thousands recordings and rewrites in its program memory.

In order to increase the functionality of the Arduino board, several companies have developed hardware for connecting additional electronic boards on the Arduino terminal. These additional electronic boards are called "Shields" and add various functions specific to the Arduino. In this article will be described the GSM Shield.

The Arduino Mega 2560, displayed in the following figure, is a microcontroller board based on the ATmega2560 microcontroller. It has 54 pins for digital inputs / outputs, 16 analog inputs, 4 UARTs (hardware serial ports), a crystal oscillator of 16 MHz , a USB connection, an input power, an ICSP connection and a reset button.



Figure 5: Arduino Mega2560 (Banzi,2012)

#### 2.5. GSM Shield

The card with GSM module SIM900 displayed in figure 6 is a board that can be easily attached on the Mega2560 Arduino board, extending the functionality of the GSM module to the resulting assembly. This board is stacked on the Arduino board, through all its pins. The GSM module SIM900 board uses this

structure to its power feeding and functioning. (Banzi, 2012).

The SIM900 GSM module, produced by SIM Com, is responsible for communication with the GSM network making possible the effecting and receiving phone calls and text messages.

The modem is a device for wireless communication. This equipment enters a new trend of technology known as M2M (Machine to Machine) in which this project is based.

GSM modems can operate even in voice mode, but their greatest uses are for data transmission that occurs through three types of technology: CSD Connection (Circuit Switch Data); GPRS Connection (General Packet Radio System) and SMS (Short Message Service).



Figure 6: SIM900 GSM/GPRS (IteadStudio, 2012)

#### 2.6. Supervisory and Data Acquisition System

The supervisory and data acquisition systems (SCADA) were created due to the need for an interface that was operator friendly to plant control. Are also known as Human Machine Interface (HMI) or 'Supervisory' only.

The main objective of the Supervisory system is to provide a simple way to monitor, control and manage multiple points of an automated plant. The data that will be processed by the SCADA can be obtained through a connection with a controlling element, equipped with sensors or even the data is entered directly by the operator.

### 3. ASPECTS OF INTERNET-INTEGRATED WIRELESS COMMUNICATION

Since the creation of the telegraph, through the creation of radio communication, reaching the creation of the phone until the creation of the cell phone, much has evolved in terms of point to point communication. However, the major event of the last century was the development of wireless and mobile communication, like the cell phone. From this, it comes the need and wish to improve the cell phone. The GSM network which was created and later enhanced to data

communication, as previously it was possible only voice communication.

### 3.1. Historical of GSM Technology

The data transmission via mobile phone has become an innovative solution for the incorporation of new values in technology, which allows to add a quality hitherto existing services in fixed-line telephony service with the mobility factor, the main distinguishing point in cellular telephony. Added to this, there is now a full integration between the mobile phone and the Web network, which forms a vast technological potential application (Wirelessbrasil, 2012).

### 3.2. GSM Network Operation

According to the schematic of Figure 7, the mobile stations (MS), that may be mobile (and portable) phones or mobile units installed in cars, talk with the Base Station System (BSS) through the air interface of Radio Frequency (RF). The Base Station System (BSS) consists of a Base Transceiver Station (BTS) and Base Station Controller (BSC). It is common for several BTS be located in one same location, creating two to four sectored cells surrounding a common antenna tower.

The BSC are often linked to the BTS by microwave links. The link to the BTS BSC is called the Abis interface. Typically, 20 to 30 BTS are controlled by a BSC. In turn, several BSS are subject to a Central Switch and Control (MSC), which controls the traffic between several different cells. Each Central Switching and Control (MSC) will have a Visitor Location Register (VLR) in which the mobile units that are out of the cells of their local area will be listed, so that the network knows where to find them. The MSC is also connected to the Location Register of Local Mobile Unit (HLR), the Authentication Center (AUC) and the Equipment Identity Register (EIR), so that the system can verify that users are subscribers and equipment legal situation. This helps preventing the use of fraudulent or stolen mobile units. There are also facilities within the system to the Operations and Maintenance Organizations and Network Management (NMC).

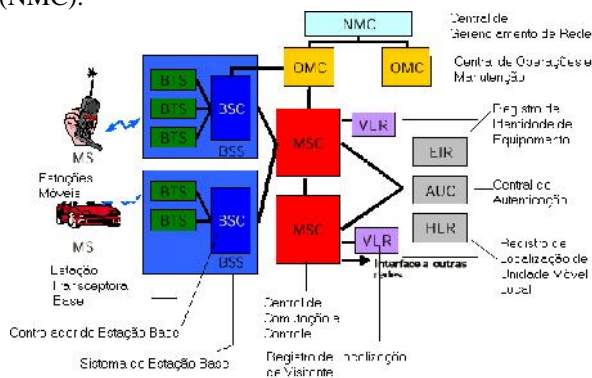


Figure 7: GSM Network Frame (Wirelessbrasil, 2012)

## 4. SOLAR ENERGY AS A SOURCE OF ELECTRICITY

Solar energy as an energy source will not have the ability to substitute other energy sources used today because its generation is seasonal not allowing a continuous feed of large systems. However, its popularity, with the implementation of new systems around the world, will make it complement other sources of generation, thus reducing the environmental impact generated by other energy sources

### 4.1. Historical of Photovoltaic Solar Energy

As a definition, solar photovoltaic energy is the energy coming from the direct conversion of light into electrical energy. In 1839, this phenomenon was first described by French physicist Edmond Becquerel, when reporting the properties of selenium, which produced an electrical current directly proportional to the incident radiation.

The subject began to receive a great attention from the '60s, during the Cold War, as, despite its high cost, this technology proved suitable to meet the energy demands in aerospace missions as well as for supplying satellites. Government programs in countries like Germany, Spain and Japan stimulated domestic application of this technology, which allowed for a production scale, further reducing costs.

Nowadays photovoltaic cells are quite popular, produced with thin films. They use even less material and energy in their manufacturing process than the polycrystalline silicon, which justifies its price more affordable (Serrão, 2010).

### 4.2. Environmental Impacts

It can be said that the environmental impact is one of the most relevant points as it comes to solar energy. The most significant environmental impacts occur at the production, construction and decommissioning of photovoltaic cells.

In constructing these plates, various types of environmental and health hazardous materials are used, besides being consumed a large amount of energy, which is related to the emission of air pollutants such as greenhouse gases. Compared with a conventional thermal power plant, solar power generation emits 20% less CO<sub>2</sub> for the same amount of energy produced (Naturlink, 2012).

### 4.3. Current Feasibility in Brazil

According to a study conducted by the Energy Research Company (EPE), residential solar power production is now economically viable for 15% of Brazilian households.

Under the new resolutions of the National Electric Energy Agency (ANEEL), published in 2012, customers who possess solar panels in their homes can not only reduce the consumption of electricity purchased from concessionaire but also generate credits for use in other months.

Under this system, the generating unit installed in a residence, for example, will produce power and what is not consumed will be injected into the distribution system, which uses the credit to subdue the consumption of subsequent months. The credits can be used for a period of 36 months and information will be provided for the consumer, so that one knows the balance of power and have control over his invoice.

In 2012 there were only eight projects of this type and some residences distributed, generating approximately 15MW of a total of 430,000 MW consumed by Brazil (BRAZIL, 2012).

The country has the option of waiting for the cost of solar power to decline to put it on auctions or to create a specific auction so there is no dispute with other cheaper sources such as wind. But that would be sold a small amount of energy not to burden the consumer (Tolmasquim, 2012)

## 5. SISTEM DESIGN

System integration was the most laborious part of the project, due to the high level of detail required for its integration. It required a high volume of surveys of sites of companies and components manufacturers.

### 5.1. Motivation

Initially, the goal was to create a monitoring system for some kind of alternative energy, to be deployed by a company that would provide installation and maintenance of equipment used in the generation of a residential customer.

The methods to achieve these results were studied and analyzed to choose the technologies which fitted best the goal. The decisions were taken at the beginning involved the following circumstances:

- Renewable Energy Generation Method: Solar or Wind Power.
- Harnessing System: Battery bank, Direct Grid Injection or Hybrid System.
- Measurement of Power Generated: Conventional Electronic Meter or Analog Voltage and Current Readers.
- Process Controller: PIC, PLC or Arduino.
- Monitoring System Communication Interface: Wi-fi, GPRS or SMS.
- Monitoring System: Webserver, Elipse ou Intouch.

### 5.2. Research and Choices

The initial stages of project development involved long research time and studies for decision making. The primary research methodology was accessing several sites on the subject, in addition to performing some technical visits, which had great influence in the paths to be followed in the project.

### 5.2.1. Renewable Energy Generation Option

To decide which method of power generation would be used, several studies were conducted to understand the advantages and disadvantages of each alternative energy source.

The price for wind power generation is significantly cheaper than the solar power generation although the techniques for wind energy are well developed and should not evolve much in the short term.

Solar power is still quite expensive to be produced, but the growing trend in production technologies is that due to government incentives these energy prices will fall enough and should equal the wind energy prices shortly. This point has greatly influenced the choice as one of the goals was to have a vision of the future solar power market and proved to be the wisest choice.

Another point that influenced the choice of solar energy was that the blades of the wind turbine have an impact on the local birds, especially due to the shock of these birds in blades and unknown effects on modifying their usual migration behaviors. Furthermore the noise impact caused by wind power generators that may reach a constant noise up to 43 dB, which can disturb the inhabitants of the residence up to 300m distances.

Then photovoltaic panels were chosen to generate alternative energy, harnessing solar energy. Two modules KD54 plates of 50Watts each were acquired from Kyocera, the most trusted brand and market leader in Brazil.

### 5.2.2. Harnessing System

At first the idea was to mount a hybrid system to harness the energy generated. Throughout the research it was noticed that there was no a need to use a battery bank, except in cases where there is no electricity grid next the residence. In the Lakes Region, location of the tests, almost all residences have access to the electricity grid, confirming the dispensability of a system with battery bank. Thus, all the energy generated by the system will be injected directly into the grid of the power supply company (concessionaire) and slaughtered the amount that has consumed the dealership.

The equipment was chosen for this Inverter Grid Tie400W of GreenEnergyStart, which is a company focusing on alternative energy and use of current technologies.

### 5.2.3. Measurement of Energy Generated

At the beginning it was thought to use voltage and current readers in order to reach the energy generated because knowing the voltage and current power which is the product of these two quantities could be calculated.

The main methods found to construct a current player were the utilization of high power resistor and precision and very low resistance, which are known as shunt resistors and the utilization of Hall effect sensors,

which are analog signal proportional to the magnetic field created by a current that flows close to the sensor.

The solution was to acquire a meter certified by INMETRO, a trust company in the Brazilian market. Through contact with the supplier decided to buy an Electronic meter company Elster (A102C).

#### 5.2.4. Process Controller

In any automated system, and monitoring, there is a necessity of a device that scan the sensors and treat the data according to the need.

One of the options for this type of process is the Programmable Logic Controller (PLC), but the factors that have eliminated the PLC project were its high cost and its high power consumption when in operation.

It was then decided by the use of a microcontroller, by owning a low cost and power consumption is extremely low.

The plate was chosen for the Arduino MEGA2560, which has more than 60 I / O and communication is based on an Atmel 2560 processor.

#### 5.2.5. Means of Communication with the Monitoring System

Currently, there is a tendency in industries increasingly frequent use of wireless technologies for communication. Taking advantage of the growth of wireless networks, it was decided to this method to make communication between field elements and the monitoring system.

GSM networks are present in almost all places in Brazil, thanks to the "boom" of the cell and increased technical capacity of operators. In the Lakes Region almost everywhere already have access to this technology. So GSM networks has become a great option for use in this project, as it currently has lower costs for sending packets through GPRS connection and sending short messages (SMS).

For these purposes, was chosen GSM Shield V1.1 Icomsat company because its integration with Arduino is fast and easy, besides it be based on SIM900D which is a Quad Band GSM SIMCOM chip company.

#### 5.2.6. Monitoring System

For the choice of supervisory system, two softwares, the Intouch and the Ellipse were considered and soon consolidated in the market to attend the need of the project.

The option was for the supervisory of the Ellipse, for being a Brazilian software and with an excellent technical support by phone, which could be useful, not to mention its lower price.

A Webserver is a complement to supervisory, because it can be accessed by mobile phones, tablets and other portable devices currently used.

The next step was the choice of communication protocol. Researching quite found out that the best solution was the Modbus TCP protocol, which is free from any type of recovery and is not complicated to use.

### 5.3. The Project Design

The project design was divided into parts, each part consisting of a different system design.

#### 5.3.1. Mounting the System Generator

The project was started with the mounting of the solar power generation system. The first step was the determination of the best place for installation of solar panels. For added security, The location must be high, to prevent theft and animal access and should not be obstructions from sun rays (shade).

Then there was the determination of a position and an inclination to the plate, so that better take advantage of solar cycles. As the location of this project is the Southern Hemisphere, it is necessary to position the plate towards the north. To determine the ideal inclination, it is necessary to verify the latitude of the installation location and the approximate formula for calculating the inclination is that this should be 130% of latitude. In Rio de Janeiro, the approximate latitude of 23° is in this way allows an inclination of 30 ° on each plate.

The binding plates were performed in parallel, which gives a sum of currents of each plate, and maintains a voltage equal to a single plate. The following figure shows the installation and positioning of the plates.

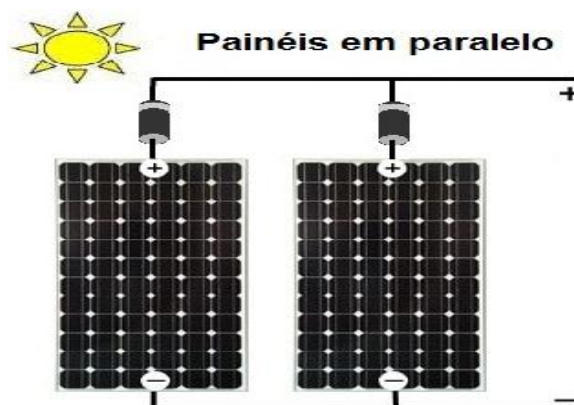


Figure 8: Panels in parallel (Mpptsolar 2012)

After the connection of the plates, the connection of the Grid-tie was made. The positive output terminal of the plates was connected at the positive input of the CC inverter and also the negative output terminal was connected to the negative CC input inverter. Soon after, was connected to the AC output of the inverter in the network, by transforming all input power direct current into alternating current.

#### 5.3.2. Installation of the Measuring System

With the generation system installed and working, the meter Elster A102C was installed between the inverter and the electric grid. The meter is installed in series with the inverter and the grid. The determined meter has 3 methods to inform the measured data. The output in



standard IRDA, which works through infrared, pulse output, is available through the auxiliary outputs, and an LCD display on the meter itself.



Figure 9: Padrão da Mensagem IRDA (Meterspec)

As the meter uses the LED emitting red light to send serial devices to reading information then was acquired a phototransistor infrared to decode the information in the standard meter IRDA. Each bit of zero value the meter wants to send, it sends a pulse of infrared LED emitter in Figure 9.

### 5.3.3. Controller

The monitoring system starts with the treatment of information measure and this function is realized by Arduino MEGA 2560.

The IRDA output was initially used to receive the data, but it was noticed that to be a serial message, the difficulty to decode the data would be much greater and more imprecise, since any loss of pulse would cause an incorrect reading. The pulse output of the meter was then used to obtain the required data.

Using a programming library that transforms the Arduino in a Modbus device. The pulse output is connected on the Arduino interrupt pin and when it detects a pulse signal, the Arduino registers and saves this information. Powering the Arduino is done by an external source or then connected directly at the output of the plate solar.

### 5.3.4. GSM communication

The Shield GSM has been connected in the Arduino board, as shown in Figure 10 and inserted a SIM card from a wireless carrier.

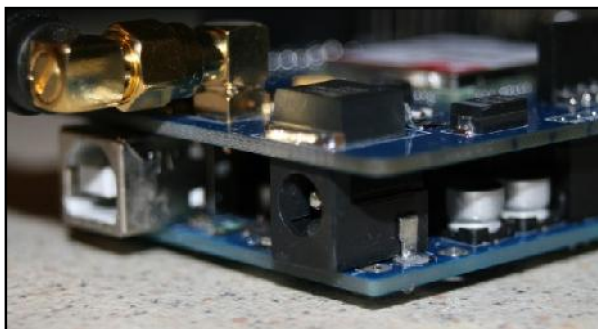


Figure 10: Mounting the GSM Shield with the Arduino (the author)

In programming, GSM is functioning as a transparent device, ie, it transmits any message that arrives at your receptor. Programming the GSM plate is made by Arduino.

The function of GSM Shield is connect to the internet through GPRS and connect to a server through port 502, which is the default port for Modbus.

Then, when the Arduino sends a message to the GSM, it transmits directly to the IP defined through the door 502. This server is the supervisory station.

### 5.3.5. Supervisory station

In this step, the process of creating screens for user interface was started. There are two options for monitoring, one for the company that provides the system and one for the user who purchases the system.

Both are password protected and the user is only able to access the monitoring of your own residence. The company operator can access the monitoring of all installed points.

The first screen, known as the Home screen, displays the logo of the company supplying of the system described in this project, the version of supervisory and rights to creators, plus a welcome message and option to start the program. Clicking the named "About" button several information about the program, such as licenses and rights, information about the current version and contact is displayed, as shown in Figure 11.

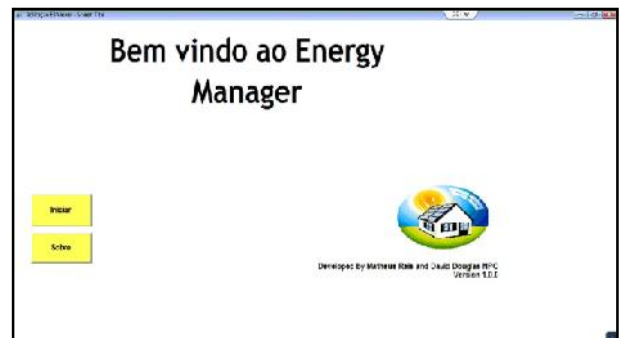


Figure 11: Home screen the supervisory (the author)

In the version of the user, after the program is started, a screen is displayed for this user in case there is more than one installation point for him, so that he can select the point that he wants to see, that is to say, the information of the installed system.

In the version of the operator, a screen is displayed where you can choose, with one click, the city to be monitored. After it is performed by the operator, he selects the point at that he wants to view the data received by the generation system of the selected location. If any the system is producing below the normal standard, the place is displayed with a yellow circle. If the system is not producing, a red circle is displayed, as shown in Figure 12.



Figure 12: Map of the city of Macaé in the highlighted points (the author)

In Figure 13, it may be perceived a common screen to the user and the operator. It shows the data processed by the Arduino, as the quantity of energy generated at the moment and in the last 12 hours.

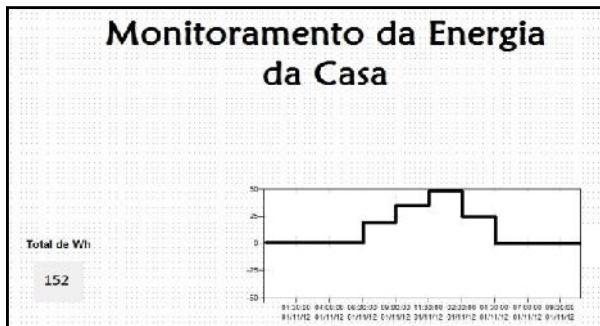


Figure 13: Screen Supervisory Control (the author)

The communication driver that was used in the supervisory is Modbus and was configured with parameters of identification of Arduino. A tag communication was created to store the information obtained by Arduino through register read function. The Ellipse performs in the range of 1 minute, reading the register that stores the information from the meter and displays on screens previously determined. The history is done by storing in a internal database values and the tag after displaying the chart.

A low alarm value has been placed in this tag communication and time associated with a solar shift, ie, in a time of solar generation a certain house is not generating a signal is displayed on the selection screen of points to be monitored and because there is a problem with the system and if the alarm persists a technical should go to the local to do the verification.

## 6. RESULTS OBTAINED

After adjustment of the positioning experiments the plates were made for sampling results on the relevance of the project. Sampling and analyzing the total energy generated per day in the second fortnight of November 2012, there was obtained an average of 240Wh per day for each plate of 50W of power, as shown in Table 1.

Each plate 50W Kyocera, used in the project, costing on average R\$500.00 and each inverter 400W

R\$600.00 costs on average and supports up to 8 plates 50W.

Table 1: Total generation Nov./2012 second fortnight

| Month of November 2012 |                          |                      |
|------------------------|--------------------------|----------------------|
| PLATE OF 50W - KYOCERA |                          |                      |
| Day                    | Total of Generation (Wh) | Condition of the Day |
| 16                     | 212                      | Cloudy               |
| 17                     | 281                      | Sun                  |
| 18                     | 231                      | Sun / Cloudy         |
| 19                     | 241                      | Sun / Cloudy         |
| 20                     | 175                      | Cloudy               |
| 21                     | 149                      | Cloudy               |
| 22                     | 206                      | Cloudy               |
| 23                     | 275                      | Sun                  |
| 24                     | 270                      | Sun                  |
| 25                     | 260                      | Sun                  |
| 26                     | 216                      | Sun / Cloudy         |
| 27                     | 225                      | Sun / Cloudy         |
| 28                     | 303                      | Sun                  |
| 29                     | 301                      | Sun                  |
| 30                     | 256                      | Sun                  |
| Total                  | 3601                     |                      |
| Average                | 240,07                   |                      |

The tariff of local dealership in the region is 0.66 R\$ / kWh. If the average is maintained throughout the month will be achieved approximately 7.2 kWh per month for each plate installed 50W. Considering the maximum charging of the inverter, which are eight plates is 57.6 kWh per month.

This value generated by eight plates 50W, means an economy in R\$ 38.02 per month for the consumer.

Therefore, each system with eight plates Kyocera 50W and 400W inverter GreenEnergy costs an average of R\$4,600.00. Following this reasoning, in 132 months or 11 years, the savings would pay the amount originally spent. Considering that the life of the plates is 30 years, the results are acceptable.

## 7. CONCLUSION

Based on this project remote monitoring it is possible to glimpse many other alternatives for use of the system by modifying and adapting specific points to the desired need.

In this way, it was possible to view part of a worldwide trend that is M2M (Machine to Machine), ie establish communication Machine to machine without human interference, and the solution that makes it possible multiple mobile machines and equipment be monitored remotely.

By means of M2M, it is possible that all the information collected by the monitoring modules are transmitted through a data network, be it wi-fi, ethernet, RF, Zigbee, GPRS, 3G, LTE, PLC and etc..

Analyzing the possibilities of using of using M2M opens up an enormous range. M2M applications involve different fields, such as fleet management (public safety, public transportation, taxis, delivery vehicles); utilities (SmartMeters - electricity meters, gas, water, and Telemetry - telemetry sensors and controllers,



navigation systems, vehicle maintenance, passenger safety and anti-theft systems), electronics (cameras and cameras, video games, ebook readers ) and others such as monitoring patients requiring constant monitoring, ATMs and self-service machines (credit card).

Thus, developing a project using newly discovered resources and on the rise causes the project to be very gratifying. It is estimated that the M2M market in Brazil should have a rate higher than the rest of the telecommunications market growth, averaging approximately 3.7% of the subscription market in 2016.

The possibility of developing new grouped the generation of clean energy technologies and add generation connected directly to the electrical grid, something that is also innovative because until a few years ago there was not the technology viable Grid Tie Inverter for minigeneration and microgeneration.

Therefore it is very satisfying to develop a project that contributes to sustainability, something the modern world will need to enhance with intense speed so that future generations can live in a nice planet as we live today.

There is a need of investment in rural residences which are not completely isolated and without opportunities for growth so that people could know a new world and approach the ones who live in large urban centers.

## REFERENCES

- ANEEL (National Electric Energy Agency). Resolution 482. Available in [www.aneel.gov.br/cedoc/ren2012482.pdf](http://www.aneel.gov.br/cedoc/ren2012482.pdf) accessed 01.nov.2012.
- Arduino. Available in <http://www.arduino.cc/> accessed 07.set.2012
- Banzi, M., "Getting Started with Arduino" book, Vol, No 1., December/2011, NovaTec.
- BRAZIL. Government Information. available in <http://www.brasil.gov.br/noticias/arquivos/2012/07/03/energia-solar-ja-e-economicamente-viavel-para-15-dos-lares-brasileiros-diz-epe> Accessed 22.out.2012
- BRAZIL, Solar - Photovoltaic Energy. Available in <http://solarbrasil.com.br> accessed 02.nov.2012.
- CRESESB (Reference Centre for Solar Energy and Wind Sérgio de Brito Saved). Studies of Solar Energy. Available in <http://www.cresesb.cepel.br> accessed 01.nov.2012
- Elster. Measurement systems. Available in [http://energia.elster.com.br/pt/eletronicos\\_monofasicos.html](http://energia.elster.com.br/pt/eletronicos_monofasicos.html) accessed 01.nov.2012.
- Energy, Solar - Photovoltaic Energy. Available in <http://www.solarenergy.com.br> accessed 02.nov.2012.
- FirstSolar. Solar Panels. available in <http://www.firstsolar.com/Projects/Projects/Projects-Under-Development/Agua-Caliente-Solar-Project/Overview> Accessed 22.nov.2012
- Iteadstudio. GSM Shield. Available in <http://blog.iteadstudio.com/tag/shield/> accessed 09.set.2012
- FSI. TCP / IP Architecture. Available in <http://fsi-ufam.blogspot.com.br/2009/11/modelo-osi.html/> accessed 21.nov.2012
- InfoWester. GSM network. Available in <http://www.infowester.com/2g.php> accessed 02.nov.2012
- Junior, Pedro Adolfo de Souza. Managed Access Control System Via GPRS. 2011. 65f. Completion of course work (undergraduate) - Department of Electrical Engineering, Federal University of Paraná. Available in <http://www.eletrica.ufpr.br/ufpr2/tccs/176.pdf> accessed 02.out.2012.
- Leitão, José Miguel, "Solar Energy: Bet or Necessity," E-topia: Electronic Journal of Utopian Studies, 11 (2010)..
- Meterspec. Elster A102C manual. Available at <http://www.meterspec.com/143.pdf> Accessed 20.nov.2012
- Mpptsolar. Solar Energy. Available in <http://www.mpptsolar.com/pt/paineis-solares-em-serie.html> accessed 02.nov.2012
- Oliveira, Victor Hugo Freitas. Development of a Remote Telemetry System. 2009. 58f. Completion of course work (undergraduate) - Department of Computer Engineering and Automation, Federal University of Rio Grande do Norte. Available in <http://www.engcomp.ufrn.br/publicacoes/TCC-2009-1-1.pdf> accessed 05.out.2012.
- Painelsolares. Solar Panels. Available in <http://painelsolares.com/o-que-e-um-painel-solar/> accessed 01.nov.2012
- Serrão, Marcons Antônio dos Santos. Sizing a Photovoltaic System For A summer house in Paraty Pouso da-Cajaíba. 2010. 89F. Project Completion of course (Undergraduate) - Department of Electrical Engineering, Polytechnic University of Rio de Janeiro. Available in [http://monografias.poli.ufrj.br/monografias/mono\\_poli10000620.pdf](http://monografias.poli.ufrj.br/monografias/mono_poli10000620.pdf) accessed 02.nov.2012.
- Sukhatme, SP, Nayak, JK 2008 Solar Energy - Solar Energy:. Principles of Thermal Collection and Storage.3rd Ed - Tata McGraw-Hill, 125-142.

# APPROACHING DEMAND OF CASH TRANSACTIONS AT BANK BRANCHES

Miguel Aguilar Zaragoza<sup>(a)</sup>, Idalia Flores de la Mota<sup>(b)</sup>

<sup>(a)</sup>National Autonomous University of Mexico

<sup>(b)</sup>National Autonomous University of Mexico

<sup>(a)</sup>[maguilarz@live.com](mailto:maguilarz@live.com), <sup>(b)</sup>[idalia@unam.mx](mailto:idalia@unam.mx)

## ABSTRACT

Trying to maintain a service level in many operations can result very expensive, especially if there are severe penalties due to a low level related with the service level objective. In such cases, we seek to achieve a goal even at the cost of significant loss in the operation efficiency. In particular, bank companies at their branches undertake transactions related to cash money as main service, so, not having available cash becomes a critical issue, which slightest consequence is the bank company prestige deterioration, and not only affectations for a specific branch.

This paper has the purpose of establish an approach for cash transactions demand at bank branches, with perspectives that best fit, which helps to control the risk of *stocking out* of cash, setting up some parameters, e. g. branches' *safety stock* of cash, and projecting horizons of cash balances, given certain scenarios.

The model we present is developed with *frequentist* and *Bayesian* perspectives, the last one is a 'recent' develop that covers in a more efficient way, the *Markov Chain Monte Carlo* (MCMC) simulations. So, by means a real evaluation, it will be estimated the accuracy level that can be got approaching the cash demand with each of these, in particular with the best.

Keywords: bank branches, Compound Process, INLA, Generalized Linear Models, simulation.

## 1. INTRODUCTION

Within their administration, banks have the ongoing task of developing their activities seeking to be efficient, i.e., fulfilling the service (level) at minimal cost. Among the main activities, banks offer services through their teller desks and automated devices that generate cash inflows and outflows in their branches.

The purpose of finding the efficient *cash management* in bank branches means to minimize money-transfer cost, opportunity cost (caused by the money storage in the vaults of the branches),

and the amount exposed to risk of theft, maintaining the operation of this entities with a desired level. All these decisions are better supported with a truthful approach of cash transactions. Since it is the most representative proportion of cash services, and therefore "more predictable", the scope of this article is only local currency, so it does not include foreign currencies.

Thus, the objective is to propose a model that helps us to know about the amount of money we need in every branch in order to facilitate, to whom are in charge of the branch vault, to manage it in an appropriate manner through an efficient handling of orders required to central vaults.

It has been written about the efficient *cash management* in bank branches, such as the paper of Julia Garcia (2013), however, it has the lack of not having completely realistic assumptions, and it could be obtained solutions with unexpected consequences, including negative *stop costs* (transfer costs). Also, some proposals have been published for the management of foreign currencies in bank branches. In his paper, Bell (1984) discusses this issue as one of *inventory management* and results are shown through the use of decision rules, for any branch with a considerable volume of transactions with foreign currencies. Nevertheless, the foreign exchange demand is very singular; furthermore, as it is mentioned, it is not easy to implement these rules. Nonetheless, it is attractive the treatment of the vault stocks as an inventory. Also, proposals have been arisen to determine the optimal stock levels that ensure available cash to cover future expenses of any entity. Nadia Makary (1968) proposes an optimal inventory policy for the case where costs of money storage and lack of it, are convex functions, and in the scenario when the decision to increase or decrease inventory levels is with no fixed cost. All these efforts have addressed their problems giving little attention to model demand with a substantial accuracy. Since we suggest that some parameters and scenarios, for branches, should be set up based on the cash transactions

demand, we need to approach it with more accuracy.

The context of the problem being addressed is described in the following section.

## 2. DESCRIPTION OF THE PROBLEM

The *cash management at bank branches* problem is addressed considering the situations contained in this section.

### 2.1. Context

Branch administrators must handle the cash flow of the vaults, according to the following diagram:

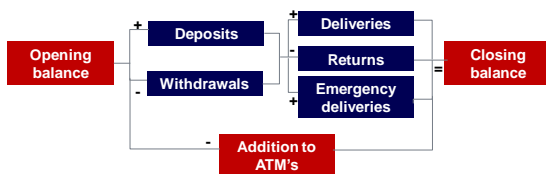


Fig. 1: Cash inflows/ outflows in vaults of bank branches

In Fig. 1 possible movements in the cash stock in a branch are represented. The *Opening balance* is the amount of the stock at the beginning of each day; *Deposits* and *Withdrawals* are the amounts of transactions with users in teller desks; *Deliveries* and *Emergency deliveries* are cash supplies from central vaults to the branch, with different sense of urgency; *Returns* are shipments of cash from the branch to the central vaults; *Addition to ATM's* represents the amount of cash coming out of the branch vault to feed their ATMs. Finally, the *Closing balance* is the amount of the stock at the end of the day, obtained by doing arithmetic operations with the prior concepts.

### 2.2. Variables and parameters

Traditional mechanisms of control for handling cash in banks, with branches that have a high level of discretionary to take decisions related to ordering *deliveries* and *returns* to central vaults, are based on establish bounds to the *maximum number of transfers* and the *minimum amount per transfer*, both for cost purposes; and to the *maximum cash storage*, for security purposes. Assuming that the first two bounds allow an efficient operation, it is easy to see that the third one owns the power to affect the good approach of the other two parameters. Also, it is easy to notice that it could be inconsistency between the cost-purposes and security-purposes bounds. That's why it is important that *maximum number of transfers* and *minimum amount per transfer* are set once established the *maximum cash storage*. Even in our case (that avoids the cost-control parameters), it is

very important to define a suitable maximum level to store cash, that allows efficiency while respecting the security parameter.

Besides the *maximum cash storage*, it is suggested to maintain a *safety stock*, both considering the available information. Consequently, in this paper it is proposed to define this interval (of amounts of money allowed for a branch) based on historical data and the feedback of the branch administrator to determine the lower bound, and the historical data and the feedback of the Risk/Control Office to determine an upper bound. In an analogous manner, other two parameters are going to be defined: the *minimum* and *maximum amount of money in the ATMs*, complementing the historical data with suggestions of the devices vendors.

Of course, *deposits* and *withdrawals* are unknown, but future demand is being estimated with a forecast model grounded in a *Compound Process*. So, the unique variables considered for this model are: number of transactions and transactions amount of money, both for *deposits* and *withdrawals*.

### 2.3. Problem objective

The main idea to solve the problem discussed is to generate an approach for cash transactions demand that allows, not only to predict *deposits* and *withdrawals* in order to ensure the *service level*, but also to establish tactics, setting up the *safety stock* and *maximum cash storage* for branch vaults, and the *minimum* and *maximum amount of money in the ATMs*, that yields more objectivity in the event and amount of cash in transfer orders, with a huge potential for cost savings.

It is going to be presented the best perspective of the model proposed, measuring the accuracy level reached in a considerable number of branches.

## 3. APPROACHING CASH TRANSACTIONS DEMAND

The demand approach, that is the base of our proposal, is being taken from the perspective that better fits of the following: *Model M1* and *Model M2*, described below. Before that, we show a brief of the needed theory.

### 3.1. Compound Process

It is called *Compound Process* to the pair  $\{N(0, t], Y_i\}$ , where  $Y_i$  is the random variable associated with the *i*th occurrence of the counting process. These two variables are independent each other.

It is defined the random sum of a Compound Process as:

$$S(0, t] = \begin{cases} \sum_{i=1}^{N(0, t]} Y_i & \forall N(0, t] \geq 1 \\ 0 & N(0, t] = 0 \end{cases}$$

Depending on the probability distribution of  $N(0, t]$ , the distribution of the random sum and the

Compound Process take their names. That is, if  $N(0, t]$  has a *Negative Binomial* distribution, the distribution of  $S(0, t]$  and the process, are called: *Compound Negative Binomial* and *Compound Negative Binomial Process*, respectively.

In the case of a Compound Poisson Process, the expected value of the random sum is:

$$E(S(0, t]) = \left( \int_0^t \lambda(s) ds \right) * E(Y_i).$$

Variance of the random sum:

$$\text{Var}(S(0, t]) = \left( \int_0^t \lambda(s) ds \right) * [E(Y_i)^2 + \text{Var}(Y_i)].$$

### 3.2 Generalized Linear Models

Generalized Linear Models (GLMs) are a family of models which response variable,  $y$ , may be quantitative or qualitative, assuming it has a distribution function that belongs to the exponential family, i.e., that its density function can be expressed as:

$$f(y|\theta) = e^{(p(\theta)y - q(\theta) + g(y))},$$

where  $p(\theta)$ ,  $q(\theta)$ ,  $g(y)$  are functions.

The components of the GLMs are:

- **Random component:** Is the response variable  $y$ . We need to define the probability distribution that has the same.
- **Systematic Component:** It specifies the variables used in the linear predictor, result of the linear combination of 1 and the explanatory variables selected for the model construction. Within these explanatory variables it could be the interaction of them.
- **Link:** The linkage between the components defined above. It relates a monotonic function of the expected value of the response variable,  $g(\mu)$ , with the linear predictor. The simplest function  $g(\mu)$  is the identity and it is called *identity link*.

#### 3.2.1 Poisson Loglinear Model

It is a model for response variables which values belong to the set of natural numbers. It assumes a Poisson distribution for the random component and it uses, as link, the log function.

Let  $\mu$  be the expected value of the Poisson variable  $y$  and  $x$  the explanatory variable, then the association between these variables in a Poisson Loglinear Model has the following representation:

$$\begin{aligned} \log \mu &= \alpha + \beta x \\ \Rightarrow \mu &= e^{\alpha + \beta x} = e^{\alpha} (e^{\beta})^x. \end{aligned}$$

#### 3.2.2 Gamma Model

It is a model for response variables which values belong to the set of non-negative numbers. It assumes a Gamma distribution for the random component and it uses, as link, the inverse function.

#### 3.2.3 Hypothesis testing for GLMs

In addition to estimating  $\beta$  parameters, it is necessary to review the veracity of the assertions

that are made regarding to some unknown population characteristics. The procedure to do this is known as *hypothesis testing*.

There are test statistics for *significance* of variables, as well as others that test the accuracy in which the systematic component can describe the random component with the selected link, i.e., *goodness of fit*. Since the last one is a measure of the overall performance of the model, it is suggested to be the criterion when choosing one perspective or other.

#### 3.2.3.1 Deviance

The *deviance* is a measure that summarizes the model adequacy. Let  $L_M$  be the maximized log-likelihood value for the model of interest and  $L_S$  the maximized log-likelihood value for the most complex model, i.e., the *saturated model*. The deviance of a model  $M$  is defined as -2 times the logarithm of the likelihood ratio to compare the  $M$  model and the saturated:

$$\text{Deviance} = -2[L_M - L_S].$$

The *deviance* has the purpose to test the hypothesis that all parameters that are in model  $S$  but not in model  $M$  equals zero. For large samples, it has approximately a chi-square distribution with *degrees of freedom* equal to the number of parameters in model  $S$  but not in the model  $M$ .

The *Null deviance* is defined as the deviance when Model  $M$  is just a constant.

To compare two models:  $M_0$  and  $M_A$ , none of them saturated but  $M_0$  a special case of  $M_A$ , we can do it through their deviances:

$$-2[L_{M_0} - L_{M_A}] = M_0 \text{ Deviance} - M_A \text{ Deviance}.$$

This test statistic is analogous to the *F-test* that compares linear regression models with Normal distribution response variables.

The following ratio is called<sup>1</sup> the *pseudo R<sup>2</sup>*, since there is no a  $R^2$  in GLMs:

$$(\text{Null deviance} - \text{Residual deviance}) / \text{Null deviance}.$$

This ratio can be interpreted as the proportion of the variation in the response variable explained by the explanatory variables.

### 3.3 INLA<sup>2</sup>

INLA (Integrated Nested Laplace Approximation) is a computational approach in the *R* software introduced by Rue et al. (2009). This approach performs Bayesian inference in the class of *Latent Gaussian Models*, i.e., models which density  $p(x|\theta)$  is assumed Gaussian with mean equals zero and with precision matrix  $Q(\theta)$ , where  $\theta$  represents the vector of hyperparameters. Thus, distributions are in the following form:

$$\begin{aligned} (\theta) &\sim p(\theta) \\ (x|\theta) &\sim N(0, Q(\theta)^{-1}) \end{aligned}$$

<sup>1</sup> Dobson (2002)

<sup>2</sup> Lingren, Finn., Rue, Havard, 2014. Bayesian Spatial Modelling with R- INLA

$$(y_i|x, \theta) \sim p(y_i|\eta_i, \theta),$$

where, as we mentioned,  $\theta$  are (hyper)parameters,  $p(\theta)$  is typically taken to be non-informative,  $x$  is a latent Gaussian field,  $\eta$  is a linear predictor based on known covariate values  $c_{ij}$  ( $\eta_i = \sum_j c_{ij} * x_j$ ), and  $y$  is a data vector. The joint distribution of the variables in the model is  $p(y,x,\theta)$ , that is function of  $(y_i|x, \theta)$ ,  $Q(\theta)$  and  $p(\theta)$ . It takes  $y$  fixed to get the posterior marginal densities of the latent variables  $p(x_i|y, \theta)$ , given a fixed hyperparameter value, then it is integrated these marginals over the approximations of the hyperparameters posterior density  $p(\theta|y)$ .<sup>3</sup>

The principal objective of the INLA approach is to get an approximation to the marginal posteriors for the latent variables as well as to the hyperparameters of the Gaussian Latent Model.

The INLA approach consists in approximating the full posterior  $p(\theta|y)$  (by using the Laplace approximation) that will be used later to integrate out the uncertainty with respect to  $\theta$  when approximating the posterior marginal of  $x_i$ . The second step computes the Laplace approximation of the full conditionals  $p(x_i|y, \theta)$  for selected values of  $\theta$ . Finally, the approximation for the marginal of the latent variables is obtained  $p(x_i|y)$ .

Summing up, INLA uses accurate deterministic approximations instead of *Markov Chain Monte Carlo* (MCMC) simulations in order to estimate posterior marginals.

### 3.4 Models description

It starts from the premise that any bank has the following current and historical information:

- Branch ID
- Date of accounting record
- Number (of transactions) of deposits
- Amount (of money) of deposits
- Number (of transactions) of withdrawals
- Amount (of money) of withdrawals

From the information, mentioned above, the following variables should be generated (the nomenclature is suggested as well):

- Branch: Branch ID (Qualitative, nominal)
- Wkindingday: Day of the week wherein the accounting movements were registered (Qualitative, nominal)  
 $I=$  Monday,  $2=$  Tuesday, ...,  $5=$  Friday
- Mnthday: Day of the month wherein the accounting movements were registered (Qualitative, nominal)
- Payday: Indicator variable for paydays, considering the majority of people. It is suggested to consider 1 or 2 consecutive days (Qualitative, nominal)

- Pstpayday: Indicator variable for working days after paydays. It is suggested to consider 1 or 2 consecutive days (Qualitative, nominal)
- Holiday: Indicator variable for public holidays (Qualitative, nominal)
- Pstholiday: Indicator variable for the working day after public holidays (Qualitative, nominal)
- Month: Month wherein the accounting movements were registered (Qualitative, nominal)
- $I=$  January,  $2=$  February, ...,  $12=$  December
- Deptxn: number (of transactions) of deposits (Quantitative, discrete)
- Depamnt: amount of money from deposits (Quantitative, continuous)
- Wthtxn: number (of transactions) of withdrawals (Quantitative, discrete)
- Wthamnt: amount of money from withdrawals (Quantitative, continuous)

The paradigm for both models,  $M1$  and  $M2$ , is to define, for each branch, the random sum of a Compound Process  $\{N(t_i, t_{i+1}], Y_{i,i+1}\}$ , where:

- $N(t_i, t_{i+1}]$  represents the number of transactions between  $t_i$  and  $t_{i+1}$ ,
- $Y_{i,i+1}$  represents the amount of money of the  $N(t_i, t_{i+1}]$  transactions performed.

At first, it is understood that the difference between  $t_i$  and  $t_{i+1}$  is one day, i.e.,  $N(t_i, t_{i+1}]$  represents the number of transactions on day  $i$ , and  $Y_{i,i+1}$  represents the average amount of the transactions performed on day  $i$ .

Importantly, we must distinguish transactions that represent deposits from which represent withdrawals. Thus, it will be a model to estimate the amount of deposits:  $\{N_d(t_i, t_{i+1}], Y_{d,i,i+1}\}$ , and one for withdrawals:  $\{N_w(t_i, t_{i+1}], Y_{w,i,i+1}\}$ .

Models,  $M1$  and  $M2$ , must be developed in an independent way, talking about branches, because of the particularity that could exist in each one, e.g., a branch placed in a zone of high commercial activity, has a distinct behavior in relation to a branch located next to residential neighborhoods: different amounts, trends, seasonalities, etc. In this way, models are simpler to treat, since it decreases the number of variables and the quantity of problems to address (e.g., cross-correlation between agents at the same point of time<sup>4</sup>, which should be reviewed in panel data).

#### 3.4.1 Model M1

$M1$  consists in approaching, punctually, every component of the processes:  $\{N_d(t_i, t_{i+1}], Y_{d,i,i+1}\}$  and  $\{N_w(t_i, t_{i+1}], Y_{w,i,i+1}\}$ , through the use of GLMs. Because of the features of data, it is natural to suggest, for the first component, to approach the future data according to a Poisson Loglinear

<sup>3</sup> Cseke, Botond., Heskens, Tom, 2011. Approximate Marginals in Latent Gaussian Model. *Journal of Machine Learning Research* 12 (2011) 417-45

<sup>4</sup> Gujarati, Damonar N., Porter, Dawn C., 2009. *Econometria*. Mc Graw Hill, pp. 591- 616



Model<sup>5</sup>, while the second component based on a Gamma Model, because of its probability distribution flexibility that allows to represent a variety of distribution forms with only two parameters<sup>6</sup>.

Thus, regressions are performed according to the following statements, for the *deposits* case:

Previous to the regression fit, there are generated indicator variables for the (originally) nominal variables, considering that the number of indicators should be the possible categories in each nominal variable minus one. Trying to sum up the regression expressions, this is, without all the indicator variables, here is presented the description for the number of transactions approach:

$Deptxn \sim \beta_0 + \beta_1 * Wkingdayfact + \beta_2 * Mnthdayfact + \beta_3 * Payday + \beta_4 * Pstpayday + \beta_5 * Holidayfact + \beta_6 * Pstholiday + \beta_7 * Monthfact.$  (Poisson Loglinear Model).

Although it could happen a frequent problem named *overdispersion* (variance > mean), we are not going to do any adjust, since our objective is to approach punctually, and the *overdispersion* correction doesn't affect it.

The amount of money approach is:

$Depamnt \sim \beta'_0 + \beta'_1 * Wkingdayfact + \beta'_2 * Mnthdayfact + \beta'_3 * Payday + \beta'_4 * Pstpayday + \beta'_5 * Holidayfact + \beta'_6 * Pstholiday + \beta'_7 * Monthfact.$  (Gamma Model).

In an analogous manner, they are approached *Wthtxn* and *Wthamnt* in order to determinate the values of the Compound Process for *withdrawals*.

### 3.4.2 Model M2

Bearing in mind randomness in the regression parameters, *M2* approaches punctually each component of the processes:  $\{N_d(t_i, t_{i+1}], Y_{d,i,i+1}\}$  and  $\{N_w(t_i, t_{i+1}], Y_{w,i,i+1}\}$ , with a Bayesian perspective. It is used the INLA function (mentioned in 3.3 section) without a specification for the prior distribution of the parameters, considering the following fits<sup>7</sup>:

$Deptxn \sim \beta_0 + \beta_1 * Wkingdayfact + \beta_2 * Mnthdayfact + \beta_3 * Payday + \beta_4 * Pstpayday + \beta_5 * Holidayfact + \beta_6 * Pstholiday + \beta_7 * Monthfact.$  (Poisson Loglinear Model).

$Depamnt \sim \beta'_0 + \beta'_1 * Wkingdayfact + \beta'_2 * Mnthdayfact + \beta'_3 * Payday + \beta'_4 * Pstpayday + \beta'_5 * Holidayfact + \beta'_6 * Pstholiday + \beta'_7 * Monthfact.$  (Gamma Model).

<sup>5</sup> Agresti, Alan., 1996 *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc., pp. 71- 97.

<sup>6</sup> Wilks D., 1990. Maximum likelihood estimation for the gamma distribution using data containing zeros. *Journal of Climate* 3, 1990 pp. 1495–1501

<sup>7</sup> Regressions are described summing up parameters in case of multiplication with indicators variables.

In the same way, it is approached *Wthtxn* and *Wthamnt* in order to determinate the values of the Compound Process for *withdrawals*.

## 4. RESULTS

Below, it is exposed the illustration of a model fit in order to show how regressions (described above) could approach real data. It is shown only the case of *number of deposits* in a determined branch. In Fig. 2 there is the actual behavior of deposits, in Fig. 3 it is illustrated the approach with *frequentist* perspective, while in Fig. 4 we drew the approach with *Bayesian* perspective.

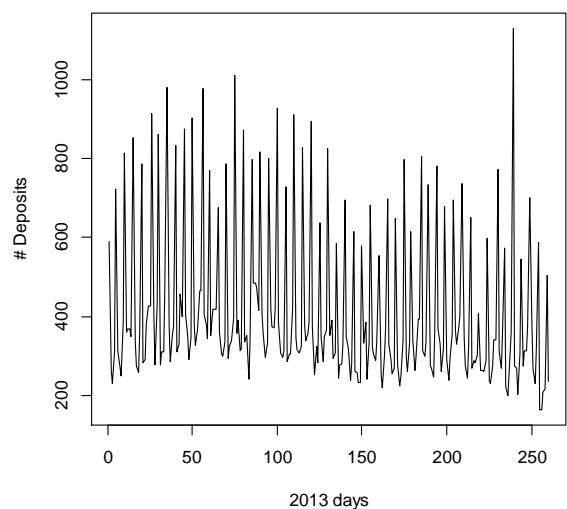


Fig. 2: Number of deposits for a branch in 2013

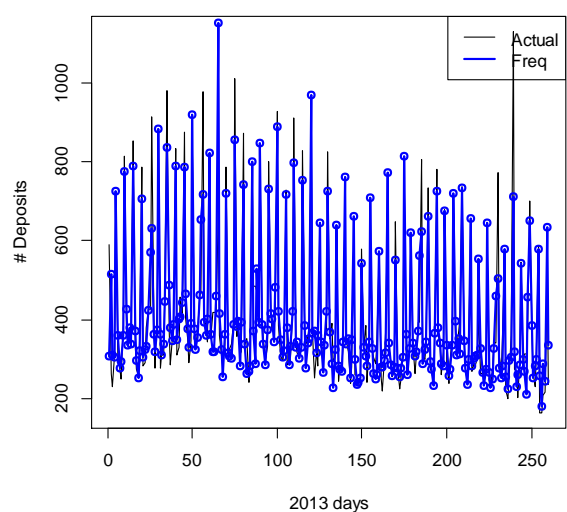


Fig. 3: Number of deposits for a branch in 2013 (black) Vs traditional GLM approach (blue)



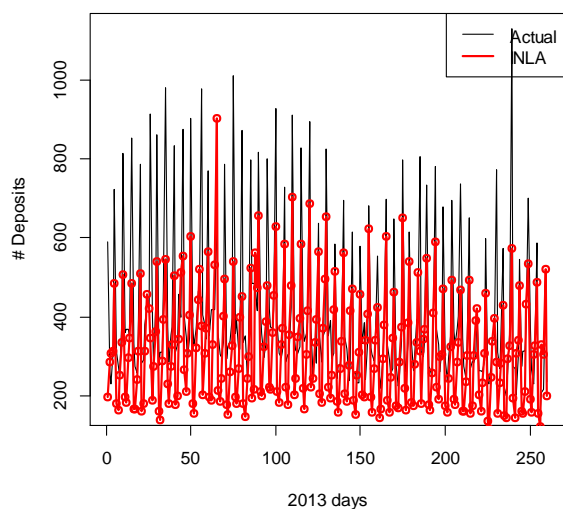


Fig. 4: Number of deposits for a branch in 2013 (black) Vs INLA approach (red)

It was fitted the model to determine the components of the Compound Process, both for *deposits* and for *withdrawals*. It was fitted the model for 30 real branches, considering 2013 data, and evaluating the performance of both perspectives: *frequentist* and *Bayesian*.

In all the cases, approaching through a conventional *Generalized Linear Model*, i.e., in a *frequentist* manner, we got a better performance of the model. The *pseudo R<sup>2</sup>* was up to .86, for  $N_d(t_i, t_{i+1}]$  regression, and up to .82, for  $Y_{d,i,i+1}$  one.

With the same position for *withdrawals*, the *pseudo R<sup>2</sup>* was up to .86, for  $N_w(t_i, t_{i+1}]$  regression, and up to .74, for  $Y_{w,i,i+1}$  one.

Trying to compare regressions fitted with INLA Vs regressions with traditional GLM, we realized that, for the 30 branches, *standardized residuals* were greater than the obtained with traditional GLMs. INLA fitted  $Y_{d,i,i+1}$  with a sum of squares of *standardized residuals* greater than 1049. Traditional GLM did it with a sum of squares of *standardized residuals* since of 1.31.

INLA fitted  $Y_{w,i,i+1}$  with a sum of squares of *standardized residuals* greater than 1077. Traditional GLM did it with a sum of squares of *standardized residuals* since of 1.48.

## 5. CONCLUSIONS

The purpose of this paper is to suggest a model approach for cash transactions demand at bank branches, with the intention to help control the risk of *stocking out* of cash and being efficient managing the cash.

The model proposed to approach the cash flow in a branch is to establish a *Compound Process*  $\{N(t_i, t_{i+1}], Y_{i,i+1}\}$ , one for *deposits* and another for *withdrawals*. Each of these components should be

estimated by a regression model that explains both variables with others that represent days in the week, days in the month, payday, holidays and months, some of them with a lag, described in section 3.

Because of the shown evidence, we suggest to do regressions in a traditional way, we mean, using GLMs in order to approach  $N$  and  $Y$ . In many areas it is used a cutting edge way to approach variables, we refer to Bayesian perspective using INLA, but to fit the model proposed it is not suggested.

We did not have access to real data for ATMs, but because of the nature of its operation, it is natural to suppose that it performs better, fitting *withdrawals*, with a *frequentist* perspective than *Bayesian*, as it happened for *withdrawals* in branches.

Approaches obtained with the *frequentist* perspective can be used to set the *maximum cash storage* and *safety stock*. Since both parameters are normally fixed once a year, we proposed to determine these as:

- For *safety stock*: Let  $P$  be a percentile (defined by the central office of the bank according to the desired *service level*) of the daily difference between *withdrawals* and *deposits*, approached for a year with the *Compound Process*, and let  $Q$  be a percentile (estimated with a percentage of *withdrawals* satisfied at the beginning of a day) of the daily *withdrawals* approached for a year, as well. Then, the *safety stock* proposed is  $\max\{P, Q\}$ . It is important to point out that if the branch administrator knows about an extraordinary *withdrawal*, it must be covered the *extraordinary demand* besides the *safety stock*.
- For *maximum cash storage*, let  $R$  be the maximum of the daily *deposits* approached for a year, and  $T$  the amount of money that Risk/Control Office approve for the branch. Then, the *maximum cash storage* proposed is  $\min\{R, T\}$ .
- *Minimum amount of money in an ATM*: let it be a percentile (estimated with a percentage of *withdrawals* satisfied in a day, previous to the replenishment) of the daily *withdrawals* approached for a year.
- *Maximum amount of money in an ATM*: let  $P_{ATM}$  be a percentile (defined by the central office of the bank with the desired *service level*) of the daily *withdrawals* approached for a year and let  $V$  be the maximum amount of money that vendors recommend to hold in a ATM of a specific model. Then, the *maximum amount of money in an ATM* will be  $\min\{P_{ATM}, V\}$ .

## ACKNOWLEDGMENTS

This research was supported by UNAM PAPIIT grant IN116012.

## REFERENCES

- Agresti, Alan. 1996. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, Inc., pp. 71-97.
- Bell, Peter C. and A. Hamidi-Noori, 1984. Foreign Currency Inventory Management in a Branch Bank, *The Journal of the Operational Research Society* Vol. 35, Num. 6: 513-525
- Bonini, C.P., Asuman, W. H., Bierman, H. Jr., 1997. *Quantitative Analysis for Management*. Chicago: McGraw Hill. IRWIN Series, 1997(9), Chapter 1.
- Cseke, Botond., Heskes, Tom, 2011. Approximate Marginals in Latent Gaussian Model. *Journal of Machine Learning Research* 12 (2011) 417-45
- García Cabello, Julia, 2013. Cash efficiency for bank branches. *SpringerPlus*. 2:334.
- Gujarati, Damodar N., Porter, Dawn C., 2009. *Econometría* Mc Graw Hill, pp. 591- 616
- Lingren, Finn., Rue, Havard, 2014. Bayesian Spatial Modelling with R- INLA
- Makary Girgis, Nadia, 1968. Optimal Cash Balance Levels, *Management Science* Vol. 15, Num. 3 Theory Series, 130-140
- Sven Axsäter, 1996. Using the deterministic EOQ formula in Stochastic Inventory Control. *Management Science* 42:830- 834
- Wilks D., 1990. Maximum likelihood estimation for the gamma distribution using data containing zeros. *Journal of Climate* 3, pp. 1495–1501

## AUTHORS BIOGRAPHY

**Miguel** holds a BCs degree in *Actuary* from ITAM, and is currently studying a Masters in *Operations Research*, at UNAM. At work, he has served to the public and private sector in many activities related to the optimization/ continuous improvement in processes that have to do with product sales; passenger arrivals coming from foreign flights, at Mexican airports; and cash management in bank branches. Nowadays, he is making an effort to implement the content of this proposal in the global firm he is working for.

Statistic inference and stochastic processes are the main topics he is passionately interested in.

**Idalia** She received her Ph.D. in Operations Research at the Faculty of Engineering of the UNAM. She graduated Master with honors and received the Gabino Barreda Medal for the best average of her generation. She has been a referee and a member of various Academic Committees at CONACYT. She has been a referee for journals such as *Journal of Applied Research and Technology*, the *Center of Applied Sciences* and

*Technological Development*, UNAM and the *Transactions of the Society for Modeling and Simulation International*. Her research interests are in simulation and optimization of production and service systems. She is a full time professor at the Postgraduate Program at UNAM.

# ON THE SEARCH FOR NOVEL SIMULATION APPLICATIONS TO SUPPORT AIRPORT OPERATIONS MANAGEMENT

Olusola Theophilus Faboya<sup>(a)</sup>, Peer-Olaf Siebers<sup>(b)</sup>

<sup>(a)</sup>Computer Science Department, College of Education, Ikere-Ekiti, Ekiti State, Nigeria

<sup>(b)</sup>School of Computer Science, University of Nottingham, Nottingham, UK

<sup>(a)</sup>[faboyaolusola@gmail.com](mailto:faboyaolusola@gmail.com), <sup>(b)</sup>[peer-olaf.siebers@nottingham.ac.uk](mailto:peer-olaf.siebers@nottingham.ac.uk)

## ABSTRACT

Airports are complex socio-technical systems with many different stakeholders which often have very different needs. Operations Research modelling tools and techniques are used successfully to support the management of such systems by helping to better understand and improve their operations. A review of the literature shows that there are many existing modelling studies related to airport operations management but it also unveils some territories that are still untapped. This might be due to the fact that the right tools or techniques for these jobs have not been identified yet. In this paper we identify some of these untapped areas and discuss how simulation modelling could be used as a decision support tool for gaining insight into systems operation in these areas. We take a closer look at one of the identified areas (airport facilities maintenance scheduling) and conduct a hypothetical proof-of-principle simulation study to demonstrate the feasibility and applicability of using simulation in this area. Our conclusion is that simulation studies are a very useful aid for this kind of investigation.

Keywords: airport, operations management, agent-based modelling, simulation, maintenance scheduling.

## 1. INTRODUCTION

An airport is a complex, large-scale socio-technical system comprising multiple stakeholders, multiple jurisdictions and complex interaction between many actors (Wu and Mengersen 2013). The large growth of air travel in the last two decades has forced many airports to increase their capacity and optimise their processes (Ma et al. 2011). There are many operation divisions in an airport which have the roles of ensuring that all operation facilities (such as aircrafts and other airport facilities) are serviceable at all times while also managing crises such as emergency situations efficiently. Increasing the performance and efficiency of existing airport facilities requires, a thorough understanding of different stakeholders needs. These stakeholders within their jurisdictions have different views and needs that have to be considered when finding solutions to various airport problems. And like

in many other systems, we cannot afford to find the right solutions by experimenting with real objects since building; destroying and making changes may be too expensive, dangerous or even practically impossible (XJ Technologies 2012). As such, simulation modelling has become one of the ways of solving problems that appear in the real world when other forms of modelling prove to be impossible to use in Operations Research (OR).

The goal of this paper is to identify and find new simulation modelling applications for airport operations and investigate the feasibility of these applications in airport operations by building hypothetical proof-of-principle models using different worldviews with appropriate simulation paradigms.

The remainder of the paper is structured as follows: Section 2 provides a literature review on existing airport operation management models and identifies areas with potential for further development. We then choose one of these areas (airport facilities maintenance scheduling) and provide a proof-of principle study, described in the following sections. Section 3 focuses on the design and the implementation of our simulation model. Section 4 covers testing through experiments. The results of the experiments are then discussed in Section 5. Finally, in Section 6 we provide our conclusions and propose further developments.

## 2. LITERATURE REVIEW

### 2.1. Simulation Methods

Besides analytical methods that are often used in airport operations management, several simulation modelling methods exists that can be used for this purpose. The main ones are System Dynamics (SD), Discrete Event (DE), and Agent-Based (AB) simulation modelling.

SD models help us to understand the behaviour of complex systems over time (Wakeland et al. 2004). They use a very high abstraction level, are deterministic, and represent a system in terms of aggregates and flows (Kirkwood 1998). In DE models, movement of passive entities (components) that make up a system can only be described based on state change and the time at which the change occurs (Albrecht 2010). In AB models, the components can be viewed as individual objects with autonomy and interactivity (Ma

et al. 2011) where detailed individual object emergent behaviours can be studied when large populations of objects interact with each other. However, AB modelling is a mindset more than a technology. This mindset consists of describing a system from the perspective of its constituent units (Bonabeau 2002) and therefore allows system components to execute various behaviours appropriate for the system they represent. This allows unanticipated behaviour to emerge, make it a good modeling method for heterogeneous, autonomous and proactive actors such as human centered systems.

## 2.2. Airport Modelling Studies

Quite a large number of airport modelling studies have been found in the literature. These focus on various areas of airport operations. Below we provide an overview of these studies subdivided into various areas of airport operations.

### Passenger Terminal Models

Existing studies in this area include the work of Tang et al. (2012); Hebert and Dietz (1997); Suryani, Chou and Chen (2010); Schultz and Fricke (2011); and Hanta and Pozivil (2010). These studies used either analytical, SD, or DE modelling technique which has limited capability in representing passengers' behaviour. Furthermore, discretionary activities such as duty free shopping and the use of bathrooms are not considered. More detailed consideration of human behaviour in these models would provide more realistic predictions of overall system performance.

### Maintenance Scheduling Models

Many modelling studies on maintenance scheduling have focused on the development of optimisation methods. These studies include Suryani, Chou and Chen (2010); Hecht et al. (1998), and Duffuaa and Al-Sultan (1997). It is observed that all of these studies used methods with high abstraction levels such as integer programming, queuing theory or SD. Also at a high abstraction level, Miller and Clarke (2007) used Monte Carlo simulation to evaluate the strategic value of air transportation infrastructure. These high abstraction levels have limitations when it comes to modelling some maintenance scheduling problems where high levels of agents' interactions and message passing between objects of the systems are relevant to study system operations. Furthermore, there is limited information about modelling maintenance of other airports' infrastructure that provide conveniences to customers.

### Human Behaviours Models

Studies on human behaviours in airport operations are looked into in two ways:

1. **Human behaviour in a normal situation:** Schultz and Fricke (2011) modelled human behaviour as flow using a mathematical model while Ma et al. (2011) and Cimler et al. (2013)

used AB models to study advanced passenger traits movement in the departure hall.

2. **Human behaviour in an extreme situation:** Galea and Galparsoro (1994) and Galea, Owen and Lawrence (1996), developed a prototype egress model to simulate the evacuation of large numbers of individuals from an enclosure such as an aircraft. Matthew et al. (2011) used AB modelling to model the evacuation of individuals with disabilities in a densely populated airport. Chow and Fong (2011) modelled emergency evacuation in the arrival hall of a crowded airport terminal using EXODUS and SIMULEX. However, the paradigms used in these studies can be considered appropriate because of their capability to support adequate representation of the systems' components and interactions, but it is observed that studies on terrorist attacks are limited.

### Airport Marketing Models

Kuhn et al. (2010) present an AB model to assist market share analysis and help the investment analyst to develop earnings forecasts for the year ahead. It is clear that there are limited existing models in this area, and the available one used AB modelling to model a macro level factor, which could best be studied using an aggregate viewing tool for better prediction of model behaviour.

## 2.3. Classification of Existing Airport Models

Based on the findings from the literature review and with the help of several experts in the field, a graphical representation that classifies the existing airport operation models has been developed which is shown in Figure 1. This classification tells us, at a glance, some of the essentials of the model structure. The horizontal axis distinguishes the levels of abstraction (macro, meso, and micro), while the vertical axis represents the queue constraints (from low to high). The approaches listed on the horizontal axis are: mathematical models (at the extreme level of abstraction) where system problems are represented as mathematical equations using various analytical methods. Continuous modelling follows which may be in the form of SD models, and then discrete modelling follows in form of DE models which focus on processes, and AB models which focus on individual entities and their interactions. These have the lowest level of abstraction.

Figure 1 shows that modelling airport terminal operations such as passenger movement in the departure area, human behaviour, baggage handling and security check-in were mostly carried out using either analytical, SD or DE techniques. It also shows that in airport infrastructure maintenance scheduling, studies were limited to using analytical techniques which have limited capabilities to represent objects' interactions and message passing - one of the key factors in maintenance scheduling. Modelling maintenance scheduling for

airport facilities such as escalators, lifts, air conditioners, television sets, power supply points, hot spots, and so on are not considered. Furthermore, studies in emergency evacuation show that not much has been done on discretionary activities and terrorist attacks in particular. On airline and airport marketing, limited works were found, and the available ones did not focus on airport or airline performance improvement.

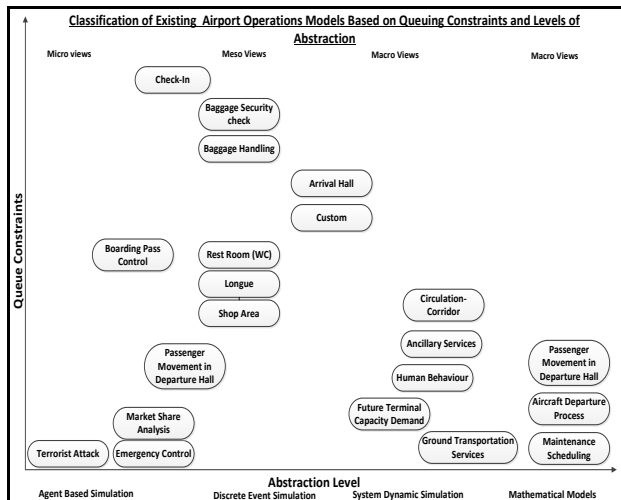


Figure 1: Classification of Existing Airport Operations Models

To support finding the gaps we have re-categorised the airport models presented in Figure 1 while considering the system operation areas and the adequacy of system component representation in the models.

### 2.4. Identifying the Gaps

The new classification has been achieved by adopting Greasley's worldview framework (Greasley 2013). It provides us with suggested techniques based on the level of detail required to represent the components of a system under study in sufficient detail, and looking into the kind of insight each of the available simulation methods used in these areas give. Using this framework we were able to identify some gaps in the existing classification as presented in red ovals in Figure 2.

By adopting Greasley's worldview framework we can see that many operations are shifted towards the AB modelling paradigm. This is a result of the system operations being non-static, considering the behaviour of the components of the systems involved.

The red ovals in Figure 2 represent areas of airport operations where gaps have been found. These areas are Airport Facilities Maintenance Scheduling and the Airline Marketing Strategy. The two areas could be of significant importance to the growth of an airport in terms of revenue generation. Airport revenues are increasingly dependent on the numbers of customers that use the airport. Moreover, customers' satisfaction and some marketing strategies will play a huge role in winning and retaining customers over time. Customers' satisfaction depends on, among other things, availability

of needed facilities in the terminals and lounge where customers spend most of their time, while winning more customers requires strategic planning. However, to achieve success in these areas, airport management needs a good decision support system to aid its decision and policy making.

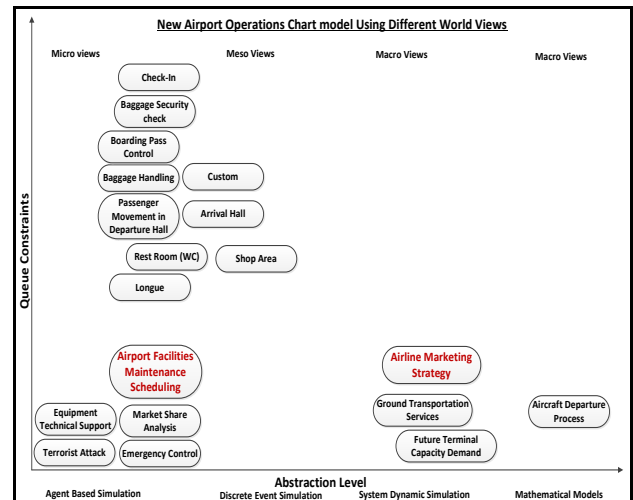


Figure 2: Classification of Airport Models Including New Areas

### 3. HYPOTHETICAL CASE STUDY

In this section we take a closer look at one of the new simulation modelling opportunities identified earlier (Airport Facilities Maintenance Scheduling) through a hypothetical proof-of-principle simulation study.

We acknowledge that (due to space constraints) some of the diagrams in this section are difficult to read. You can download larger versions of these from our website (Siebers 2014).

#### 3.1. Problem Description

Here is a description of our hypothetical problem: A group of airport facilities are distributed geographically within the airport mostly in the terminal area where passengers and well-wishers spend most of their time. Facilities such as escalators, lifts, air conditioners, television sets, power supply points, hot spots, and so on are provided for customers' convenience while at the lounges and terminals. Also there is an airport hangar that houses a fleet of aircrafts that need routine maintenance and various levels of service. These airport facilities and aircrafts contribute to the growth of the airport in form of revenue and customer satisfaction while working, so they need to be up and running for optimum daily operations. However, they sometimes break down and need to be maintained, repaired or replaced.

To service these facilities, there are facilities maintenance crews (facilities-technicians) and aircraft maintenance experts (aircraft-technicians) in the airport estates office who service the facilities and aircrafts based on certain management policies (replacement and maintenance policies). But the airport management has

limited capability to employ a maximum of five technicians each for the aircraft and other airport facilities.

### 3.2. Conceptual Model

For defining our conceptual model we have employed Robinson's conceptual modelling framework (Robinson 2004).

The specific objectives of this case study are "to determine the number of technicians required to ensure that 90% of the airport facilities and 95% of the aircrafts are in good working condition", and "to ensure that 80% of airport facilities and 95% aircrafts are available all of the time", and "to determine the replacement policy that will give the optimum result at minimum cost". Our constraints are that at most five technicians can be employed for both, airport facilities and aircrafts.

Assumptions are that: airport facilities problem identification takes zero time; technicians never fail and always have all necessary spare parts and tools on board; technicians do not optimise their routes; and technicians are equipped with a radio and can take a new assignment while in motion.

Simplifications are that: distance travelled by the technicians before getting to the site of work is not modelled; all technicians work 24 hours a day; the airport facilities parameters are the same; the aircraft maintenance check types (A, C, or D) are also not modelled; and the parameters are the same.

We decided to use an AB approach for modelling this system as we are dealing with individual entities that are communicating with each other.

The Sequence Diagram in Figure 3 shows the event ordering and message exchange between the Airport Facility Object (airport facilities) and the Technician Object (maintenance crew). It is a result of discussions

with several airport operations managers and has been validated by them.

### 3.3. Model Implementation

To implement the conceptual model we used AnyLogic, a Java based multi-paradigm simulation modelling IDE (AnyLogic 2014).

In our implementation the active object classes are the AirportFacilities (aircraft and other facilities) and the MaintenanceCrew (facilities technician and aircraft technician). The AirportFacilities agent types are the same considering states and transitions but have different parameter values. They are autonomous and can respond to changes in their internal state such as failed, repaired, and maintained. The same is true for the MaintenanceCrew agent types.

#### 3.3.1. AirportFacilities Agents

The state chart representing all AirportFacilities agent types is presented in Figure 4.

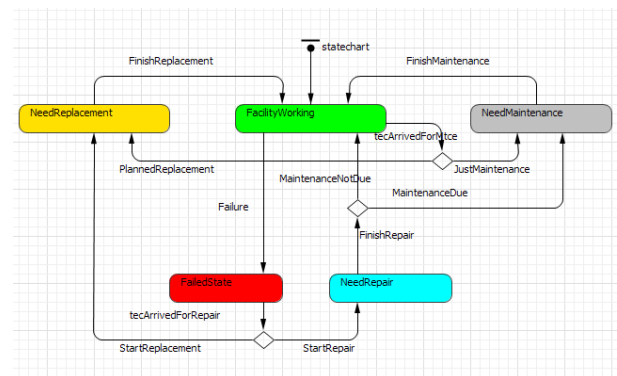


Figure 4: AirportFacilities Agent State Chart

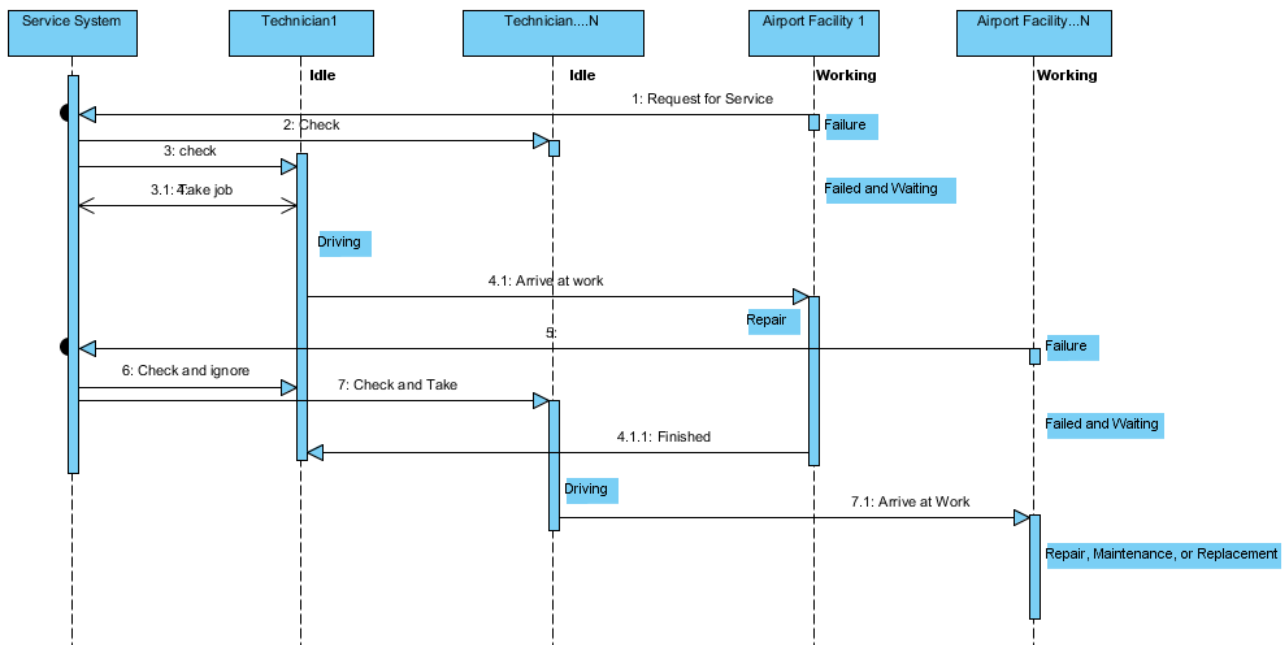


Figure 3: Airport Facilities Maintenance Scheduling Sequence Diagram



All airport facilities and aircrafts are assumed to be working initially i.e. in "FacilityWorking" state.

- Transition to "FailedState" state is triggered by a stochastic event failure rate. Request for service is made, and broadcast notification is sent to all technicians.
- One of the technicians takes the job and drives to the failed facility while the others return back to their "Idle" state.
- The transition from "NeedReplacement" state to "FacilityWorking" state is triggered by a timeout (triangular distribution).
- The transition from "NeedRepair" state to "FacilityWorking" is deterministic having two branches "MaintenanceNotDue" and "MaintenanceDue".
- The transition from "NeedMaintenance" state to "FacilityWorking" state is triggered by a timeout (triangular distribution).
- A working facility may request for scheduled maintenance or being on a planned replacement after a given number of maintenance periods, therefore, the transition from "FacilityWorking" has two branches which depend on whether the "ReplaceOldFacility" policy is active or not.

### 3.3.2. MaintenanceCrew Agents

The state chart representing all MaintenanceCrew agent types is presented in Figure 5.

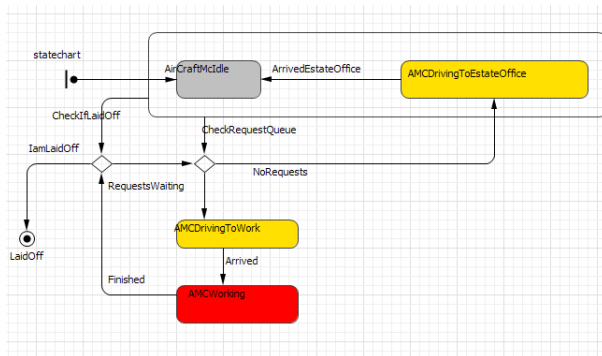


Figure 5: MaintenanceCrew Agent State Chart

All technicians are at the "Idle" state in the estates office at the initial stage. The agent implementation is based on the following transitions among states:

- The transition "CheckRequestQueue" sent by service system (dispatcher) to all technicians has two branches: "RequestsWaiting" and "NoRequests".
- The transition "Arrived" always indicates that a technician has arrived at the failed facility for work.

- The transition "Finished" sent by the facility to the technician indicates "FacilityWorking" and then releases the facility.

However, a technician can be employed or laid-off. Therefore, each technician in the collection has an index that is used to check its employability status. After completing a job, the technician checks if the index is still within the "Service Capacity". If the technician is idle or driving to the estate office, a special message "CHECK IF LAID OFF" is sent to the technician to check the same condition.

### 3.4. Model Verification and Validation

When developing simulation models it is crucial to gain credibility through verification and validation. This is particularly important for real world case studies related to airport operations management.

The model discussed in this paper, however, is purely academic and is based on a hypothetical situation due to non-availability of real world data. It has been thoroughly verified (e.g. code debugging) and model design and implementation have been validated by domain experts (face validation). This ensures that the model design is a reasonable imitation of the real world system and that our model implementation produces reasonable outputs.

### 4. TESTING THROUGH EXPERIMENTATION

In the experimentation presented here we are interested in the number of technicians and replacement policy that give optimum results for both, airport facility and aircraft availabilities, at low cost. Therefore, our focus is on technician utilisation, facility and aircraft availabilities, and a combination of the number of technicians and replacement policies that give optimum results in terms of profit. The replacement policy is the amount of maintenance that a facility or an aircraft must have before it can be replaced, whether it is still working or not.

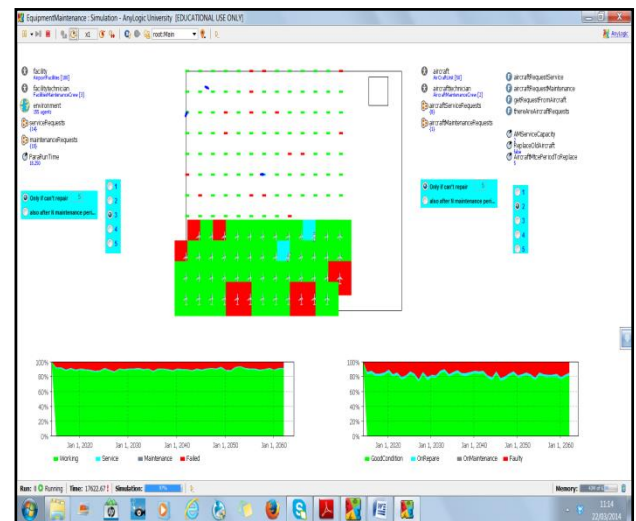


Figure 6: Screenshot for Simulation Model run

Three different scenarios were simulated and the output from each of the scenario is presented as a screenshot to gain a quick overview.

**A: Base Scenario**

The base scenario uses all of the default values as shown in the Table 1. The output from the base scenario simulated model for airport facilities and aircrafts is presented in Figure 7 and the results of the simulation run are shown in Table 2.

Table 1: Base Scenario: Experimental Factors

|                       | Airport Facilities                                   | Aircrafts                                            |
|-----------------------|------------------------------------------------------|------------------------------------------------------|
| Number of technicians | 3                                                    | 2                                                    |
| Replacement policy    | Replace only failed facility that cannot be repaired | Replace only failed aircraft that cannot be repaired |

This scenario runs on default technicians' values of 3 for airport facilities and 2 for aircraft. It gives 89% and 80% availability, 11% and 20% unavailability for airport facilities and aircrafts, respectively. And 96% and 95% technician utilisation with idle levels of 4% and 5%, respectively. The costs of operating the system are £7,500 for airport facilities and £8,000 for aircrafts, while the respective profits are put at £18,500 and £25,000.



Figure 7: Base Scenario Output View

Table 2: Base Scenario Results

|                       | Airport Facilities | Aircrafts |
|-----------------------|--------------------|-----------|
| Number of technicians | 3                  | 2         |
| <b>Facilities</b>     |                    |           |
| Availability (%)      | 89                 | 80        |
| Unavailability (%)    | 11                 | 20        |
| <b>Technicians</b>    |                    |           |
| Utilised (%)          | 96                 | 95        |
| Idle (%)              | 4                  | 5         |
| <b>Cost/Benefit</b>   |                    |           |
| Profit (in 1000s)     | 18.5               | 25        |
| Cost (in 1000s)       | 7.5                | 8         |

**B: Scenario 1**

Scenario 1 was observed based on varying the number of technicians working on the facilities while the replacement policy remained at default value as shown in Table 3 below.

Table 3: Scenario 1 Experimental Factors

|                       | Airport Facilities                                             | Aircrafts                                                      |
|-----------------------|----------------------------------------------------------------|----------------------------------------------------------------|
| Number of technicians | Options A,B,C<br>3,4,2                                         | Options A,B,C<br>2,3,1                                         |
| Replacement policy    | Replace only failed facility that cannot be repaired (default) | Replace only failed aircraft that cannot be repaired (default) |

In this scenario, there are three options A, B, C representing changes made to the number of technicians at default replacement policy. The output from this run is shown in Figure 8 and the results in Table 4.

Option A: The simulation starts with default technicians' values for the two cases as in the base scenario. This gives the same result as the base scenario shown in Table 2.

Option B: The number of staff is increased from 3 to 4 for airport facilities and 2 to 3 for aircrafts. The result is 97% and 96% facility and aircraft availabilities, 3% and 4% non-availability, respectively; technician utilisation of 92% and 90% and idleness of 8% and 10% for facilities and aircrafts, respectively, and the costs of operations are £7,600 and £8,000, while the respective profits are put at £20,000 and £28,000 for facilities and aircrafts.

Lastly, in option C the number of technicians is reduced from 4 to 2 for airport facilities and from 3 to 1 for aircrafts. This gives 22% and 39% availability with non-availability of 78% and 61%, respectively; however, technician utilisation of 100% and idleness of 0% are recorded for both airport facilities and aircrafts, respectively. The costs of this option are £4,000 and £3,000 while the profits are £4,000 and £12,000, respectively, for airport facilities and aircrafts.

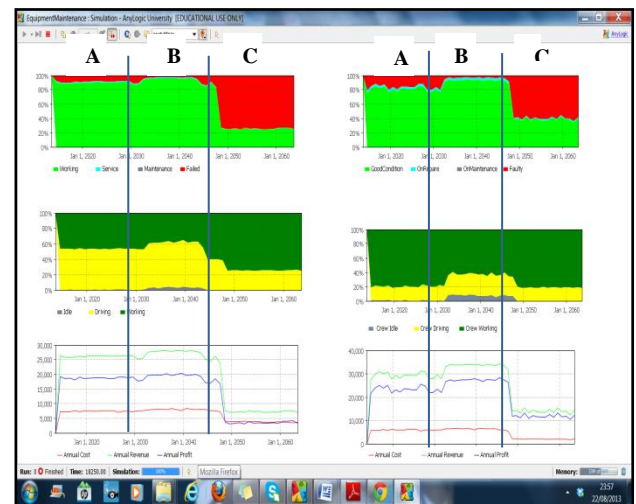


Figure 8: Scenario 1 Output View

Table 4: Scenario 1 Results

| Option                | Airport Facilities |     |     | Aircrafts |    |     |
|-----------------------|--------------------|-----|-----|-----------|----|-----|
|                       | A                  | B   | C   | A         | B  | C   |
| Number of technicians | 3                  | 4   | 2   | 2         | 3  | 1   |
| Facilities            |                    |     |     |           |    |     |
| Availability (%)      | 89                 | 97  | 22  | 80        | 96 | 39  |
| Unavailability (%)    | 11                 | 3   | 77  | 20        | 4  | 61  |
| Technicians           |                    |     |     |           |    |     |
| Utilised (%)          | 96                 | 92  | 100 | 95        | 90 | 100 |
| Idle (%)              | 4                  | 8   | 0   | 5         | 10 | 0   |
| Cost/Benefit          |                    |     |     |           |    |     |
| Profit (in 1000s)     | 18.5               | 20  | 4   | 25        | 28 | 12  |
| Cost (in 1000s)       | 7.5                | 7.6 | 4   | 8         | 8  | 3   |

**C: Scenario 2**

Scenario 2 was observed based on varying the number of technicians working on the facilities while varying the replacement policy as shown in Table 5. The output for all possible combinations of experimental factors (number of technicians and replacement policy) for Scenario 2 are shown in Figure 9 as options A–E and the results from this run are shown in in Table 6.

Table 5: Scenario 2 Experimental Factors

|                       | Airport Facilities                                | Aircraft                                          |
|-----------------------|---------------------------------------------------|---------------------------------------------------|
| Number of technicians | A,B,C,D,E<br>3, 3, 4, 4,2                         | A,B,C,D,E<br>2, 2, 3,3,1                          |
| Replacement policy    | Replace after<br>(default,3,3,2,2)<br>maintenance | Replace after<br>(default,4,4,3,3)<br>maintenance |

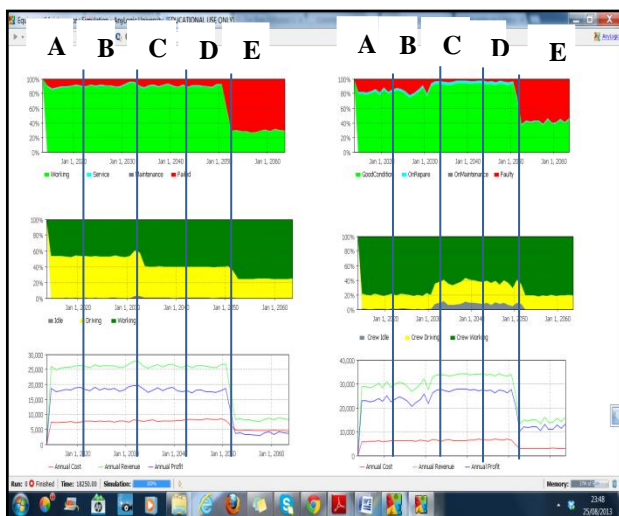


Figure 9: Scenario 2 Output View

Options are defined as follows:

Option A: This presents the default settings, and gives the same results as the base scenario, see Table 2.

In Option B the replacement policy is changed from "replace only failed facility" to "replaced airport facility after 3 maintenance periods" and "replace aircraft after 4 maintenance periods". This option gives facility and aircraft availabilities of 91% and 82; and non-availability 9% and 18% for both facilities and aircrafts, respectively; with technician utilisation of 96% and 97% and idleness of 4% and 3% for facilities and aircrafts, respectively. The costs are £7,500 and

£6,200; while the profit is £27,000 and £24,000, respectively, for facilities and aircrafts.

Option C: there is an increase in the number of technicians from 3 to 4 for airport facilities and from 2 to 3 for aircrafts, while replacement policy settings remain unchanged from the previous run in both cases. This gives availability 93% and 97% for facilities and aircrafts, respectively; and non-availability of 7% and 3%, respectively; the technician utilisation shows 93% and 89% for facilities and aircrafts with idleness of 7% and 11%, respectively. Also, the costs of running this option gave £7,600 and £6,200 for facilities and aircrafts, respectively, with profit of £28,000 in both cases.

Option D: There is a change in the replacement policy years settings from 3 maintenance periods for airport facilities to 2 maintenance periods and also from 4 maintenance periods for aircraft to 3 maintenance periods while the numbers of technician remain the same as in previous run. This option gives facilities availability of 94% and 96% for aircraft, and non-availability of 6% and 4%, respectively, technician utilisation of 93% and 90% respectively and idleness of 7% and 10% respectively. While the costs remain the same as in the previous option for the two cases, the profit is at £27,000 for both cases. In Option E, there is further reduction in the number of technicians to 2 for facilities and 1 for aircrafts, while the replacement policy is the same as the last run. This option produces facilities availability of 25% and 40% aircrafts availability, and non-availability of 75% and 60%, respectively, and gives technician utilisation of 100% in both cases. The costs of this option are £5,000 and £4,000 for facilities and aircrafts with profit of £4,000 and £12,000, respectively.

Table 6: Scenario 2 Results

| Option                | Airport Facilities |     |     |     |     | Aircrafts |     |     |     |     |
|-----------------------|--------------------|-----|-----|-----|-----|-----------|-----|-----|-----|-----|
|                       | A                  | B   | C   | D   | E   | A         | B   | C   | D   | E   |
| Number of technicians | 3                  | 3   | 4   | 4   | 2   | 2         | 2   | 3   | 3   | 1   |
| Replacement policy    | 5                  | 3   | 3   | 2   | 2   | 5         | 4   | 4   | 3   | 3   |
| Facilities            |                    |     |     |     |     |           |     |     |     |     |
| Availability (%)      | 89                 | 91  | 93  | 94  | 25  | 80        | 82  | 97  | 96  | 40  |
| Unavailability (%)    | 11                 | 9   | 7   | 6   | 75  | 20        | 18  | 3   | 4   | 60  |
| Technicians           |                    |     |     |     |     |           |     |     |     |     |
| Utilised (%)          | 96                 | 96  | 93  | 93  | 100 | 98        | 97  | 89  | 90  | 100 |
| Idle (%)              | 4                  | 4   | 7   | 7   | 0   | 2         | 3   | 11  | 10  | 0   |
| Cost/Benefit          |                    |     |     |     |     |           |     |     |     |     |
| Profit (in 1000s)     | 26                 | 27  | 28  | 27  | 4   | 23        | 24  | 28  | 27  | 12  |
| Cost (in 1000s)       | 7.5                | 7.5 | 7.6 | 7.6 | 5   | 6         | 6.2 | 6.2 | 6.2 | 4   |

**5. DISCUSSION**

Comparing with the base scenario (see Table 2), Scenario 1 (see Table 4) showed that an increase in the number of technicians by 1 for both, facilities and aircrafts, has a significant impact on the availability of working facilities (9%) and aircraft (20%) with a corresponding increase in profit of 8% and 12%, respectively, during this period; while the costs of operating the system slightly increased by 1.3%. This setting actually provides a solution that complies with the objective of this study - to increase the level of

facility and aircraft availabilities to 90% and 95%, respectively. However, the increase in the cost is expected because additional technicians would increase the running cost. There is also an increase in the technicians' idle time by 100% (i.e. from 4% to 8%) as shown in Table 4. This shows that engaged technicians are more than what is expected to give an optimum results. Furthermore, reduction in the number of technicians by 1 from the default value returns very low level facility and aircraft availabilities despite 100% technicians' utilisation. This in turn results in a significant reduction in profit.

It can be seen from Table 6 option B that reduction in the replacement policy from 5 (default) to 3 maintenance periods increases the availability of facilities and aircrafts slightly; while the technicians utilisation remain the same hence, costs remain the same, there is also a slight increase in the profit. Table 6 options C and D show that the same number of technicians with different replacement periods gives different facility and aircraft availabilities, almost the same technician utilisation and different profit.

Furthermore, it is good to note as shown in Option E that choosing low replacement policy with fewer technicians (very low) will not improve availability but rather increase facilities failure. This is expected because few technicians regardless of replacement policy will not be adequate to give optimum service even with a low replacement period. But choosing appropriate replacement policy allows fewer technicians to be engaged for the same number of facilities and hence, reduced cost of maintaining technicians.

The results also show that the cost of facility and aircraft replacements always increase the overall cost of running the system, but the benefits accrued from this option always out-weight the costs in terms of facility and aircraft availabilities, reliability and increase in profit. This is evident in the simulated model results (see Table 2, 4 and 6) that the effectiveness of replacement policy depends largely on the number of technicians engaged. Therefore, appropriate number of technicians should be combined with a certain replacement policy in order to get optimum results.

However, there are some issues that need further investigation. The effect of replacement policy is observed to be less significant even with an adequate number of technicians as shown in Table 6 options C and D, this is against our expectation. Though the cause of this is not clear, it is believed to be as a result of non-availability of real life data.

## 6. CONCLUSIONS

Proffering solutions to problems involving complex socio-technical systems such as airports cannot be found by experimenting with the real object. Simulation modelling provides an alternative means for an easier and better prediction of a systems performance to aid decision making. We think that it provides a transparent model development process that can be used even by non-experts in simulation modelling, and that it requires

less intellectual effort to interpret the results and communicate the model internally to others than alternative (e.g. mathematical) modelling approach.

In this paper we have identified some new application areas for simulation modelling in an airport operations management context and selected one of them for further investigation. We have demonstrated with a proof-of-principle simulation study how one could build a simulation model for "Airport Facilities Maintenance Scheduling" and what kind of outputs one can expect from such a simulation model. The simulation study helped us to come up with a solution that fulfilled all of the criteria listed in the simulation study objectives while minimizing the operational costs.

To support effective decision making through our simulation study we decided to use an AB approach, as we are dealing with a human centric system. Such an approach allows us to take an object oriented worldview in order to address the problems of individuality, autonomy and interactivity among the entities within this complex, non-linear system. It also supports the study of individual entities' emergent and unanticipated behaviours.

In the end we have achieved our objectives by finding new opportunities for simulation modelling and by demonstrating the feasibility for one of the identified areas. The next steps will involve looking more closely at the other areas we have identified, and providing proof-of-principle models for those areas. Also, we are interested in looking at some of the existing traditional simulation applications within an airport context which currently employ SD or DE simulation to see if AB simulation can provide any advantages or additional insight in these cases. We believe there are many more opportunities of applying this method – which have yet to be seized.

## ACKNOWLEDGMENTS

We would like to thank Dr Jason Atkin and Dr Geert De Maere for their help with designing and validating our hypothetical models and Dr Simon Miller for his proof reading support.

## REFERENCES

- Albrecht, M.C., and AZ, P.E., 2010. *Introduction to Discrete Event Simulation*. Available from: <http://www.albrechts.com> [accessed 7 Aug 2013]
- Alfares, H., 1999. Aircraft maintenance workforce scheduling: A case study. *Journal of Quality in Maintenance Engineering*, 5(2), 78-88.
- AnyLogic (2014) AnyLogic Software. Available from: <http://www.anylogic.com> [accessed 6 June 2013]
- Bonabeau, E (2002). Agent-based modelling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of United State of America*. 99 (3). pp 7280-7287. October 4-6 2001, Irvine, CA
- Chow, W.K., Fong, N., (2011). Simulation of Emergency Evacuation in the Arrival Hall of a Crowded Airport. *Fire Safety Science-Proceedings*



- of the Tenth International Symposium. pp 613-626. June 19-24, Maryland, USA.
- Cimler, R., Kautzka, E., Olsevicova, K., Gavalec, M. (2013). Agent-based model for comparison of aircraft boarding methods. *Proceedings of 30<sup>th</sup> International Conference Mathematical Methods in Economics*, pp 73-78. September 11-13 2012, Karvina, Czech Republic.
- Duffuaa, S., and Al-Sultan, K., 1997. Mathematical programming approaches for the management of maintenance planning and scheduling. *Journal of Quality in Maintenance Engineering*, 3(3), 163-176.
- Galea, E., and Galparsoro, P., 1994. A computer-based simulation model for the prediction of evacuation from mass-transport vehicles. *Fire Safety Journal*, 4(22), 314-366.
- Galea, E., Owen, M., and Lawrence, P., 1996. A Computer modelling of human behaviour in aircraft fire accidents. *Toxicology*, 115, 63-78.
- Greasley, A., 2013. The Simulation Modelling Process. In: Taxman, F.S., and Pattavina, A., ed. *Simulation Strategies to Reduce Recidivism*, New York: Springer.
- Hanta, V., and Pozivil, J., 2010. Modelling of a Conveyors system as a tool for testing of a baggage handling process. *The International Journal of Transport and Logistics*, 7, 331-334.
- Hebert, J., and Dietz, D., 1997. Modeling and Analysis of an Airport Departure Process. *Journal of Aircraft*, 34 (1), 43-47.
- Hecht, M., Handal, J., Czekalski, L., and Rosin, A., 1998. Impact of maintenance staffing on availability of the U.S. air traffic control system', *Proceedings of the Annual Reliability and Maintainability Symposium (IEEE)*, pp. 113-119. Anaheim, CA.
- Kirkwood, C., 1998. *System Dynamics Methods: A Quick Introduction*. Available from: <http://www.public.asu.edu/~kirkwood/sysdyn/SDIntro/preface.pdf> [accessed: 28 June 2013]
- Kuhn, J., Courtney, J., Morris, B., and Tataru, E., 2010. Agent-based analysis and simulation of the consumer airline market share for Frontier Airlines. *Knowledge-Based Systems*, 23(8), 875-882
- Ma, W., Kleinschmidt, Trisan, Fookes, Clinton and Yarlagadda, P. K.D.V., 2011. Checking-in processing: simulation of passengers with advanced traits. In: Fu, Micheal, and White, K., ed. *Proceedings of the 2011 Winter Simulation Conference*, IEEE, Phoenix, Arizona, USA.
- Matthew, M., Yong, S., K., Keith, C., Antony, C., 2011. Modelling Emergency Evaluation of Individual with Disabilities in a Densely Populated Airport. *Journal of Transportation Research Board*, 32-38.
- Miller, B., and Clarke, J., 2007. The hidden value of air transportation infrastructure. *Technological Forecasting & Social Change*. 74, 18-35.
- Robinson, S., 2004. *Simulation: The Practice of Model Development and Use*. England: John Wiley & Sons Ltd.
- Schultz, M., and Fricke, H., 2011. *Managing Passenger Handling at Airport Terminals. Individual-based Approach for Modelling the Stochastic Passenger Behavior*. Ninth USA/Europe Air Traffic Management Research and Development Seminar. Available from: <http://www.atmseminar.org/>
- Siebers (2014) All figures, tables, and the simulation model presented in this paper: Available from: <http://www.cs.nott.ac.uk/~pos/publications.html>
- Suryani, .E., Chou, S., and Chen, C., 2010. Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamic framework. *Expert Systems with Applications*, 37, 2324-2339.
- Tang, T., Wu, Y., Huang, H., and Caccetta, L., 2012. An aircraft boarding model accounting for passengers' individual properties. *Transportation Research Part C: Emerging Technologies*, 22, 1-16.
- Wakeland, W., Gallaher, E., Macovsky, L., and Aktipis, C., 2004. A comparison of System Dynamics and Agent-Based Simulation Applied to the study of Cellular Receptor Dynamics. In: *Hawaii International Conference on System Sciences*, vol 3.
- Wu, P., and Mengersen, K., 2013. A review of models and model usage scenarios for an airport complex system. *Transportation Research Part A: Policy and Practice*, 47, 124-140.
- XJ Technologies, 2012. Big Book of AnyLogic. Available from: <http://www.anylogic.com> [accessed 6 June 2013]

#### AUTHORS BIOGRAPHY

**Olusola Theophilus Faboya** is a lecturer in Computer Science, College of Education, Ikere-Ekiti, Nigeria. He has obtained a BSc. in Computer Science from Ondo State University, Ado-Ekiti, Nigeria and an MSc. from the University of Nottingham, UK. He is a member of many professional bodies, which include IEEE, IEEE Computer Society, Nigeria Computer Society (NCS) and International Association of Computer Science and Information Technology (IACSIT).

**Dr Peer-Olaf Siebers** is a lecturer in Computer Science at the University of Nottingham, UK. His main research interest is the application of data driven computer simulation to study human-centric complex adaptive systems. More information about his research can be found on this website: <http://www.cs.nott.ac.uk/~pos/>.

# ANALYSIS OF OBESITY MODELLING

Maja Atanasijević-Kunc<sup>(a)</sup>, Tina Sentočnik<sup>(b)</sup>, Jože Drinovec<sup>(c)</sup>

<sup>(a)</sup>University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

<sup>(b)</sup>University of Ljubljana, Faculty for Medicine, Ljubljana, Slovenia

<sup>(c)</sup>University of Maribor, Faculty for Medicine, Maribor, Slovenia

<sup>(a)</sup>[maja.atanasijevic@fe.uni-lj.si](mailto:maja.atanasijevic@fe.uni-lj.si), <sup>(b)</sup>[tina.sentocnik@siol.net](mailto:tina.sentocnik@siol.net), <sup>(c)</sup>[joze.drinovec@lj-kabel.net](mailto:joze.drinovec@lj-kabel.net)

## ABSTRACT

Obesity has become very frequent. It increases the risk for developing diabetes type 2, hyperlipidemia and hypertension, and among other also several cardiovascular diseases. This was the reason it was recognized by World Health Organization as a chronic disease. This means that it should be treated and if possible prevented. Regarding mentioned facts it is not surprising that economic and social burden of this preventable problem is tremendous. In the paper some of recent publications describing modelling results of mentioned problems are analyzed to identify the explanatory power of problem interpretations and the efficacy of proposed control strategies at the level of an individual patient and at the population level. For this also proposed healing strategies and their efficacy were studied. Very promising solution is proposed for closed-loop treatment control of individual integrative treatment. The complexity of problem reduction at the population level asks for further research.

Keywords: overweight, obesity, mathematical modelling, simulation

## 1. INTRODUCTION

Overweight and obesity have reached pandemic levels. As a consequence it has become clear that excessive fat accumulation may impair health. This was a reason why has World Health Organization (WHO) in 1997 declared obesity being a disease (WHO 2000). This means that such patient's condition needs to be treated and if possible, prevented. Once considered a problem only in high-income countries (North America, Europe), overweight and obesity are now dramatically on the rise also in low- and middle-income countries, particularly in urban settings (WHO 2013, 2014).

WHO uses the gradation of obesity regarding the so called body mass index (BMI) which is defined as the weight in kilograms divided by the square of the height in meters ( $\text{kg/m}^2$ ) (WHO 1995). It provides very useful population-level measure of overweight ( $25 < \text{BMI} < 30$ ) and obesity ( $\text{BMI} > 30$ ) as it is the same for both sexes and for all ages of adults. However, it should be considered as a rough guide because it may not

correspond to the same degree of fatness in different individuals. This information is frequently complemented with a waist circumference as it is also in correlation with the accumulation of visceral adipose tissue (VAT) which has additional negative influence to patients' health. Visceral fat is a strong predictor of insulin resistance in obese subjects, but not in normal weight individuals. Unlike subcutaneous fat, increased visceral fat is associated with increased hepatic glucose production and reduced glucose disposal (Carroll et al. 2008). Health is more threatened in the case of higher BMI and if fat accumulation is long lasting. In Slovenia (being an average country of EU regarding obesity prevalence) for example, around 55% of population has BMI greater than 25 (Atanasijević-Kunc and Drinovec 2011a).

Increased BMI represents an enlarged risk for developing numerous diseases. First we have to mention diabetes type 2 (D2). This chronic disease has also reached epidemic extensions. Statistical data proves (Atanasijević-Kunc and Drinovec 2011b) that over 85% of D2 patients are also overweight or obese. WHO forecasts that in the next decade the number of people with D2 is going to increase for 50% (WHO 2014). Chronic diseases which are very frequent among obese people are also hyperlipidemia and hypertension.

Obesity increases also the development of cardiovascular diseases (CVD) (mainly heart disease and stroke). If at the same time patient has also D2, increased cholesterol and increased blood pressure, health is threatened even more as the influences of these diseases are synergistic (Scott et al. 2004). CVD were the leading cause of death in 2012 (WHO 2014).

Among the consequences it is necessary to mention also musculoskeletal disorders, especially osteoarthritis and some cancers (endometrial, breast, and colon) in addition to psychological and social problems.

Pandemic extensions of overweight and obesity which call for corresponding treatment, numerous consequences in the form of chronic diseases and serious health complications which introduce addition expenses show that this situation does not represent a stress only for an individual patient but is also social



and economic burden for a wider community which cannot be neglected.

Modelling and simulation have proven to be very useful in analysis, explanation, and prediction also in the field of medicine, pharmacy, and other bio-processes, where the problems connected with obesity are no exception (Atanasijević-Kunc et al. 2008a, Levy et al. 2011). In this sense the following questions are important:

1. How processes connected with obesity development are modelled and what is the informative level of the existing models?

2. What modelling approaches and what sorts of models are used frequently?

3. Can models help in a quantitative prediction of economic burden of overweight and obesity regarding certain community, wider region, country or even a group of countries, like for example European Union?

4. How such problems can be interpreted from the viewpoint of insurance companies, hospitals, but also individual patients and their medical doctors?

5. Can mathematical models help in the explanation what can be done and how these problems can be decreased? Is in solving such problems important to take into account only individual patient, or is more efficient to address a wider community, or perhaps corresponding combination would be the best approach?

6. Is it possible to include also control or regulation principles to improve explanation models' level and the level of problem solving?

7. Can modelling results explain why obesity treatment is so insufficient or is efficient so rarely?

Presented questions have on one side motivated the present work and have on the other also helped in development the criterions which can be used in models' classification and in formulation of some problems which are not solved or are not solved satisfactory. In models' analysis which is presented in the paper the following aspects and criterions are taken into account:

1. Models' message(s)

2. Model classification (static, dynamic, continuous, discrete, hybrid, linear, nonlinear, ...)

3. Model complexity (description specter is very wide, from simple to very complex, which cannot be described in such a way that the reader can repeat the presented experiments, or information is hid because of economic aspects)

4. The level of description or granulation level (mentioned problems can be addressed at the epidemiological level where larger population is observed, but then again the obesity can be presented at the level of individual person or even at the level of some important chemical processes)

5. Are psychological aspects also taken into account?

6. Are addressed also social and/or economic burdens?

7. Are control or regulation mechanisms which influence the disease development and treatment, taken

into account (such weakened or even damaged mechanisms are usually important in disease development but are not always described or taken into account through modelling description)?

8. Multi-model problem description and combination of obesity models with the models of other related processes

9. Data analysis on the basis of which the model was developed and their reliability (always important, but not always correspondingly presented).

The paper is organized in the following way. When searching for the solution of this immense trouble it is first important to understand the problem at the level of an individual patient. Some of the crucial modelling results important for understanding dynamic system properties are addressed in the first part of the next section. But for problem reduction it is of crucial importance to take into account also psychological analysis of patient's capabilities, as the efficacy of long-term body mass reduction programs is otherwise incredibility miserable.

In the third section models are first analyzed at the epidemic level due to the fact that obesity was recognized to be a disease when it has reached serious extensions. Such "distant" problem observation can indicate also correlations with other chronic diseases, social processes, and life styles which were rare in the previous centuries. In addition it is shown that epidemic models can be combined with the population models uncovering other processes among which populations' aging is very important. Part of this section is devoted also to the investigation of connections of different-level or multi-model problem descriptions.

In the fourth section the observation is extended to include some of economic, social and health care consequences.

Finally, in the conclusion identified answers to posed questions are summarized and some ideas for future investigations are described.

## 2. MODELS AT PATIENT'S LEVEL

Biological systems are in general very complex multivariable processes influenced by specific properties of individuals and/or random environmental influences. In such situations *evidence based observations* and *statistical descriptions* are typical starting tools for ordering the information which is at the same time also a basic goal of modelling. Consequently not only the disease descriptions but also the problem interpretations can be adapted through time as realized by (Chang and Christakis 2002). They have observed that through the 20<sup>th</sup> century obesity shifted in ontological status from being the product of something that individuals *do* to something that they *experience*.

Problem observation through the usage of statistical tools indicated also several important connections among the variables which influence the development of overweight and obese person. Important variables are for example age, body mass, BMI, sex (Jackson et al. 2002), but also environmental or/and social networks

(Christakis and Fowler 2007, El-Sayed et al. 2012, Bruzzone et al. 2012).

## 2.1. Energy equilibrium

When searching for the main reasons of overweight and obesity and for the main solutions of these problems we have to start with the observation of an individual person as all problems actually start at this level. The relations between eating and drinking or energy intake, physical activity or energy expenditure and weight change have been studied and published extensively (Westerterp et al. 1995, Hall 2006, Chow and Hall 2008, Thomas et al. 2009, Hall 2010a, Hall 2010b, Hall et al. 2011, Hall et al. 2013). Through the last decade one of the most active research group in the field is the one of dr. Kevin Hall from National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland. They are developing mathematical models of very different complexity describing the problem, but one which should be mentioned explicitly (Hall et al. 2011) is the three compartment continuous nonlinear model consisting of fat mass (fm), lean mass (lm) and extracellular fluid compartments (ecf), each demanding for description first order ordinary differential equation. Inputs to the model are energy intake in the form of carbohydrate, fat, protein intake, the change in sodium, and energy expenditure which is the result of physical activity, thermic effect of feeding and resting metabolic rate (the energy needed to maintain the basic physiological processes). Model parameters are evaluated regarding statistical estimations depending on age and sex of observed person, while initial conditions are for each compartment evaluated from the average regression equations for men and women using the information of body mass, BMI and sex (Jackson et al. 2002). Block diagram of this model is illustrated in Fig. 1.

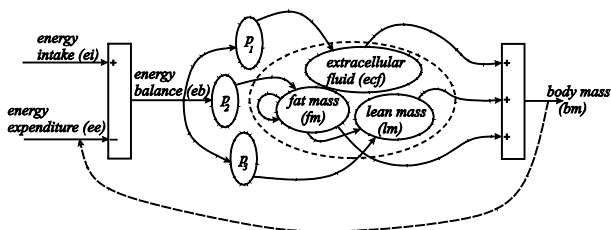


Figure 1: Schematic representation of the three-compartment dynamic model describing body mass changes in an individual person

Basic information of this model is very clear. If energy balance is negative body mass is expected to decrease (genetic and individual properties are not crucial regarding the main system dynamics). To help people understand this better, the authors realized an interactive version of the model available through internet (Hall et al. 2011). The goal is to show that body mass changing is rather slow process and the patient has to persist for longer time to reach new steady state after

life style is changed. If energy balance is increased also body mass is going to react correspondingly.

In spite of the fact that basic information is very clear and simple it usually does not work in practice, even in the cases when people are not pleased with their appearance and are at the same time also informed of potential negative risks of obesity. In (Mokdad et al. 2001) it is reported that among the US adults participating in programs for losing or maintaining weight, only 17.5% were following the recommended guidelines for reducing calories and increasing physical activity. As duration of such programs is limited it can be expected that long term efficacy of body mass reduction is essentially worse.

Additional problem, which have to be mentioned and asks for further extensions of presented model (and will be described in the next subsection), is the observation that even after successful dieting process people often regain their body mass which is frequently even higher as before the starvation. It was recognized that repeated starvation episodes influence resting metabolic rate (Astrup, et al. 1999.). This means that described model (see Fig. 1) should be used also in wider time ranges with which it is important to take into account also parameters' time-varying properties (they change with patient's age, body mass) and patient's history which additionally influence metabolic rate.

As mentioned in the Introduction the information of BMI is not the only one when obese patient is examined. An important parameter is also waist circumference (which should be 80cm or less in women and 94cm or less in men (Lean et al. 1995)) because it is also in strong correlation with the accumulation of visceral adipose tissue (VAT) which has additional negative influence to patients' health (Carroll et al. 2008). From this an important question arises: Which factors are contributing to the relative change of visceral versus subcutaneous abdominal fat (SAT) during weight loss? In (Chaston and Dixon 2008) it was reported that there is no convincing evidence that any investigated weight loss intervention selectively targets the reduction of VAT to a greater extent than another. This problem was investigated also by (Hall and Hallgreen 2008). Very interesting result, which can be used to extend the presented model in Fig. 1 to four-compartment model (fm-compartment can split into VAT and SAT compartments), is that changes of VAT mass are allometrically related to changes of total body fat mass (fm) during weight loss, regardless of gender or weight loss intervention and can be described as:

$$\frac{dVAT}{VAT} = k \frac{dfm}{fm} \quad (1)$$

where  $k=1.3 \pm 0.1$ . Greater weight loss will cause a greater absolute reduction of VAT mass and may thereby result in improved metabolic health. This observation is very important from two additional viewpoints:

- patients should be evaluated as rather successful even if their weight loss is relatively low (this problem will be addressed later again),
- increased VAT in obese patients represents an additional health problem if this condition results in fatty liver (Phillips and Barton 2014) which can be detected for example by the ultra-sound examination.

We have investigated this problem in obese and extreme obese patients but the allometric connection of fatty liver phenomenon with other allometric parameters is still under investigation.

## 2.2. Psychological influences to problem solution

As indicated in the previous subsection obese people would like to decrease their body mass but are very rarely successful. In general they are not capable to change their life style which is necessary for efficient, long-term and with this also for healthy body mass reduction. Taking into account also psychological processes is therefore of crucial importance for problem reduction at the level of an individual patient.

In very interesting work of (Navarro-Barrientos et al. 2011) it is suggested to extend energy balance description with the behavioral model as schematically presented in Fig. 2. It is developed regarding the extension of the theory of planned behavior (Ajzen and Madden 1986) to dynamic description.

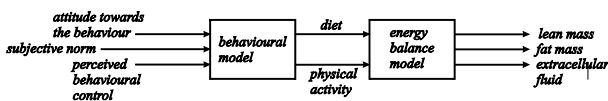


Figure 2: Block diagram of behavioural influence to body mass dynamics as proposed by (Navarro-Barrientos et al. 2011)

For this purpose they have proposed a fluid analogy which mimics the concepts of inventory management in supply chains. The proposed result consists of five reservoirs, where the level of final one – the so called *behavior* is influenced by the *intention* and *PBC* (*perceived behavioral control*), while the *intention* depends on the level of *attitude*, *subjective norm* and *PBC* too. Input variables to these fluid reservoirs are three exogenous variables (strength of beliefs about the outcome  $x$  evaluation of the outcome it entering *attitude* reservoir, strength of normative beliefs  $x$  motivation to comply is input signal to *subjective norm*, strength of control belief  $x$  perceived power of the control factor is input signal to *PBC*) and five endogenous variables, one for each tank representing disturbances. The proposed fifth order dynamic behavioral structure can influence the dieting program as well as physical activity and enable different control strategies.

In (Dong et al. 2013) described structure is extended to closed-loop operation where authors proposed hybrid model predictive control to decrease excessive gestational weight gain.

Different modelling result of dynamic behavioral model as well as precisely defined closed-loop adaptive control approach is described in (Sentočnik et al. 2013) and is schematically presented in Fig. 3. It is suitable for all patients with self-control difficulties. Authors have defined controllability of the patient as one of the crucial properties for the successful integrative patient's treatment. It can be achieved by regular patient's examination by a team of experts. The advantage is simpler and objective estimation of patient's motivation to follow the prescribed behavioral pattern. Patient's efficacy can be evaluated only from body mass changes through treatment period in spite of the fact that also psychological part of treatment, dieting, and physical activity are very important. Medical expert must be capable to adapt treatment to patient's abilities of life style changes so that patient feels he/she is relatively successful. Such satisfaction in general motivates patient to endure in treatment. Proposed approach is very successful and has recently essentially decreased a drop out which is now less than 10%.

Mentioned modelling results also incorporate situations where due to treatment interruptions and as a consequence uncontrollability patient regain the body mass.

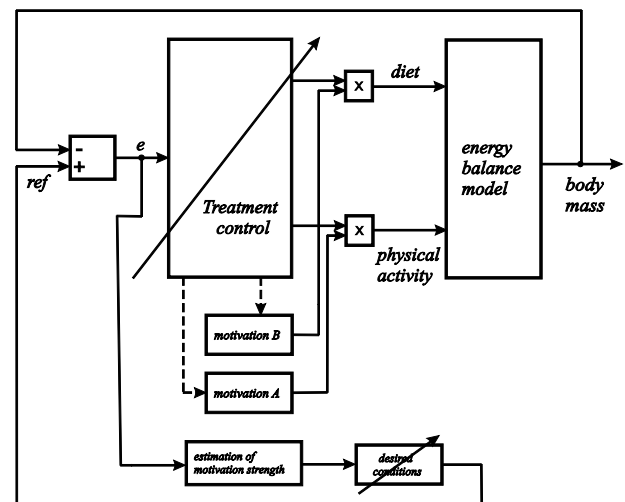


Figure 3: Block diagram of adaptive closed-loop patient control (Sentočnik et al. 2013)

## 2.3. Other modelling results

A huge number of other results have been published which are important for obesity problem understanding at the level of an individual patient, but cannot be mentioned here due to paper limitation space. We would like to add only two representatives dealing with an important patient state, namely *metabolic syndrome* (MS) which refers to a clustering of specific cardiovascular disease (CVD) risk factors influencing the insulin resistance. The risk factors include obesity, insulin resistance, dyslipidemia, and hypertension and it is known to increase the risk for CVD and D2. MS helps to identify individuals at high risk for both D2 and CVD. Results presented by (Jeong et al. 2014) show

that the proposed model can quantify the risk of MS (regarding fasting glucose, waist circumference, HDL-cholesterol, triglycerides and blood pressure) and effectively identify a group of subjects who might be classified into a potential risk group for having MS in the future by the usage of the so called radar charts.

Very interesting contribution on MS is presented in (Angelova 2013) where laboratory animals are used in experimental modelling approach and results statistical evaluation. The advantage of the *animal models* is the opportunity of precise control of diet and motor activity and the opportunity to carry out histological studies of MS, which are difficult to perform in humans.

### 3. EPIDEMICS MODELLING

For centuries, communicable diseases were the main causes of death around the world. Life expectancy was often limited by uncontrolled epidemics. As a consequence first *epidemic models* have become an important tool for a simplified description of transmission of communicable or infectious diseases through individuals. Epidemic mathematical models of communicable diseases can be divided into *top-down approaches* (Zauner et al. 2012, Miksch 2012) which describe aggregated groups of people or/and patients and are in general simplistic problem descriptions. Typical and often used representatives of this group are different statistical descriptions (among which Markov models are frequently used) and (nonlinear) ordinary differential equations which are organized into the so called *compartment models* (they are actually nonlinear state-space descriptions) which consist of two or more compartments. For example, *SIS-models* consist of those who are susceptible -  $S(t)$  to the disease and those who have been infected -  $I(t)$  with the disease and are capable of spreading the disease to those in the susceptible category. In *SIR* and *SIRS-models* additional compartment  $R(t)$  is used for those individuals who have been infected and then removed from the disease, either due to (temporal) immunization or due to death. Those in this category are not able to transmit the infection to others. The *SEIS-model* takes into consideration the exposed or latent period of the disease, giving an additional compartment,  $E(t)$ . There are several diseases where an individual is born with a passive immunity from its mother and in such cases compartment  $M(t)$  is added into model description (*MSIR* or *MSEIR-models*) (Anderson and May 1991).

In contrast to *top-down approaches* *bottom-up models* describe single individual elements with certain properties in observed environment and from their interactions and evolving processes overall system emerges which is usually of interest. Typical representative from this group are agent-based models (Šalamon, 2011) and sometimes also cellular automata are used.

#### 3.1. Obesity epidemic modelling

With medical research achievements in terms of vaccination, antibiotics and improvement of life

conditions, non-communicable diseases (NCDs) started causing major problems in industrialized countries. One of the main reasons, contributing to such situation, is of course obesity. This epidemic was started by the overproduction of food in the United States (Dreifus, 2012). The underlying biological, psychological, and social mechanisms of obesity epidemics have fascinated a broad range of scientific audience.

Here two important questions arise. Is obesity, (partly at least) contagious or not? Is it possible to extend the usage of models describing epidemics of infectious diseases to NCDs? What kinds of adaptations are needed, if any?

As already mentioned, several authors have reported that social networks can represent the possibility of spreading social epidemics (El-Sayed et al. 2012). In (Christakis and Fowler 2007), for example, correlation among social ties and obesity is investigated. It was concluded (on the base of longitudinal statistical study) that a person's chance of becoming obese is increased if he or she has a friend or sibling who has become obese. These effects were not observed among neighbors. Cohen-Cole and Fletcher 2008, on the other hand, presented several doubts of these conclusions. They have realized that presented results may have difficulties in distinguishing social network effects from environmental confounds of weight gain. Though they advise caution in interpreting the available evidence of a social contagion in weight. It is therefore obvious that the variables that may influence increase of body weight in an individual are very problematic to be identified exactly. Sometimes even the main influences to a group of persons is unclear.

Regarding mentioned doubts recently a very interesting paper was published by (Ejima et al. 2013) indicating that a scientific approach to comparatively assessing the control programs has still to be considered. The authors are proposing two forms of *SIR-models* and so answering one of the posed question with *yes*. They have first used the classical form of *SIR-model*, where independent variable is time. Then this form was transformed so that they are observing the spread of this disease regarding patients' age, which enables to present how this chronic disease is developing through life-time. In addition they have taken into account the fact that obesity is caused by both non-contagious as well as by contagious paths. Authors proposed to include into the group  $S(t)$  all susceptible, in this case they are never-obese, in the group  $I(t)$  infected, which mean obese, and  $R(t)$  are recovered or ex-obese. Model parameters were evaluated observing available epidemiological data. As the most important practical finding, they have identified that the optimal choice of intervention programs considerably varies with the so-called transmission coefficient (model parameter) of obesity,  $\beta$ . When  $\beta$  is small, the transmission cannot be maintained by social contagion alone and their model suggests that preventing weight gain among never-obese individuals would be the most

effective option. This seems to be expected result as prevention is usually the best solution also from a viewpoint of life quality and economics. When transmission coefficient is large enough to sustain the transmission of obesity through the social path, dietary restriction among obese individuals could potentially be the most effective. Here it is important to point out that in most cases for effective body mass reduction in obese patients specialized integrative treatment is necessary (Sentočnik et al. 2013). But it is very important to accent, that despite the dependence of optimal interventions on  $\beta$ , it should be noted that the transmission potential of obesity in community setting has yet to be explicitly estimated. From Ejima et al. 2013 it is also not explicitly clear how corresponding control actions should be designed.

Problem observation regarding patients' age was proposed also in (Atanasijević-Kunc et al. 2012) where model-combination was developed taking into account also population dynamics and optimal healing dynamics at the individual level. These results, being a mixture of *top-down and bottom-up models* illustrate the effect of population aging which is observed in European countries with which also the relative burden of NCDs is becoming even more unpleasant.

Multi-model or modular problem observation is proposed also in (Glock et al. 2012). In this case three-layer structure is developed starting with a population model, at the second level disease is taken into account which is finally overlaid by economics evaluation. Used modelling approach is System Dynamics and modelling result is therefore in a form of ordinary differential equations.

As realized in (El-Sayed et al. 2012) through past decade the interest is growing in systems approaches of epidemiologic research where social networks analysis and agent-based models (ABMs) are two approaches that seems to be very promising regarding the epidemiologic literature. This methodology was reported also in (Ramirez-Nafarrate and Gutierrez-Garcia 2013) and (Bruzzone et al. 2012) for simulation of obesity epidemic.

Ramirez-Nafarrate and Gutierrez-Garcia observed the prevalence of child obesity. Each agent (child) was characterized with eight parameters (age, gender, weight, height, BMI, weight status category (underweight, healthy weight, overweight, obese), caloric intake, and energy expenditure) while very simplified equation for energy balance was used at each computational cycle. Simulation results suggest that the fraction of overweight and obese children at the end of elementary school can be reduced by moderate physical activity.

The approach of Bruzzone et al. 2012 bases on a very complex modelling result regarding adult population where used data origin from two very different environments, USA and Italy. The authors analyzed the influence of different cultural and social conditions on modelling results. They have taken into account age, gender, social and employment status,

education, ethnicity, religion, marital status, and numerous other parameters. Several statistical data evaluations proved (which is in agreement with previously mentioned publications) that it is difficult to estimate in quantitative manner relative influence of each parameter to increased body mass. Developed intelligent agent-based simulator named BACCUS (which is stochastic discrete event simulator) enables to generate a social network including family relationships, friendships, and working relationships. Relative advantage of such modelling result is of course the possibility to extend it to include large scale health care interventions. However, modelling presumptions which have been taken into account (for example: children, having an obese parent increases the risk of obesity by 30%; not-married, being single increases the risk of obesity, obese people tend to have obese friends, ...) need to be reevaluated and confirmed also by further investigations.

#### 4. ECONOMICS, HEALTHCARE SYSTEMS AND POLICYS

From pandemic extensions of overweight and obesity it is clear that the burden of this chronic disease and detected consequences (social and economic) is enormous and exact evaluation is in reality not possible. Only estimations are available.

Taking into account prevalence of obesity and population distribution in Slovenia and Austria, for example, it can be expected that in average in similar EU countries over € 300 million is needed each year due to obesity in the population of one million people (Atanasijević-Kunc et al. 2012). Similar result was evaluated by (Keaver et al. 2013) regarding the situation in Republic of Ireland where this number was estimated to be in 2009 over € 250 million. With forecasted increasing prevalence of this chronic disease (Ejima et al. 2013, Keaver et al. 2013) the burden is expected to grow at least proportionally.

As pointed out in several mentioned publications the greatest potential lies in prevention. Due to indicated problem complexity no simple solution is available. Behavioral changes cannot be sustained if important drivers of behavior are not considered in correct manner. Therefore a systems-oriented, multilevel framework has to be taken into account as is described for example in (Huang et al. 2009). Most concerning is the situation presented with block-diagram in Fig. 4 illustrating (in simplified manner – for detailed description see (Huang et al. 2009)) the local and global influences to obesity prevalence.

It is clear that no general intervention policy nor mathematical modelling result can be complex enough to resolve this immense trouble. As indicated in (Huang et al. 2009) individual behavior, regarding eating and physical activity are influenced by socio-environmental factors and by powerful biological processes. Behavior change cannot be sustained if these drivers of behavior are not considered in appropriate manner. The solutions need interdisciplinary consideration where

mathematical and simulation results can help in evaluation hypotheses addressing part of the whole problem.

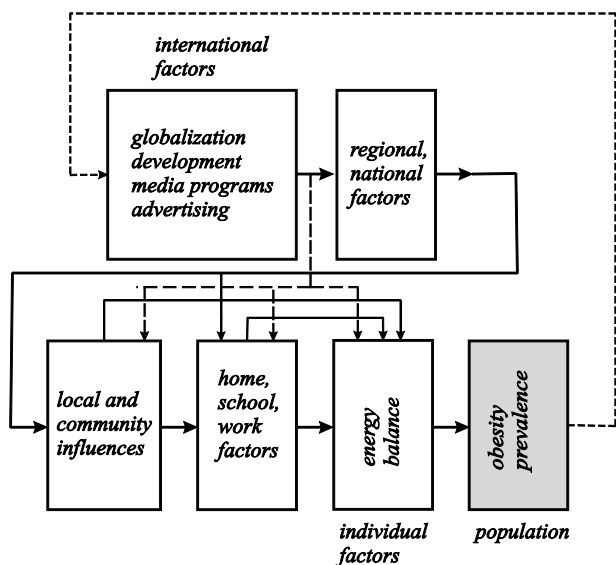


Figure 4: Influences to prevalence of obesity

Similar structure is addressed in (Van Koperen et al. 2013). Authors are describing in qualitative way the logic model used for implementation of strategies to prevent childhood obesity.

Childhood obesity prevention is, as already mentioned, addressed also in (Ramirez-Nafarrate and Gutierrez-Garcia 2013) using agent-based modelling approach advising to increase activity.

When investigating efficient healing and prevention programs an interesting result presented in (Preston et al. 2013) have to be mentioned. Described modelling results suggest that obesity in early adulthood is an important risk factor for mortality above age 50 and therefore an increased attention should be devoted on reducing the incidence of obesity at young adult ages.

## 5. CONCLUSIONS

In the paper we have analyzed some of recently published works addressing modelling results of obesity and related processes and problems. Main findings are the following:

At the level of an individual (person/patient) energy balance influence is well understood and efficiently described by ordinary differential equations.

Obese patients are rather unsuccessful in long-term body mass reduction due to psychological disturbances which are the result of very complex processes and individual descriptions are to be identified.

In spite of the fact that the nature of mentioned disturbances is complex, very good closed-loop control results can be expected if patient controllability is achieved. For this the integrative treatment must be adapted to patients' capabilities regarding the life style transformation and with this it is expected that the patient increases the treatment time to needed duration.

At the population level several modelling structures developed for contagious diseases were extended to the cases of non-communicable diseases, including obesity. Compartment models indicate that obesity prevention is eventually the most promising solution.

Due to computer technology development also discrete event and agent-based simulation results are very promising.

At the population level general closed-loop approaches are at the moment not possible because of problem complexity as all social, economic and political environment in addition to personal characteristic can influence the prevalence and incidence of disease development. So systematic problem solving asks for interdisciplinary approaches where partial modelling results should contribute to decrease of this preventable burden.

## REFERENCES

- Ajzen, I., and Madden, T., 1986. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control, *J. Exp. Soc. Psychol.* 22, pp. 453–474.
- Anderson, R.M, May, R.M., 1991. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, University Press.
- Angelova, P., Boyadjiev N., 2013. A Review on the Models of Obesity and Metabolic Syndrome in Rats, 2013. *Trakia Journal of Sciences*, No 1, pp 5-12, Trakia University, Available from: <http://www.uni-sz.bg> [accessed 12 May 2014]
- Astrup, A., Gøtzsche, P.C., Van de Werken, K., Ranneries, C., Toubro, S., Raben, A., and Büemann, B., 1999. Meta-analysis of resting metabolic rate in formerly obese subjects, *The American Journal of Clinical Nutrition*, Vol. 69, pp. 1117–1122.
- Atanasijević-Kunc, M., Drinovec, J., Mrhar, A., 2008a. Usage of modelling and simulation in medicine and pharmacy. *Slovenian Medical Journal*, 77(1), 57–71.
- Atanasijević-Kunc, M., Drinovec, J., 2011a. Social and Economic Burden of Obesity – Multi - Model Analysis. *Proceedings of the 33rd International Conference on Information Technology Interfaces ITI*, pp. 509–514. June 27-30, Cavtat / Dubrovnik, Croatia.
- Atanasijević-Kunc, M., Drinovec, J., 2011b. Burden of diabetes type 2 through modelling and simulation. Topics in the prevention, treatment and complications of type 2 diabetes. Rijeka: InTech, cop. 2011, str. 3-28, ilustr. [http://www.intechopen.com/source/pdfs/22092/In\\_Tech-Burden\\_of\\_diabetes\\_type\\_2\\_through\\_modelling\\_and\\_simulation.pdf](http://www.intechopen.com/source/pdfs/22092/In_Tech-Burden_of_diabetes_type_2_through_modelling_and_simulation.pdf). [COBISS.SI-ID 8750932]
- Atanasijević-Kunc, M., Sentočnik, T., Tomažič, S., Drinovec, J., 2012. Motivation problems in the process of mass reduction through modelling and simulation. *Proceedings of the 9<sup>th</sup> International*



*Multidisciplinary Modelling & Simulation Multiconference*, pp. 282–292, September, Vienna (Austria).

- Bruzzone, A. G., Novak, V., Madeo, F., 2012. Agent Based Simulation Model for Obesity Epidemic Analysis, Proceedings of the International Workshop on Innovative Simulation for Health Care, pp. 209–217, September, Vienna (Austria).
- Carroll, J.F., Chiapa, A.L., Rodriguez, M., Phelps, D., R., Cardarelli, K.M., Vishwanatha, J.K., Bae, S., Cardarelli, R., 2008. Visceral Fat, Waist Circumference and BMI: Impact of Race/ethnicity. *Obesity*, 16, 600–607. doi:10.1038/oby.2007.92
- Chang, V.W., Christakis, N.A., 2002. Medical modelling of obesity: a transition from action to experience in a 20<sup>th</sup> century American medical textbook. *Sociology of Health & Illness*, 24(2), 151–177.
- Chaston, T.B., Dixon, J.B., 2008. Factors associated with percent change in visceral versus subcutaneous abdominal fat during weight loss: findings from a systematic review. *International Journal of Obesity*, 32, pp. 619–628.
- Chow, C.C. and Hall, K.D., 2008. The dynamics of human body weight change, *PLoS Computational Biology*, 4(3), pp. 1–13.
- Christakis, N.A., and Fowler, J.H., 2007. The Spread of Obesity in a Large Social Network over 32 Years, *The New England Journal of Medicine*, Massachusetts Medical Society, pp. 370–379.
- Cohen-Cole, E., and Fletcher, J.M., 2008. Is Obesity Contagious? Social Networks vs. Environmental Factors in the Obesity Epidemic. *Journal of Health Economics*, 27, 1382–1387.
- Dong, Y., Rivera, D.E., Downs, D.S., Savage, J.S., Thomas, D.M., and Collins, L.M., 2013. Hybrid Model Predictive Control for Optimizing Gestational Weight Gain Behavioral Interventions, Proceedings of American Control Conference (ACC), Washington, DC, USA, pp. 1970–1975.
- Dreifus, C., 2012. A Mathematical Challenge to Obesity. *The New York Times*, Available from: [http://www.nytimes.com/2012/05/15/science/a-mathematical-challenge-to-obesity.html?\\_r=0](http://www.nytimes.com/2012/05/15/science/a-mathematical-challenge-to-obesity.html?_r=0) [accessed 12 May 2014]
- Ejima, K., Aihara, K., Nishiura, H., 2013. Modeling the obesity epidemic: social contagion and its implications for control. *Theoretical Biology and Medical Modelling*, 10(17), 1–13. Available from: <http://www.tbiomed.com/content/pdf/1742-4682-10-17.pdf> [accessed 12 May 2014]
- El-Sayed, A.M., Scarborough, P., Seemann, L., Galea, S., 2012. Social network analysis and agent-based modeling in social epidemiology, *Epidemiologic Perspectives & Innovations*, 9(1), 1–9. Available from: <http://archive.biomedcentral.com/content/pdf/1742-5573-9-1.pdf> [accessed 12 May 2014]
- Huang, T.T., Drewnowski, A., Kumanyika, S.K., Glass, T.A., 2009. A Systems-Oriented Multilevel Framework for Addressing Obesity in the 21<sup>st</sup> Century, *Preventing Chronic Disease*, Vol. 6, No. 3, A82, pp. 1–10. Available from: [http://www.cdc.gov/pcd/issues/2009/jul/09\\_0013.htm](http://www.cdc.gov/pcd/issues/2009/jul/09_0013.htm). [accessed 12 May 2014]
- Glock, B., Einzinger, P., Breitenecker, F., 2012. A Modular Architecture for Modelling Obesity in Inhomogeneous Populations in Austria with System Dynamics – First Step: A Population Model and How to Integrate it in a Disease Model, *Proceedings of the 9th International Multidisciplinary Modelling & Simulation Multiconference*, pp. 513–518, September, Vienna (Austria).
- Hall, K.D. 2006. Computational model of in vivo human energy metabolism during semistarvation and refeeding, *Am. J. Physiol. Endocrinol. Metab.* 291, pp. E23–E37.
- Hall, K.D. and Christine E. Hallgreen, C.E., 2008. Increasing Weight Loss Attenuates the Preferential Loss of Visceral versus Subcutaneous Fat: a Predicted Result of an Allometric Model, *International Journal of Obesity*, 32(4): 722, doi:10.1038/ijo.2008.14.
- Hall, K., 2010a. Mechanisms of metabolic fuel selection: Modeling human metabolism and body-weight change, *IEEE Engineering in Medicine and Biology Magazine*, pp. 36–41, doi: 10.1109/MEMB.2009.935465
- Hall, K. 2010b. Predicting metabolic adaptation, body weight change, and energy intake in humans, *Am. J. Physiol. Endocrinol. Metab.* 298 (3) (2010), pp. E449–E466.
- Hall K.D., Sacks G., Chandramohan D., Chow C.C., Wang Y.C., Gortmaker S.L., Swinburn B.A., 2011. Quantification of the effect of energy imbalance on bodyweight. *Lancet*. Aug 27; 378(9793), 826–37, (PMID:21872751), <http://bwsimulator.niddk.nih.gov> [accessed 7 July 2014]
- Hall, K.D., Butte, N.F., Swinburn, B.A., Chow, C.C., 2013. Dynamics of childhood growth and obesity: development and validation of a quantitative mathematical model. *Lancet Diabetes Endocrinology*, 1, pp. 97–105.
- Jackson, A.S., Stanforth, P.R., Gagnon, J., Rankinen, T., Leon, A., Rao, D.C., Skinner, J.S., Bouchard, C., and Wilmore, J.H., 2002. The effect of sex, age and race on estimating percentage body fat from body mass index: The heritage family study, *International Journal of Obesity & Related Metabolic Disorders*, Vol. 26, pp. 789–796.
- Jeong, S., Jo, Y.M., Shim, S.O., Choi, Y.J., and Youn, C.H., 2014. A Novel Model for Metabolic Syndrome Risk Quantification Based on Areal Similarity Degree, *IEEE Transactions on Biomedical Engineering*, Vol. 61, No. 3, pp. 665–679.
- Keaver, L., Webber, L., Dee, A., Shiely, F., Marsh, T., Balanda, K., Perry, I., 2013. Application of the UK

- Foresight Obesity Model in Ireland: The Health and Economic Consequences of Projected Obesity Trends in Ireland, *PLoS ONE* 8(11): e79827. doi:10.1371/journal.pone.0079827.
- Van Koperen, T.M., Jebb, S.A., Summerbell, C.D., Visscher, T.L.S., Romon, M., Borys, J.M., and Seidell, J.C., 2013. Characterizing the EPODE logic model: unravelling the past and informing the future, *Obesity Reviews*, Vol. 14, pp. 162–170
- Lean, M.E.J., Han, T.S., Morrison, C.E., 1995. Waist circumference as a measure for indicating need for weight management, *British Medical Journal*, Vol. 311, pp. 158–161.
- Levy, D.T., Mabry, P.L., Wang, Y.C., Gortmaker, Huang, T.T.-K., Marsh, T., Moodie, M., Swinburn, B. 2011. Simulation models of obesity: a review of the literature and implications for research and policy. *Obesity reviews*, 12, 378–394.
- Miksch, F., 2012. Mathematical Modelling for New Insights into Epidemics by Herd Immunity and Serotype Shift, Dissertation, Vienna University of Technology.
- Mokdad, A.H., Bowman, B.A., Ford, E.S., Vinicor, F., Marks, J.S., and Koplan, J.P., 2001. The continuing epidemics of obesity and diabetes in the United States, *JAMA*, 286 (10), pp. 1195–1200.
- Navarro-Barrientos, J.E., Rivera, D.E., Collins, L.M., 2011. A dynamical model for describing behavioural interventions for weight loss and body composition change, *Mathematical and Computer Modelling of Dynamical Systems*, 17 (2), pp. 183–203.
- Phillips, G., and Barton, A. 2014. Diet Therapies for the Treatment of Nonalcoholic Fatty Liver Disease (NAFLD): A Systematic Style Review, *European Journal of Nutrition & Food Safety*, 4(3): pp. 189–190.
- Preston, S.H., Mehta, N.K., Stokes, A., 2013. Modeling Obesity Histories in Cohort Analyses of Health and Mortality, *Epidemiology*, Vol. 24, No., doi:10.1097/EDE.0b013e3182770217.
- Ramirez-Nafarrate, A. and Gutierrez-Garcia, J.O., 2013. An Agent-Based Simulation Framework to Analyze the Prevalence of Child Obesity, *Proceedings of the 2013 IEEE Winter Simulation Conference*, pp. 2330–2339.
- Scott, G., Brewer, H., Jr, Cleeman, J., Sidney, S. and Lenfant, C., 2004. Definition of the Metabolic Syndrome: Report of the National Heart, Lung, and Blood Institute/ American Heart Association Conference on Scientific Issues Related to Definition. *Circulation*, 109: 433–43.
- Sentočnik, T., Atanasijević-Kunc, M., Drinovec, J., Pfeifer, M., 2013. Efficacy Analysis of a Body-Mass-Reduction Treatment Using Mathematical Modelling, *Mathematical and computer modelling of dynamical systems*, 1–24, <http://www.tandfonline.com/doi/full/10.1080/13873954.2013.814680>, doi: 10.1080/13873954.2013.814680.
- Šalamon, T., 2011. *Design of Agent-Based Models, Developing Computer Simulations for a Better Understanding of Social Processes*, Academic Series, Czech Republic.
- Thomas, M.D., Ciesla, A., Levine, J.A., Stevens, J.G., Martin, C.K., 2009. A mathematical model of weight change with adaptation. *Mathematical Biosciences and Engineering*, 6(4), 873–887.
- Westerterp, K., Donkers, J., Fredrix, E. and Boekhoudt, P., 1995. Energy intake, physical activity and body weight: A simulation model, *British Journal of Nutrition*, 73, pp. 337–347. [dvoprostorni model, parametric temeljijo na statistiki]
- WHO - World health organization, 1995. Expert Committee Physical Status: The Use and Interpretation of Anthropometry. *WHO Technical Report Series* no. 854. Geneva.
- WHO - World health organization, 2000. *WHO Technical Report Series*, No. 894, Geneva. Available from: [http://whqlibdoc.who.int/trs/WHO\\_TRS\\_894.pdf?ua=1](http://whqlibdoc.who.int/trs/WHO_TRS_894.pdf?ua=1) [accessed 3 June 2014]
- WHO - World health organization, 2013. Global Health Observatory (GHO). Available from: [http://gamapserver.who.int/gho/interactive\\_charts/ncd/risk\\_factors/overweight\\_obesity/atlas.html](http://gamapserver.who.int/gho/interactive_charts/ncd/risk_factors/overweight_obesity/atlas.html) [accessed 29 March 2014]
- WHO - World health organization (Reviewed May 2014). Obesity and overweight, Fact sheet No. 311. Available from: <http://www.who.int/mediacentre/factsheets/fs311/en/> [accessed 3 May 2014]
- Zauner, G., Popper, N., Einzinger, P., Miksch, F., Urach, C., Zechmeister, I., Endel, G., Breitenecker, F., 2012. Hybrid Combination and Modular Modelling Techniques in Health Technology Assessment – An Example Based Explanation for Communicable Diseases, Preprints of the 7<sup>th</sup> Vienna Conference on Mathematical Modelling, pp. 365.

# REAL TIME INTERACTION AND HANDLING TOOL OF HISTORICAL DATA ON TIME-LAPSE RECORD FOR CONSTRUCTION MANAGEMENT

Livia Fernandes, MSc.<sup>(a)</sup>, Gerson G. Cunha, DSc.<sup>(b)</sup>, Celia Lopes, DSc.<sup>(c)</sup>, Luiz Landau, DSc.<sup>(d)</sup>

<sup>(a)</sup><sup>(b)</sup><sup>(c)</sup><sup>(d)</sup> Federal University of Rio de Janeiro – LAMCE/COPPE

<sup>(a)</sup>[lfaria@lamce.coppe.ufrj.br](mailto:lfaria@lamce.coppe.ufrj.br), <sup>(b)</sup>[gerson@lamce.coppe.ufrj.br](mailto:gerson@lamce.coppe.ufrj.br), <sup>(c)</sup>[celia@lamce.coppe.ufrj.br](mailto:celia@lamce.coppe.ufrj.br), <sup>(d)</sup>[landau@lamce.coppe.ufrj.br](mailto:landau@lamce.coppe.ufrj.br)

## ABSTRACT

This paper introduces a novel tool for interaction and visualization of real-time time-lapse videos, which can be applied to small, medium, and large-sized civil construction projects. The tool called Timelapse Player Bidimensional enables the user to visualize and analyze high-resolution photographs from a large remote database catalogued chronologically using the Quadtree technique. The usage of the Timelapse Player Bidimensional supports the user with a detailed temporal analysis in many aspects, such as process identification, productivity, delivery of material, accident investigation, disputes, delays, and monitoring the progress of construction sites.

Keywords: Photographic Recording, Time-Lapse, Collaborative Decision Making, and Tangible Interface

## 1. INTRODUCTION

New methods for managing projects through visualization technologies are becoming a standard, especially in outdoor scenarios and in large enterprises (Portalés et al., 2010). Photographic images are being used to monitor and assess construction operations for managerial issues. Companies are using pictures that gather information on activities related to the construction site. These records can be of great value for the identification and solution of management problems in constructions, as shown in Abudayyeh (1997).

With the advancement of technology and high performance computing, it is possible to use photographic records databases for tool condition monitoring (Dutta et al, 2013) and as parts of tools for decision-making. Image processing techniques are being applied to interpret and analyze captured data using an approach that can be manual or automated.

Time-Lapse is a technique that first captures a sequence of images at lower frequency, then the events it's registering, and afterwards reproduces the content faster than the original event. It is also important to consider the artistic relevance of photographic recording in time-lapse production and how it alters the perception of project development, by allowing decision makers to view weeks, months, or years of work in a compact time frame.

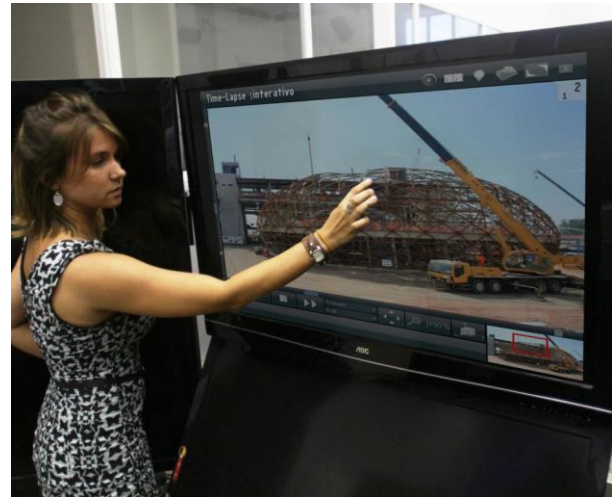


Figure 1: Visualization tool being used with a touch table device.

This paper presents an engineering project analysis system based on photographic event recording.

We provide a way of interacting with these records through an interactive media player capable of navigating through the results using a two-dimensional timeline controlled by the user in real time.

The system can be treated as a visualization tool (see Figure 1) and is particularly interesting for analyzing works from a remote location. Since it is an interactive and intuitive tool, it allows temporal analysis from the visualization of a selected period.

### 1.1. Related work

Project management through continuous photographic records to generate videos is not a system commonly employed for managing construction supervision, although the concept of using this feature as a monitoring tool is not entirely new. Few works discuss the relationship between fast image access and tangible interaction. Abed and Arditi (2002) presented such a proposal, discussing storage space issues and video production, but without considering access possibilities beyond simple video playback.

Usually this kind of image treatment and manipulation can be found in microscopic photography and satellite images; little has been discussed on this subject in the field of civil engineering. The work of Yamaoka, Doerr et al. (2011) overviews the display wall option for viewing high-resolution images.

## 2. OVERVIEW

The purpose of this paper is to present an approach to interactively visualize a large collection of high-resolution images in a display environment for construction managers. Timelapse Player measures and highlights critical points in the construction monitoring, identifying processes, safety, problems with suppliers, disputes and activity follow-up.

This project intends to create a series of visualization solutions capable of interacting with collected data through different physical environments. The concepts applied can be used in other tools (as shown in Figure 2).

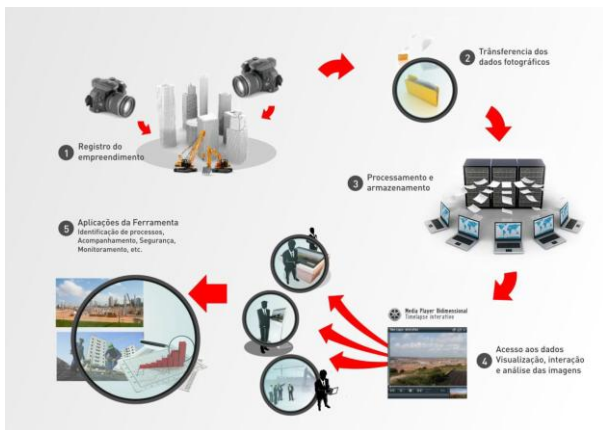


Figure 2: Scheme of the proposed system

Timelapse Player Bidimensional manipulates high-resolution digital images that, when displayed in sequence, create a movie with images from the construction site. The software runs on a remote computer, and was designed to operate with an undetermined number of cameras collecting these images through a video server located at the construction site.

The tool requires the use of several components, including a computer, at least one high-resolution camera to take pictures, and a video camera for real-time visualization (live view), a video server, internet connection, coaxial cables, and accessories.

The tool is a solution comprising five steps:

1. Data collection;
2. Data transfer;
3. Cataloging, storage and processing;
4. Description of images composition;
5. Data visualization, interaction and image analysis.

### 2.1. Data capture

Step 1 (Figure 2) represents an external environment of construction in which the solution for tool operation must be installed. At this stage, it is important to choose the installation judiciously so that the quality of photographic records is not compromised.

It is essential to perform a study of the construction site to evaluate if a structure could grow in front of the capture point. Climate conditions, sunrise and sunset hours should also be taken into consideration. The development of this project was based on construction, however, the tool can be used in other fields, including indoor environments.

The image capturing system defines the quality and frequency of the images to be analyzed. As the main interest in this case is the analysis of remote construction sites, a device capable of capturing high definition images was chosen, allowing a better perception of the details of the events.

It was necessary to create a physical metal structure to fix the devices and protect them against weather events, since the capture system is fixed mostly in outdoor locations (see Figure 3).

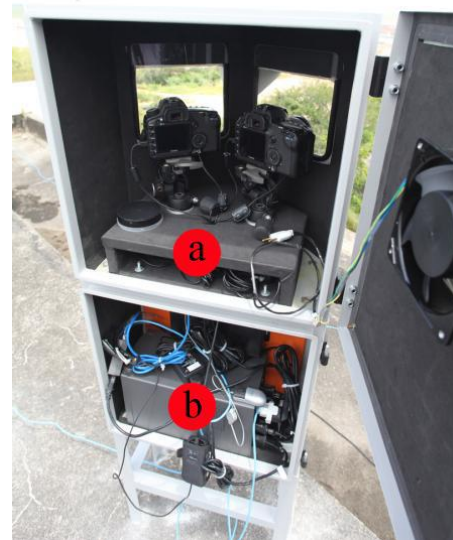


Figure 3: Example of the capture system. a) Cameras; b) Local machine for storing data.

The idea is to fixate the capture point and interact with the system remotely. Thus, it is necessary to create a tool that allows the manipulation of cameras and automatically stores captured images into a disk. This tool must have versions for client and server, and a local network must be created for the communication between them.

The server version should support the handling of files obtained by the camera, making local copies and sending them through the network. The client version must support the control of camera properties (such as aperture, focus, zoom, etc.) and also be able to remotely capture images and change the frequency of recordings.

The periodicity of the image capture is critical for the alternation of results. It will determine not only the amount of data to be stored, but also control the effect of continuity in time/space in a sequential viewing of the images. For this project, we used a capture interval of five minutes, but this interval may change depending on the speed in which the events take place in the project to be analyzed.

## 2.2. Data Transfer

Step 2 (Figure 2) is responsible for transferring data collected at the construction site to a local HD. Then, it will be transferred to a remote database. The transmission of the material can be made by wireless or cable internet. It is important to note that the internet connection should have a good throughput to avoid data loss or delay in the process.

## 2.3. Cataloging and data processing

In step 3 (Figure 2), the images are received from the construction site, indexed and catalogued according to the order of creation of the files. The time when the image was taken is used to create its identifier according to the format:

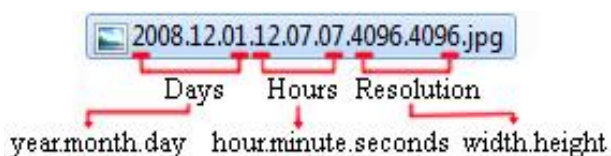


Figure 4: Identifier format.

The transferred photographic material is on a hard drive with online access for an eventual research. The photo storage location must be decided by the project manager. Data replication from the remote area to the storage station must be constantly supervised to prevent data loss. The backup physical area must be safe and in stable conditions for preserving the material, specifications such as HD type and storage space may vary according to the necessity of the project.

## 2.4. Description of image composition

To support the process of cataloging, an array of registers is created containing all existing intervals during the time of capture. The size of the array is determined as follows:

- no. of columns = number of days;
- no. of rows =  $\frac{1440 \text{ (no. of minutes in the day)}}{\text{time interval (in minutes)}}$

Thus, each array element is a binary number (see Figure 5), which indicates whether or not there is a photographic recording relative to the pair (minute/day) in time. Considering an element  $i$  of the array, the moment in time can be calculated as follows:

- minute = time interval \*  $L_i$ ;
- day =  $C_i$ ,

where  $L_i$  is the line of element  $i$ , and  $C_i$  is the column.

|                |      |             |   |   |   |   |   |   |   |   |   |    |    |    |    |
|----------------|------|-------------|---|---|---|---|---|---|---|---|---|----|----|----|----|
| MINUTOS eixo Y | 1440 | 0           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 1320 | 0           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 1200 | 0           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 1080 | 0           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 960  | 0           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 840  | 0           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 720  | 0           | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 600  | 1           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 480  | 1           | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 360  | 1           | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 240  | 1           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 120  | 1           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                | 0    | 1           | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1  | 1  | 1  | 1  |
|                |      |             | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|                |      | DIAS eixo X |   |   |   |   |   |   |   |   |   |    |    |    |    |

Figure 5: Example of an array of record with 13 days and predefined interval between captures.

The image processing of this project refers to necessary adjustments to align the images to data segmentation operations. In these operations, pyramid-based methods are applied to all images in order to generate representations of intermediate images, depending on the desired level of approximation (for expansion). This process aims to resize the image and recursively divide it to form a pyramid of images with different resolutions and number of partitions at each level. This process takes place until the latter reaches a certain level of resolution and a linear interpolation chain is created.

The tilling process allows quick access to section images without the need to open each high-resolution image in every manipulation. This allows the image shown (by zoom and pan) to be seamlessly manipulated in real time, in a very efficient manner.

## 2.5. Data visualization, interaction, and image analysis

Stage 4 (Figure 2) is the manager manipulation using material collected by the visualization tool (Timelapse Player Bidimensional), which promotes temporal visualization of the time-lapse selected by the user. The manager, while using Timelapse Player (Figure 6), can choose the day, month, and hour of a specific past event or from current images. After choosing, it is possible to analyze the image sequence using any frame per second rate preferred. And, finally, the enterprise manager can analyze diverse situations that could compromise the continuity of the work.





Figure 6: Visualization tool in use.

Recorded images are associated with a position in the Euclidean space, and their manipulation will be possible by switching coordinates in this space. The x-axis will be responsible for the alternation of day, the y-axis by the alternation of minutes during the day, and the z-axis remains fixed.

This step uses the array of the registers generated in the previous step to see if a registry has an associated image. In other words, the array shows the periodic intervals in which the captures were made.

At this moment, it is possible to create a playlist (containing system coordinates) to generate real-time video from captured data. The images associated with the coordinates from this list are displayed sequentially, creating a video that shows project changes during the time interval defined by the list in real-time.

The user can also interact with images through panning and zooming operations. To do this, the pyramid-based method is critical.

As the captured images are often of high resolution, the process of loading, viewing, and switching between them would become too expensive and time consuming, making it a nonviable approach for generating an interactive video.

To solve this limitation, the display system allows the use of zooming and panning through the location of the segmented images, and the display of the visible portion of the image in the display window.

### 3. INTERACTION

System interaction is performed through an intuitive interface that suggests system use as a media reader, having two dimensions. Such dimensions can be represented by displacement bars in the interface, simulating a coordinate system with two axis. The manner of interaction and system features can be explained by the interface, pictured in Figure 7.



Figure 7: Timelapse Player Bidimensional Interface

#### 3.1. Data Loading

As numbered in Figure 7, functions are divided as follows:

- Item 3 of Figure 7, registration matrix reading and storage location;
- Item 8 and 9 of Figure 7, Image visualization: Choice of coordinates (minutes, days) to identify the images that will be displayed.
- Item 17 and 19 of Figure 7, Zoom and panoramic vision: Give a more detailed analysis of happenings;
- Creation of video flows: Defines the sequence of images that should be shown;
  - Vertical: Time changes and days are fixed (item 14 and 18 of Figure 7);
  - Horizontal: Time is fixed and days change (item 15 and 18 of Figure 7);
  - Mixed: Time and days alternate, creating a time line (item 7 of Figure 7).
  - Standard player control that can be seen on items 10, 11, 12, 13 of Figure 7;
- Item 3 of Figure 7, Save video button for future evaluation and follow-up;
- Item 2 of Figure 7, Notes can be made on the video and compared to follow-up reports;
- Item 1 and 6 of Figure 7, overlapping of layers for comparison between electronic plans/models and the work evolution;
- Item 20 of Figure 7, LiveView real-time follow-up using an auxiliary security camera. Item 4 is the maximize and minimize screen. Item 16 is movement (from left, right, up or down). Item 5 is the close button.

Timelapse Player Bidimensional was a tool based on studies on civil construction to help managers in their decision-making process, in addition to optimizing activities that require documentation from past events, since quick image location can be easily performed. The



user can get the material for analysis with month, date, and time precision. Such interaction is possible to the catalogue system and tool search.

#### 4. APPLICATION

This paper shows an analysis of the use of the tool being discussed, and how it can help civil engineering in small, medium, and large ventures, helping organizational management and activity control for making decisions.

##### 4.1 Flaw identification and assemblage and activity delays

The moving point, called NVC-Móvel, was installed with the purpose of following-up the assemblage of the Núcleo de Visualização Colaborativa – NVC structure (Figure 8).



Figure 8: Image sequence from the Núcleo de Visualização Colaborativa (NVC)

Flaws and pathologies in metal structures were identified in records from almost 1 year. The NVC-Móvel camera was placed for a certain amount of time in a specific location, decided by the construction manager.

During the assemblage, it was observed that the metal structure moved and was excessively deformed in 2 different points. The Timelapse Player screen pictured in Figure 9 shows the metal structure falling. In order to avoid the structure was put at risk, a crane was rented to sustain it until the problem was solved.



Figure 9: Timelapse Player screen showing the NVC metal structure.

During the NVC construction video, it is possible to see the image sequence that shows the structure slowly falling over the course of several days. A detailed analysis can be done using the tool, as pictured in Figure 10. The zoomed image shows structure flaws in detail.



Figure 10: Close image of the NVC metal structure falling.

Another discovery took place during the observation of the installation of the pre-made plates, which were delivered having deformities, creating problems during the assemblage. The plates were irregular, which made it difficult for them to fit in the structured and compromised the finishing process. The deformity of the plates (Figure 11) added extra costs to the venture's budget, in addition to compromising the delivery date.

From the images collected by the NVC-Móvel point, it was possible to observe, with help from the tool, the flaws in the material delivered for covering and finishing the NVC. The zoom performed by the tool (Figure 12) proves the lack of quality for assembling the project.



Figure 11: Timelapse Player screen showing the defective plates during the NVC assemblage.



Figure 12: Close image of NVC's plates.

The analysis also identified the downtime of equipment and human resources in the construction site. Some of the problems during the NVC assemblage forced equipment and personnel to be moved, making a negative contribution for the delay of the task sequence in the construction site.

During the image capture process in both points, Torre-Pan and NVC-Móvel, some unexpected events took place, caused by server failure, Internet, or cameras. Analysis results were not affected by these problems, however, it should be mentioned that maintenance needs to occur frequently so that the nature of records and the loss of relevant construction information are not lost.

## 5. CONCLUSION AND FUTURE WORK

Photographic evidence from the venture give a variety of new manners of analysis. Photography-associated video is a great source of research that surpasses the operational environment since the strategies used can be perfected, and new manners of analysis and applications can be defined.

Typical applications include productivity analysis, accidents, investigations, litigations, delays, construction operation remote monitoring, educational videos, etc. Such applications require the capture and reproduction of the image sequence through a tool that

displays the construction operation video for a detailed analysis.

Timelapse Player Bidimensional presents a completely new approach for visualizing a remote, chronologically organized, high-definition image bank. The tool exhibition system consists in the use of two techniques: tiles bounds, which are regular divisions that reduces image processing, making loading lighter, and *Quadtree*, which allowed quick access to images due to space subdivision.

Aside from digital images being a proof with legal value, they can be directly consulted when searching for process identification information and activity follow-up.

This paper can be extended to integrate a project management system such as MS Project and Primavera. They are commonly used for managing projects and allow a more detailed control of events, creating spreadsheets and diagrams showing the progress of activities.

## ACKNOWLEDGMENTS

This research was supported by Federal University of Rio de Janeiro – LAMCE/COPPE.

## REFERENCES

- Abed, J. and D. Arditi (2002). "Time-lapse digital photography applied to project management." Journal of Construction Engineering and Management-Asce **128**(6): 530-535.
- Abudayyeh, O. (1997). "A Multimedia Construction Delay Management System." Microcomputers in Civil Engineering 12: 183-192.
- Portalés, C., J. L. Lerma, et al. (2010). "Augmented reality and photogrammetry: A synergy to visualize physical and virtual city environments." ISPRS Journal of Photogrammetry and Remote Sensing **65**(1): 134-142.
- Yamaoka, S., K. U. Doerr, et al. (2011). "Visualization of high-resolution image collections on large tiled display walls." Future Generation Computer Systems-the International Journal of Grid Computing and Esience **27**(5): 498-505.
- S. Dutta et al. (2013). "Application of digital image processing in tool condition monitoring: A review". CIRP Journal of Manufacturing Science and Technology (6):212-232.

# MAPPING OF WIRELESS TECHNOLOGIES TO SUPPORT REAL TIME LOCATION SYSTEMS FOR TRACKING RESOURCES OF LARGE ENTERPRISES

Larissa P. M. Cruz <sup>(a)</sup>, Luiz Landau <sup>(b)</sup>, Gerson Gomes Cunha <sup>(c)</sup>, Maria Celia S. Lopes <sup>(d)</sup>

Universidade Federal do Rio de Janeiro/COPPE <sup>(a)</sup> <sup>(b)</sup> <sup>(c)</sup> <sup>(d)</sup>

<sup>(a)</sup> larissapmc@gmail.com, <sup>(b)</sup> landau@lamce.coppe.ufrj.br, <sup>(c)</sup> gerson@lamce.coppe.ufrj.br, <sup>(d)</sup> celia@lamce.coppe.ufrj.br

## ABSTRACT

**This study aims to present a methodology that helps to select the most appropriate technology for resource tracking in large projects such as refineries, offshore, mining, among others. The method used is the correlation between the metric performance parameters and technical characteristics of the sensors that Real Time Location System (RTLS) are based, such as: Infrared, Ultrasound, Bluetooth, UWB, RFID, Zigbee and GPS. Thus, is obtained a table with qualitative indicators that are subsequently aligned in ascending order of operation, generating a radar chart to view evaluative rating clearly. The development of this method and its application aims to minimize mistakes in planning and achievement, enabling the development of RTLS's systems with more appropriate technologies.**

**Keywords:** Real time Locating Systems, wireless, metric performance, appropriate technologies

## 1. INTRODUCTION

It's possible to notice in the current Brazilian panorama of civil construction some large technological advances. Some significant investments were spent in the last few years, mainly in the industrial base, such as refineries, shipyards, mining and hydroeletrics. These large ventures have as a characteristic on the construction or even the operational phase, the capacity to encompass large areas divided by activities. All of these activities are naturally dynamic and encompass a lot of processes which must follow protocols and rules that are generally monitored based on results. But, having the knowledge about all the practices and measuring the evolution of the activities opposing the investment and the operational chronogram may be a difficult task depending on how big is the enterprise.

So, the possibility to identify and track points, which are critical and essential to the enterprise through automated processes, can be a powerful resources management tool. Specially, detecting failures on the production and risk areas may not only affect directly the enterprise's budget but also represent a significantly decreasing in the number of job accidents.

As a result, the calculation of the activities and resources' statistics has to be extremely detailed and studied as well as the aggregated values of the actives must have priority, aiming the production chain's dimension.

Through this study, it's intended to indicate one actives classification base in order to facilitate the identification of critical moments on the processes. Thus, it'll be indicated through a table the behavior of the tracking sensors related to the nature of the objects' materials which will be contained in the devices or will possibly act as an obstacle in the system's application environments.

Moreover, taking the GPS and active capture sensors as main line, a table comparing the techniques used by the

different tracking systems will be offered. This table will indicate what is possible to achieve regarding precision, easiness to install, cost and other characteristics.

By the use of the table with indicators measuring the tracking systems' operation capacity together with visualization of the generalized classification of the sensors' operation, it will be generated a mesh of the positioning related to the operation of the system.

Therefore, the main goal is to obtain a mapping of the most adequate technology through the correlation of the sensors and also prioritize their performance. With the help of the qualitative indicators, which were elaborated to the Selection Table RTLS, it's intended that the user, having or not having technical knowledge, can visualize the necessities that the ideal system may has in order to solve the great demand of the project, minimize mistakes regarding the project and the installation. This will create safer choices.

## 2. HOW THE TRACKING AND REAL-TIME LOCATION SYSTEMS WORK: PRINCIPLES AND TECHNIQUES

According to TAPIA et al (2011), a location system is basically constituted by three entities, fixed sensors, mobile modules and an element that will manage the devices or sensors network, depending on the chosen technology. The mobile modules are monitored by fixed unities that are positioned on higher levels compared to the rest of the system's components.

According to MALIK (2009), the transmitters operate through tags and fixed landmarks and they can act as transmitters, receivers or both, generating many technological combinations.

The sensors are part of the positioning tool, which the RTLS solutions are based. They can have different types with different operating principles, such as geo-positioning by GPS, cellphone tracking or systems that use tracking through RFID passive tag's. The information about location, usually don't include speed, direction or spatial orientation.

### 2.1 Technical characteristics

RTLS are used most of the times, in confined areas. The tags are attached to items that are going to move with the purpose of knowing their movement, tracing the route, or merely monitoring them. The landmarks RTLS, called AP- Access point, can be either transmitter or receiver, distributed in the defined area of range for data collection. Most of the cases, the more landmarks are installed, the better will be the accuracy of the localization until the performance find some technological restrictions (HJELM, 2006).

There are some models of projects of real-time localization systems. However, there are two primary tracking models: Location by shock points and relativity of coordinates. The first consists of fixed networked readers that will find identification signals transmitted by short-

range "tags" in motion. And the latter, on which location signals of a tag are received by multiple readers connected on the same network, estimating the position by the calculation of location algorithms such as trilateration, multilateration or triangulation (MALIK, 2009).

## **2.2- Techniques of tracking and tracing.**

The tracking systems are categorized according to the techniques used to estimate the location of mobile devices. Following, there is a summary of the location techniques and then the operation of the algorithms used in location systems.

### **a. Direct identification**

These are devices that are only detected when they are close to the reader and may be attached to antennas or even portals. As an example, we can observe infrared-based systems, in which information is transmitted when the sender and receiver establish eye contact. Another example is the system based on "radio-frequency identification" (RFID) present in the access cards to restricted areas.

### **b. Difference of the time of emission and perception of the sign**

This technique consists of the perception of the difference of time between the spread of signal, considering the emission and the arrival at the receiver of the network components. It's necessary to previously know the velocity that the signal spreads, together with the measure of time, in order to estimate the distance of the device in relation to a lot of raised points. Among the techniques based on this notion, the most used are: Time-of-Arrival (ToA), Times-Difference of Arrival (TDoA) e Time-of-Flight (ToF) Pags-Zamora, 2006).

### **c. Angle of arrival - AOA**

Through prior knowledge of the distance between the various receptors and angles of arrival detected in each landmark, it is possible to estimate the object's position via trigonometric functions (HJELM.J., 2006).

### **d . Received-Signal-Strength-Indication (RSSI)**

RSSI are values that make a sensor able to determinate the location of a mobile device, when there is an approximation between them, by analyzing the intensity of the received signal. From these RSSI values, there is also the possibility of knowledge of the location through the interpolation of three reach balls from three sensors calculated from their centers. The area that matches will be the likely location of the device. This method is called Triangulation.

### **e. Location by pattern - Fingerprinting**

In order to have a bigger precision, there is the technique of "Fingerprinting" or location by pattern. It calculates the position through the intensity of the received signal and its default behavior, mainly when it comes to its location, this way, forming a database. This technique is divided between the "online" and "off-line" phases. In the "off-line" phase, through the values of the RSSI the points are mapped in an area of interest. During the "online" phase, this process is used to create parameters of the propagation model. According to KAEMARUNGS (2005), the location of the mobile station is determined by the identification of the closer points, present in the "Fingerprinting".

## **3. DEFINITION OF THE ENVIRONMENT, ASSETS AND PROCEDURES THAT ARE GOING TO BE TRACKED**

We can take "indoor" as the environments that have a delimitation of space, with previously known obstacles such as buildings subdivided between walls or between floors and

fixed points of entry and exit. For this study, it's possible to classify environments "outdoors" as uncovered areas, without assuming to be free of obstacles. In the case of big ventures, it's a large area, that is able to present subdivisions and mainly uncontrollable interferences that can't be measured most of the times. These environments present the biggest complexity, when it comes to the control of the relation between devices that form the wireless network, tracking targets and installing antennas in strategic points for monitoring, due to the coexistence of a diverse range of materials and processes.

## **3.1. Processes and methods of definition of the actives**

Processes and methods of definition of the assets defining a method of evaluation and hierarchy of labor input can be an essential step so that it becomes possible to implement monitoring tools and process automation. Once the nature of the environment of work, of the materials, of the tools and the own routine of work of each enterprise can be diverse and even unstable, we propose a kind of roadmap with guidelines that will assist an essential phase for the implementation of a monitoring system: data collection. Considering that the review of the literature doesn't present an automatic way for a definition of guidelines and to make the scope of the project more consistent by defining critical points and identifying productivity we suggest an approach that will transform parameters in indirect indicators of need. The assessment should follow these hints:

1. Are there indirect parameters<sup>1</sup>?
2. If yes, how can they be measured?
3. Once collected and classified, how can they be converted to direct parameters<sup>2</sup> in order to be analyzed and represent future effective improvements in the project?

## **3.2 Definition of the object and indirect parameters: materials, machinery and people.**

During the parameters definition phase It's interesting to see the integration with the existing project management systems (PM's - Project Management), which approach performance standards such as Project Management Pyramid.

At this moment, the actives which will be tracked according to their priority order in the enterprise operating context will be enumerated.

### **a. Materials: Inventory and Quality Control**

About the elements that comprehend the big enterprises industry, according to CHENG et. Al. (2011), construction materials, installation and base tools are estimated to be between 50 and 60% of the construction total cost. During the project phase, these materials are classified among three categories: Off-the-shelf: easy to replace, Long-lead Bulks: items to weigh, Engineered items: hard to replace. These different categories vary in cost, waiting time for supplying and waiting time for replacing.

<sup>1</sup> It is possible to understand indirect parameters as actives such as materials, equipment, tools and people that may have their paths and activities tracked and delimited.

<sup>2</sup> Direct parameters are the result indicators of this monitoring that come in the shape of data. This data will serve as base to automated routines and enterprise's processes.



Generally, engineering items have a high cost and they are offered in a small quantity. These items have unique properties and require not only a long time to be replaced but a bigger plan for their field use.

The nature of the material can be an obstacle to some kind of sensors. Metallic, transparent and liquid materials cause interference, spoiling the operation of some technologies. We will present further a table that classifies the technologies' performance related to the material.

Both the active's value in the process' context and the easiness to replace them are factors that will also guide the choice of the sensors, because many times, the investment in the technology may be bigger than the profit generated.

The storage requirements, the handling and the transportation will also define the structure's design, which will be built to the RTLS operation. As an example, it's necessary to relate the storage way to the packing ways: pallet, metal boxes, cardboard boxes.

The ways of transportation, regarding the materials that are moved, may also define the structure. As example, it's necessary to measure the size of the trucks and the forklifts in order to make them adequate to the installation of the monitoring portals.

Based on the technical restrictions of the active sensors and the GPS, which works as base of the tracking systems, a table was built. This table establishes a relationship between the material's nature of the possible trackable targets and sensors' operating limitations showed against these materials. On the table 1 – Indicator of sensors operation related to the nature of tracked objects' material, it's possible to understand that the X (red color) is used for situations when there is no good operation, or even when there are blocking failures. In the case of V signal (green color), there is a possibility to install the system and its operation won't be connected to the inter-related conditions in the table. However, it's necessary to highlight the fact that in some situations represented by X, there's a possibility to avoid this operating difficulty using hybrid hardware solutions, or even increasing the number of antennas to transcend the obstacles.

These information were taken from tests presented in varied studies that served as theoretical base to this study.

**Table 1:** Indicator to the operation of the sensors related to the nature of the tacked objects' materials.

| SENSORS | IV | US | BT | UWB | RFID | ZIGBEE | GPS |
|---------|----|----|----|-----|------|--------|-----|
| LIQUID  | V  | V  | V  | X   | X    | V      | V   |
| GLASS   | V  | V  | V  | V   | X    | V      | V   |
| METAL   | X  | X  | X  | X   | X    | V      | V   |
| WOOD    | X  | X  | V  | V   | V    | V      | X   |
| PLASTIC | X  | V  | V  | V   | V    | V      | V   |

**b- Machinery**

It's possible to attach sensors to the equipment, allowing the localization of small tools which are not only used by different teams but also necessary to the activities.

These sensors can store information regarding the use of these tools, such as time, local and who used them. These data may indicate to the responsible technicians if the equipment is worn out, defining a correct moment to start the maintenance. The sensors may still contain information

about permission of use related to the tool that the operator is allowed to use.

**c- People: monitoring safe areas**

To begin the planning it's necessary to make a calculation of operators' behavior general statistics. Both the safety standards and the possible restrictions in the work place are information that must be inter-related in order to make possible having a map of the risks and their correct evaluation. The occurrence of risks increases when the workers either do their job away from the safety inspectors or out of the control zones or near the work places boundaries.

It's important to elicit that the activities prediction with their routines and risks prevention must be planned during the project phase (PTD – Prevention through Design). So, it's essential to have the measurement of both the area and the resources' path. The monitoring of the workers and the analysis of their routine and the management of the actives must be shaped to make available the correction of the project. This can lead to improvements in both the safety and the performance of the activities.

As an example, it's possible to permit or deny the access of people from the group to restricted areas. It would avoid the exposure of people to areas of risk or even the contact with unauthorized people in places that valuable objects or even content with contamination are stored.

- Monitoring of the activities

In case of people, the sensor is attached to objects that are going to be part of the employees' uniforms, as badges or tags attached to helmets.

These accessories will be able to record information about schedule, for example, allowing that the journey of work to be automatically monitored. They can also work as keys of access to areas where the managers are going to determine as appropriate or not for each employee. This automation would increase the control over the lost objects, avoiding the embarrassment of those employees who may not have any relation with what happened.

The badges can be keys of access to some tools, stricting the use to employees who are trained.

Another feature of this track is to generate data as execution time of tasks and locations on which employees are passing by. These are factors that will serve as tools for analyzing team performance.

In Picture 1, it's possible to notice a representation of the field implementation of the monitoring system for employees, where 1 indicates a fixed portal signal reading (receiver) and the number 2 represents the transmission signal tag with information carried by the employees.



**Picture 1:** Representative scheme of monitoring workers

**3.3. Definition of project strategy: taking decision for the adoption of the system at the organizational level of the enterprise.**

The cycle of movement of assets can pass through the following cycles: Open and Closed.

#### a- Solutions to open cycles

The solutions to open cycles comprehend multiple organizations and tend to benefit multiple industry sectors as control of retail items and tracking of raw material.

The “Business Case” for open cycles is more complex to develop because it’s necessary to identify problems in the whole supply chain. The value of information at various points and of different interested parts should be quantified by evaluating the benefits of different natures.

#### b- Solutions to closed cycles

The solutions to closed cycles solve specific problems as lost, strayed or stolen assets, manufacturing flaws, storage that generates waste and lost time due to lack of knowledge in the implementation of activities. This “business case” presents less difficulties than the later, because the problems will be identified in only one organization. Therefore, quantification of losses that relate to assets becomes more direct and controllable. In terms of investment, the same organization that pays for deployment experiences the benefits.

### 3.4. Site analysis: lifting restrictions of the deployment environment RTLS (Site Survey)

A crucial activity to the proper application of the tracking system is the analysis of the place where the field monitoring will work. Through the observation of the routine of performance of tasks, it should delimit favorable conditions and the limits that the space offers. This way, the installation of means of capture of data should be well performed, so that the next stage together with the technical information sensors obtain the final choice of the system. The collected information will help, for example, in the choice among the structures of capture as portals or handheld readers.

This step influences directly in the chance of success, since the good adaptation of the team in relation to the system is directly connected to the little change in operating habits. In Picture 2, we can see an illustration of the performance of a portal in areas where there is flow of trucks and other transportations.



**Picture 2:** Scheme representative portal for capturing identification signals. Part of the localization system. 1 - Portal with antennas / 2 - Truck tagged RTLS

### 4. DECISION MAKING: RATING FOR SENSORS AND EXPOSURE OF METRIC PARAMETERS OF ASSESMENT.

The systems can be classified according to the kind of sensor used. In active systems, the signal emitted by the sensor determines the distance between the mobile devices, that in the case of tracking individuals is an accessory and the sensor is generally fixed. From the moment in which the positioning of the sensors is known, the location can be computed.

This study aims to gather information and evaluate the application of RTLS's based on assets sensors, that present

lots of developed solutions for the area of big ventures, besides the possibility of association between sensors, forming hybrid solutions for a better range in situations of severity of conditions. The GPS, which is a passive sensor, is included in the evaluation’s selection, because it has robust long-ranged and well-spread commercial solutions.

Active sensors may have many different types: based on communication through network such as 802.11 – specification of the type of network or in dedicated sensors which use others electromagnetic bands. In this study we will approach the following sensors: Infra-red, ultrasound, UWB (ultra wide band), RFID and Zigbee.

#### 4.1 Active Sensors

##### a- Infrared sensors

Short range transmission of a mobile device’s identification to a fixed receiver with a known location. Because of the limited range of infra-red transmission, the optical spread needs to be in the line of sight (LOS – line of sight).

The cost of the components is basically cheap, yet the development of the transceiver is expensive. Besides that, the interference rate is high in places where there are machines operating and equipment that emit infra-red light. Another interference factor is the solar beams, which block the infra-red beams’ communication.

##### b- Ultrasound Sensors

A positioning system based on Ultrasound technology has as main advantage the precision of the obtained positioning calculation. The Ultrasound promotes a refined measure with the smallest precision unities through the application of a technique that calculates the arrival time of the signals to the sensors (TOS – time of sight). The calculation of the positioning generates three-dimensional information and it’s obtained through the distance between the transmitters and the receivers which change their fixed position among themselves. The definition of the access points and the antennas’ positioning which will need to have their coordinates configured or sent to the database manually is made in the mapping phase of the area that will be covered. This previous positioning knowledge is necessary to the system.

To minimize interferences, a lot of antennas need to be placed in close spaces, which can become a big obstacle to the installation of this system.

Therefore, apart from the sensors provide an extremely high calculus precision, the usage of this system may be discarded when facing the prediction of investments in infrastructure and manual adjusts. The system can also present the disadvantage of failing by the trivial interferences, easily found in varied environments such as strident noise, intermittent fluorescent lights or even people talking noise. Because of this, the system uses sophisticated algorithms to filter these errors. This makes the investment even more expensive.

A consolidated example of the usage of these sensors is the MIT Cricket system, which encompass a fixed infrastructure of nodes that emit signals and through the trilateralization technique it obtains the mobile nodes’ positioning with decentralized dependency.

##### c- Sensor based on radio transmission: UWB - Ultra Wide Band

Ultra Wide Band - broadband is based on sending ultra-short pulses (smaller than 1ns) in a low cycle (1:1000). The frequency range that the UWB system uses is higher than 500MHz. Unlike the conventional RFID systems which operate on individual wavebands in the radio spectrum,



UWB transmit a signal along many wavebands simultaneously, between 3,1 and 10,6 GHz. UWB signals are transmitted faster than RFID signals. UWB device not only consume less power than RF tags but also operate in large areas with a bigger signal range.

UWB neither suffer nor cause interference if used near other radio signals, because of the different type of spectrum used. UWB short length pulses are easily filtered with the purpose to define correctly the signals that are generated from multiples platforms. The signal is able to transcend easily walls, furniture and clothes. On the other hand, both metal objects and liquids cause interference. This disadvantage can be avoided by positioning more UWB readers.

The short pulse of the wavebands allows not only defining precisely the arrival time (TOA - Time of Arrival) but also orienting, through the time of flight (TOF), the correspondent receivers that will receive a volley of pulses (GEZICI, 2005).

Currently, some systems were created using UWB. The Ubisense system, as an example, is a system that uses an unidirectional platform with control through a bidirectional channel TDMA - "time division multiple access". The devices transmit signal to the network's receivers and they are located with the application of the angle detection and the signal time of arrival techniques (AOA and TDOA). The Ubisense network operates creating cells, which require at least four sensors or readers. Through the buildings or even through a group of buildings, an unlimited number of readers can be connected over the network, looking like how a cellphone network operates. The readers receive the emitted data from the devices, coming from 45 meters maximum and pass them to the "Ubisense Smart Space" platform.

The UWB microwaves frequencies are used by Siemens positioning radars (Gulden 2003). This radar use the RTOF technique, which measure the reflection time through the frequency modulated continuous wave (FMCW) among transponders and base stations. This system was installed in industrial equipment to monitor their activities, such as positioning of a forklift or even a crane. However, this system is feasible only in environments that allow having a direct view LOS (Line-of-sight).

#### **d- Sensor based on radio transmission: Bluetooth**

Bluetooth uses a communication protocol by short range radio frequency, which allows establishing a communication with many mobile devices if they are compatible. Being able to establish a communication with multiple points (point-to-multipoint), the Bluetooth operates freely on industrial, scientific and medical frequency bands (Industrial Scientific Medical - ISM band) at 2.4GHz. It has a propagation spectrum in a variable frequency, oscillating among a range up to 1600 channels per second.

As a consequence of using the short range protocol, the Bluetooth devices connect through the network can estimate their positioning through measuring the variation of the signal's magnitude (RSS - Received Signal Strength) among devices that have defined this location information. So, each device embraces a position of either master or slave, being possible to change it later. The latest Bluetooth versions have a propagation spectrum in an adaptable varying frequency (AFH - Adapt Frequency-hopping) which allows having a coexistence with other technologies from different frequencies of the same band.

The short range and the time a Bluetooth device takes to detect another similar device isn't fast enough to track an object or people moving.

#### **e- Sensor based on radio transmission: RFID - Identification by Radio Frequency**

RFID is an integrated circle that stores and recovers data by electromagnetic transmission compatible with the radio frequency (RF). This technology is reaching improvements on handling and processing of data. A RFID system is composed by some elements. Among them we have RFID readers, tags and the mean of communication among them.

A reader transmits the signal through radio frequency. When it reaches the tags, the signal is reflected by modulation with the added information. However, it's necessary to into account that the investment on readers is relatively big compared to the tags. They also present a quite limited reading range, varying from 1 to 2 meters. Passive RFID systems usually use four frequency bands.

RFID active tags are small transceivers that can transmit actively their respective identifications (ID) or other additional data answering to transmitted question.

The frequency bands used are similar to the ones used for the passive tags, except the low and high frequencies. The advantage of using an active system is the range, which through the use of antennas can be extended to meters away. The active tags are designed to track objects with a great aggregated value which move in mounting processes at severe environments.

The UHF passive tags transmit through the frequency between 860MHz and 960MHz. They don't need batteries as power source. On the other hand, they have an active transmitter and use backscatter.

#### **f- ZigBee**

Zigbee follows the low cost pattern and the low power waste pattern with communication through wireless networks that can increase the signal range through the combination of different topologies. The low cost allows the technology to get its usage spread in wireless monitoring applications. The small request for energy increases the battery life in at least 2 years.

The Zigbee network is composed by three types of different devices, as it's illustrated in picture 10: coordinator (ZC), router (ZR) and final device (ZED). The coordinator composes the base of the network as tree and it's allowed to store information. The routers not only run the application but also are capable to act as intermediary leading the data to other devices. The router is an optional component that can be associated to the coordinator and be part of a multiple points connection (multi-hop routing) and message distribution.

Summing up, the final devices have only the function to establish a communication among themselves and the similar devices, being unable to retransmit data from others devices. Their operating is optimized to operate with a low energy consume, connecting just one coordinator or router each.

The Zigbee networks may be configured among the following topologies: star, bus, tree, mesh and ring. We usually see Zigbee networks configured to operate as mesh, opposing the other topologies, this one offer options to substitute router that eventually may fail. Because of this, in thousandths of a second, new nodes will be recognized and associated to the network. The allying and router's changing activities is operated by the managers of the network, called access points (APs). The flexibility proposed by this

organization manner of the network makes a scalability of up to 65.536 nodes feasible.

The applications encompass usages in the industrial control area, reception of medical data, smoke alert and burglary, automation of buildings and houses, among others.

#### **g- Passive Sensors: GPS - Global Positioning System**

The Global Positioning System is an outdoor system of positioning more spread, supplying the geographic localization of mobile devices anywhere in the world. To determine a precise positioning of a device, traditionally technologies based on GPS require a visual contact with the satellites which circle the Earth from a distance about 20000 kilometers. With the help of 24 satellites and a redundant backup, the information is transmitted with great accuracy.

To get this positioning, the local devices receive signals from at least 3 satellites. After that this data is processed locally.

The GPS has a solid operation in open and large areas, but, without the LOS with the satellites, the signals become weak, or even inexistent. The majority of the reflected signals in close environments are significantly smaller than the necessary distance to establish communication.

#### **4.2. Measurements Parameters**

It's intended to evaluate the presented technologies through the conceptual parameters that measure the technological performance of the sensors. The parameters' concepts chosen for this study will be list ahead:

##### **a- Safety and privacy**

These parameters are focused exclusively on the user needs related to privacy and control of the collected information. Factors that measure the privacy of the system are related to the control of the access to data that inform the localization history of a selected active and its distribution. Investments in applied software architecture can present effective gains on safety and privacy.

##### **b- Cost**

The cost of a positioning system encompasses some variables such as: the cost of the components that will compose the infrastructure, the positioning devices used by each user and the cost of the installation and maintenance of the system.

Some systems need a big and complex infrastructure which requires more investment and time. Other systems reuse the infrastructure that may be already installed, eliminating hardware costs. The calculation of the cost to the user may consider not only the investment in hardware but also the maintenance, which is a big part of the final cost. It's still necessary to consider the useful life of the battery and its requirement of consumption, which will lead to smaller necessity to change it, decreasing the maintenance cost. A device that is able to calculate its own positioning, may have a bigger investment because of the hardware complexity, but the benefits with the increasing in the useful life and the safety gains while transferring data may compensate the money spent.

##### **c- Accuracy**

It's possible to consider one the most important selection standard. The accuracy makers reference to an average of the distance mistakes and can present a variation on the range that can vary from millimeters to kilometers, depending on the technology used. As long as the accuracy is increased, the positioning system acquires a performance improvement in its operation. This can be the main parameter to use as base for the choice.

##### **d- Precision**

The exactness of the positioning system is the ability to reproduce a same result facing different tests. The accuracy indicates an average of the distance mistake and the exactness indicates the probability of success in the approximation made from the average calculated before.

##### **e -Delay**

It's possible to understand by delay, the time that the system takes to calculated the positioning and return the information to the base until the positioning become formalized. The delay in the duration may happen because of two reasons: either the target/device tracked moves very quickly or there are external interferences and the environment present dynamic changes.

##### **f- Scalability**

The scalability is defined by the number of objects that the system can locate related to the infrastructure of the devices (network) within a defined period.

##### **g- Complexity**

The complexity makers reference to the structure of the components of the hardware, software and the operational factors which encompass the installation of a particular system. An aspect that influences the complexity level is the intervention the user will need to do in order to make the system work. The following factors influence the system complexity: the difficulty level of the system's interface, the number of configuration the user will need to do and the quantity of hardware that will be installed. These activities will be evaluated during the installation and the maintenance.

Considering the restriction of the CPU capacity and the energy that a battery is able to store, a system that uses less complexity to calculate must be evaluated better.

##### **h- Responsiveness**

The responsiveness of a system is indicated by the location rate. The difference of the tracked object's moving time between two points and time the system take to recognize the new positioning provide the system location rate.

In conditions of use to big areas, it will be required by the system a rescaling encompassing the increase of the covered area's range. In another words, the sensibility rate of the devices that compose the system must be enough to absorb a quick change in complex scenarios, which is the environment applied to the theme of this dissertation.

##### **i -Robustness and tolerance to absence**

Robustness is connected to the capacity of the system to keep operating even if it's facing some absences such as theft/ failure in some of the network sensors or obstruction of the communication channel.

##### **j- Commercial availability (market)**

Among the existing positioning systems, some are available on the market but others are developed through academic researches. Most of the developing companies maintain the essential characteristics to the operation of the system in secret. The already developed systems and tested through researches, expose clearly the project details, which can be an advantage to the implementation of future improvements.

##### **l-Limitations**

In spite of having effective evolutions on performance, the positioning systems still have restrictions. Some technologies that can present low cost installation advantages, offer a short range on coverage. Others may initiate interferences if used together with other wireless communication systems. So, these restrictions have to be pondered when choosing the system.

To evaluate the system's performance it's necessary to take into consideration which parameters will be necessary to obtain the positioning are offered as base for the applications. Yet, this evaluation must be aggregated to the goal and to the necessity of the tracking answer.

### 5. METHOD'S PRESENTATION

It's proposed a construction of a qualitative mesh establishing a correlation between the measurable parameters and the technical characteristics, with applications aiming a situation that encompass the necessities and peculiar demands to an environment with big enterprises, previously classified.

Through the table 2- RTLS selection table, it's possible to obtain a mapping of the indicators extracted from the realization of a theoretical listing that had as goal, review the main RTLS usages existing in the market nowadays and through academic researches, which somehow presented results and situations arising from field tests.

**Table 2: RTLS selection table**

| ACTIVE SENSORS E GPS                       | INFRARED                                                                   | ULTRASOUND                                                                                                                                          | BLUETOOTH                                                                                                                                                    | UWB                                                                                                                                                                  | RFID                                                                                                                                                               | ZIGBEE                                                                                                | GPS                                                                                     |
|--------------------------------------------|----------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|
| PERFORMANCE PARAMETERS                     |                                                                            |                                                                                                                                                     |                                                                                                                                                              |                                                                                                                                                                      |                                                                                                                                                                    |                                                                                                       |                                                                                         |
| <b>COST</b>                                | MEDIUM                                                                     | HIGH                                                                                                                                                | LOW                                                                                                                                                          | MEDIUM TO HIGH                                                                                                                                                       | LOW TO MEDIUM                                                                                                                                                      | LOW                                                                                                   | MEDIUM                                                                                  |
| <b>ACCURACY</b>                            | 10CM                                                                       | 16CM                                                                                                                                                | 2M                                                                                                                                                           | 15CM                                                                                                                                                                 | 2-3m                                                                                                                                                               | 5M                                                                                                    | 5M - 50M                                                                                |
| <b>PRECISION</b>                           |                                                                            | 95% - 9CM                                                                                                                                           | 95% - 2M                                                                                                                                                     | 50% - 0,3M                                                                                                                                                           | 50% - 1M                                                                                                                                                           |                                                                                                       | 50% - 25M                                                                               |
| <b>RANGE</b>                               | 30M                                                                        |                                                                                                                                                     | 10 to 20M                                                                                                                                                    | Better than RFID                                                                                                                                                     | up to 30m (active RFID)                                                                                                                                            | 10 to 200M                                                                                            | global                                                                                  |
| <b>LIMITATIONS</b>                         | Do not penetrate walls or other opaque materials                           | Noises / Lights/ Metallic materials/ Irradiation causes malfunctioning                                                                              | Long delay during the network update                                                                                                                         | The systems use the same frequency of the others authorized systems. The usage of UWB is prohibited in some countries.                                               | Presents malfunctioning with metallic and liquid materials. Short reading range. It's necessary the use of portal or hand collectors.                              | The end device gets low range because their low-battery consume.                                      | Line of sight Covered areas provide low GPS signals                                     |
| <b>ROBUSTNESS AND TOLERANCE TO ABSENCE</b> | Signal gets blocked and performs poorly in the presence of direct sunlight | The systems uses good outlier detection algorithms to filter out many such sources of error.                                                        | AFH-adaptive frequently hopping Technology once the communication channel lost his signal because of the congestion, the devices start a new channel search. | Imunidade MULTIPATH It goes trough walls. Do not suffer interference from another signals of RF.Ultrapassa paredes. Não sofrem interferência de outros sinais de RF. | Utilizes techniques of spectral spreading to obtain a better rejection of the interferent signals originated from other systems acting in the same frequency range | Using a mesh network the system do not stop working if some device stop.                              | Some Gps models such a RTK, are rugged enough to resist a bad environmental conditions. |
| <b>SCALABILITY</b>                         | Scales poorly due to the limited range of IR                               | not scalable                                                                                                                                        | Each sensor could recognize 200 more devices                                                                                                                 | Up to 128 devices works at the same network. Transmits signal in different frequency bands simultaneously. Increasing the abrangency zone for data transfer          | unlimeted                                                                                                                                                          | Scalability of 65.536 devices.                                                                        | Global scale coverage                                                                   |
| <b>RESPONSIVIDADE</b>                      | Update com mais de 60Hz e uma latência de menos de 10 milisegundos.        | .Determining the positions of and interactions between moving objects requires each to push its location to a central service at regular intervals. | Slow mechanism that provides a short number of samples within an determined time period.                                                                     | Very fast. Transmission capacity of 100 Mbits/sec. Delay- 50msec Wide bandwidth.It's capable of transmitting a lot of data simultaneously spending low energy.       | The tags can be interrogated by the readers up to 1500 times per second. Delay- less than 100msec.                                                                 | To save energy the device hibernates. The system works better when is at 1% of the function capacity. | To reach their localisation the device gets at least 03 different satellite signals     |
| <b>COMPLEXITY</b>                          | Incurs significant installation, reconfiguration, and maintenance costs.   | The system require excessive manual intervention                                                                                                    | The system have to be restarted everytime a new link is created.                                                                                             | Uses pre-existing network infrastructure                                                                                                                             | Manually configurated tags. It's necessary a reconfiguration to set a new position.                                                                                | Low complexity of configuration Open source. No need for additional hardware and use a free frequency | High complexity of configuration                                                        |

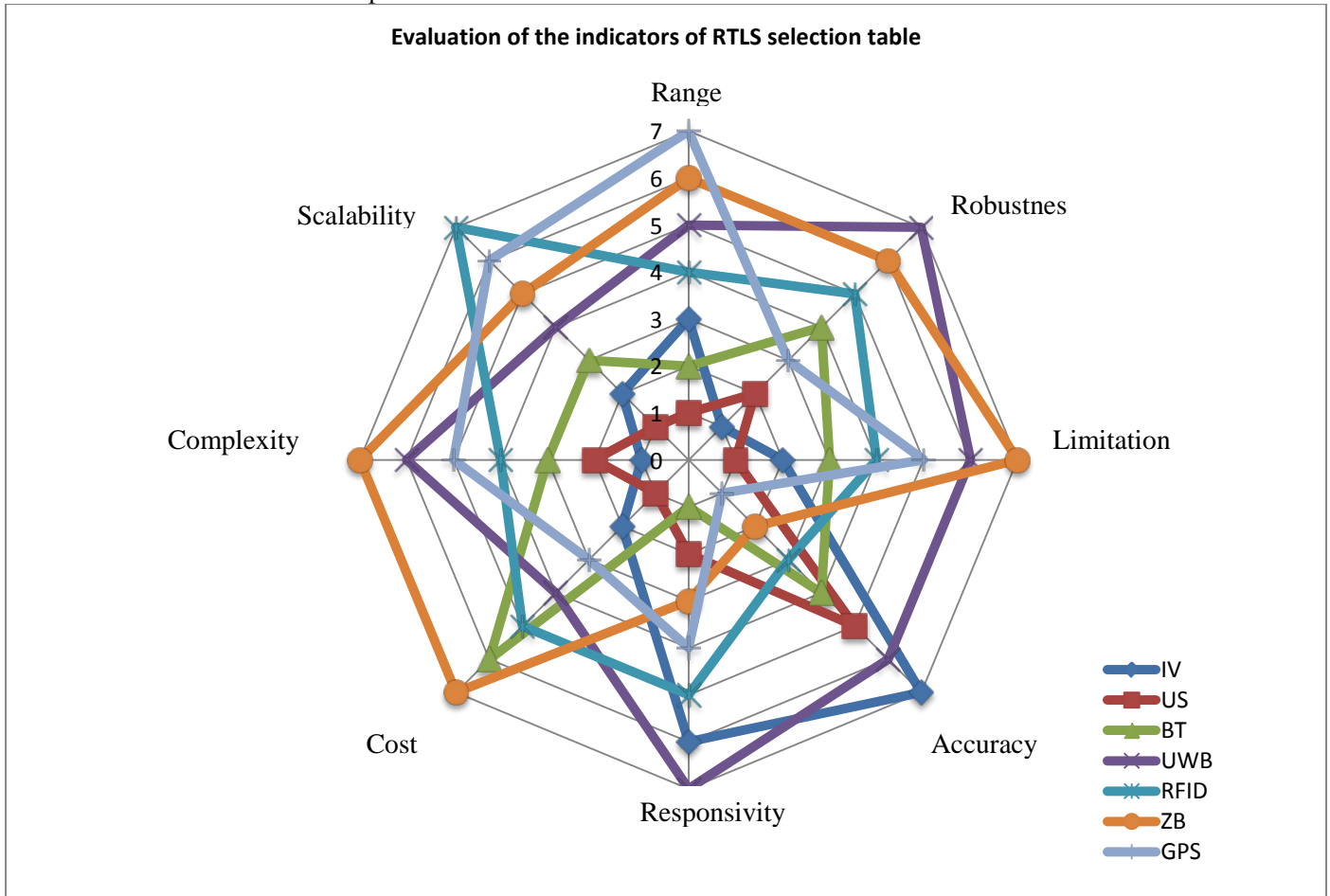
### 5.1. Evaluation of indicators of RTLS

Selection Table In order to generate a relative comparison between the performance of the sensors. The evaluations between the indications for each parameter were ranked with an evaluating grade between one and seven, that is the total number of valued sensors, in crescent order, being 1 the worst performance. This classification was represented through graphical radar, in order to make possible to see more clearly the performance of the sensors,

applied to a situation that was the closest to the environment with severe and extreme conditions. It's generally the case of large ventures.

We can see, in the graphic the evaluation of the indications of the RTLS selection table and following, the description of the logical assessment of 5 of the 8 parameters starting from the last parameter of the first columns of the RLTS selection table following until the first parameter.

Graphic 1 / Radar: Evaluation of the indicators of RTLS selection table



**Accuracy:** The values on the RTLS selection table indicate a better result obtained by each system. Usually these cases occur under propitious and controllable conditions. Without relation to the other's performance evaluations, the order from the worst result to the best result is the following: **GPS, Zigbee, RFID, Bluetooth, Ultrasound, UWB and Infrared.**

The systems based on UWB obtain less accuracy as long as they reach a large range. In case of auxiliary applications to GPS, such as the D-GPS which through the obtainment of the difference regarding the positioning based on terrestrial stations can provide accuracy in the positioning of 10cm.

**Range:** the indicated values in the RTLS selection table are the best obtained by each sensor. The organization of the system's infrastructure is directly related to this parameter. For example, a RFID system can increase its range by positioning antennas diagonally among themselves. So, the signal transmits further, increasing the communication range value.

If the configuration of a **Zigbee** network doesn't take as priority the obtainment of a greater energy autonomy, the final device which will be connected to the router or

gateway, it will be possible to increase the range up to 200m.

Related to the **Bluetooth**, the better range result is assigned to those from class 1 in maximum energizing conditions. On the other hand, the major quantities found on market devices are the ones from class 2, which have a maximum range of 10 meters. Bluetooth has three classes and 4 versions and in every case, the range depends on the transmitting conditions, enclosure material of the sensor, antenna's configuration and battery conditions.

The RFID active Tags have an internal battery that last 1 to 3 years and can reach a 30 meters range. The passive tags have a good interaction and transmission quality. However their range is less than 1 meter or up to 10 meters to the semi-passive tags. The passive tags don't operate very well with metals and liquids and the semi-passive ones require a high power from the hand readers.

To make feasible the installation of these systems, the environment's layout must be studied because the positioning of the system's infrastructure can't interfere on the local movement. This parameter is directly connected to the restriction specifications, since the possible interferences

must be solved, mainly with the system infrastructure's positioning combinations.

**Limitations:** Most of the interferences that hit the **Ultrasound** operation are present in the environment. So these system restrictions may be many ones, such as noises, lights, irradiations from other equipment and metals.

**Infrared** sensors may be uncertain since they present failures on the emitted signals in the presence of sunlight. Considering this restriction, the application of a RTLS based on infra-red becomes unfeasible for outdoor environments.

**Bluetooth** present a delay in the network update, the positioning information is sent only when it's required by the system and not automatically, spoiling the real time communication.

**RFID** sensors present interferences against liquids and metals, but some active tags models are already using techniques to neutralize these effects.

In spite of having great long range solutions, the restrictions of a closed environment create a barrier to the **GPS** signals. Yet, this restriction was already solved in conformity with the evaluation regarding robustness and tolerance to absence that will be presented afterwards.

**UWB** uses the same frequency band as other licensed systems, so, to avoid interference in some countries, its use is prohibited or as an alternative, the short range transmissions are used.

The **Zigbee** restrictions aren't connecting to environment conditions, but to the infrastructure configuration that restrict the range in order to get a better energy autonomy.

**Robustness and tolerance to absence:** the evaluation about this parameter measure the technical capacity the system has to solve the difficulties and keep running.

The **infrared** doesn't have any specific technique. The replication of the infrastructure's components is enough to solve the interferences. In this case, the chances of being incompatible with the environment increase because physically, the system will take more space. To keep operating against the many types of interferences that hit the **Ultrasound** operation, the system applies sophisticated algorithms, which make the investment more expensive. GPS applications to outdoor environments are well solid and resistant to interferences, but indoor environments with walls and ceilings weaken the signal reception from the satellites. To solve this situation there's a technique known as A-GPS (Assisted-GPS) that uses additional links with radio communication to generate more information about the local positioning.

Through the AFH technology, the **Bluetooth** increases its acting field because of the compatibility with a lot of channels. It does it until find a free band.

The make possible working with other devices solving the possible interferences, the **RFID** uses the spectral mirroring technique.

If there's a malfunction in one of the **Zigbee** sensors, the mesh topology allows the others sensors to keep operating.

**UWB** has the multipath immunity, generating more resistance to interferences and a greater capacity to penetrate opaque materials.

**Scalability:** Systems based on **Ultrasound** are considered not scalable, in other words, they focus all the positioning information in one component with the purpose to locate all the others. So, in this evaluation they have the

worst result. Followed by **Infrared** sensors, which have a configuration quite similar to Ultrasound. They present a small scalability rate because the limited range and specially for presenting a lot of failures facing physical barriers.

The next on the list is the **Bluetooth**. Its network is composed by 7 active slave devices and p to 200 inactive devices when the network is in parked mode. As a conclusion, within the range that Bluetooth accepts, each sensor is capable to detect other 200 devices.

The **UWB** acts in varied simultaneous bands and allows many devices to stay interconnected through a network. Followed by: **Zigbee** with an expansion capacity up to 65536 devices, **GPS** with a global scale and finally, **RFID** which doesn't have a limited number of tags per network.

Therefore, for large environments, RFID should be the most indicated because of the expansion capacity.

**Complexity:** infra-red needs to have a constant maintenance. The access points of the system based on Ultrasound will need to be configured manually every time when the positioning changes.

Every time a new **Bluetooth** communication link is created, the device must be rebooted.

On the same complexity level of the previous ones, but with simpler configuration steps, all RFID tags belonging to the system must be reprogramed if there's a positioning change. These four cases require time and investment in training the team.

The process of localization detection of a **GPS** device goes through detection of satellite signals, reading, acquisition of data and then processing an outcome occurs.

**UWB** uses existing infrastructure that are already installed for other purposes.

**Zigbee** has a low complexity with procedures of simple configurations without the need of network, hardware already add to the deployed network, besides the use of the frequency band without the need of a license and an open protocol.

## 6. METHOD'S APPLICATION: SELECTION TABLE RTLS + GRAPHIC 2/RADAR

Evaluating the presentation of 2/Radar graphic, we can notice by analyzing the ranking the following relationship: the more externally positioning technology, the better will be its performance.

Thus, the search to identify the most appropriate technology can be associated with the analysis of the graphic. However, it is worth noting that not always the best performance of the technology can be considered the most suitable for the occasion of the system application.

Nevertheless, the table lookup is essential to obtain the indication of the technology part, because the indicators show the physical and real and individual capability of the components that make up the tracking systems.

Therefore, together with the analysis of the table, the method consists in using the chart to have a view of the performance rating, since the guidelines for implementing the system are well defined the user can visually notice the comparative relationship between technologies.

### 6.1. Prioritization of indices for evaluating performance according to the nature of the assets and the environment.

Regarding the priority of the user's projects necessities, the characteristics of assets are directly related, such as equipment, cost and the environment, as the flow of assets and delimitation, being outdoor or indoor.

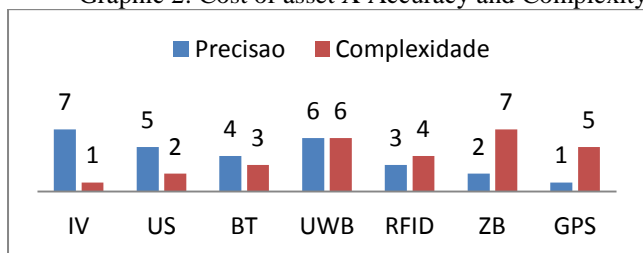


The graphics from 2 to 4 are going to be presented with an indication of some of the possible direct relationships between variables relating to the assets and the environment, as previously noted, with the definition of the requirements of evaluation that directly influence the performance of the RTLS's for such situations.

The sensors discussed are listed in the left column for each metric parameter, the sensors receive a review ranging from 1 to 7, 1 being the worst performer and 7 the best performance in a given situation generally. The main point for the evaluation of sensors is based on qualitative indications in determined in the RTLS Selection table.

**a- The cost of the asset X - accuracy and complexity**

Graphic 2: Cost of asset X Accuracy and Complexity

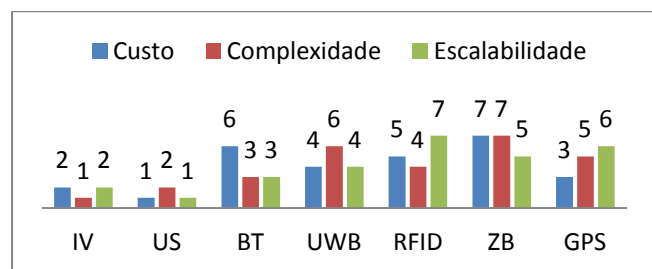


Greater complexity, higher cost of deployment of the system. If the cost of the asset is high and representative within the project budget, the investment justifies the purpose: have greater control and monitor more assertively such assets. If they get lost, misrouted or break, that mean a considerable loss. Likewise the high cost of the asset may require a more refined screening accuracy, and is also justified as in the previous case. We understand the high cost, not only as the investment in kind which involves the assets, but also the added value that can be their indispensability to the functioning of the project activities of the enterprise.

P.s: Because the range does not affect this case - the cost of the tool will not determine that it is necessarily tracked in extensive areas. When it comes to large enterprises the largest range can be definitive for different situations

**b-Number of Asset X Cost, Complexity and Scalability**

Graphic 3: Number of Asset X Cost, Complexity and Scalability



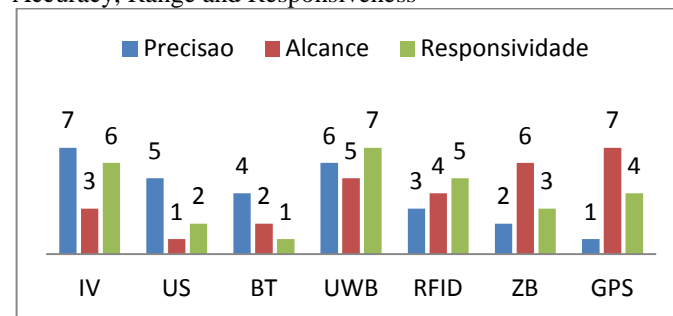
The amount of the asset directly influences the final cost of RTLS deployment, it is considered basically the amount of tags / devices that comprise the solution will increase.

Some systems have the complexity directly proportional to the level of difficulty of setting up the infrastructure solution. Certain settings point to point are operated manually. Either devices, or the network, or the access points and stations may require constant maintenance and reconfigurations. Thus the system comprehends a larger number of components; the demand for labor may be increased.

Finally, the ability of scalability that is the definition of the number of objects that the system can find is an important direct factor.

**c- Default flow in the environment X Accuracy, Range and Responsiveness**

Graphic 4: Default Flow in the environment X Accuracy, Range and Responsiveness



The default flow of people, materials and machines defines the layout that the system will be able to install its structure. In this way the range is directed to the space within the system's range in order to collect all the necessary movement routine. The accuracy will define if the actives are or not in specific places such as rooms and aisles. In case this is a relevant requirement, the accuracy will need to have good results. The responsiveness determines the tracking capacity in real-time. If the question about the positioning identification of the active is immediate, the systems which have low responsiveness rate will be eliminated from the selection.

**7- CONCLUSION**

At the moment that is decided to use the automatization of the data the control of the processes by the entrepreneur, the conception of the system's project begins. Many times, occur that the RTLS consultants required to suggest the most adequate system are representatives from certain technologies. So the solution is guided to the commercial interests and not to the best application through adequacy and implementation studies. To minimize the consequences of these situations and to help future improvements on research and development, it's suggested to use the RTLS selection table presented in this study.

The intention of elaborating the RTLS selection table was the gathering of the main indicators that could qualify each sensor according to the performance evaluation parameters that influence more in the final choice. Obtaining with this, a mesh that serves as base to the analysis of the feasibility and the restrictions applying the RTLS technology in big enterprises.

Moreover, the order by grade defines the positioning technology in the ranking. However, it's possible to see that the grades are too much absolute many times, regarding numerous application variables. The regular gaps for ordering, in other words, grades from 01 to 07 with uniform spacing in some situations don't measure the real performance difference among the technologies.

These particularities of the classification intensity must be compensated with the table indicators though. In other words, each item described in the table was filled with technical characteristics inherent to the technologies, with its restrictions and advantages against the physical restrictions of the devices the system may use, independently of the conjecture about the application.



## REFERENCES

TAPIA, D. et al, "Implementing a Real-Time Locating System Based on Wireless Sensor Networks & Artificial Neural Networks to Mitigate the Multipath Effect". In: 14th *International Conference on Information Fusion*, Chicago, Illinois, USA, July 5-8, 2011.

MALIK, A., (2009). *RTLS For Dummies*. Wiley. p. 384. ISBN 978-0-470-39868-5.

HJELM, J., KOLODZIEJ, K. W., *Local Positioning Systems: LBS Applications and Services*, Boca Raton, FL: CRC Press , 2006. 463p.

URRUELA, A.; PAGES-ZAMORA, A. ; RIBA, J., "Divide-and-Conquer based closed form Position Estimating for AOA and TDOA Measurements", ICASSP 2006, IEEE Proceedings on; Volume IV, May 2006, pp.: 921-924.

KAEMARUNGSU, K. "Efficient design of indoor positioning systems based on location fingerprinting," *Wireless Networks, Communications and Mobile Computing, 2005 International Conference on* , vol.1, pp. 181- 186 and vol.1, 13-16, 2005.

CHENG, T., VENUGOPAL, M. ; TEIZER, J., et al, "Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments." *Automation in Construction* 20, Elsevier, 2011, doi:10.1016/j.autcon.2011.05.001

GEZICI, S. et al , "Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 70–84, Jul. 2005.

BOHN, J.S., TEIZER, J., "Benefits & barriers of construction project monitoring using hi-resolution automated cameras", *ASCE Journal of Construction Engineering and Management* 136 (6) (2010) 632–640.

# PROBABILISTIC APPROACH AS A SUPPORT SYSTEM FOR SAFETY-CRITICAL ENVIRONMENTS BY KNOWLEDGE DRIVEN BAYESIAN NETWORKS

Ciro D'Elia<sup>(a)</sup>, Fabio De Felice<sup>(b)</sup>, Paola Mariano<sup>(c)</sup>, Antonella Petrillo<sup>(d)</sup>, Simona Ruscino<sup>(e)</sup>

<sup>(a)</sup><sup>(c)</sup><sup>(e)</sup> Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, Cassino (Italy)

<sup>(b)</sup> Department of Civil and Mechanical Engineering, University of Cassino and Southern Lazio, Cassino (Italy)

<sup>(d)</sup> Department of Engineering - University of Naples "Parthenope", Naples (Italy)

<sup>(a)</sup> [delia@unicas.it](mailto:delia@unicas.it), <sup>(b)</sup> [defelice@unicas.it](mailto:defelice@unicas.it), <sup>(c)</sup> [p.mariano@unicas.it](mailto:p.mariano@unicas.it), <sup>(d)</sup> [antonella.petrillo@uniparthenope.it](mailto:antonella.petrillo@uniparthenope.it), <sup>(e)</sup> [s.ruscino@unicas.it](mailto:s.ruscino@unicas.it)

## ABSTRACT

Probability estimation is an integral part of risk analyses. This work intends to propose a probabilistic approach as a support system for risk assessment in order to establish a deeper understanding of accident causation pathways as a means for proposing improved preventive strategies, especially at the level of organizational and structural factors. This study addresses the problem of "damaging event" probability estimation with few statistics by the use of Knowledge Driven Bayesian Network (KDBN), that models the a priori knowledge of the risk context dynamics. Moreover the proposed approach aims at providing a quantitative methodological technique useful to monitor, prevent, and evaluate, and assess the risks at workplace.

Keywords: Knowledge Driven Bayesian Network, safety, probabilistic risk analysis, risk assessment.

## 1. INTRODUCTION

Safety-critical environments are those domains in which hardware failure or late decision-making by operators could result in loss of life, significant property damage (Naderpour et al., 2014, a; Del Rio *et al.*, 2013). Since the beginning of the industrial revolution, many serious accidents at large-scale technological systems that have had grave consequences (Falcone et al., 2007; Bruzzone *et al.*, 2014).

Promoting situation awareness is an important design objective for a wide variety of domains, especially for process systems where the information flow is quite high and poor decisions may lead to serious consequences (Naderpour et al., 2014, b). In the present work we use an approach based on Bayesian Network (BNs) to describe the circumstances and relationships between circumstances due to accidents at workplace.

In fact, in our opinion BNs are powerful conceptual, mathematical and software tools used to model complex problems with variables related by probabilistic links (Washington and Oh, 2006). Furthermore BNs are used for professional and academic applications in a great variety of domains:

medical, financial, industry, security, space, artificial intelligence.

Definitely, BNs provide an attractive means for modeling and communicating complex structures and BN form the basis for efficient algorithms, both for propagating evidence and for learning about parameters.

This study wants to give three important contributions. First, it proposes a situational network modeling process which is used to model risks situations at workplace. Second, it presents a situation assessment model that exploits the specific capabilities of Bayesian networks and risk analysis. The proposed situation assessment model can be applied to different scenario. Third, it develops a model for managing risks situations in safety critical environments in which the degree of automation and complexity continues to increase and the number of operators decreases, and where each operator must be able to comprehend and respond to a growing amount of risky status and alert information. More in detail, the model exploits the knowledge of experienced operators for building the network structure, thus overcoming the lack of statistics usually needed for learning conventional Bayesian nets and compensating it with the use of a contextual risk model. Moreover, as better explained in Section 3, the proposed approach joins the low complexity of a the most widely used technique for risk evaluation, namely the risk matrix method, with the completeness and repeatability of the AISS approach, much less used due to its complexity.

According to our previous work (D'Elia *et al.*, 2013) in which we presented general features for the Knowledge Driven Bayesian Network (KDBN) model, in the present research our aim is to furnish a deeper analysis of the model from a practical point of view. Our preliminary work took into account the mathematical theory underlying the model. Now a complementary analysis is presented. To this purpose, it was decided to apply our model to a real case study which was specifically developed for use in the investigation and analysis of occupational accidents within industrial settings, in order to evaluate in practice the ability of KDBN models to catch real risky situations. Definitely, the model represents an

*innovative decision support system* to manage risks situations in safety-critical environments in which the effect of situational complexity on human decision-makers is a concern.

The paper is organized as follows. Section 2 presents literature overview. In section 3 problem statement is analyzed; section 4 presents methodological approach and finally results and conclusions are presented.

## 2. LITERATURE OVERVIEW

As stated by Jacinto (Jacinto et al., 2009) the first aim of accident research is to understand accidents; this is generally done through the search of their causes, so that adequate prevention measures can be designed.

Analysis of accidents is not well supported by conventional and qualitative methods due to data scarcity (Khakzad et al., 2014) as for example by the *Risk Matrix method*. This method is a well-known qualitative method that is used during risk assessment to define the various levels of risk as the product of the harm probability categories and harm severity categories (Encyclopædia Britannica, 1911; Fabiano et al., 2004). However, Risk Matrix Method is a very subjective method that sometimes could also be arbitrary because it depends on the experience of the expert (De Felice and Petrillo, 2012).

Thus, in the absence or shortage of adequate tools the evolution of investigative methods over time reveals various trends, showing a gradual shift from searching for a single/immediate cause, to the recognition of multiple causes, in which management and organization failures and their interactions with the working activities became an important issue for the understanding of accidents (Roed-Larsen, 2003). One of this method is the *AISS Method* proposed by Association Internationale de la Sécurité Sociale that takes into account different parameters such as the material, work equipment; the environment; the organization of work, etc. (ISSA, 2001). The AISS method presents several weaknesses because it is quite complicated to use it due to several parameters to be considered.

From this point of view considerable research have been conducted to develop safety performance assessment using the Bayesian technique (Yu, and Abdel-Aty, 2013). There has been a steady growth of interest in the application of Bayesian Network (BN) to risk analysis due to its capability to model complex system (Lu et al., 2011). BNs have been applied in several knowledge areas, such as medicine, environmental assessment, business risk and product life-cycle analysis, etc. In the workplace risk area some applications are from Galán et al. (2007) that applied a canonical probabilistic test (based on Bayesian models) to the analysis of nuclear system safety; Papazoglou et al. (2006) that applied functional block diagrams and event trees to quantify the risk of falls. More specifically for construction and mining accidents,

Matías et al. (2008) compared the predictive capacity of BNs with other expert systems.

The Bayesian framework provides a complete and coherent way to balance the empirical data and prior expectations, which would be very promising to be applied in safety analyses (Jang et al., 2010). As stated Ahmed et al., (2011) one key advantage of the Bayesian inference method compared to the conventional frequent inference approach is that extra knowledge and experience about the data can be used as prior information in the analyses.

This study fills the gap in formulating a Bayesian model in safety at workplace.

## 3. PROBLEM STATEMENT

The classical approach used for risk assessment is based on the Risk Matrix method, which provides a statistical evaluation for risk given by eq. 1, where  $P$  represents the frequency of occurrence of the damage  $D$  and  $D$  represents the damage value.

$$R = P \cdot D \quad (1)$$

The risk expressed by eq. 1 is associated to a given activity, and its value can be eventually reduced through the application of some safety regulations. In this context the risk can be indicated as *original* ( $R_o$ ), if no regulations are applied, and as *dynamic* ( $R_t$ ) otherwise. In this paper, a Bayesian model for dynamic risk assessment is proposed. More formally, the expression of the dynamic risk is given by eq. 2, where  $x$  is the activity under observation,  $\underline{DB}$  is a vector of safety regulations applied and  $D$  is the damage occurred.

$$R_t(x) = P(D|x, \underline{DB}) \cdot D \quad (2)$$

As introduced in section 2, another method to the risk evaluation has been proposed by AISS (ISSA, 2001), but it is not much used due to its complexity. This method uses in fact a more complex approach in respect to the risk matrix method: it starts by the consideration that accidents on working sites depend on many different factors, such as bad design, bad conditions for equipment, insufficient cleaning, floor encumbrances, environment illumination, noise and so on. All of these factors can be grouped into three main categories: job material and equipment; environment; job human organization. The AISS method associates to each of these classes a risk evaluation, identifying four operational phases: (1) evaluation of the global risk related to the job place, given by eq. 3, where  $M_a$  represents the risk factor related to the material or the equipment used, and  $Env$  represents the effect of the environment; (2) estimation of the worker ability to dominate the risky situation; (3) evaluation of the risk of accidents, given by eq. 4, where  $P$  is the worker ability evaluated during phase 2, weighted through the factor  $k$  related to the influence of material/equipment;

(4) contextualization of risk evaluation in acceptable ranges.

$$R_g = M_a \cdot Env \quad (3)$$

$$R_{acc} = R_g \cdot kP \quad (4)$$

The approach proposed in this paper can be considered as placed in the middle of the two previously described methods, joining in some way the low complexity (hence the simple applicability) of the risk matrix approach with the greater completeness and repeatability of the AISS approach, which considers the risky different factors more in detail during the evaluation process. The proposed method is based on the use of Bayesian networks to calculate the joint probability given by eq. 2 by taking into account information related to the entire working context, namely organization, material, equipment, workers training, environment conditions, application of safety regulations, relationships among dangerous events and damages occurred.

#### 4. METHODOLOGICAL APPROACH

##### 4.1. Basic principles

As already stated, the probability  $P$  of eq. 2 is very complex to calculate, since it depends on many parameters. *Bayesian Networks* can be viewed as a valid tool to describe this probability, because they allow to structure a complex problem as a combination of less complex sub-problems, hence easier to solve, built by considering the cause-effect relationships between the entities involved in the problem itself.

More specifically, in the risk assessment model addressed by this paper the statistical-physical entities considered are:

- Damages  $D_i$

- Dangerous Events  $E_m$
- Causes  $C_n$
- Duties and Bans  $DB_j$

These domains conceptually corresponds to the variables of the AISS method. The choice of these kinds of entities responds to the necessity of respecting the indications of the reference Legislative Decree 81/08 and of exploiting the information provided by the operators experience on field, in order to compensate the lack of statistics needed by conventional BN learning algorithms. The structure of the net should be indeed derived and built up through a learning process; however, since the problem involves a huge number of variables, a significant number of statistics would be needed to properly learn the net. In order to overcome this difficulty, the proposed approach, called *Knowledge Driven Bayesian Network (KDBN)*, exploits the knowledge of experienced operators for building the network structure, allowing them to transfer part of their know-how in the system.

After the problem variables have been identified, a phase of evaluation and review of the relationships among them can be performed, in order to understand in which way they co-operate to determine the risk. Finally, the results of the risk evaluation process are analyzed, deciding if they can be accepted as they are or if a new and improved evaluation should be carried out.

From the Bayesian network point of view, the elements of these four domains corresponds to the nodes of the DAG (Direct Acyclic Graph) defining the network structure: to each of them a random variable is associated, and the edges between these nodes represents the conditional probabilities (i.e. the cause-effect relationships) which relate them.

In Figure 1 is shown the conceptual scheme of the proposed KDBN approach.

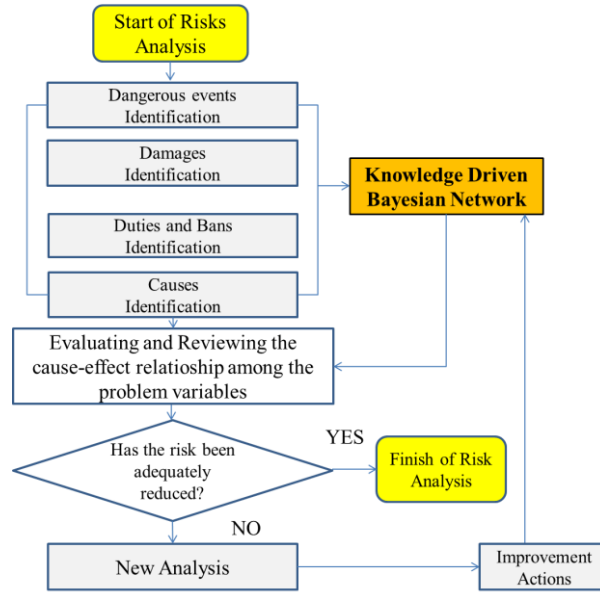


Figure 1: Conceptual scheme of the proposed KDBN approach

#### 4.2. Knowledge Driven Bayesian Network

Figure 6 shows the general structure of the proposed Knowledge Driven Bayesian Network: damages are depicted in red, dangerous events in yellow, the causes are displayed in green, and finally the duties and bans are depicted in white/grey. The duties and bans can directly influence the damages and the cause, through which their effect reaches the dangerous events; finally dangerous events, produce an immediate consequence on damages. This approach significantly reduces the number of parameters defining the problem, making the KDBN method usable also in the case of having a small number of statistical data available for the risk evaluation.

The probability of occurrence of a certain damage  $D_i$ , related to a particular job activity  $x$ , given that a number of safety regulations  $\underline{DB}$  is applied, namely  $P(D_i|x, \underline{DB})$ , is given by eq. 5 ( $x$  is omitted for shortness), where the sums are over all the possible dangerous events and over all the possible combinations of causes.

$$P(D_i|x, \underline{DB}) = \sum_m \sum_n P(D_i, E_m, C_n | \underline{DB}) \quad (5)$$

Each node of the KDBN net is influenced only by the nodes directly connected to it. This leads to the decomposition of the joint probability  $P(D_i, E_m, C_n | \underline{DB})$  in three terms, as expressed by eq. 6.

$$P(D_i, E_m, C_n | \underline{DB}) = P(D_i | E_m, \underline{DB}) \cdot P(E_m | C_n) \cdot P(C_n | \underline{DB}) \quad (6)$$

Given the network structure, to further reduce the complexity of the problem the three terms of eq. 6 can be structured in turn and calculated on the basis of the experts knowledge, available in terms of relative frequencies of incidence between the couples of entities, namely  $P(E_m | C_n)$ ,  $P(D_i | E_m)$ ,  $P(D_i | \underline{DB}_j)$ , and  $P(C_n | \underline{DB}_j)$ . More specifically, these terms represents

the cause-effect relationships among the couples of variables involved in the KDBN.

The statistical surveys about the accidents occurred in the work sites provide an information which can be inserted in the KDBN system in the form of global parameters, namely  $\alpha, \beta, \gamma, \delta', \phi$ . These parameters can be possibly estimated during the learning phase, and are involved in the calculation of each element of eq. 6.

For a deeper description of these parameters see (D'Elia *et al.*, 2013).

- $P(D_i | E_m, \underline{DB})$

The term  $P(D_i | E_m, \underline{DB})$  represents the probability that a certain damage  $D_i$  occurs given that a dangerous event  $E_m$  happened and given that some safety regulations  $\underline{DB}$  have been applied. The calculation of this probability takes into account two contributions:  $P(D_i | E_m)$ , which expresses how the dangerous event  $E_m$  can lead to a certain damage  $D_i$ ;  $P(D_i | \underline{DB}_j)$ , which conveys the influence of the safety regulation  $\underline{DB}_j$  on the damage  $D_i$ . More specifically, the duty and bans  $\underline{DB}$  have the effect of reducing the amount of the damage  $D_i$  through a “discount” mechanism: if a DB is applied, it introduces a probability mass which moves from the more serious damage (the considered one  $D_i$ ) to the immediately less serious damage (namely  $D_{i+1}$ ), hence reducing the probability of occurrence of  $D_i$  and contemporarily increasing the probability of occurrence of  $D_{i+1}$ . Contemporarily  $D_i$  receives the percent discount coming from the immediately more serious damage  $D_{i-1}$ . This leads to the formulation of eq. 7 and 8, where  $P_R$  indicates the residual probability remained after the application of the discount mechanism.

$$P(D_i | E_m, \underline{DB}) = \begin{cases} (1 - \phi)P' & \text{if } D_i \neq \text{NotDamage} \\ \phi + (1 - \phi)P' & \text{if } D_i = \text{NotDamage} \end{cases} \quad (7)$$



$$P' = P(D_i|\varepsilon_m)P_r(D_i|DB) + P(D_{i+1}|\varepsilon_m)[1 - P_r(D_{i+1}|DB)] \quad (8)$$

- $P(E_m|C_n)$

The term  $P(E_m|C_n)$ , given by eq. 9, represents the probability that a given combination of causes  $C_n$  brings to a dangerous event  $E_m$ .  $E_m$  has its own frequency of occurrence, namely  $P_o(E_m)$ , which is grown by the influence of the causes  $C_n$ , in particular each cause of the combination moves a statistical weight, namely  $P_L(E_m|C_j)$ , from the situation in which no dangerous event  $E_{NotD}$  occurs in favor of the occurrence of  $E_m$ . This behavior is expressed by the sum of eq. 9, which is performed over all possible causes.

$$P(E_m|C_n) = P_o(E_m) + \sum_j C_j \cdot P_o(E_{NotD})P_o(E_m)P_L(E_m|C_j) \quad (9)$$

- $P(C_n|DB)$

The term  $P(C_n|DB)$  represents the probability of having a certain combination of causes  $C_n$  (determining the dangerous events) given that a number of safety regulations  $DB$  has been applied. The application of a each duty&ban affects the frequency of occurrence of a particular cause by introducing a statistical *discount* (as seen for the influence of DBs on the damages), represented by a probabilistic weight moving from the considered cause to the not contemplated cause. The expression of this probability is given by eq. 10, which refers to eq. 11, 12 and 13, where  $P_R$  indicates the residual probability of the cause after the application of the discount mechanism, and the product is done over all the  $L$  possible duties and bans.

$$P(C_n|DB) = (1 - \delta') \cdot P'' \quad (10)$$

$$P'' = \prod_{i=1}^K P_c \quad (11)$$

$$P_c = \begin{cases} P(C_i|DB) & \text{if } C_i = 1 \\ 1 - P(C_i|DB) & \text{if } C_i = 0 \end{cases} \quad (12)$$

$$P(C_i|DB) = p \cdot \prod_{j=1}^L P_R(C_i|DB_j) \quad (13)$$

For a more theoretical description of the proposed model see (D'Elia *et al.*, 2013).

## 5. CASE OF STUDY

As case of study for the risk evaluation, in this work the use of the **circular saw** has been considered. The values for damages, dangerous events, causes and duties and bans considered are described in the following.

- **Damages:** three levels of damages have been used, namely *serious*, *slight* and *very slight*.

- **Dangerous Events:** *instability, impact, slipping/tumble, pricks/cuttings, electrical danger, chemical agent, noise.*
- **Causes:** five causes have been considered, namely *not-adequate training, organization, not-adequate working area, behavior, and mechanical malfunctioning.*
- **Duties and Bans:** *safety gloves, safety glasses, safety visors, cleaning of the surrounding area, protection of the transmission organs, pieces-pushing devices, controls protection, masks for the protection of the breathing apparatus, dust extractors, ears protection, grounding of metal bodies, IP 54 electrical protection.*

The global parameters used for the risk evaluation are then described in Table 1.

Table 1: Parameters used for the risk evaluation in the case of the circular saw

| Parameters | Value |
|------------|-------|
| $\alpha$   | 0.1   |
| $\beta$    | 0.3   |
| $\gamma$   | 0.98  |
| $\delta'$  | 0     |
| $\Phi$     | 0.8   |

In order to show how not to apply some safety regulations can negatively influence the risky situation, thus increasing the risk value, three evaluation tests have been taken in consideration: (1) no safety regulation applied; (2) all safety regulations applied; (3) two meaningful safety regulation not applied among all. For all the tests the results will be shown in terms of graphs and tables reporting the evaluation statistics for risks and damages.

### 5.1. No safety regulations applied

As evidenced in the problem statement, in the case of no safety regulations applied the risk is called *original risk*. The first test has been performed considering all the duties&bans as turned off. Figure 2 shows the probability of the damages related to the considered activity (circular saw), while Figure 3 shows the risk values estimated for each type of damage. A quantitative description of the results achieved is reported in Table 2. It can be noticed that, if no duty&ban is applied, the original risk associated to the serious damage is more relevant than the other kind of risks.



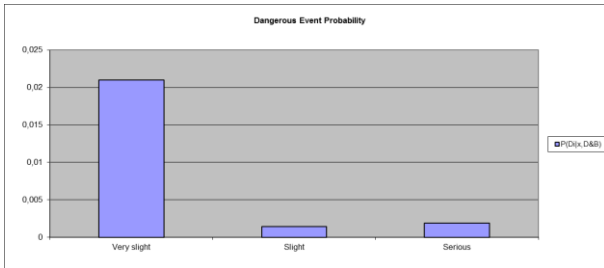


Figure 2: Probability of damages if no DB is applied in the case of circular saw

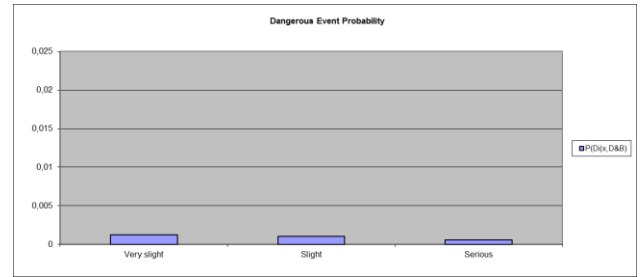


Figure 4: Probability of damages obtained for each damage when all DBs are applied, in the case of circular saw

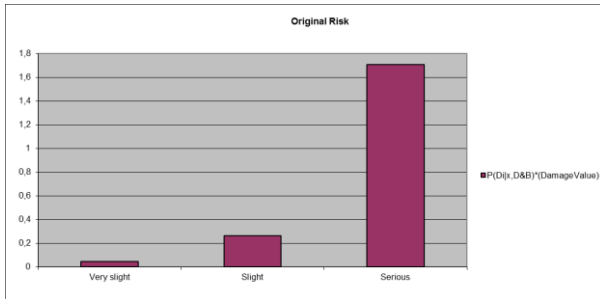


Figure 3: Original Risk in the case of circular saw

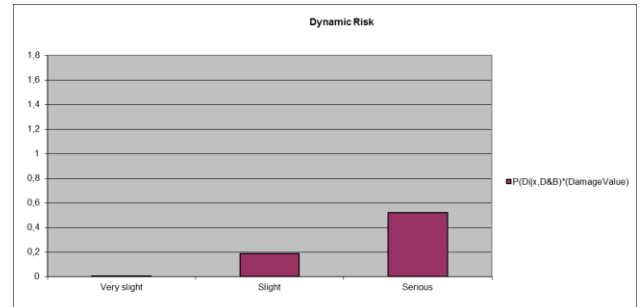


Figure 5: Risk evaluation when all DBs are applied, in the case of circular saw

Table 2: Values of the probability of damages and associated risks in the case of no DB applied

| Damages     | $P(D_i x, DB)$ | Risk        |
|-------------|----------------|-------------|
| No Damage   | 0,97564652     | 0           |
| Very slight | 0,021004691    | 0,047260556 |
| Slight      | 0,00145032     | 0,26432082  |
| Serious     | 0,001898469    | 1,708621714 |

Table 3: Values of the probability of damages and associated risks when all DBs are applied, in the case of circular saw

| Damages     | $P(D_i x, DB)$ | Risk        |
|-------------|----------------|-------------|
| No Damage   | 0,997139489    | 0           |
| Very slight | 0,001250616    | 0,002813886 |
| Slight      | 0,001030433    | 0,187796446 |
| Serious     | 0,000579462    | 0,521516094 |

## 5.2. All safety regulations applied

When all the safety regulations provided for a specific job activity are applied, the risk should assume the minimum value. Figure 4 and Figure 5 respectively show the probability of the damages and the distribution of risk in the case of all duties and bans applied, while Tab. 3 reports the values associated to the graphs.

In respect of the original risk, it can be seen that the application of all possible safety indications reduces the probability associated to each kind of damage, with a consequent decrease of the amount of the associated risks. The percentage of reduction depends in general on how each DB influences each damage. For the case of study proposed in this paper, the influence of all DBs determined that the risk associated to the very slight damage decreases from 0.04 to 0.002, the risk associated to the slight damage decreases from 0.26 to 0.18, and the risk associated to the serious damage decreases from 1.70 to 0.52.

## 5.3. Some safety regulations not applied

An intermediate situation occurs when some of the safety regulations provided are not applied. The choice about what DBs are preferable to be considered in the task of reducing risk can be done in reference of the influence that such DBs have on damages. For this third kind of test two DBs which among all influence more significantly the serious damage have been chosen, namely *pieces-pushing devices* and *controls protection*. The results are shown in Table 4. As expected, it can be seen that the removal of these DBs mainly affects the risk associated to the serious damage, which increases from 0.52 to 0.82. The DBs *discount* mechanism applies percent reductions that propagate through each kind of damage; hence also the values achieved for the slight and very slight damage are influenced, even if this happens in a minor way. In particular, the value of the risk associated to the very slight damage remains almost unvaried, while the risk associated to the slight damage decreases, since the *discount* introduced by the considered DBs to the serious damage, and weighting on the slight damage, is missing. Moreover, the results achieved for this third experimental test show that the proposed model has a predictive ability, since it allows

to determine which safety regulations are more decisive in the task of reducing risk, thus possibly assisting the expert operators in designing and managing the job context and activity.

Table 4: Values of damages and associated risks in the case of two DB not applied

| Damages     | $P(D_i x, DB)$ | Risk        |
|-------------|----------------|-------------|
| No Damage   | 0,997092465    | 0           |
| Very slight | 0,001285025    | 0,002891306 |
| Slight      | 0,000710522    | 0,129492701 |
| Serious     | 0,000911988    | 0,820789194 |

#### 5.4. Comparison with statistical surveys

The results obtained by the risk evaluation process can be compared with the statistical surveys in order to understand the ability of the model to describe real situation of risk in the workplace. More specifically the national statistics data on the accidents occurred in the case of circular saw, in a period of one year, were available. These statistics considered the situation in which all the safety regulations are respected, hence the comparison with the results obtained for the second experimental test was possible (Table5). The statistics shows that, among all the registered accidents, the 45% led to a very slight damage, the 30% led to a slight damage and the 25% led to a serious damage; in the experimental test the obtained percentages have been 43.72% for the very slight damage, 36.02% for the slight damage and 20.26% for the serious damage. From the comparison it can be noticed that the risk evaluation results are quite adherent to a real situation such as the circular saw, showing the effectiveness of the proposed approach.

Table 5: Comparison of the risk evaluation with the statistical surveys

| Damages | Statistical Surveys | KDBN approach |
|---------|---------------------|---------------|
|---------|---------------------|---------------|

|                    |     |        |
|--------------------|-----|--------|
| <b>Very Slight</b> | 45% | 43,72% |
| <b>Slight</b>      | 30% | 36,02% |
| <b>Serious</b>     | 25% | 20,26% |

## 6. CONCLUSIONS

In this paper a new probabilistic approach based on Knowledge Driven Bayesian Network to support process operators in risks situations has been proposed. The methodology presented in the study comprised two main stages: (1) define an analytical framework for the analysis of systemic factors, and (2) identify the systemic/organizational contributing factors through an in-depth investigation and analysis of a number of accidents. A quantified detailed logical model based on a real case study was defined in order to assess the relevant contribution of the various risks factors. In this way a rational prioritization of the available risk reducing measures were identified. The results of this research indicate that Bayesian networks are very useful in explaining the causes of falls.

## ACKNOWLEDGEMENTS

The first application of this research to real context of safety on construction sites has been possible thanks to FILAS financing (<http://www.filas.it/>) to ISA (Ingegneria Sicurezza Ambiente s.r.l.) for a project called DCSS (Distributed Control Safety System, <http://www.lit.ing.unicas.it/?q=ProjectDCSS>).

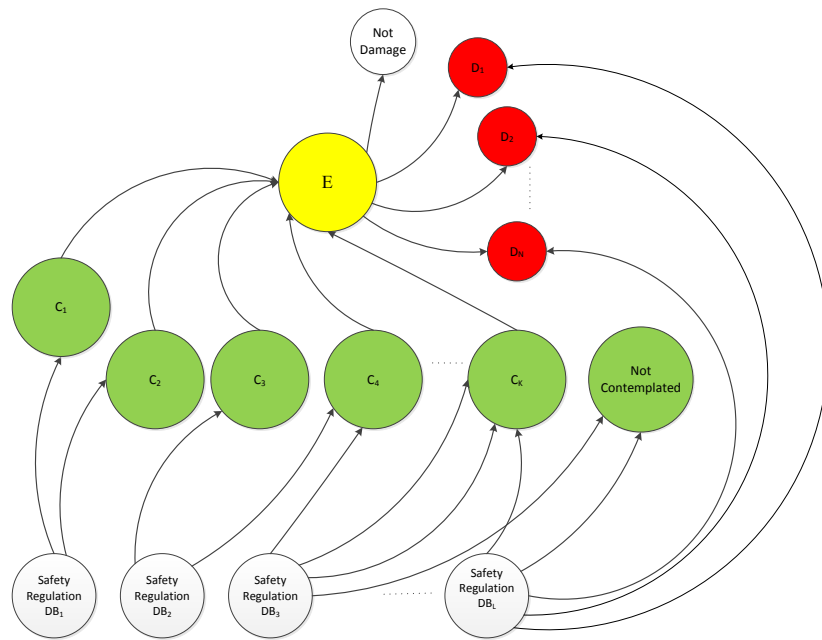


Figure 6: Generic model of the proposed Knowledge Driven Bayesian Network

## REFERENCES

- Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. *Accident Analysis and Prevention*, 43, 1581–1589.
- 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*, Volume 46, Issue 2, 2014, Pages 71-80.
- D’Elia, C., De Felice, F., Mariano, P., Petrillo, A., Ruscino, S., 2013. A Bayesian network analysis for safety management. *Proceedings of the 25<sup>th</sup> European Modeling and Simulation Symposium, EMSS 2013*. Pages 446-455. 25 - 27 September 2013. Athens; Greece.
- De Felice, F., Petrillo, A., 2012. Methodological approach to reduce train accidents through a probabilistic assessment. *International Journal of Engineering and Technology (IJET)*. Vol. 4 No 6 Dec 2012-Jan 2013. pp. 500-509. ISSN: 0975-4024.
- Encyclopædia Britannica, 1911.
- Del Rio Vilas, D., Longo, F., Monteil, N.R., 2013. A general framework for the manufacturing workstation design optimization: A combined ergonomic and operational approach. *Simulation*, vol. 89, Issue 3, March 2013, Pages 306-329.
- Fabiano, B., Currò, F., Pastorino, R., 2004. A study of the relationship between occupational injuries and firm size and type in the Italian industry. *Safety Science*, 42 (7), 587–600.
- Falcone, D., Di Bona, G., Duraccio, V., Silvestri, A., (2007). Integrated hazards method (IHM): A new safety allocation technique. *Proceedings of the IASTED International Conference on Modelling and Simulation*. pp. 338-343. Montreal, QC; Canada
- Galán, S.F., Mosleh, A., Izquierdo, J.M., 2007. Incorporating organizational factors into probabilistic safety assessment of nuclear power plants through canonical probabilistic models. *Reliability Engineering and System Safety*, 92, 1131–1138.
- ISSA Prevention Series 2034, 2001.
- Jacinto, C., Canoa, M., Guedes Soares, C., 2009. Workplace and organisational factors in accident analysis within the Food Industry. *Safety Science*, 47 (2009) 626–635.
- Jang, H., Lee, S., Kim, S., 2010. Bayesian analysis for zero-inflated regression models with the power prior: applications to road safety countermeasures. *Accident Analysis & Prevention*, 42, 540–547.
- Khakza, N., Khan, F., Paltrinieri, N., 2014. On the application of near accident data to risk analysis of major accidents. *Reliability Engineering and System Safety*, 126, 116–125
- Lu, S., Wu, D., Lu S.C., Zhang H.P., 2011. A Bayesian Network Model for the Asian Games Fire Risk Assessment. *Proceedings of China located International Conference on Information Systems for Crisis Response and Management*.
- Matías, J.M., Rivas, T., Martín, J.E., Taboada, J., 2008. A machine learning methodology for the analysis of workplace accidents. *International Journal of Computer Mathematics*, 85 (3), 559–578.
- Naderpour, M., Lu, J., Zhang, G., 2014 (a). An intelligent situation awareness support system for safety-critical environments. *Decision Support Systems*, 59 (2014) 325–340.

- Naderpour, M., Lu, J., Zhang, G., 2014 (b). A situation risk awareness approach for process systems safety. *Safety Science*, 64, 173–189.
- Papazoglou, I.A., Aneziris, O., Post, J., Baksteen, H., Ale, B.J.M., Oh, J.I.H., Bellamy, L.J., Mud, M.L., Hale, A., Goossens, L., Bloemhoff, A., 2006. Logical models for quantification of occupational risk: falling from mobile ladders. *Proceedings of International Conference on Probabilistic Safety Assessment and Management*, May 13–19, 2006. New Orleans
- Roed-Larsen, S., 2003. Accident investigation practices – results from a European survey. *Proceedings of the JRC/ESReDA Seminar on Safety Investigation of Accidents*, Petten, Holland, 12–13 May 2003, pp. 9–14.
- Washington, S., Oh, J., 2006. Bayesian methodology incorporating expert judgment for ranking countermeasure effectiveness under uncertainty: example applied to at grade railroad crossings in Korea. *Accident Analysis and Prevention*, 38, 234–247.
- Yu, R., Abdel-Aty, M., 2013. Investigating different approaches to develop informative priors in hierarchical Bayesian safety performance functions. *Accident Analysis and Prevention*, 56, 51–58.

# A MULTIMODAL OPTIMIZATION METHOD FOR SIMULATION SYSTEMS

Zhizhao LIU <sup>(a)</sup>, Wei LI <sup>(b)</sup>, Ming YANG <sup>(c)</sup>

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

<sup>(a)</sup>[liuzhizhao2007@163.com](mailto:liuzhizhao2007@163.com), <sup>(b)</sup>[fleehit@163.com](mailto:fleehit@163.com), <sup>(c)</sup>[myang@hit.edu.cn](mailto:myang@hit.edu.cn)

## ABSTRACT

Considering multimodal optimization problem and potentially run times for continuous simulation systems, Particle Swarm Optimization (PSO) method based on Support Vector Machine (SVM) and Cluster Analysis (CA) is proposed. SVM, a global metamodel of simulation system, is built and PSO algorithm is used to search for local optimal points based on SVM. To determine local optimums, the population individuals of PSO are classified by CA. Consequently, six typical multimodal function optimization problems are selected to verify the optimization performance.

Keywords: simulation system, multimodal optimization, SVM metamodel, cluster analysis

## 1. INTRODUCTION

Simulation optimization is optimization of performance measures based on outputs from simulations (Fu 2001). Optimization problems exist widely in aviation, aerospace, shipbuilding, information and other fields, which is also the main research direction of simulation (Fu 2002, Yang Zhang and Wang 2004, Kleijnen 2007, Min Ma and Yang 2007, Fu and Chen 2008). Therefore, simulation optimization is an important part of simulation theory research.

In this paper, we consider the potentially extensive run times and multimodal optimization problem for continuous simulation system. Multiple extreme values usually exist in real systems, which is the same as simulation systems. The essence of this optimization problem is multimodal function optimization. In the field of simulation, there is little research on multimodal optimization problems. At present, intelligent optimization algorithms are applied for multimodal optimization problem, getting optimization solution by the iteration of population. However, expensive simulation limits simulation run times, leading the result that intelligent optimization algorithms can not be applied directly (Barton and Meckesheimer 2006, Barton 2009, Jin Chen and Simpson 2001). On the other hand, the last generation individuals are located around local optimums. The individuals near the same local optimum cannot be guaranteed to be coincident completely. How to determine the number of local optimal solutions, and obtain the corresponding values is also a valuable problem to study.

Based on the two problems above, PSO optimization method based on SVM metamodel and CA is proposed. The method is applied to multimodal function optimization problems, which is aimed to valid that the method proposed in this paper can get all the effective local optimal solutions in different situations of multimodal optimization.

## 2. RESEARCH ON THE KEY PROBLEMS

The multimodal optimization problem and determination of the local optimal points will be discussed, Section 2.1 for the first one and Section 2.2 for the second. As the evaluation of multimodal optimization is different from the general optimization, the evaluation indexes are given in Section 2.3.

### 2.1. Multimodal Optimization Problem

In view of the limitations of optimization method based on gradient, intelligent optimization algorithms are selected for multimodal optimization problem. Simulation optimization based on metamodel has advantages of rapid and practical. In order to avoid too many run times for simulation system, metamodel is introduced to the multimodal optimization method. In this paper, PSO based on SVM metamodel is proposed.

#### 2.1.1. PSO method

In 1995, J. Kennedy and R. Eberhart proposed an intelligent algorithm named PSO based on the foraging behavior of birds, with many advantages such as clear meaning of parameters, small computation of iteration, breaking away from local optimum and so on (Cui and Zeng 2011). The main iterative equation is given as Eq. (1). From the equation, what can be seen is that the population individual is affected by current velocity, the best position of this individual in history and the best position of all the individuals in history. If acceleration  $C_1$  is increased and  $C_2$  is decreased properly, population individual will be stable at/near local optimal points, which can achieve the purpose of local optimum searching.

$$\begin{aligned}
v_{ij}(t+1) &= w \cdot v_{ij}(t) + C_1 r_{1j}(t)(P_{ij}(t) - x_{ij}(t)) \\
&\quad + C_2 r_{2j}(t) \cdot (P_{gj}(t) - x_{ij}(t)) \\
x_{ij}(t+1) &= x_{ij}(t) + v_{ij}(t+1)
\end{aligned} \tag{1}$$

where,  $v_{ij}(t)$  denotes the  $j$  th dimensional velocity of individual  $i$  in the  $t$  th generation.  $w$  is a inertia constant.  $C_1$  and  $C_2$  are acceleration constants.  $r_{1j}(t)$  and  $r_{2j}(t)$  are random numbers independently.  $P_{ij}(t)$  is the best position in  $j$  th dimensional of individual  $i$  until the  $t$  th generation.  $x_{ij}(t)$  is the  $j$  th position of individual  $i$  in the  $t$  th generation.  $P_{gj}(t)$  is the best position in  $j$  th dimensional of all population individuals until the  $t$  th generation.

### 2.1.2. SVM metamodel

SVM is based on statistical learning theory, which uses structure-risk-minimum criterion instead of empirical risk minimization criterion and transforms low dimensional nonlinear problem into high dimension linear problems. It has the advantages of simple mathematical formula, intuitive geometric meaning, good generalization ability and so on (Pasolli Melgani Tuia et al 2014).

Given a training sample set:  $S = \{(x_1, y_1), \dots, (x_l, y_l)\}$ ,  $x_i \in R$ ,  $y_i \in R$ ,  $i = 1, 2, \dots, l$ . The Regression function  $f(x) = \omega \cdot x + b$  satisfies the following conditions (Nello and John 2004, Mountrakis Im and Ogole 2011):

$$\begin{aligned}
\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\
s.t. \quad f(x_i) - y_i \leq \xi_i + \varepsilon \\
y_i - f(x_i) \leq \xi_i + \varepsilon \\
\xi_i, \xi_i^* \geq 0 \quad i = 1, 2, \dots, l
\end{aligned} \tag{2}$$

where  $\xi_i, \xi_i^*, i = 1, 2, \dots, l$  are slack variables.  $\varepsilon$  - insensitive function is the loss function. By introducing Lagrange function, using kernel function  $K(x, y)$  instead of inner product  $\langle \psi(x), \psi(y) \rangle$  and combined with the KKT condition, the regression model expression is as shown in Eq. (3).

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b \tag{3}$$

### 2.2. Determination of the local optimal points

Since the population individuals of PSO algorithm are stable near the local optimal points of metamodel, each of the local optimal points gathers many individuals. Fig. 1 is an example for single input single

output system where  $\bullet$  denotes the PSO individual and  $-$  denotes the relationship of input and output variable.

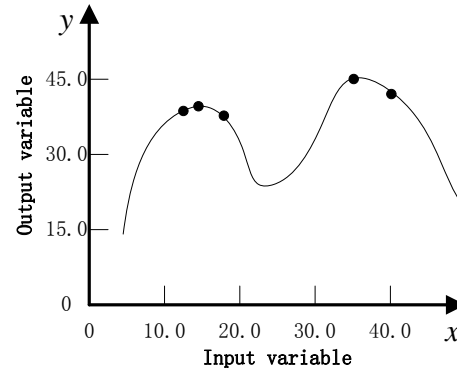


Fig. 1 PSO individuals of last generation

From Fig. 1, it is easy for human to determine how many local optimums the curve has from the curve or the PSO individuals. But for computer, it need specific algorithm. If there are multiple input variables, the input/output relationship could not be shown in a figure and even for people it is hard to find all the local optimal points from dozens or even hundreds of individuals.

In order to solve the problem, CA method is applied to cluster the individuals automatically. CA is a prevail method for data analysis (Yang 2005). It is widely applied in machine learning, data mining and statistics, etc. The basic process is mentioned which consists of 4 basic steps: 1) every sample is a cluster and compute similar measure; 2) the most similar two clusters are incorporated. 3) compute similar measure for new clusters and incorporate the most two clusters again; 4) iteration is done again and again until all the samples are summed up to one cluster.

Due to the advantages of local population gathered, we cluster individuals according to the distance of input variables and get the best output value in each cluster considered as the local optimal points. Suppose the input variable of population individual  $A_i = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{im}]$ , where  $m$  is the dimension of input. The distance of input variable is selected to be the similar measure as in Eq. (4).

$$s_{ij} = \sqrt{\sum_{k=1}^m (\alpha_{ik} - \alpha_{jk})^2} \tag{4}$$

Since the population individuals are near the local optimum, similar measure values called linkage distances near the same local optimum are little and the opposite are very large relatively. So the number of clusters can be determined by linkage distances in hierarchical diagram. Suppose that linkage distances between the two clusters are  $d_1, d_2, \dots, d_k$ . We use a threshold seen in Eq. (5). If the linkage distance is lower than threshold, the linked two clusters are incorporated to one cluster.



$$C_{sc} \cdot \max_{i=1,2,\dots,k} (d_i) \quad (5)$$

where,  $\max_{i=1,2,\dots,k} (d_i)$  is the maximum of  $d_i$  and  $C_{sc}$  is a constant.

### 2.3. Evaluation indexes of optimization results

Four indexes are used to evaluate the ability of searching local optimal points (Yang 2004). As bad local optimums have no meaning for the decider, only the local optimums close to the optimization one on the numerical value are required to consider. The detail content of five indexes is as bellow:

**Definition 1:** Number of Effective Peaks (NEP) for  $h$ : suppose that optimization algorithm gets  $n$  peak values, denoted by  $p_i$ ,  $i=1,2,\dots,n$ . If  $p_i$  meets  $p_i/p > h$  (in this paper  $h = 0.9$ ), where  $p$  is the corresponding peak value of  $p_i$ . Then,  $p_i$  is an effective peak for  $h$ . For  $p_i$ ,  $i=1,2,\dots,n$ , the number of effective peaks is NEP for  $h$ .

**Definition 2:** Maximum Peak Ratio (MPR): suppose that the algorithm gets  $n$  peaks, denoted by  $p_i$ ,  $i=1,2,\dots,n$  and the corresponding peak values are  $q_i$ ,  $i=1,2,\dots,n$ . Then MPR is equal to  $\sum_{i=1}^n p_i / \sum_{i=1}^n q_i$ .

**Definition 3:** Degree of Precision (DP): suppose that the algorithm gets  $n$  peaks, denoted by  $p_i$ ,  $i=1,2,\dots,n$  and the corresponding testing function's peak values are  $q_i$ ,  $i=1,2,\dots,n$ . Then DP is equal to  $\sum_{i=1}^n (q_i - p_i)^2$ .

**Definition 4:** Relative Maximum Absolute Error (RMAE): suppose that the algorithm gets  $n$  peaks, denoted by  $p_i$ ,  $i=1,2,\dots,n$  and the corresponding peak values are  $q_i$ ,  $i=1,2,\dots,n$ . Then RMAE is  $\max |(q_i - p_i)/q_i|$ .

## 3. PSO METHOD BASED ON SVM AND CA

In this study, we focus on the ability of searching effective local optimums for the optimization method. We consider the simulation system as a black box and that the number of local optimums is unknown. It is assumed to get the maximum value while the opposite situation is similar. The detailed steps for the method are given as follows:

### (i) Sampling and Metamodeling.

Select a proper experiment design method for sampling, and then build a SVM metamodel of simulation system. The SVM metamodel is as shown in Eq. (3).

### (ii) Metamodel evaluation.

There are three common performance measures to evaluate metamodel: Multiple Correlation Coefficient ( $R^2$ ), Relative Average Absolute Error (RAAE) and Relative Maximum Absolute Error (RMAE). They are shown in Eq. (6-8). In engineering application, the main evaluating index,  $R^2$ , should be more than 0.9 and the assistant evaluating indexes are good as small as possible. Under ideal condition,  $R^2$  equals to 1 while RAAE and RMAE equal to 0. If the evaluating indexes are satisfied, Step (iii) is executed. Otherwise, implement (i).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

$$RAAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i - \bar{y}|} \quad (7)$$

$$RMAE = \frac{\max(|y_i - \hat{y}_1|, \dots, |y_i - \hat{y}_n|)}{\frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}|} \quad (8)$$

where  $y_i$  is the  $i$  th output value.  $p_i, i=1,2,\dots,n$  is the  $i$  th predictive value.  $\bar{y}, i=1,2,\dots,n$  is the mean of outputs.

### (iii) Multimodal optimization.

Suppose that the PSO individual  $A_i^k = [\alpha_{i1}^k, \alpha_{i2}^k, \dots, \alpha_{im}^k]$  and the predicted fitness from SVM metamodel  $\hat{F}^k = [\hat{\beta}_1^k, \hat{\beta}_2^k, \dots, \hat{\beta}_n^k]$ , where  $k=1,2,\dots,n$ ,  $m$  is the number of input variables and  $n$  is the size of PSO population. The parameters of PSO algorithm is modified properly, making individuals could not skip from local optimum. After repeated iteration as in Eq. (1), all population individuals are stable near several local optimal points.

### (iv) Preliminary Screening.

Some individuals at the last generation will be rejected if their fitness is too bad. The screening equation is shown in Eq. (9).

$$\hat{\beta}_i^\xi \geq C_{ps} \cdot \hat{\beta}_{\max}^\xi \quad (9)$$

Where  $\hat{\beta}_i^\xi$  is the predicted fitness of individual  $A_i^\xi$  at the  $\xi$  th generation.  $\xi$  is the number of last generation.  $\hat{\beta}_{\max}^\xi$  is the maximum value of individual fitness until the  $\xi$  th generation.  $C_{ps}$  is a constant and its maximum value equals to 1. Go to step (v) directly if there is no need to screen.

**(v) Cluster Analysis.**

Confirm the number and values of local optimal points as follows:

1) In consideration of magnitude difference, the population individuals of PSO algorithm at the last generation are standardized as shown in Eq. (10).

$$\tilde{\alpha}_{ij}^{\xi} = \frac{\alpha_{ij}^{\xi} - b_j}{a_j - b_j} \quad (10)$$

Where  $\alpha_{ij}^{\xi}$  is the  $j$ th variable of individual  $A_i^{\xi}$ .  $a_j$  is the maximum value of the  $j$ th variable and  $b_j$  is the minimum one.

2) Every individual  $A_i^k$  can be seen as a cluster, so there are  $n$  clusters, denoted by  $G_1, G_2, \dots, G_n$ . The distance  $s_{ij}$  can be given in Eq. (4) and the distance matrix  $S = \{s_{ij}\}$ .

3) Get the minimum value from the matrix  $S$ , denoted by  $s_{ij}$ . The cluster  $G_j = \{A_{j,1}^{\xi}, A_{j,2}^{\xi}, \dots, A_{j,t}^{\xi}\}$  is incorporated into the cluster  $G_i = \{A_{i,1}^{\xi}, A_{i,2}^{\xi}, \dots, A_{i,r}^{\xi}\}$ . We get the new cluster  $G_i = \{A_{i,1}^{\xi}, A_{i,2}^{\xi}, \dots, A_{i,r}^{\xi}, A_{j,1}^{\xi}, A_{j,2}^{\xi}, \dots, A_{j,t}^{\xi}\}$ . Go to the step 2) until only one cluster exist.

4) Hierarchical diagram can be drawn and the number of clusters can be determined from the linkage distance. The clusters are denoted by  $G_1, G_2, \dots, G_v$ .

5) From the cluster  $G_k = \{A_{k,1}^{\xi}, A_{k,2}^{\xi}, \dots, A_{k,t}^{\xi}\}$ , we can get the corresponding fitness set  $H_k = \{\hat{F}_{k,1}^{\xi}, \hat{F}_{k,2}^{\xi}, \dots, \hat{F}_{k,t}^{\xi}\}$ . The maximum value of set  $H_k$  is the local optimal value denoted by  $\hat{F}_{k,\max}^{\xi}$ . The corresponding input variable value is  $A_{k,\max}^{\xi}$ . Through the  $v$  kinds of clusters, we can get  $v$  local optimal values denoted by  $\hat{F}_{1,\max}^{\xi}, \hat{F}_{2,\max}^{\xi}, \dots, \hat{F}_{v,\max}^{\xi}$  and the corresponding input variable values denoted by  $A_{1,\max}^{\xi}, A_{2,\max}^{\xi}, \dots, A_{v,\max}^{\xi}$ .

**(vi) Verification.**

Take local optimal points  $A_{1,\max}^{\xi}, A_{2,\max}^{\xi}, \dots, A_{v,\max}^{\xi}$  to the simulation model for achieving the actual output values  $F_{1,\max}, F_{2,\max}, \dots, F_{v,\max}$ .

**4. NUMERICAL EXPERIMENTS AND RESULTS**

To analysis the performance of proposed method quantificationally, six well known continuous deterministic multimodal optimization problems are tested. They indicate the input-output relationship. The expressions are as follows:

**Example 1** (Lu Liang and Zhang 2008)

$$f_1(x) = \sin^6(5.1\pi x + 0.5), x \in [0, 1]$$

**Example 2** (Lu Liang and Zhang 2008)

$$f_2(x) = e^{-4 \ln 2 \times (x-0.0667)^2 / 0.64} \sin^6(5.1\pi x + 0.5), x \in [0, 1]$$

**Example 3** (Liu Wang and Wang 2004)

$$f_3(x) = x(x+1)\sin[(2x-0.5)^2\pi-1], x \in [-1.5, 1]$$

**Example 4** (Zhang and Shao 2008)

$$f_4(x, y) = 200 - (x^2 + y - 11) - (x + y^2 - 7)^2, x, y \in [-6, 6]$$

**Example 5** (Bi and Wang 2011)

$$f_5(x, y) = 10e^{(-0.03(x-3)^2 - 0.03(y-3)^2)} + 8e^{(-0.08(x+5)^2 - 0.08(y+5)^2)} + 7e^{(-0.08(x-4)^2 - 0.04(y+7)^2)}, -10 \leq x, y \leq 10$$

**Example 6** (Zhang and Shao 2008)

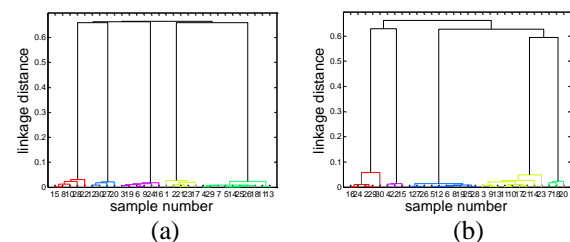
$$f_6(x, y) = \left( \frac{a}{b + (x^2 + y^2)} \right)^2 + (x^2 + y^2)^2, -5.12 \leq x, y \leq 5.12$$

Support Vector Machine (SVM) is chosen to be the metamodel of the six examples. The index values are shown in Table 1. The SVM index  $R^2$  is greater than 0.9.

Table 1: index values of metamodels

| NO. | $R^2$     | RAAE      | RMAE      |
|-----|-----------|-----------|-----------|
| 1   | 0.9991751 | 0.0283195 | 0.0719181 |
| 2   | 0.9969014 | 0.0576008 | 0.2498982 |
| 3   | 0.9996853 | 0.0270918 | 0.0481523 |
| 4   | 0.9918178 | 0.0220666 | 1.0733524 |
| 5   | 0.9999872 | 0.0033617 | 0.0274754 |
| 6   | 0.9238889 | 0.1059319 | 2.8144157 |

PSO algorithm is applied on the SVM metamodel where population size is 50, iteration number is 100,  $w=1$ , constant  $C_1=10$ ,  $C_2=0.1$ . So the information of last generation individuals can be got. Cluster analysis is applied according to the distance of inputs. The Hierarchical diagrams are shown as in Fig. 1. The threshold constant  $C_{sc}=0.1$ , so the numbers of clusters are respectively 5, 5, 8, 4, 3, 5. The local optimal points can be obtained as in Fig. 2.



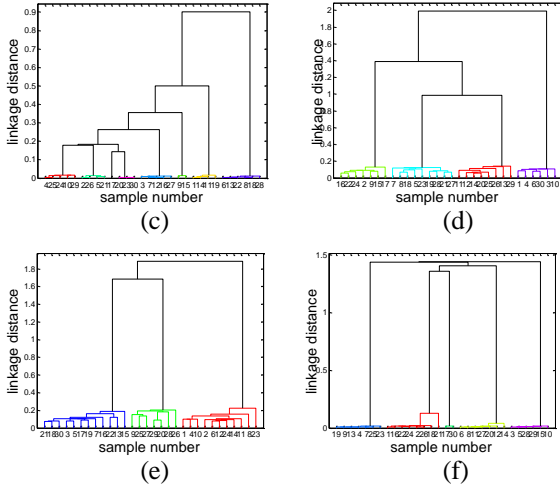


Fig. 1. Hierarchical diagrams (a) function 1 (b) function 2 (c) function 3 (d) function 4 (e) function 5 (f) function 6

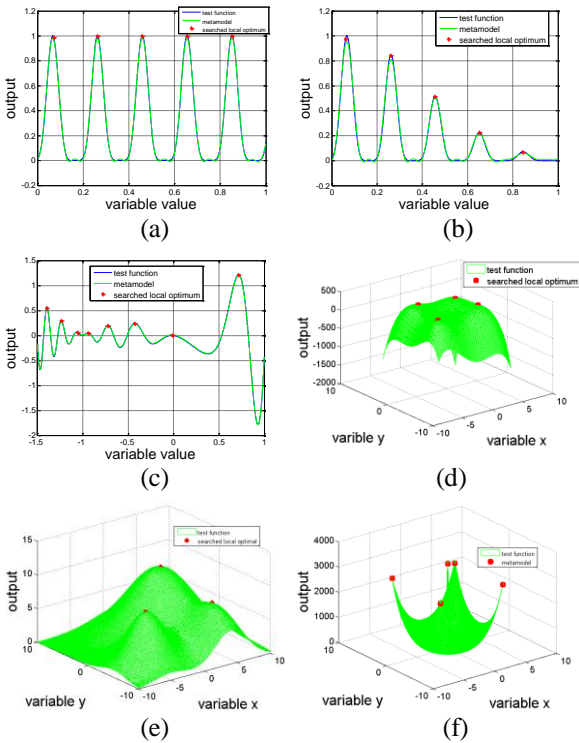


Fig. 2. Multimodal optimization results (a) example 1 (b) example 2 (c) example 3 (d) example 4 (e) example 5 (f) example 6

The ability index values of searching local optimal points are shown in Table 2. The EPN of each example is the same as the true number of examples, the MPR greater than 0.94, and the RMAE is lower than 0.08.

Table 2: Evaluation index values of searching local optimums

| NO. | Number of Peaks | NEP | MPR   | DP       | RMAE     |
|-----|-----------------|-----|-------|----------|----------|
| 1   | 5               | 5   | 0.994 | 0.000272 | 0.013320 |

|   |   |   |       |           |          |
|---|---|---|-------|-----------|----------|
| 2 | 5 | 5 | 0.998 | 0.000009  | 0.005434 |
| 3 | 8 | 8 | 0.996 | 0.000038  | 0.084723 |
| 4 | 4 | 4 | 0.988 | 45.509    | 0.031813 |
| 5 | 3 | 3 | 0.995 | 0.00691   | 0.008312 |
| 6 | 5 | 5 | 0.949 | 113112.47 | 0.066122 |

We evaluate the efficiency by  $s$ , the time ratio of proposed method and the direct optimization, which is good as small as possible. As the SVM modeling time and the optimization time of PSO can be ignored compared with that of simulation system run, the evaluation index  $s$  can be seen in Eq. (11).

$$s = \frac{N^* \times T}{N \times T} = \frac{N^*}{N} = \frac{N_{sample}}{N_{pop} \times (N_{gen} + 1) \times N_{times}} \times 100\% \quad (11)$$

where  $N^*$  is the run times of simulation system by the proposed method.  $N$  is the run times of simulation system by the direct optimization.  $T$  is the time of every simulation run.  $N_{sample}$  is the times of sampling.  $N_{pop}$  is the size of population.  $N_{gen}$  is the iteration number.  $N_{times}$  is the Monte Carlo times of PSO. In this paper,  $N_{pop} = 50$ ,  $N_{gen} = 100$  and  $N_{times} = 5$ . From the Table 3, the results show that the proposed method can largely shorten the time and increase the efficiency of optimization.

Table 3: Evaluation index values for efficiency increased

| NO. | $N_{sample}$ | $s$   |
|-----|--------------|-------|
| 1   | 101          | 0.40% |
| 2   | 101          | 0.40% |
| 3   | 251          | 0.99% |
| 4   | 169          | 0.67% |
| 5   | 441          | 1.76% |
| 6   | 529          | 2.12% |

## 5. CONCLUSION

PSO optimization method based on SVM and CA is proposed for multimodal optimization problem of simulation systems, which consider the multimodal optimization problem and the potentially run times. The number and values of local optimal points can be obtained automatically by computer without too much human labor. Through six typical problems, it can be shown that for the circumstances of equal and unequal peaks this method can find all the effective peaks with a high precision and largely shorten the optimization time. Future work will include the applications of the method to simulation systems and multimodal optimization method to provide robust solutions for discrete-event simulations.

## ACKNOWLEDGMENTS

The material is based upon work supported by the "Fundamental Research Funds for the Central Universities"(Grant No. HIT.NSRIF.2015035).

## REFERENCES

- Fu M.C., Simulation Optimization. Proceedings of the 2001 Winter Simulation Conference, pp.53-61(2001).
- Yang M., Zhang B., Wang Z. C., The Analysis of Modeling and Simulation Development Direction. Journal of System Simulation 16(9), 1901-1904+1913(2004).
- Fu M.C., Optimization for Simulation, Theory vs. Practice. Inform Journal on Computing 14(3), pp.192-215(2002).
- Fu M.C., Chen C.H., Some Topics for Simulation Optimization. Proceeding of the 2008 Winter Simulation Conference (WSC2008), pp.27-38(2008).
- Min F.Y., Ma P., Yang M., A knowledge-based method for the validation of military simulation. Proceeding of the 2007 Winter Simulation Conference (WSC2007), pp. 1374-1381 (2007).
- Kleijnen J.P.C., Design and Analysis of Simulation Experiments. Springer Science and Business Media (2007).
- Barton R.R., Meckesheimer, M., Metamodel-Based Simulation Optimization. Handbook in OR&MS, pp. 535-574 (2006).
- Barton R.R., Simulation Optimization Using Metamodels. Proceedings of the 2009 Winter Simulation Conference, pp.230-238(2009:).
- Jin R.C., Chen W, Simpson T.W., Comparative studies of metamodeling techniques under multiple modeling criteria. Structural and multidisciplinary optimization, 23(1), pp. 1-13(2001).
- Pasolli E.; Melgan F.; Tuia D.; et al, SVM Active Learning Approach for Image Classification Using Spatial Information, IEEE Transactions on Geoscience and Remote Sensing, 52(4), pp. 2217-2233(2014).
- Nello C., John S.T., An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press(2004).
- Mountrakis G.; Im J.; Ogole C., Support vector machines in remote sensing: A review, Isprs Journal of Photogrammetry and Remote Sensing, 66(3), pp. 247-259 (2011).
- Cui Z. H., Zeng J. C., Particle Swarm Optimization. Science Press (2011).
- Yang X. B., Research of Key Techniques in Cluster Analysis. Ph.D. degree of Zhejiang University (2005).
- Lu Q., Liang C.Y., Zhang E. Q., A dynamic sharing scheme based multimodal niche genetic algorithm. Proceeding of the 7th World Congress on Intelligent Control and Automation, pp.5333-5338(2008).
- Liu H. J., Wang X. F., Wang Z. B., Adaptive genetic algorithm for multi-peak searching. China

Academic Journal Electronic Publishing House, pp. 143-146(2004).

- Bi X. J., Wang Y. J., Niche artificial bee colony algorithm for multi-peak function optimization. Systems Engineering and Electronics, 33(11), pp. 2564-2568(2011).
- Zhang M.F., Shao C., Niche artificial fish swarm algorithm for multimodal function optimization. Control Theory & Applications, 25(4), pp. 773-776(2008).
- Yang H.Q., a study on techniques of improving genetic algorithm and its application. Master degree of Zhejiang University of technology (2004).

## AUTHORS BIOGRAPHY

**ZHIZHAO LIU** is a Ph.D. student at HIT, and received the M.E. from HIT in 2012. His research interests include simulation optimization and operational effectiveness evaluation. His email is liuzhizhao2007@163.com.

**WEI LI** is a Lecture at Harbin Institute of Technology (HIT), and received the B.S., M.E. and Ph.D. from HIT in 2003, 2006 and 2009 respectively. His research interests include simulation evaluation, simulation data analysis, and distributed simulation. His email is 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.

# FEM ANALYSIS OF RADIAL-AXIAL PROFILE RING ROLLING PROCESS

Zhengkun Feng<sup>(a)</sup>, Henri Champliaud<sup>(b)</sup>

Department of Mechanical Engineering,  
Ecole de technologie supérieure  
1100 Notre-Dame Street West,  
Montreal, Quebec, H3C 1K3, Canada

<sup>(a)</sup>[zhengkun.fengca@etsmtl.ca](mailto:zhengkun.fengca@etsmtl.ca), <sup>(b)</sup>[henri.champliaud@etstmtl.ca](mailto:henri.champliaud@etstmtl.ca)

## ABSTRACT

Profile radial-axial ring rolling is widely used for producing critical structural components in many industries. This complex bulk forming process is characterized by high nonlinearity, unsteady three-dimensional deformation and dynamic contact boundary conditions caused by the rotations of ring and rolls. In this paper, a numerical model based on finite element approach with non-uniform mesh of mixed cubic and tetrahedral elements is proposed to simulate and analyze the process. A large expanded ring with T-profile section has been obtained by numerical simulations.

Keywords: Profile radial-axial ring rolling, Finite element analysis, Residual stress, Bulk forming

## 1. INTRODUCTION

Ring rolling is widely used to produce seamless rings for critical structural components due to the advantages over the alternative processes (Allwood, Tekkaya, and Stanistreet 2005a). However, the trial and error technique still used in manufacture is intensive energy, time and labor consuming. As the fast evolution in computing technology has significantly reduced CPU time, more accurate analyses of the process can be performed by modeling and simulations.

Radial-axial ring rolling is characterized by unsteady three-dimensional deformation and high nonlinearity that have significant effects on the deformation behaviors of the process, as well as the mechanical properties and microstructure of the expanded rings. Although ring rolling process exists in manufacture more than one hundred years (Eruç and Shivpuri 1992a and 1992b), the process was mainly studied with experimental and theoretical methods prior to the 1990's (Johnson and Needham 1968). However, the process modeling based on finite element method becomes increasingly an important research subject due to the evolution of computing capacity. Utsunomiya et al. (2002) analyzed a cold ring rolling process with a two-dimensional finite element model because three-dimensional models required large computational time. Song et al. (2002) established also a two-dimensional model to simulate hot ring rolling process. Joun et al. (1998) established a viscoplastic model to investigate

the process. Moon et al. (2008) analyzed the origin of the polygonal-shaped defects in hot ring rolling process. Casotto et al. (2005) evaluated the sensitivity of cooling parameters to the ring defects of hot ring rolling process and the geometry of hot rolled rings at room temperature after cooling with a two-dimensional finite element model. Hua et al. (2009) established a finite element model to investigate the behaviors of cold ring rolling process. Yea et al. (2003) investigated the roll force, pressure distribution and contact length using a three-dimensional finite element model. The unsteady three-dimensional behaviors, high nonlinearity and the dynamic contact boundary caused by the continuous rotations of the ring and rolls consume large computing time. In order to reduce CPU consuming, Davey and Ward (2003) used the ALE formulation, and Ranatunga et al. (2004) used the Upper Bound Elemental technique. Forouzan et al. (2003a and 2003b) used a thermal spoke method to simulate the idle roll without considering the nonlinearity to reduce computational time. The explicit finite element model has been used by many investigators, including Wang et al. (2006), to increase computational efficiency.

Profile radial-axial ring rolling is more complex, particularly due to the geometry of the sections. A number of attempts have been made to study this process. In the 1970's, Mamalis et al. (1976) investigated experimentally the process. They analyzed the effects of mandrel feed rate, the initial section and the thickness reduction of the ring on the metal spread and the flow patterns. Recently, researches on hot radial-axial ring rolling process for large and profile rings and analyses of the microstructure, strain and stress distributions of rolled rings have increased: Yeom et al. (2007) investigated an optimum hot ring rolling process for a large-scale ring of Ti-6Al-4V alloy using FE analysis; Hua et al. (2007) established an analytical model for the motion of a guide roll in a profile ring rolling process; Kim et al. (2007) developed a profile ring rolling process to simulate the strain distributions in the deformation zone of large slewing rings; Hua et al. (2009) determined the relationship between deformation behaviors and the feed amount during cold ring rolling process for a small ring with L profile

section; Allwood et al. (2005b) evaluated the influence of process asymmetry on the stability of incremental ring rolling; Rytter et al. (2010) investigated the relationship between the deformation and the evolution of the microstructure and the texture of the ring in a cold ring rolling process; The influence of the ratio of feed rate per revolution at the main roll side to that at the idle roll side on the evolution of stress and strain was investigated by Wang (2011). However, publications are still rare on the subject of comprehensive analyses for hot radial-axial ring rolling process with large diameters and with profile sections.

Since the process is complex and characterized by highly nonlinear behaviors, unsteady three-dimensional deformation, continuous local forming, this research is focusing on establishment of a comprehensive model to predict the applied forces and the effects of process parameters on the shape of the expanded rings. The objective of this research is to obtain seamless rolled rings with uniform quality, smooth surface, short production time and less material loss. Investigation of the process behaviors will be carried out to analyze the geometry defects that relate to process parameters, the stress and strain, and the residual stress analyses. This paper provides preliminary investigation of radial-axial ring rolling process with profile sections based on finite element analysis with non-uniform mesh of mixed cubic and tetrahedral elements. Firstly, a numerical model based on finite element approach using the well-known ANSYS/LS-DYNA software for the process is built. Then, the simulation results are represented and discussed.

## 2. GEOMETRICAL SET UP AND OPERATING SEQUENCE

A profile radial-axial ring rolling machine may consist of a main roll, a mandrel, two conical rolls, and a doughnut-shaped blank used as initial ring (Fig. 1). There is a circumferential groove in the mid height of the mandrel to make a profile surface for the internal surface of the ring as illustrated on the top right corner of this figure. The conical rolls, with contacts on the upper and lower surfaces of the ring, locate at the opposite position to the contact surfaces of the radial rolls. The main roll is a driving roll which rotates around its stationary axis at a constant angular velocity, and drives the ring to rotate by the friction on their contact surfaces. The mandrel is an idle roll which advances at a constant feed rate towards the main roll. The radial pass between the main roll and the mandrel reduces the ring thickness by compression in the radial direction. Consequently, the ring diameter expands. Meanwhile, the two conical axial rolls rotate in opposite directions around their stationary axes and press towards each other for the purpose of creating an axial compression pass to reduce the ring height. In addition, they provide the stability and prevent from fishtail defects on the upper and lower surfaces of the expanded rings.

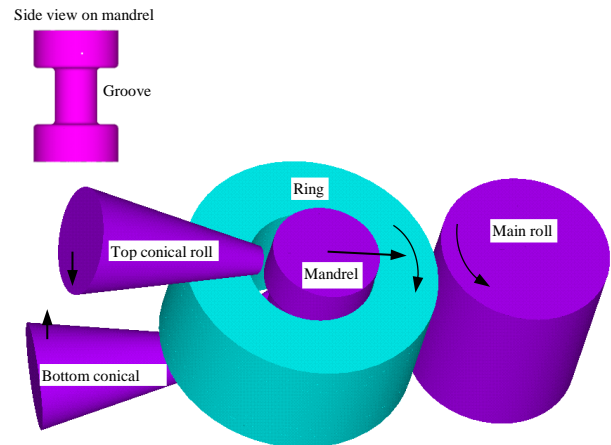


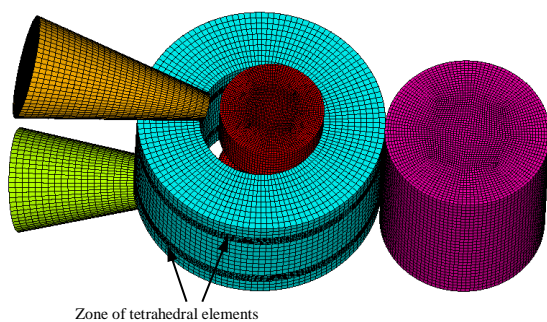
Figure 1: Geometrical set up

## 3. PROCESS MODELING BASED ON FINITE ELEMENT METHOD

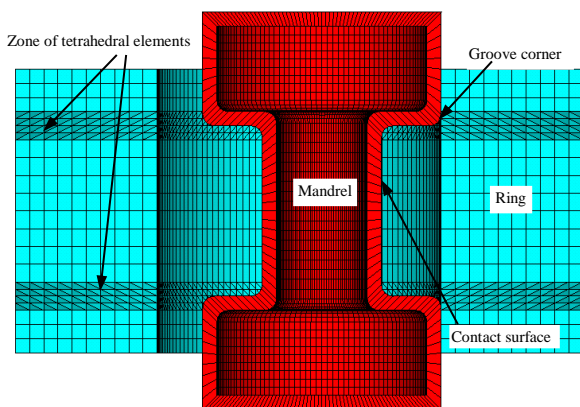
The dynamic explicit three-dimensional model for simulating the profile radial-axial ring rolling process was illustrated in Figure 2. The numerical model was created on ANSYS/LS-DYNA platform. The height, the internal and external radii of the ring were 1.0 m, 0.5 m and 1.0 m, respectively. The radii of the main roll and the mandrel were 0.665 m and 0.42 m, respectively and their heights were 1.4 m. The groove width and depth were 0.6 m and 0.5 m, respectively. The groove was axisymmetrical to the mandrel axis and symmetrical to the plane perpendicular to this axis which located at the mid height of the mandrel. Each conical roll had top radius of 0.112 m, bottom radius of 0.42 m and height of 1.5 m. The ring, the mandrel and the main roll were meshed with 8-node cubic elements. The mesh in the contact zones with the corners of the mandrel was refined and meshed with linear 4-node tetrahedral elements (Fig. 2b) because the larger deformation of the ring occurred at these locations (See the following section of Simulation results and discussions). The tetrahedral elements were derived from cubic elements. Each cubic element to be derived was divided into six tetrahedral elements without additional nodes. The six vertexes of each cubic element to be derived were the vertexes of the six tetrahedral elements. In addition, Flanagan-Belytschko stiffness form with exact volume integration for cubic elements was used for hourglass control (Flanagan and Belytschko 1981). The model was assumed as isotherm and the mechanical properties at temperature of 800 °C were used. Elastic perfectly plastic property was used for the material model of the ring. The elastic modulus, yield stress, Poisson's ratio and density of the ring were 92 GPa, 134 MPa, 0.3 and 7850 kg/m<sup>3</sup>, respectively. The mandrel and the main roll were meshed with the same element type as the ring. There was only one cubic element through the thickness for the mandrel. For the ring, there were 79360 elements of which 61440 elements were tetrahedral elements. There were 32384 nodes in total and 10 cubic elements through the thickness. The mandrel was



assumed as elasto-plastic with very high yield stress (1800 GPa) to provide better contacts with the ring surfaces during the process. The conical rolls and the main roll were assumed as rigid bodies and the conical rolls were meshed with 4-node shell elements. There was only one 5-integration-point element through the thickness for the conical rolls and one cubic element through the thickness for the main roll. The static and dynamic friction coefficients on the contact surface of ring with main roll and with mandrel were assumed constant as 0.6. No friction was assumed on the contact surfaces of ring with the conical rolls so that the conical rolls had no rotation. Automatic surface-to-surface contact type was selected for these contact surfaces. The speed of the main roll and the feed rate of the mandrel were assumed as constants.



(a) FEM model



(b) Sectional view on contact surface between the ring and the mandrel

Figure 2: Three-dimensional profile radial-axial ring rolling model

#### 4. SIMULATION RESULTS AND DISCUSSIONS

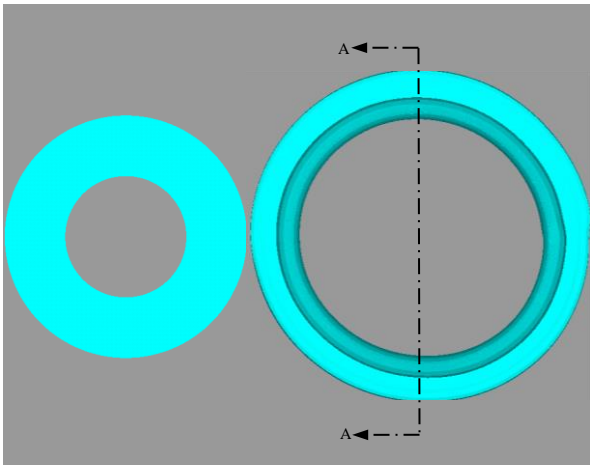
The total simulation time was 1000 seconds in the numerical simulation with the model described in the previous section. The time step was selected as  $10^{-4}$  seconds. The speed of the main roll and the feed rate of the mandrel were 0.17 rad/s and 0.75 mm/s, respectively. For simplicity, a small feed rate was applied to the conical rolls during the simulation to keep the stability of the system. At the end of the simulation, an expanded profile ring on the right side in Fig. 3a was

obtained from the initial ring with rectangular profile shown on the left side. Figure 3b shows the expanded profile ring with internal radii  $R_1$  and  $R_2$  of 0.725 m, 1.17 m and external radius  $R_3$  of 1.37 m as shown on the cutting section A-A in Fig. 3a. Figure 4 shows the Von Mises strain distribution on a top view of the expanded ring surface. The maximal strains were located in the narrower portion of the radial compression pass and near the contact surface with the mandrel. As the ring had clockwise rotation due to the friction on the contact surface with the main roll, the strain on the lower part in the figure was higher than that on the upper part of ring. It was much more significant on the ring internal surface which had the radius of  $R_2$  (See Fig. 3b). It was also observed from the top view in this figure that the strain decreased along the ring rotational direction from exit to entrance due to the progressive compression through the radial compression pass. Figure 5 shows the three principal strain distributions on the cutting section A-A in Fig. 3a. All of the three distributions and the deformed ring shape were symmetrical about the plane perpendicular to ring axis which is located at the mid height of the expanded ring. The first principal strain was the hoop strain which was positive at every point in the section (Fig. 5a). The second principal strain was the axial strain which was negative in the zone near the mid height of the ring and positive near the zones near the ends of the ring (Fig. 5b). The third principal strain was the radial strain which was negative at every point in the section (Fig. 5c). Figure 6 shows another cutting sectional view of the three principal strain distributions of the expanded ring. The cutting section was the symmetrical plane which was the plane perpendicular to ring axis and located at the mid height of the expanded ring. The first principal strain which presented the hoop strain was extension strain everywhere in the section (Fig. 6a). The second principal strain which presented the axial strain was compression strain everywhere in the section (Fig. 6b). The third principal strain which presented the radial strain was also compression strain everywhere in the section (Fig. 6c). As the feed rate of the conical rolls was very small, the axial strain was less than those in the other two directions and it is also observed in Fig. 5.

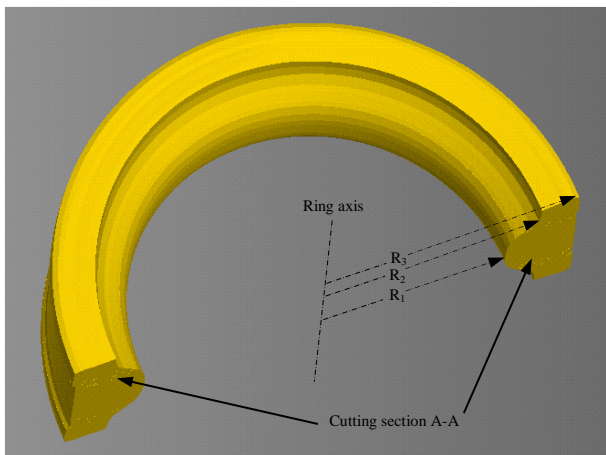
#### 5. CONCLUSIONS

A numerical model for profile radial-axial ring rolling has been established for analyzing the process behavior. Almost well expanded ring has been obtained by numerical simulations. The non-uniform mesh with linear 8-node cubic elements and 4-node tetrahedral elements has been used. The Flanagan-Belytschko stiffness form with exact volume integration has been used to control hourglass of cubic elements. The results show that the strain distributions on the expanded ring have much different patterns from radial-axial ring rolling for rectangular sections (Feng and Champlaud 2013). The future work will focus on establish a model

for expanded ring with more accurate profiles and validate the model with experimental results.



(a) Initial blank (left) and expanded ring (right)



(b) Sectional view on cutting section A-A  
Figure 3: Obtained expanded ring

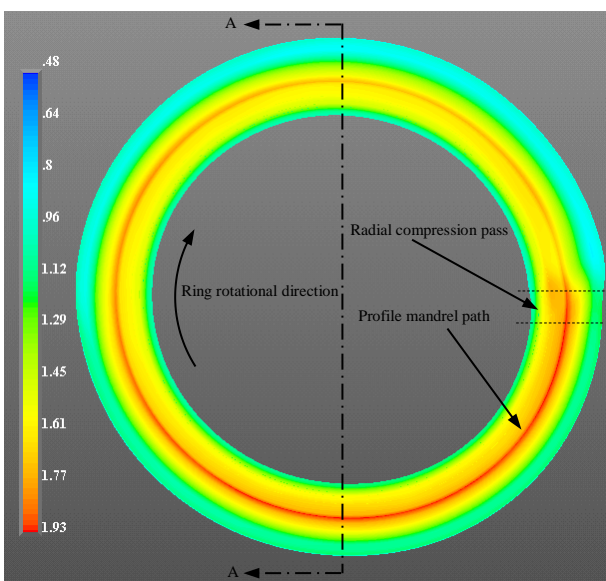


Figure 4: Distribution of Von Mises strain on the top surface of the expanded ring

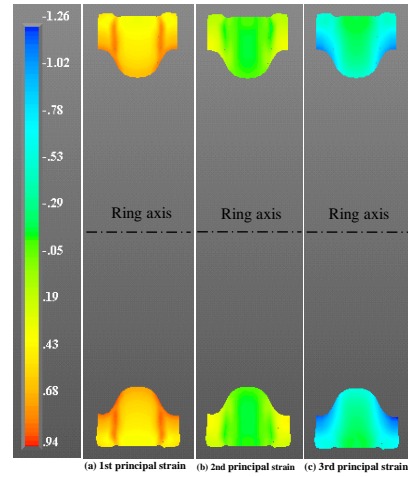
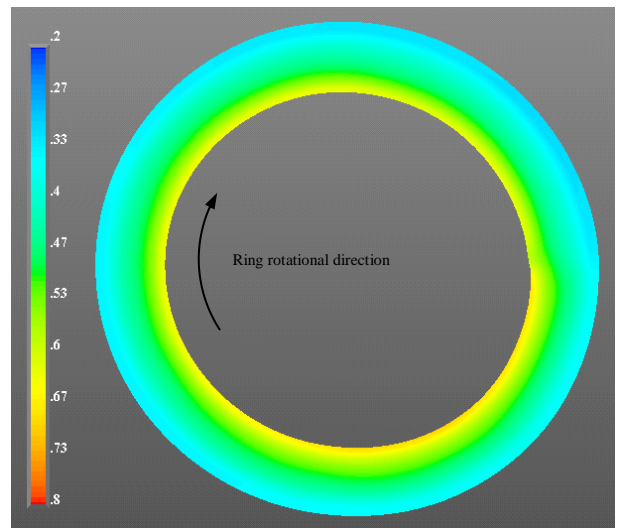
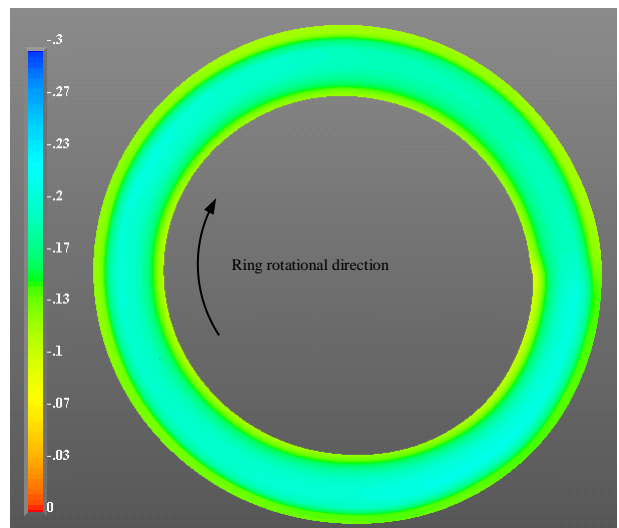


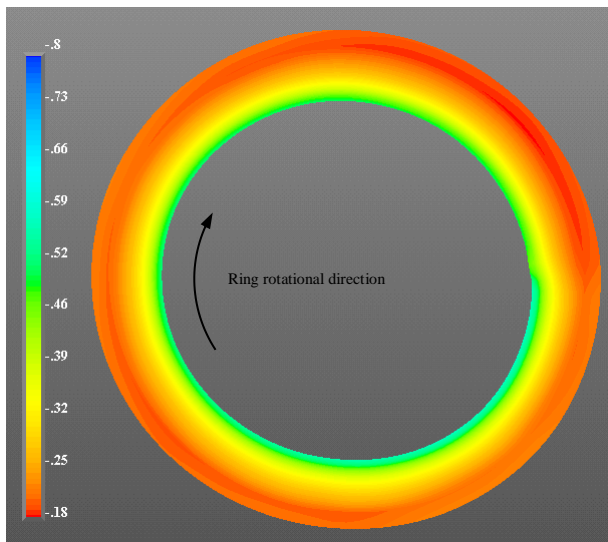
Figure 5: Distributions of the three principal strains on the cutting section A-A of the expanded ring



(a) Distribution of the first principal strain



(b) Distribution of the second principal strain



(c) Distribution of the third principal strain

Figure 6: Distributions of the three principal strains on the cutting section A-A of the expanded ring

#### ACKNOWLEDGMENTS

The authors thank the Natural Sciences and Engineering Research Council (NSERC) for the financial support to this research.

#### REFERENCES

- Allwood, J.M., Tekkaya, A.E. and Stanistreet, T.F., 2005a. The development of ring rolling technology. *Steel Research International*, 76(2-3), 111-120.
- Allwood, J.M., Kopp, R., Michels, D., Music, O., Oztop, M., Stanistreet, T.F., Tekkaya, A.E., Tiedemman, I., 2005b. The Technical and Commercial Potential of an Incremental Ring Rolling Process. *CIRP Annals - Manufacturing Technology*, 54(1), 233-236.
- Casotto, S., Pascon, F., Habraken, A.M., Bruschi, S., 2005. Thermo-mechanical-metallurgical model to predict geometrical distortions of rings during cooling phase after ring rolling operations. *International Journal of Machine Tools and Manufacture*, 45(6), 657-664.
- Davey, K. and Ward, M.J., 2003. An ALE approach for finite element ring-rolling simulation of profiled rings. *Journal of Materials Processing Technology*, 139(1-3), 559-566.
- Eruç, E. and Shivpuri, R., 1992a. A summary of ring rolling technology - I. Recent trends in machines, processes and production lines. *International Journal of Machine Tools and Manufacture*, 32(3), 379-398.
- Eruç, E. and Shivpuri, R., 1992b. A summary of ring rolling technology - II. Recent trends in process modeling, simulation, planning, and control. *International Journal of Machine Tools and Manufacture*, 32(3), 399-413.
- Feng, Z. and Champiaud, H. 2013. Modeling and simulation of hot radial-axial ring rolling process

for producing power plant parts. *Materials Science and Technology Conference and Exhibition*, 2013, MS and T 2013, pp. 677-686. October 27-31, Montreal (Quebec, Canada).

- Flanagan, D.P. and Belytschko, T., 1981. A Uniform Strain Hexahedron and Quadrilateral With Orthogonal Hourglass Control. *International Journal for Numerical Methods in Engineering*, 17(5), 679-706.
- Forouzan, M.R., Salimi, M., Gadala, M.S., Aljawi, A.A., 2003a. Guide roll simulation in FE analysis of ring rolling. *Journal of Materials Processing Technology*, 142(1), 213-223.
- Forouzan, M.R., Salimi, M. and Gadala, M.S., 2003b. Three-dimensional FE analysis of ring rolling by employing thermal spokes method. *International Journal of Mechanical Sciences*, 45(12), 1975-1998.
- Hua, L., Qian, D.S. and Pan, L.B., 2009. Deformation behaviors and conditions in L-section profile cold ring rolling. *Journal of Materials Processing Technology*, 209(11), 5087-5096.
- Hua, L., Zuo, Z.J., Lan, J. and Qian, D.S., 2007. Research on following motion rule of guide roller in cold rolling groove ball ring. *Journal of Materials Processing Technology*, 187-188(0), 743-746.
- Johnson, W. and Needham, G., 1968. Experiments on ring rolling. *International Journal of Mechanical Sciences*, 10(2), 95-113.
- Joun, M.S., Chung, J.H. and Shivpuri, R., 1998. An axisymmetric forging approach to preform design in ring rolling using a rigid-viscoplastic finite element method. *International Journal of Machine Tools and Manufacture*, 38(10-11), 1183-1191.
- Kim, K.H., Suk, H.G. and Huh, M.Y., 2007. Development of the profile ring rolling process for large slewing rings of alloy steels. *Journal of Materials Processing Technology*, 187-188(0), 730-733.
- Mamalis, A.G., Hawkyard, J.B. and Johnson, W., 1976. Spread and flow patterns in ring rolling. *International Journal of Mechanical Sciences*, 18(1), 11-16.
- Moon, H.K., Lee, M.C. and Joun, M.S., 2008. Predicting polygonal-shaped defects during hot ring rolling using a rigid-viscoplastic finite element method. *International Journal of Mechanical Sciences*, 50(2), 306-314.
- Ranatunga, V., Gunasekera, J.S., Vaze, S.P., De Souza, U., 2004. Three-Dimensional UBET Simulation Tool for Seamless Ring Rolling of Complex Profiles. *Journal of Manufacturing Processes*, 6(2), 179-186.
- Ryttberg, K., Knutson Wedel, M., Recina, V., Dahlman, P., Nyborg, L., 2010. The effect of cold ring rolling on the evolution of microstructure and texture in 100Cr6 steel. *Materials Science and Engineering: A*, 527(9), 2431-2436.

- Song, J.L., Dowson, A.L., Jacobs, M.H., Brooks, J., Beden, I., 2002. Coupled thermo-mechanical finite-element modelling of hot ring rolling process. *Journal of Materials Processing Technology*, 121(2-3), 332-340.
- Utsunomiya, H., Saito, Y., Shinoda, T, Takasu, I., 2002. Elastic-plastic finite element analysis of cold ring rolling process. *Journal of Materials Processing Technology*, 125-126(0), 613-618.
- Wang, M., Yang, H., Sun, Z., Guo, L., Ou, X., 2006. Dynamic explicit FE modeling of hot ring rolling process. *Transactions of Nonferrous Metals Society of China*, 16(6) 1274-1280.
- Wang, M., 2011. Comparison of evolution laws of stress and strain fields in hot rolling of titanium alloy large rings with different sizes. *Transactions of Nonferrous Metals Society of China*, 21(7), 1611-1619.
- Yea, Y., Ko, Y., Kim, N., Lee, J., 2003. Prediction of spread, pressure distribution and roll force in ring rolling process using rigid-plastic finite element method. *Journal of Materials Processing Technology*, 140(1-3), 478-486.
- Yeom, J.T., Kim, J.H., Park, N.K., Choi, S.S., Lee, C.S., 2007. Ring-rolling design for a large-scale ring product of Ti-6Al-4V alloy. *Journal of Materials Processing Technology*, 187-188(0), 747-751.



# SEARCHING AND INDEXING DISTORTED DATA COLLECTIONS

Tomáš Kocyan, Jan Martinovič, Michal Podhorányi

VŠB - Technical University of Ostrava

IT4Innovations

17. listopadu 15/2172, 708 33 Ostrava, Czech Republic

[tomas.kocyan@vsb.cz](mailto:tomas.kocyan@vsb.cz), [jan.marinovic@vsb.cz](mailto:jan.marinovic@vsb.cz), [michal.podhoranyi@vsb.cz](mailto:michal.podhoranyi@vsb.cz)

## ABSTRACT

Success of many models and artificial intelligence methods strongly depends on ability to quickly and precisely search input data collection. Despite the existence of many algorithms for faster searching, the most of them fail while processing distorted input. Unfortunately, the distortion is natural for many types of data collections, especially for measurements of natural phenomena such as precipitations, river discharge volume etc. In this type of collections, there are no exact levels for generated values. This paper discusses possibilities of indexing and searching such distorted inputs and also proposes an alternative approach for their indexing. The proposed approach utilizes the Voting Experts algorithm for splitting the input regarding statistical indicators, the Dynamic Time Warping for dealing with distorted inaccuracies and hierarchical clustering for grouping similar sequences. Finally, the sample result of proposed algorithm applied on data collections consisting of measured river discharge volumes is shown.

Keywords: time series, indexing, dynamic time warping, voting experts, symbolic approximation

## 1. INTRODUCTION

Searching time series data collections is a usual task in applications of many domains and in some of these applications, speed and accuracy of searching can be very crucial. This is mainly true for machine learning methods such as Case-Based Reasoning (Watson, 2007), where ability of precise searching similar situations strongly reflects overall success and usability of whole system.

There are many information technologies infrastructures which support hydrologic science. Common possibility is to storage hydrologic observations in a relational database. The purpose for such database is to store hydrologic observations data in an optimizing system. Main task for such system is to provide information in standard formats, to afford sufficient auxiliary metadata about data values and to design whole system for fast searches. A relational database format is used to provide querying capability to allow data retrieval supporting diverse analysis. Very

important part of such system or database is organizing data into clear departments such as where they have been measured, at which locations and for what period of time (Horsburgh, 2008). Logical structure of system facilitates searching of observed data or different kinds of patterns (e.g. rain or discharge patterns). Searching in such type of database is usually done by means of commonly used algorithms in hydrology. Spatial patterns in catchment hydrology are often extreme events and therefore developing of programs which might reveal them is very significant point of view in hydrologic community.

Many papers are focused on indexing and searching relatively short time series (called episodes) of the same length (e.g. Chen, Huang, Wang, and Wang 2009, Keogh and Pazzani 2000) stored separately, which were either collected in the form of partial outputs (e.g. of modeling software) or created by splitting a long-standing time series in equidistant intervals. In such cases, the indexing is pretty simplified. Despite the possible distortion, the only thing to be done is to choose a suitable metric or method defining similarity between two episodes. Then, common indexing methods can be used. For dealing with eventual distortions or inaccuracies, the Euclidian distance is usually replaced by Dynamic Time Warping (Keogh and Pazzani 2000), which will be described in Section 2.3.

However, a serious problem with indexing arises when the data collection is not split, e.g. it has a form of a single long-terming time series. In such cases, at first, it is necessary to meaningfully split the time series into the particular episodes. It can be done either by cutting the time series into equidistant intervals, or by analyzing its behavior and trying to identify the meaningful parts. The first approach is much easier and faster, but it has an obvious disadvantage – there is absolutely no control over the cutting process. By this way, a meaningful episode may be divided into the two parts, which causes loss of the important information. For this reason, we decided to follow the second approach using the Voting Experts (VE) algorithm introduced in Section 2.2. Since this algorithm is originally designed for processing categorical time series, the input has to be first converted into suitable format. It is done using the Symbolic Aggregate Approximation (SAX) described in

Section 2.1. Tests showed (Kocyan, 2012), that the combination of SAX and VE splits the input correctly, but very roughly (high precision, low recall). It means that the found splits were correct, but not complete. To solve this, we extended the splitting process by a refining part. Whole process of the fine splitting including creation of an index file can be seen in Figure 1: First of all, categorical time series is roughly split into the episodes. The episodes are then clustered by hierarchical clustering into the reasonable number of clusters of similar episodes. Using the DTW, the clustered episodes are split into shorter episodes. If the fineness is not sufficient, the process of clustering and splitting can be repeated. On the other hand, if the required fineness is reached, clusters' representatives can be chosen and the index file can be built.

This paper is organized as follows: First, used tools will be introduced in Section 2. Second, the proposed approach briefly introduced earlier will be described in detail. In the Experiments section, results of proposed approach will be demonstrated and compared with usual methods for indexing. At the end, both advantages and disadvantages of the approach will be summarized and the future work will be outlined.

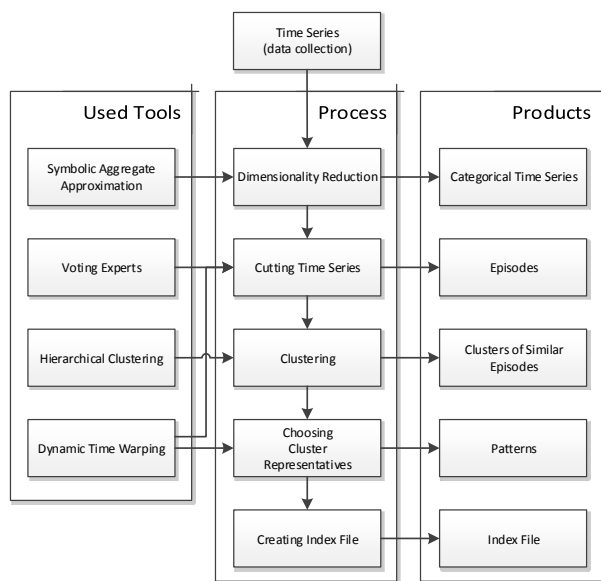


Figure 1: Proposed Approach Schema

## 2. USED TOOLS

In this section, used tools mentioned earlier will be described in necessary detail. For eventual deeper exploration of particular tools, please follow the corresponding references.

### 2.1. Symbolic Aggregate Approximation

Symbolic Aggregate approxImation (Lin, Keogh, Wei, and Lonardi 2007) is a simple dimensionality reduction method for transforming a time series  $T$  of length  $n$  into the string of length  $l$ , where  $l \ll n$ . The resultant string is composed from an alphabet  $A$  of size  $a > 2$ . The algorithm works in three steps: First, the time series is Z-Normalized (Goldin and Kanellakis 1995), i.e. transformed into the time series with approximately zero

mean and standard deviation close to 1. Then, in the second step, the Piecewise Aggregate Approximation (Lin, Keogh, Wei, and Lonardi 2007) is applied in order to reduce time dimension. At the end, the reduced time series is converted into a string. An example of original, z-normalized, PAA and SAX transformed time series is shown in Figure 2.

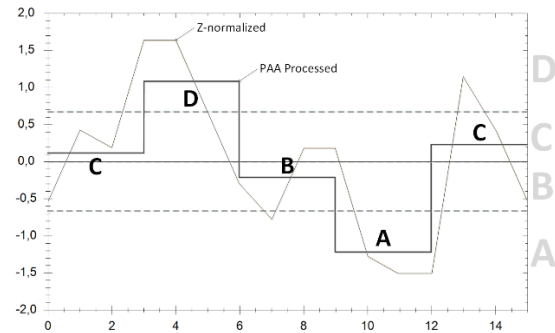


Figure 2: Example of PAA and SAX

### 2.2. Voting Experts Algorithm

The Voting Expert Algorithm is a domain-independent unsupervised algorithm for segmenting of categorical time series into meaningful episodes. It was first presented by Cohen and Adams (2001). Since this introduction, the algorithm has been extended and improved in many ways, but the main idea is always the same. The basic Voting Experts algorithm is based on the simple hypothesis that natural breaks in a sequence are usually accompanied by two statistical indicators (Cohen, Adams, and Heeringa 2007): low internal entropy of episode and high boundary entropy between episodes. The basic Voting Experts algorithm consists of the following three main steps:

- Build an nGram tree from the input, calculate statistics for each node of this tree (internal and boundary entropy) and standardize these values in nodes at the same depth.
- Pass a sliding window of length  $n$  over the input and let experts vote. Each of the experts has its own point of view on current context (current content of the sliding window) and votes for the best location for the split. The first expert votes for locations with the highest boundary entropy, the second expert votes for locations with a minimal sum of internal split entropy. By this way, the votes are counted for each location in the input.
- Look for local maximums which overcome selected threshold. These points are adepts for a split of sequence.

For detailed explanation of each of mentioned steps see (Cohen, Adams, and Heeringa 2007). Tests showed that the algorithm is able to segment selected input into meaningful episodes successfully. It was tested in many domains of interest, such as looking for words in a text (Cohen and Adams 2001) or segmenting of speech record



(Miller, Wong, and Stoytchev 2009). There are several ways how to improve the basic Voting Experts algorithm. Simply we can divide these improvements into the two main groups. On the one hand, a custom “expert” can be added to voting process (for example Markov Expert by Cheng and Mitzenmacher (2005)) and receive additional point of view on your input. On the other hand, there are methods based on repeated or hierarchical segmenting of the input (Miller and Stoytchev 2008, Hewlett and Cohen 2009).

### 2.3. Dynamic Time Warping

Nowadays, searching 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 episodes is significantly difficult, and almost excluded while using standard functions for similarity (distance) computation. Examples of such signals are presented in Figure 3. A problem of standard functions for similarity (distance) computation consists in sequential comparison of opposite elements in both episodes (comparison of elements with the identical indexes).

DTW is a technique for finding the optimal matching of two warped episodes using pre-defined rules (Muller 2007). 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 episodes from Figure 3 can be seen in Figure 4. This approach was used for example for comparison of two voice commands (Rabiner 1993). The main goal of DTW method is a comparison of two time dependent episodes  $X$  and  $Y$ , where  $X = (x_1, x_2, \dots, x_n)$  of the length  $n \in \mathbb{N}$  and  $Y = (y_1, y_2, \dots, y_m)$  of the length  $m \in \mathbb{N}$ , and to find an optimal mapping of their elements. A detailed description of DTW including particular steps of the algorithm is presented in (Muller 2007).

### 2.4. 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 exact pattern in some signal database. This case is demonstrated in Figure 5, where sequence  $s_1 = '3456'$  exactly matches the corresponding subsequence in sequence  $s_2 = '123456789'$ . However, in real situations exact patterns are not available because they are surrounded by additional values (Figure 5a), or even repeated several times in the sequence (Figure 5b).

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, own DTW modification was created. This modification is able to find the longest common time warped subsequences under selected restrictions. Because this paper’s topic is not focused on modifications of the DTW, see (Kocyan, 2012) for detail explanation.

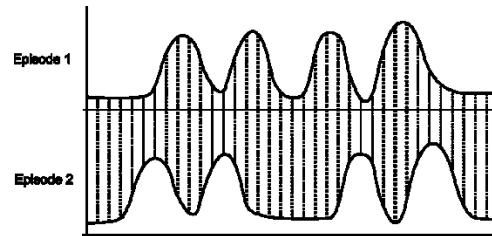


Figure 3: Standard Metrics Comparison

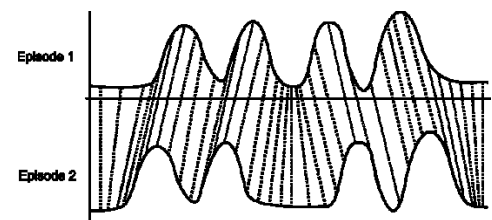


Figure 4: DTW Comparison

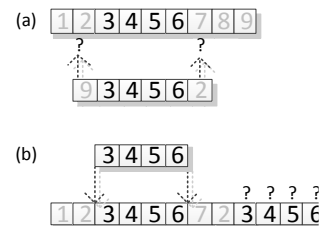


Figure 5: Basic DTW inaccuracies

### 2.5. Using DTW For Finding Cluster Representative

The DTW can be also utilized in cases, where it is necessary to gain the most suitable representative of the set of similar episodes. For this purpose, sometimes it is possible to use simple average of episodes  $X$  and  $Y$ , which means that for a representative episode  $R$  is valid, that:

$$R_i = \frac{x_i + y_i}{2}, \forall i = 1, \dots, o, \text{ where } o = |X| = |Y|. \quad (1)$$

However, this approach is not sufficient in cases, where we have data with distortion. Examples of such episodes are presented in Figure 6a and 6b. If only we used simple average presented in Equation 1, we would achieve an episode showed in Figure 6c. As we can see, this episode absolutely is not a representative and all the information about the episode course is lost. For this purpose, we designed our own DTW modification resulting in representative in Figure 6d. For detailed explanation how to derive representative using the DTW, see (Kocyan, 2012).

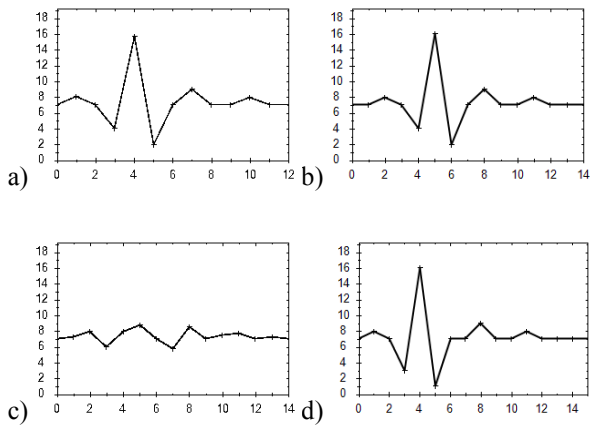


Figure 6: Sample Figure Caption

## 2.6. Voting Experts DTW Post-Process

Proposed solution for Voting Experts improvement takes the task of using Dynamic Time Warping algorithm (introduced above) and high precision cuts from Voting Experts as a starting point for looking for typical patterns located in the input. The basic idea is to refine the sparse set of high precision cuts into regular sequences as correctly as possible. The principle is very simple (as shown in Figure 7). If there are high precision cuts in the input (such as cuts A, B, C and D in Figure 7) and if the shorter sequence (bounded by cuts C and D) is a subsequence of the longer one (bounded by cuts A and B), we can deduce new boundaries E and F by projecting the boundaries of common subsequence to the longer sequence. In this very simplified example the sequences were composed by definite number of values and limited length, so the evaluation is quite straightforward.

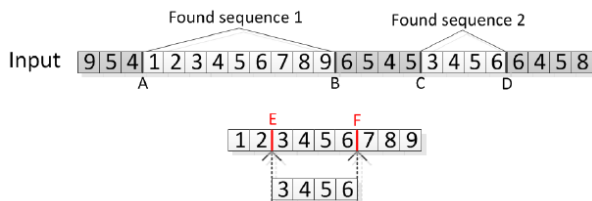


Figure 7: Refinement of sequences

In the case of application of previously mentioned process on distorted data, it is necessary to slightly modify it. Typical episodes of measurement of natural phenomena (such as precipitations, measured discharge volume etc.) are, unfortunately, subject to distortion in both time and value axes. For this reason, the Dynamic Time Warping (DTW) algorithm is used for this purpose. For our purposes, the DTW algorithm will be used as a tool for finding the longest common subsequence of two sequences (described in (Kocyan, 2012) in detail).

The main idea of the Voting Experts DTW post-process is summarized into the following steps:

1. First of all, the high precision (but not complete) cuts are created by splitting the input with high level of threshold by the Two-Way Voting Experts method.

2. Let's suppose that there are  $m$  unique sequences which have been created according to the cuts from step 1.
3. A  $m \times m$  distance matrix is build.
4. For each pair in this matrix, where the length of sequence  $s_1$  is bigger than length of sequence  $s_2$ :
  - (a) The optimal mapping of shorter sequence  $s_2$  to the longer sequence  $s_1$  is found by using DTW modified.
  - (b) If the mapping cost does not overcome selected threshold, the longest sequence  $s_1$  stores the shorter sequence  $s_2$  into its own list of similar sequences. By this way, every sequence gets its own list of the most similar shorter sequences.
  - (c) Each of the shorter sequences points to positions in the longer sequence, where it should be split. Because there is usually more than one similar shorter sequence, it is pointed to several locations whereas many of these locations are duplicated. For this reason, the votes are collected into internal vote storage.
  - (d) After these votes are collected, the local maximums are detected. These places are suggested as new cuts in original input.
5. The granted votes from step 4(d) are summed with votes of frequency and entropy experts in the input. Subsequently, the local maximums of votes are searched again. The cuts are made in locations where the number of granted votes is higher than the specified threshold.
6. Algorithm ends or it can continue with step 2 for further refinement.
7. After each post-process iteration, all found patterns can be received from the 4(b) step. Actually, the found patterns are groups of similar chunks. For building an index file, it is necessary to choose a representative in the same way as described in (Kocyan, 2013).

## 3. EXPERIMENTS

This part is focused on applying proposed algorithm on data collection containing measured river discharge volume. For this purpose, an artificial collection using analytic rainfall-runoff model and covering all possible situations was created.

Our process of creating data collection is done in three steps: First of all, the precipitations input containing real and artificial episodes is constructed. The input was sampled with one hour step for the total length of 10 years. Moreover, the partial episodes were distorted in both axes using predefined rules with respect to the normal distribution. After that, the input was split into short episodes and each episode was used as an input to the analytic rainfall-runoff model. At the end, the model's outputs were concatenated into the long-standing time series. As an environment for simulation run, the Floreon+ system was selected (Martinovič and Kuchař 2010). The input was composed from rainfall

episodes generated for both the left and the right tributary separately. Inserted episodes are chosen from one of the following categories:

1. Artificial Flash Flood precipitations.
2. Artificial Regional precipitations.
3. Real precipitations received from database.
4. Empty episode with no raining for decreasing the discharge volume.

The type of each episode was chosen randomly, but with a specific probability. By adjusting these probabilities, the final character can be strongly influenced. Once the episodes are selected, they have to be distorted to get closer the reality. The distortion is done in two stages: First, the selected rainfall episode is significantly distorted and used as a precipitation pattern for a single tributary. Then, this significantly distorted episode is finely distorted separately for each station belonging to the current tributary. Thanks to the distortion process described below, we get a random unique precipitation episode for both left and right tributary, but slightly varied for particular station. For detailed information see (Kocyan, 2014).

### 3.1. Post-Process adjustment

Adjusting the post-process part consists of several parameters. Despite the fact that the algorithm is built as domain independent, some domain knowledge is needed. First of all, it is necessary to set up the Voting Experts algorithm providing the rough segmentation. This algorithm needs categorical data on its input, so the measured data have to be converted first. It is done using the SAX described in paragraph 2.1. During the SAX conversion, a dimension reduction using PAA (described in Section 2.1) can be performed. Tests showed the PAA size of 6 sufficiently reduces the dimension with no loss of important information, regarding the nature of river discharge data. Data was converted in raw format, using its first differences and coefficients of growth. The number of output levels was set in all cases to 7. A sample of SAXed data can be seen in figures 8, 9 and 10. From the histogram in Figure 11 it is evident that using the first differences spreads the categorical data most symmetrically. For this purpose, this type of data was used.

Once the data is converted into the suitable format for Voting Experts, the sliding window size and threshold have to be determined. In order to take the necessary context into account, the size of sliding window have to be bigger than expected patterns. In our case of searching river discharge patterns, at least a “week” window has to be used. Since the discharge volume is measured with an hour step, a sliding window with a length of at least 28 (168 hours of measurement, PAA of 6 values) have to be considered. At the end, the window size was set to 30 and the threshold to 10.

Using this settings, cuts described as “VE CUT” in Figure 12 were made. By applying the post-process described above, additional cuts were progressively

added (marked with PP and cycle number). Since it is impossible to visualize whole data collection with increasing number of cuts, only increasing number of cuts and decreasing average length of sequences during the partial post-process cycles are shown in Figure 13.

Once enough short episodes are received, they are clustered into the clusters of similar episodes. For each of this similar episodes, a representative is derived as shown in Figure 14.

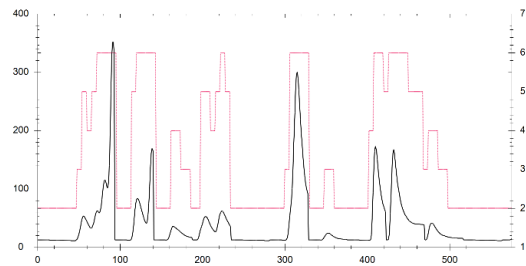


Figure 8: SAXed Raw Data

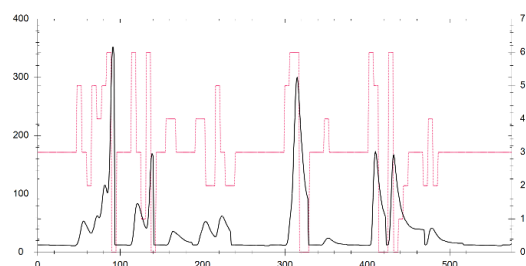


Figure 9: SAXed First Differences of Data

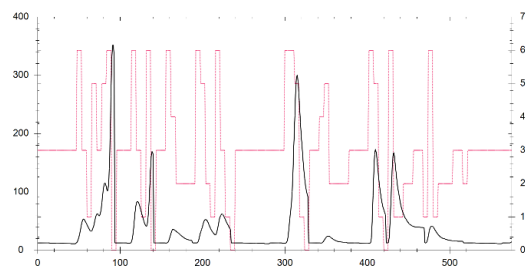


Figure 10: SAXed Growth Coefficients of Data

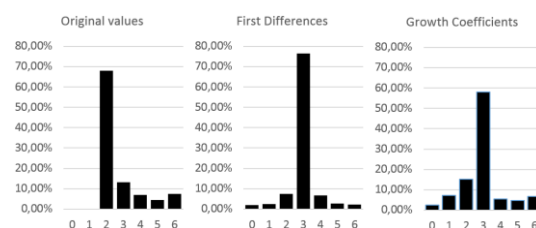


Figure 11: Histograms of SAXed Data

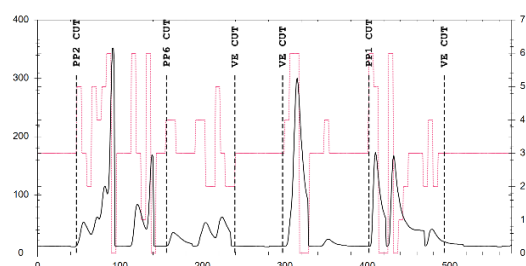


Figure 12: Voting Experts and Post-Process Cuts

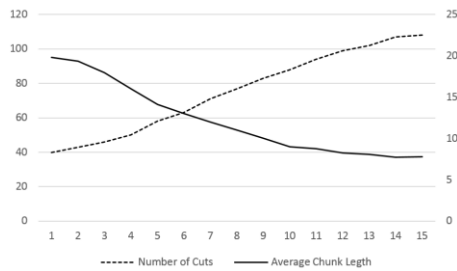


Figure 13: Voting Experts and Post-Process Cuts

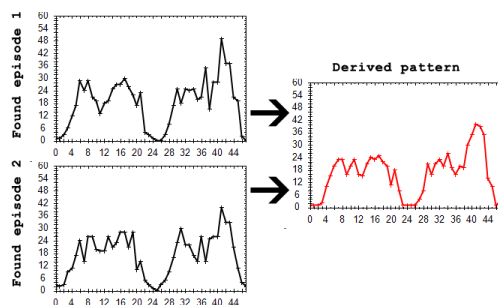


Figure 14: Found Episodes and Derived Pattern

## CONCLUSION

Tests showed that the proposed algorithm is able to find characteristics patterns in distorted data collection and build an index file for faster and more precise searching. Since the algorithm is mostly based on DTW and mutually independent parts, it is very easy to parallelize the computations and rapidly speed up the pattern search. 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 co-financed by the European Social Fund.

## REFERENCES

- Cheng, J., Mitzenmacher, M., 2005. Markov Experts. *Proceedings of the Data Compression Conference (DCC)*.
- Cohen, P. R., Adams, N., 2001. An Algorithm for Segmenting Categorical Time Series into Meaningful Episodes, *Proceedings of the Fourth Symposium on Intelligent Data Analysis, Lecture Notes in Computer Science*.
- 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*.
- 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.
- Hewlett, D., Cohen P., 2009. Bootstrap Voting Experts. *Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI)*.
- Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I. 2008. *A relational model for environmental and water resources data. Water Resources Research Hydrological Sciences* 44 (5).
- 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., Kuchař, Š., and Dvorský J., 2012. Unsupervised Algorithm for Post-Processing of Roughly Segmented Categorical Time Series. *Proceedings of Databases, Texts, Specifications, and Objects*.
- 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., Podhoranyi, M., Fedorčák, D., and Martinovič, J., 2014. Generating Rainfall-runoff Data Collection For Calibration Of Machine Learning Driven Models. *Proceedings of SGEM Multidisciplinary Scientific Conference*.
- 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
- Maheras, P., Kolyva-Machera, F., 1990. Temporal and spatial characteristics of annual precipitation in the twentieth century. *Int. J. Climatol.* 10: 495-504.
- Martinovič J., Kuchař, S., 2010. Multiple scenarios computing in the flood prediction system FLOREON+. *24th European Conference on Modelling and Simulation, Kuala Lumpur*.
- Miller, M., Stoytchev, A., 2008. Hierarchical Voting Experts: An Unsupervised Algorithm for Hierarchical Sequence Segmentation. *Proceedings of the 7th IEEE International Conference on Development and Learning (ICDL)*.
- 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.
- Watson, I., 1997. *Applying Case-Based Reasoning: Techniques for Enterprise Systems*, 15-38.

# TRAIN MOVEMENT DYNAMICS WITHIN ANYLOGIC TOOL

Roman Diviš<sup>(a)</sup>, Antonín Kavička<sup>(b)</sup>

<sup>(a)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice

<sup>(b)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice

<sup>(a)</sup> [roman.divis@student.upce.cz](mailto:roman.divis@student.upce.cz), <sup>(b)</sup> [antonin.kavicka@upce.cz](mailto:antonin.kavicka@upce.cz)

## ABSTRACT

The AnyLogic simulation tool serves for the simulation of railway traffic. The dynamics of train motion, however, are modelled in a simplified way based on acceleration invariables. When used for industrial purposes, such calculations can be inaccurate and the simulation thus does not yield valid results. In order to make the calculations of dynamics more accurate, the TrainDyn Library was created. It implements the train movement dynamics according to reality and includes the impact of forces acting during the train's motion.

Keywords: train movement dynamics, railway traffic simulation, AnyLogic tool

## 1. ANYLOGIC

The *AnyLogic* simulation represents a highly universal modelling tool enabling the utilisation of discreet simulation methods and agent-oriented approach. The creation of simulation models is supported by component libraries, which include various objects for the simulation of technological processes and social or economic systems. (Borschev 2013)

### 1.1. Enterprise Library

The *Enterprise* library represents an elementary object library for the creation of discreet simulation. The simulation model creator can use components for the creation of entities, operational processes and decision-making logic.

The *AnyLogic* simulation tool can be very well adjusted to the needs of a particular simulation. Utilisation of the *Java* programming language gives the possibility to influence the behaviour of the whole simulation. Individual library objects include events (e.g. creating an entity, processing an entity within a particular object, etc.) that enable to execute one's own *Java* code. The simulation model creator thus has the possibility to create his/her own entity classes or auxiliary classes and methods to be used within the simulation model. It is then to some extent possible to influence the behaviour of individual objects using the available *API* of the *AnyLogic* tool.

Apart from the *Enterprise library*, it is also possible to use objects from other libraries that can be

arbitrarily connected, which enables the creation of a complex simulation model.

### 1.2. Rail Library

*AnyLogic* also includes a library for the simulation of rail traffic. The library allow to propose and create railway traffic simulation within the *AnyLogic* environment. The simulation enables the creation of trains, their placement within the railway infrastructure and the management of their movement on the line.

The library's basic element, *TrainMoveTo*, serves for the setup of the train movement. Either the train has a set destination and the journey is automatically calculated or the train travels freely according to the set-up points on the track. By using other parameters it is possible to set up the acceleration or deceleration prior to the termination of the train's journey.

The dynamics of the individual trains' movement are, however, modelled in a simplified way on the basis of invariables defining the train's acceleration or deceleration. Many factors influencing the train's movement are not included. Majority of the simulation studies focusing on the microscopic simulation of the railway traffic, however, require a precise simulation of the rolling stock's movement. Precision may be crucial especially in industrial utilisation so that the yielded results can be considered valid. Therefore, it is necessary to adjust the simulation model and add an alternative way of dealing with the train movement dynamics.

## 2. TRAIN MOVEMENT DYNAMICS

Describing the movement of a system of objects requires the definition of the system's state at each moment. Mechanical system can be replaced by a point mass. Point mass is a particle, whose dimensions are negligible in relation to the other parameters. Phase coordinates of the point mass define its position within the three-dimensional space and its velocity at each moment. (Garg 1984)

Changes of phase coordinates are usually described by means of differential equations. A free object of  $n$  degrees of freedom requires  $n$  differential equations for the description of its movement in space. When a system consists of  $m$  mutually interconnected objects, it is necessary to set  $m.n$  differential equations.



Describing each railway vehicle in the system would be rather problematic. Experiments showed that a sufficiently precise calculation can be achieved by making a number of assumptions.

The rolling stock's movement can be mathematically described as the movement of a point mass with one degree of freedom. When exploring the movement of the rolling stock, it is possible to consider only the external forces, which act along the movement. The movement can be calculated according to the input and output without considering the dynamics of the processes taking place within the system. To simplify the calculation of the movement's dynamics it is further possible to use linear approximation of some non-linear continuous functions of the acting forces.

## 2.1. Mathematic Model

On the basis of the previous assumptions, the increase of work of the external forces can be defined as:

$$dA = (F - B - O) \cdot dL \quad (1)$$

in which  $F$  is the traction force,  $B$  breaking force,  $O$  resistance force and  $dL$  represents the gain in distance. Further adjustment of the equation requires the use of the relationship for the calculation of kinetic energy during translation and rotation:

$$E = \frac{1}{2} \cdot m \cdot v^2 + \sum \frac{1}{2} \cdot I \cdot \omega^2 \quad (2)$$

in which  $I$  is a moment of inertia,  $\omega$  is angular velocity,  $m$  weight of the system and  $v$  velocity of the system. The following equation holds for the moment of inertia:

$$I = m_r \cdot i^2 \quad (3)$$

where  $m_r$  represents the weight of the rotating parts and  $i$  is the distance from the rotation axis. The following equation holds for the angular velocity:

$$\omega = v \cdot \frac{u}{R} \quad (4)$$

where  $R$  is the wheel radius and  $u$  the velocity ratio of the wheel's revolution and the respective rotating part. After the equations are entered into the formula for the calculation of kinetic energy, kinetic energy is expressed as:

$$E = \frac{m \cdot v^2 \cdot \rho}{2} \quad (5)$$

where  $\rho$  is the coefficient of the rotating parts in:

$$\rho = 1 + \frac{\sum m_r \cdot i^2 \cdot u^2}{m \cdot R^2} \quad (6)$$

The change of kinetic energy under any change of state can be expressed in a differential form as:

$$dE = m \cdot \rho \cdot v \cdot dv \quad (7)$$

The work of the force acting on the system during a particular route equals the increase of the system's kinetic energy. Thus,  $dE = dA$ , and therefore, after entering into the previous equations (1,7):

$$m \cdot \rho \cdot v \cdot dv = (F - B - O) \cdot dL \quad (8)$$

The formula  $dL = v \cdot dT$ , in which  $dT$  is a gain in time, is valid for the gain in distance. After entering into the equation and the necessary adjustments, we arrive at the resulting formula:

$$F - B - O = m \cdot \rho \cdot \frac{dv}{dT} \quad (9)$$

## 2.2. Analysing the Acting Forces

The following section describes the individual forces acting in the movement of the rolling stock and their possible calculations and modelling for the purposes of computer simulation. (Iwnicki 2006, Dukkupati 2000)

### 2.2.1. Resistance Forces

Resistance forces represent a group of forces acting against the motion of vehicles, with the exception of braking forces. On the whole, resistance forces can be referred to as rolling resistance. It can be divided into two groups (i) vehicle resistance and (ii) track resistance.

Vehicle resistance depends on the construction of each individual vehicle. The physical basis of vehicle resistance rests in the point of contact of the vehicle with the track and the external environment. Vehicle resistance includes (i) resistance of the bearings, (ii) rolling resistance, and (iii) air resistance. Individual resistance forces can be expressed by means of explicit formulas, but their practical calculations are rather complex. Practical calculations make use of equations for calculating the coefficient of vehicle resistance instead of the specific value of vehicle resistance. The coefficient of vehicle resistance is defined as the quotient of vehicle resistance and the vehicle's weight. The coefficient of vehicle resistance can be expressed as a degree two polynomial with the velocity of the vehicle's movement as a variable:

$$o_v = a + b \cdot v + c \cdot v^2 \quad (10)$$

the polynomial's coefficients  $a$ ,  $b$ ,  $c$  were gained by empirical measurements and are available in the vehicle's technical documentation.

Resistance of the track depends on the track's state and construction. It consists of (i) inclination resistance, (ii) curve resistance and (iii) tunnel resistance.

Inclination resistance comes into play during the uphill or downhill journey of the rolling stock. Such movement incites a component of the gravitational force which is parallel to the track's inclination. During an uphill movement it creates resistance against the



acting traction force. Inclination resistance can be calculated as:

$$O_{skl} = G \cdot \sin \alpha \quad (11)$$

where  $G$  is the weight of the vehicle and  $\alpha$  is the angle of ascension or descending. The angle of ascension or descending is not used in practice; instead there are measurements of height gained during one-thousand-metre long sections of the vehicle's journey. Such gain in height is referred to as inclination and is defined as:

$$s = 10^3 \tan \alpha \quad (12)$$

In railway traffic, the value of the track inclination angle is very small (around  $4^\circ$ ). With small angles, the difference between functions  $\tan \alpha$  and  $\sin \alpha$  is rather small, which allows the use of  $\tan \alpha = \sin \alpha$  and we can proceed to:

$$O_{skl} = G \cdot \sin \alpha = G \cdot \tan \alpha = \frac{G \cdot s}{10^3} \quad (13)$$

The coefficient of inclination resistance can thus be expressed as:

$$o_{skl} = \frac{10^3 \cdot O_{skl}}{G} = \frac{10^3 \cdot s}{10^3} = s \quad (14)$$

Curve resistance occurs during the vehicle's journey through a curve during which the friction between the track and the vehicle's wheels increases. The general function for the calculation of the curve resistance coefficient is rather complex and involves a number of factors (track gauge, skid friction, track's degree of incline, curve radius, etc.). Due to the complexity of the issue, practical calculations use empirical formula such as:

$$o_{obl} = \frac{a}{R-b} \quad (15)$$

with  $R$  as the curve radius and  $a$ ,  $b$  as the defined coefficients for the given track type.

Tunnel resistance is caused by the air inside a tunnel. It is caused by the increase of drag during the vehicle's passage through a tunnel. The vehicle pushes an air column which is unable to escape and it causes a massive whirl. The tunnel resistance coefficient has its defined values of  $2 \text{ N.kN}^{-1}$  for a single-track tunnel and  $1 \text{ N.kN}^{-1}$  for a double-track tunnel.

### 2.2.2. Braking Forces

Braking force acts against the direction of the vehicle's movement in a similar way as the resistance forces. Contrary to the resistance forces thought, braking force is created artificially with the aim of slowing down or halting the vehicle. Mechanical brakes use the friction resistance. The calculation of the coefficient of friction caused by the braking process can be based on the force, created by the brake on a wheel, and the vehicle's

velocity. Practical calculations can assume that braking causes a constant deceleration of vehicles.

### 2.2.3. Traction Forces

Traction force acts in the direction of the vehicle's movement and is caused by a tractor. Traction force can be further divided according to the point of action into (i) indicated traction force, (ii) traction force on the wheel circumference and (iii) traction force on the coupling. As to the calculation of dynamics, we are most interested in the traction force on the wheel circumference, which represents the indicated traction force with the deduction of losses within the engine and losses caused by the transmission of force from the engine to the wheels.

The extent of the traction force, which the tractor can cause, can change according to operational velocity. The dependence between the traction force and velocity under specific output of the tractor cannot be expressed analytically. It is thus expressed graphically in a square grid system and is referred to as traction characteristics.

### 2.3. Movement Equation of the Rolling Stock

The previous analysis of the individual forces acting during the vehicle's journey enables the formulation of the differential equation for tractors as well as pulled vehicles:

$$f_o(v) - b(v) - o_{vh}(v) - o_t(v) = \rho \cdot \frac{10^3}{g} \cdot \frac{dv}{dt} \quad (16)$$

where  $f_o$  is the specific traction force on the wheel circumference,  $b$  specific braking force,  $o_{vh}$  the coefficient of vehicle resistance of the train,  $o_t$  the coefficient of track resistance,  $\rho$  the coefficient of rotating parts,  $v$  velocity and  $t$  time.

Solving the differential equation on computer requires one of the numerical methods. The simplest is the Euler numerical method, which solves the calculation in steps and the approximation derivation is calculated as a function value in each respective step.

In order to find out the dependence between the velocity and the distance within time  $T_0$  to  $T_n$ , the initial velocity of the movement  $V_0$  must be known. The equation can be expressed as:

$$k \cdot \frac{dv}{dt} = \varphi(v, t) \quad (17)$$

in which the right side of the equation is considered a constant in each iteration step. Time  $T_0$  to  $T_n$  is divided into  $n$  equal sections, the calculation step  $h$  is then defined as:

$$h = \frac{T_n - T_0}{n} = \Delta T \quad (18)$$

Derivation in each point of the curve  $v(t)$  is substituted by the quotient of final differences and the equation becomes:

$$k. \frac{\Delta v}{\Delta t} = \varphi(v, t) \quad (19)$$

and for each step it is valid that:

$$\begin{aligned} k. \Delta v_0 &= k. (V_1 - V_0) = \varphi(T_0, V_0). h \text{ for } v(T0) = V0 \\ k. \Delta v_1 &= k. (V_2 - V_1) = \varphi(T_1, V_1). h \text{ for } v(T1) = V1 \\ &\dots \\ k. \Delta v_n &= k. (V_n - V_{n-1}) = \varphi(T_{n-1}, V_{n-1}). h \text{ for } v(Tn - 1) = \\ &V_n - 1 \end{aligned} \quad (20)$$

Thus the equations for the calculation of velocity and covered distance in step  $i$ . can be expressed as:

$$V_i = V_{i-1} + \frac{h \cdot (f_0(V_{i-1}) - b(V_{i-1}) - o_{vt}(V_{i-1}) - o_t(L_{i-1}))}{k} \quad (21)$$

$$L_i = L_{i-1} + \frac{V_{i-1} + V_i}{2} \cdot h \quad (22)$$

The presented formulae enable the calculations of velocity and covered distance at each moment. The calculation's precision depends on the size of the time step  $\Delta T$ .

### 3. TRAINDYN LIBRARY

To enable the simulation of train movement using dynamics, it is necessary to adjust the simulation model. On the other hand, it would be demanding and ineffective to adjust each model and complete it with dynamics calculations. The *AnyLogic* environment allows the creation of a simulation model that can be further used through other simulation models as a library. The *TrainDyn* library was created in the same way and includes the implementation of the train journey dynamics, as described in the previous chapter.

The library implements its own calculation core, which lists detailed information on railway traffic and individual trains within the simulation model. Such information is then processed and used for the calculation of train movement dynamics.

The library uses the principle of activity scanning. The calculation of dynamics sets off periodically and the newly calculated velocity are set up for the respective train entities. This connection is carried out by the methods on the *API* interface of *AnyLogic*. The simulation model creator can adjust the sampling period as required.

The advantage of this solution is its rather simple realisation for the simulation model creator and the possibility to change the precision/complexity of calculation according to the selected sampling period. The disadvantage is that the train's velocity between two runs of the dynamic calculations remains constant. Another shortcoming of the *AnyLogic* tool is the inability to stop the train completely. It is possible to decrease the velocity to a very small value, but a complete halt of the train is achievable only by means of a well set up *TrainMoveTo* feature.

#### 3.1. Structures of the Railway and Trains

To carry out the dynamics calculations it is necessary to have more information about the train, which do not

belong to the standard *Train* class within the *AnyLogic* tool. Information about the train's characteristics is contained in the *TrainCharacteristics* class and includes the weight of the tractors as well as pulled vehicles, braking deceleration, traction characteristics of the locomotive, coefficients for the calculation of vehicle resistance and maximum train velocity.

To enable the work with train entities, which include the above-specified information, class *DynTrain* was created. It inherits from the standard *Train* class, which is an entity used in railway simulations. *DynTrain* includes train characteristics and other parameters needed for dynamics calculations (time lapsed from the last carried out calculation, required velocity and state of movement).

Railway traffic also lacks the parameters needed for dynamics calculations. It concerns mainly the track inclinations, location and radius of curves and location of tunnels. These parameters are passed on within the initialisation code of the simulation directly into the simulation core of the library. The simulation core is realised by the *TrainDynamics* class, which uses the *Singleton* design pattern. Therefore, a class instance can be gained from any part of the model and can be communicated with.

#### 3.2. API Interface of the TrainDyn Library

The classes *DynTrain* and *TrainDynamics* represent the basic functional blocks for the library's functioning. Their communication with the simulation model is carried out by means of a simple *API* interface.

The *DynTrain* class contains the following public methods:

- `setTrainCharacteristics(TrainCharacteristics tch)` – a method which sets up the train characteristics (parameters needed for the journey dynamics calculation);
- `accelerate(double speed)` – sets up new required velocity of the train and the dynamics calculation starts to accelerate or decelerate the train as required;
- `stop()` – sets up the braking of the train until a complete halt.

The *TrainDynamics* class includes the following public methods:

- `static getInstance()` – returns the instance of the *TrainDynamics* class (design pattern *Singleton*);
- `addArc(ShapePolyLine polyline, int radius)` – marks the information on a curve on the line located on the *polyline* section with the radius of *radius* metres;
- `addArc(ShapePolyLine[] polyline, int radius)` – similar to the previous method for the collective information adding;
- `addTrain(DynTrain train)` – adds the *train* to the dynamics calculation;
- `removeTrain(DynTrain train)` – removes the *train* from the dynamics calculation;

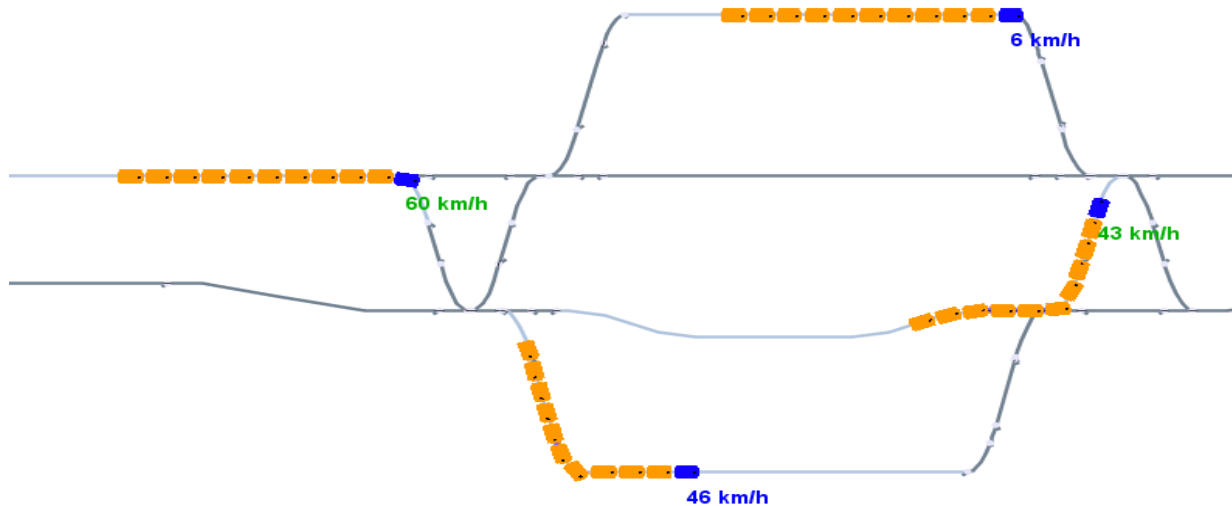


Figure 1: Testing topology simulation

- `calculateSpeeds(double time)` – carries out the dynamics calculation for all added trains, the *time* value provides the actual simulation time;
- `processTrains(ActionTrain actionTrain)` – allows for the *actionTrain* callback for all added trains within the calculation core.

The above-mentioned API enables the communication with the library and the use of its possibilities for managing train dynamics. The processes necessary for the library's utilisation are described in the following section of this article.

### 3.3. Using the Library

To be able to calculate the train's movement dynamics using the library, the simulation model creator must take several basic steps in preparing the model and then connecting the library with the simulation model. The following description presents a proposed model including the railway infrastructure and the *TrainSource*, *TrainMoveTo* and *TrainDispose* features for managing the simulation.

Prior to its use, the library must be filled into the simulation model. In the simulation's model set-up in the *Dependencies* section, the creator must add a reference to the JAR file of the dynamics library and also set up the elementary time unit of the simulation to one second. The following step involves adding information about the railway infrastructure. The *Startup code* of the simulation model must be completed by an `addArc()` call and the other railway parameters must be set up.

Then it is necessary to adjust the *TrainSource* features for creating the entities of *DynTrain* type (or the related class) and add the method call `addTrain()` in the *On exit* code. In a similar way, the individual trains must be removed within the *TrainDispose* features, where the method call `removeTrain()` is added into the event *On enter*.

The actual calculation is realised on the basis of a periodical method call `calculateSpeeds()`. Therefore, the *Event* feature must be created in the simulation model and must be set up to the *Trigger type: Timeout* a *Mode: Cyclic*. The time of repeating the event defines the precision, but also complexity of the calculation. The starting value can be set up to one second and then adjusted as required. Action must be completed with the actual

call:

```
TrainDynamics.getInstance().calculateSpeeds(time());
```

For train movement, the features *TrainMoveTo* and *TrainSource* can be completed with a call `accelerate()/stop()`, which sets up a new mode of train movement.

### 3.4. Case Study

Some functionality tests and validation of the calculation results were carried out already during the implementation of the TrainDyn library. Validation was carried out in form of verification of the actual tachograph records and their comparison with the calculations.

A case study was compiled during the library's first testing process. It includes the infrastructure of three simplified train stations. One-direction train traffic was generated within this infrastructure with a stop on a randomly chosen track at the first station. The topology is presented in figure 1. The trains are marked with their current velocity. Blue colour means that the train is braking to stop at the station, green means acceleration or constant speed. Displaying the description of each train is based on the `processTrains()` call, which moves the description to the front of the train and updates the train's velocity.

## 4. CONCLUSION

In order to use the train movement dynamics calculation for the purposes of industrial railway simulations it was necessary to adjust the *AnyLogic* simulation tool. That

was the reason for designing and implementing the *TrainDyn* library which realises its own dynamics calculation core and enables the simulation model creator rather simple connection of the library with new and existing standard models. The user sets up the movement of individual trains through several basic *API* calls and the actual dynamics calculations are carried out by the library.

Currently, the library just needs the implementation of several values of track resistance, which are not yet included in the calculations. In case higher precision is required, the sampling period of dynamics calculation can be adjusted or alternative way of numerical integration can be implemented. The testing and validation of results with the current method yielded satisfactory results and thus, the implementation of more complex methods was not necessary.

#### **REFERENCES**

- Garg, V., 1984. *Dynamics of Railway Vehicle Systems*. Burlington: Elsevier Science.
- Iwnicki, S., 2006. *Handbook of railway vehicle dynamics*. Boca Raton: CRC/Taylor.
- Borshchev, A., 2013. *The Big Book of Simulation Modeling. Multimethod Modeling with Anylogic 6*. Lisle, IL: AnyLogic North America.
- Dukkipati, R.V., 2000. *Vehicle dynamics*. Boca Raton: Narosa Publishing House.

# SELF-OPTIMIZING REAL-TIME RECOMMENDATION MODEL DESIGN BASED ON YARN

Tao liu   Shuang wang   Peng wu

School of Computer Science and Technology  
Wuhan University of Technology  
Wuhan, China  
e-mail: [1272348068@qq.com](mailto:1272348068@qq.com)  
e-mail: [ws123xhx@gmail.com](mailto:ws123xhx@gmail.com)  
e-mail: [425664759@qq.com](mailto:425664759@qq.com)

## ABSTRACT

In real-time personalized recommendation systems, aimed at some existing issues such as timely system performance requirements, large data processing capacity, the existence of bidirectional cold-start, the difficulties of spotting the potential hot news, the deficiencies of system self-optimization and etc, we propose enhanced self-learning optimization model based on the YARN platform, which illustrates algorithm scheduling framework and self optimization system revised by users' feedback. Furthermore, to fully exhibit the superiority of the YARN platform, the label propagation algorithm here used exemplifies the polymerization process of raw large data. Thus, a highly prompt recommendation system overall is achieved.

Keywords: Real-time recommendation model , Self-optimizing, Scheduling algorithm framework, YARN platform

## 1. INTRODUCTION

In the personalized recommendation system, algorithm scheduling framework as a link connecting recommendation algorithms with product features, will directly determine the content recommended for users and the costs required to achieve the desired effects. For strongly real-time articles, such as news designed for specific subscribers, huge amount of information along with short life cycle makes the traditional collaborative filtering algorithm unavailable[1], since it lacks the ability of accumulating sufficient data in a short period of time, and overly relies on similarity calculation. Or no good results will be gained. In the paper, scheduling algorithm framework proposed is a kind of reinforcement learning[2], to maximize their cumulative returns by series of activities in uncertain circumstances.

In the model, a system self-evaluation mechanism is introduced that makes it possible to automatically select desirable algorithm from the algorithm pool according to users' feedback. In this way, we can achieve the purposes of self-optimization, which leads to a situation of fittest algorithm, so that recommended results gradually are refined. When the feedback is large enough, covering sufficient user types, algorithm diversity increasingly becomes rich. Some of them meet

the universal needs of users, while some satisfy only a small fraction of users' needs. Constantly replenished by the algorithm to explore the parameter space in theirs, the system will greatly appeal to its users and optimize with the expansion of its users' scale.

Data processing approach based on YARN platform provides a very good solution to parallelism problems during the process of big data. YARN is the upgraded generation of Hadoop computing framework, which not only supports the original programming model MapReduce, and supports a wide variety of distributed computing frameworks such as MPI, Spark, Tez and so on. In this paper, data aggregation process is illustrated by label propagation algorithm[3] used under the YARN platform to show the superiority of the data processing.

## 2. REAL-TIME RECOMMENDATION MODEL

In real-time system of optimization recommendation, users' browsing history and residence time as well as feedbacks of provided content will be timely collected by the server. On server side or cloud, initial data will be preprocessed, including data cleaning, noise reduction, and then data aggregation[4], the processing of miss data[5]. In this way, we would take a collection of available data. Scheduling algorithm framework first will be calculated based on recommendations provided by the optimization algorithm. But when there is feedbacks, scheduling framework will choose a suitable algorithm based on feedbacks to select an appropriate algorithm combined with personal weight-vector model in order to present results closer to individual users. The evaluation system will in turn determine appropriate algorithm based on the feedbacks. Such non-stop loop, we form a system with the ability of self-evolution. In this way, the personalized recommendation is more accurate. Meanwhile recommendation system involved in large amounts of data processing will be tackled by taking advantage of a new generation of big data computing framework that supports multiple computing paradigms. We have different choices on different applications. The Process above shown in Figure 1.

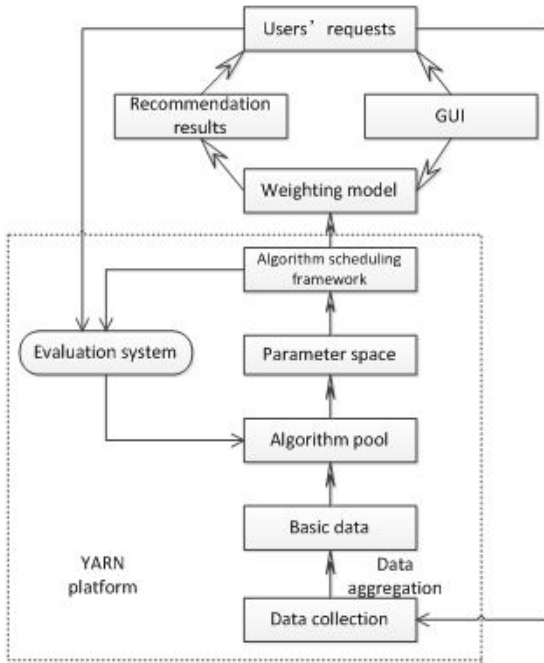


Figure 1: Recommendation System Model

Next, three major components of the system, respectively algorithm scheduling framework, evaluation system and data processing on YARN, are introduced.

### 3. SCHEDULING ALGORITHM FRAMEWORK

First, we have to admit environment is uncertain and information is insufficient, with the knowledge of highly dynamic real-time recommendation system. However, long-term interaction between users and recommendation system is to constantly optimize the process. In simple terms, if the system recommends an article to a user, the user does not click or does not like, and then we can conclude the recommendation system fails, but this process will be regarded as a small part of decision-making process with users' feedback as payoffs, where the dislike is treated as a negative return while like as a positive return. This framework is to solve the problem of how to get a greater return at less expense.

In classical model of gambling armed context [6], the enhance learning scheduling framework is adopted to solve the problem of return maximization effectively. On assumption that news recommendation system takes whether a piece of news is clicked as a return, its frame is defined as follows. A proceeds in discrete trials  $t = 1, 2, 3 \dots$ . In trial  $t$ :

1. The algorithm observes the current user  $u_t$  and a set  $A_t$  of arms or actions together with their feature vectors  $x_{t,a(a \in A_t)}$  that summarizes information of both the user  $u_t$  and arm  $a$ , and will be referred to as the context.

2. Based on observed payoffs in previous trials, A chooses an arm  $a_t \in A_t$ , and receives payoff  $r_{t,a_t}$  at whose expectation depends on both the user  $u_t$  and the arm  $a_t$ .
3. The algorithm then improves its arm-selection strategy with the new observation  $(x_{t,a_t}, a_t, r_{t,a_t})$ . It is important to emphasize here that no feedback (namely, the payoff  $r_{t,a}$ ) is observed for un-chosen arms  $a \neq a_t$ .

In the process above, the total T-trial payoff of A is defined as  $\sum_{i=1}^T r_{t,a_i}$ . Similarly, we define the optimal

expected T-trial payoff as  $E[\sum_{i=1}^T r_{t,a_i^*}]$ , which is the

arm with maximum expected payoff at trial  $t$ . Our goal is to design A so that the expected total payoff  $\sum_{i=1}^T r_{t,a_i}$  is maximized. Equivalently, we may find an algorithm so that its regret with respect to the optimal arm-selection strategy is minimized. Here, the T-trial regret of algorithm A is defined by

$$R_A(T) = E[\sum_{i=1}^T r_{t,a_i^*}] - E[\sum_{i=1}^T r_{t,a_i}] \quad (1)$$

In the context of article recommendation, we may view articles in a pool as arms. When a presented article is clicked, a payoff of 1 is incurred; otherwise, the payoff is 0. With this definition of payoff, the expected payoff of an article is precisely described as its click through rate (CTR), and choosing an article with maximum CTR is equivalent to maximizing the expected number of clicks from users, which in turn is the same as maximizing the total expected payoff.

Algorithm strategy is improved to minimize the difference between overall return and the optimal return when implementing the scheduling framework. Different demands of access speed and computation capacity for on-line rapid responses and training models should be compromised. For example, the feature vector calculation and iteration requires offline models, whereas comparing payoffs caused by different behaviors needs an online one.

In portal, because of its total large browse amounts, and a great number of new users who do not browse history records. it is difficult for traditional offline methods to mine the characteristics of a user so as to meet the needs of time-sensitive product, a condition called cold start [7]. In the scheduling algorithm, its essence is to balance exploration and exploitation. When the system running, it will gradually get



accustomed to new environment through digestion and absorption, so cold start problem in such a scheduling framework becomes a process of progressive improvement.

News Consulting, generally have strong social transmissibility, if the trend of news outbreak is able to be expected, which can be made the best use of, and then it would make a big difference to deliver the potential interesting news to the users and further affect subsequently news propagation. In the scheduling framework, we combine a variety of needs-adjusted return functions. For hot news, with good returns in the short term for all types of users, may be a potential hotspot.

#### 4. EVALUATION SYSTEM

In personalized recommendation systems, the establishment of data-evaluation system is the key. In this model, the self-test system can effectively evaluate the merits of the algorithm to achieve self-optimization functions. When the user receives feedbacks, scheduling algorithm framework will choose different evaluation systems, depending on the circumstances. Then the results will be submitted to the algorithm pool for algorithm selection. Figure 2 show that

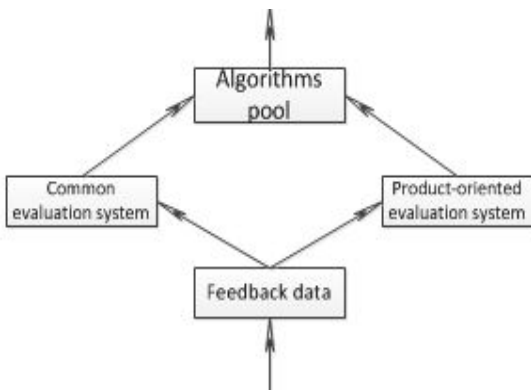


Figure2:Evaluation System

##### 4.1. Common evaluation system.

In common evaluation system, ROC(Receiver Operating Characteristic) curves is used to measure dynamic capabilities of model and to test the effect of recommendation algorithms[8]. The common confusion matrix used in model shows the distribution of predicted results, shown in Figure 3

|                    |    |                |                |  |  |
|--------------------|----|----------------|----------------|--|--|
|                    |    | Actual Value   |                |  |  |
|                    |    | p              | n              |  |  |
| Predictive Outcome | p1 | True Positive  | False Positive |  |  |
|                    | A1 | False Negative | True Negative  |  |  |
|                    |    | p              | n              |  |  |

Figure3:Confusion Matrix

Based on the difference between predicted and actual values, there are four possible results: true positive, false positive, false negative and true negative. Of those, true positive rate (TPR) and false positive rate (FPR) is an indicator concerned. ROC curve is drawn to show the relations difference TPR and FPR in different parameters. showed in Figure 4

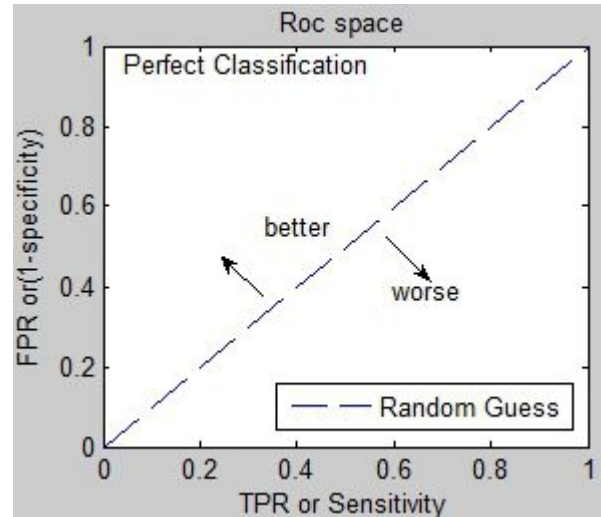


Figure4:ROC Curve

In the ROC curve, the vertical axis is TPR, characterizing the sensitivity of the model, and the horizontal axis is FPR, characterizing model's discrimination. Upper left corner of the graph features means model has perfect coverage and discrimination, and the diagonal line indicates a completely random classifier better than random classification model, the ROC curve in the upper left diagonal, while the area under the ROC curve (AUC) presents dynamic predictive ability of the model.

ROC curve can comprehensively reflect the ability to adapt to different model parameters and the external environment changes. In the personalized recommendation system, ROC curve ,as an evaluation indicator, can also be applied to optimization goals improved slightly. For example, score prediction, when prediction error rate is lower than a certain threshold, is classified correctly, otherwise considered wrong. Top-K recommendation problem is equivalent to direct classification of given data sets [9].

##### 4.2. product-oriented evaluation system.

The common evaluation above has two major problems when applied to specific product application. First, from a technical point of view, it ignores the difference between offline and online measurements, which may result in model over-fitting [10]. Secondly, from the product point of view, it fails to consider the core goal of recommendation algorithm, ie, creating long-term value for users because even the advancement, but only turns out to be a short-term optimization course. Based on the considerations mentioned above, a optimization framework of product-oriented personalized

recommendation system is proposed. Shown in Figure 5. The algorithm evaluation measurement compactly combined with products' core indicators optimizes the system to resolve problems in reality.

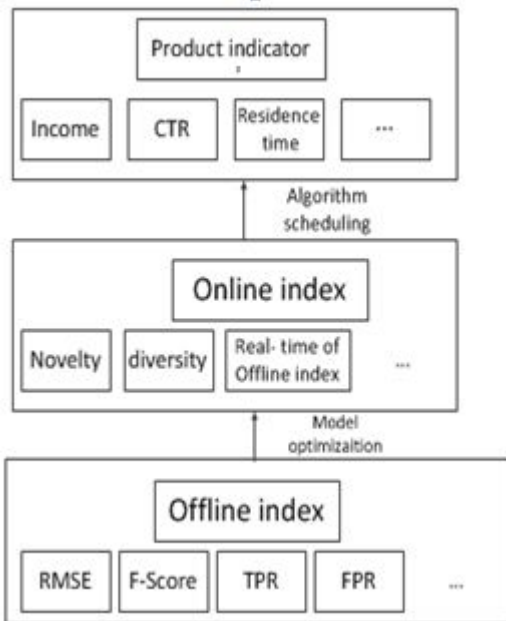


Figure5: Product-oriented Evaluation System

We want to solve the excessive reliance on offline criteria used to evaluate the merits of the algorithm. On the one hand, off-line indicators are required to be prompted in real time. By regular online calculation observe and evaluate the quality of recommendation algorithms, and make the appropriate adjustments; on the other hand, a reasonable gray on-line or A/B testing system is established to ensure the improvement of personalization algorithm. Whether an improvement or a new algorithm is to be used on line, depends on eligible performance in offline environment, and then under the online environment stability and preliminary effects tested by a weight, further formulating improvement approach applied to corresponding fields.

If there are several parallel improvements or trials are aimed at the same user or the same algorithm function, then try to conduct rigorous A/B test to identify the final decision by comparing effects on the randomly selected groups of test users under the premise of constant environment variables. In contrast, most previous work focused on offline assessment, Our work make a shift to online assessment to optimize the recommendation system, which also improves workflow and working patterns.

### 5. DATA AGGREGATION PROCESSING USING LPA BASED ON YARN PLATFORM

YARN is the next generation of computing framework Hadoop, a huge reconstruction of the first-generation MapReduce framework (MPv1). YARN in architecture and design is more flexible, more extensive than MRv1. The calculation framework shown as figure 6:

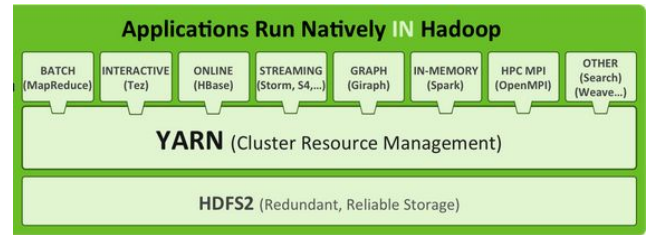


Figure 6 :Yarn Framework

In this section, we describe in detail how to use the LPA algorithm uses MapReduce calculation method to achieve data aggregation process data from a large collection data to basic data, the process is shown in Figure 7



Figure7: Data Aggregation Processing

#### 5.1. LPA algorithm.

Label propagation algorithm is a standard graph mining algorithm [12], in which the object is a undirected graph made of nodes and edges. LPA is widely used in polarity classification[13] and social network community discovery[14]. In this paper, to model mining keyword from news or articles, each node represents a user in the diagram, where each node's ID uniquely identify itself and nodes'edges represents relationships between users, such as the same area, similar preferences, etc. For label propagation algorithm, each node needs initially labeled "tag" information. In such applications, the label of each node is the user's browsing history. For a new user, the corresponding node is labeled null, which initializes the construction of graph for LPA.

According to the executing process of LPA, node labels automatically get updated during iterations to complete the label propagation. The rounds of iterations can be specified in advance or the process does not terminate until node labels change little. In each round of iteration, nodes take turns to perform the following steps to update labels.

1. Locate the neighbor nodes with edges linked to node A, and the neighboring nodes set S, representing users associated with the user A, is comprised of these neighbors and A itself;
2. Count the frequencies of different labels S. Node A's initial label with f as its frequency, that is, a user's own label is given a larger initial frequency ;
3. The weight of each A's label deduced from the second step then is calculated by the formulation like  $f \cdot IDf$ , and these labels ranking the highest K in the order of their weights form A's new label collection.

Through the process above, node A's label collection is updated.

## 5.2. Based YARN's MapReduce parallel data processing

Considering the large amount of users on portal websites, and the parallel feature of certain calculation, we use distributed computing framework MapReduce. The original data for the subsequent calculation derives from two sources, namely neighboring data associated with the user in the form  $\langle uid, list(uid) \rangle$  to present a list of other users associated with uid, and initial tag information in the form  $\langle uid, list(word) \rangle$  to represent the initial labels relevant to the uid. In order to facilitate the subsequent process, firstly, these two types of data are combined in a format  $\langle uid, list(uid), list(word) \rangle$ . pseudo-code for this process is following:

```
map(string key , string value): //key: uid ; value:
list(uid)+list(word)
/* parse value*/
list uidList=parse(value)
list wordList=parse(value)
/*PLA*/
for each u in uidList :
EmitIntermediate(u,AsString(wordList));
If u==key:// initial label's frequency of each node
int i=0
While i<f;
EmitIntermediate(u,AsString(wordlist));
i=i+1
reduce(string key,Iterator values) : //key : uid
values : list(word)列表
/*count label's frequency*/
map<string ,float> tfCounter;
for each v in values:
list wordList=parse(v)
for each w in wordList:
tfCounter[w]+=1
/*count each label's weight*/
For each lable in tfCounter.keys():
float weight=Tfddf(lable)
tfCounter[lable]=weight
/*take Top K*/
List topLabelList=GetTopK(tfCounter,K)
Emit(AsString(topLabelList));
```

LPA algorithm processing data completes the update process of a set of user tags. The user can specify the number of iterations, the calculation process can be completed. Finally obtain an ID and a keyword.

## 6. CONCLUSION

Real-time optimization system based on YARN platform can be effectively self-optimized and conduct data parallel processing, to achieve well-performing effects in recommending the personalized real-time news to users. This model still has some room for improvement. Future researches focus on the following aspects:

In the context of algorithm scheduling framework, algorithm optimization processing from the algorithm pool should give consideration to both personalized

recommendation effects and accepted efficiency. Sometimes the system has to be designed to strike a balance between the two. From the perspective of computational complexity and universality, in-depth analysis and improvement of the performance of algorithms is needed. But also the independence requirements of different algorithms within its parameter space should be guaranteed performing data processing. At the same time to get the best results, the results of each independent parameter space can be self-exchanged.

Taking better methods to solve data collection preprocessing work is expected. For Problem with large basic data, deficient useful data, incomplete data information, a better solution helps provide a better structure, more convenient treatment, more qualitative basic data. This is conducive to the efficiency and accuracy of subsequent system processing.

More work should be done to improve the robustness and accuracy of the self-regulating system with feedback. It's essential to promptly and effectively respond to users' request based on the previous feedback. In despite of there being gradual improvement during the feedback process, how to ensure a certain degree of accuracy at the outset will influence the performance of the whole system.

## ACKNOWLEDGEMENTS

This work was funded by National College Students Innovation and Entrepreneurship Training Program .The program ID is 20131049710003.The authors also wish to thank the instructors for several helpful comments and suggestions.

## REFERENCES

- Sarwar, B. Karypis, G., Konstan, J., & Riedl, J. 2001,Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web, pp. 285-295. ACM.
- Barto, A. G. ,1998. Reinforcement learning: An introduction. MIT press.
- Zhu, X. & Ghahramani, Z. 2002. Learning from labeled and unlabeled data with label propagation, Technical Report CMU-CALD-02-107, Carnegie Mellon University.
- Han J, Kamber M, Pei J. 2006. Data mining: concepts and techniques,Third Edition.Singapore Elsevier Inc.
- Miller, B. N., Konstan, J. A., & Riedl, J. 2004. PocketLens: Toward a personal recommender system. ACM Transactions on Information Systems (TOIS), 22(3), pp.437-476.
- Lihong Li,Wei Chu,John Langford,Robert E.Schapiro, 2010. A Contextual-Bandit Approach to Personalized News Article Recommendation ,World Wide Web Conference 2010 .
- S.-T. Park,D. Pennock, O.Madani,N.Good ,and D.DECoste, 2006. Naive filterbots for robust cold-

- start recommendations .In Proc.of the 12th ACM SIGKDD International Conf .on Knowledge Discovery and Data Mining ,pp. 699-705.
- Hanley, J. A., & McNeil, B. J. 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases.Radiology, 148(3), 839-843.
- Babcock B, Olston C.2003. Distributed top-k monitoring[C],Proceedings of the 2003 ACM SIGMOD international conference on Management of data. ACM, pp. 28-39.
- Faber, N. M., & Rajko, R , 2007. How to avoid overfitting in multivariate calibration—The conventional validation approach and an alternative.Analytica Chimica Acta, 595(1), pp.98-106.
- Arun Murthy.2012,Introducing Apache Hadoop YARN. <http://zh.hortonworks.com/blog/introducing-apache-hadoop-yarn/>. [accessed 10 March 2014][12]
- Inokuchi, A., Washio, T., & Motoda, H. 2000. An apriori-based algorithm for mining frequent substructures from graph data. In Principles of Data Mining and Knowledge Discovery pp.13-23. Springer Berlin Heidelberg.
- Speriosu, M., Sudan, N., Upadhyay, S. & Baldridge, J. 2011. Twitter polarity classification with label propagation over lexical links and the follower graph. In Proceedings of the First workshop on Unsupervised Learning in NLP .pp53-63. Association for Computational Linguistics.
- Ruan, J., & Zhang, W. 2007. An efficient spectral algorithm for network community discovery and its applications to biological and social networks. In Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on . pp. 643-648. IEEE

# NON-HLA DISTRIBUTED SIMULATION INFRASTRUCTURE

Jan Voracek <sup>(a)</sup>, Jiri Penzes <sup>(b)</sup>, Antonin Kavicka <sup>(c)</sup>

<sup>(a)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice

<sup>(b)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice

<sup>(c)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice

<sup>(a)</sup> [jan@voracek.net](mailto:jan@voracek.net), <sup>(b)</sup> [jirkapenzes@gmail.com](mailto:jirkapenzes@gmail.com), <sup>(c)</sup> [antonin.kavicka@upce.cz](mailto:antonin.kavicka@upce.cz)

## ABSTRACT

Distributed Simulation Infrastructure (DISIS) represents a general layer specialized in communication of distributed simulation engines. The layer supports a multiplatform approach, facilitates the synchronization and communication of simulators (involved in a distributed simulation model) and it is easy to use. The current version of DISIS supports discrete event simulation with conservative synchronization method. Applications of DISIS seem to be quite convenient for industrial applications because of high degree of flexibility.

Keywords: distributed simulation, Non-HLA communication, discrete-event simulation

## 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 a number of benefits, e.g., the simulation calculation can be easily scaled or, in some cases, more cost-effective resources can be used, etc. The major disadvantage of distributed simulation is the requirement to keep causality which means that the communication between simulators needs to be implemented. The communication becomes nontrivial when the logical processes in the simulation model are implemented in different programming languages / platforms. Each such platform usually supports different high-level communication protocols (like Java RMI, Windows Communication Foundation, etc.) and interconnecting logical processes can become very difficult to implement when a low-level communication (like TCP) needs to be used. In addition, it is also necessary to agree on a synchronization algorithm to preserve time causality of interconnected logical processes. The main goal of DISIS is to minimize these problems. (Fujimoto 2000, Topper 2002)

## 2. HLA AND DISIS

The proposed general DISIS layer can remind us of High-Level Architecture (HLA). Standard HLA, which

was originally developed for military purposes (Kuhl and Dahmann 2000), allows the interconnection of several simulations into a larger simulation unit. It is indeed a very powerful and complex tool. The biggest asset of HLA lies in its universality, which is compensated by strict adherence to standards. The main idea of HLA is a federation – a system that encapsulates all simulators. In the federation, there are described, inter alia, all objects that participate in simulation, or defined methods of data exchange. Every member of the federation – a federate – has to be designed according to Simulation Object Model (SOM), which describes the functionality of the member (Kuhl and Dahmann 2000). Federate also has to be able to receive information from other members, manage its local virtual time and time of other members of federation and also take responsibility for data ownership transfers. This is just a short list of rules of HLA architecture requirements for the federation and its members. The federation has to contain Runtime Infrastructure (RTI) backend layer, which must lead in providing of services needed for running of distributed simulation to all its members. The comparison of HLA and DISIS terminology is in the following table (Kuhl and Dahmann 2000).

Table 1: Terminology of HLA and DISIS

| HLA                     | DISIS                        |
|-------------------------|------------------------------|
| Federation              | Distributed simulation model |
| Federate                | Simulator                    |
| RTI                     | DISIS infrastructure         |
| Federation Object Model | Not required                 |
| Object Model Template   | Not required                 |
| Simulation Object Model | API                          |

Compared to HLA, the main idea of DISIS is not to support various combinations of discrete and continuous simulation, different synchronization algorithms etc. DISIS is in this respect way more restrictive, and therefore much easier for implementation. Because DISIS is not so strictly tied to the simulation engine, the software designer of a logical



process has much more freedom with regard to the implementation. There is no need to implement some standard or adjust the implementation of a simulation engine to a specific communication layer. That enables to completely replace the communication layer with another. Regarding to their implementation, simulators are not directly dependent on the DISIS layer. Among other things, this feature allows us an easy switch from the monolithic engine to the distributed one with a minimal intervention in the implementation. For example the complexity of a simulated problem may not be well-known in industrial applications in advance. Thus, with DISIS we can start with monolithic simulation and when we need a higher performance or a higher degree of decomposition, we can easily change the standing model to the distributed one. DISIS only provides basic API for message exchange and takes full responsibility for communication. There is no need to design the very simulator according to some standard (see Implementation) (McLeod Institute of Simulation Sciences 2001).

### 2.1. Other alternatives

Beside HLA there are also other solutions that solve the problem of communication. For example the DDS (Data Distribution Service) or DIS (Distributed Interactive Simulation). Like HLA, they are both very complex. The main disadvantage may be that the serialization is done by the infrastructure, not the simulator. Therefore, the user has to describe all objects exchanged in the simulation by the layer specification. Consequently, such implementation becomes quite difficult (Joshi and Castellote 2006).

### 3. DISIS

DISIS is implemented as a network of services that are exchanging messages. Simulators of distributed simulation are supposed to be connected to the mentioned services. Each simulation engine communicates only with DISIS service through which it sends messages to other logical processes (Figure 1). This communication runs via given API (Application Programming Interface). The network infrastructure is hidden from the logical processes. Thus, the logical process has no information about the physical location of other processes which it communicates with. When all DISIS services are ready (initialized and paired), each logical process has to obtain a command in order to start the simulation.

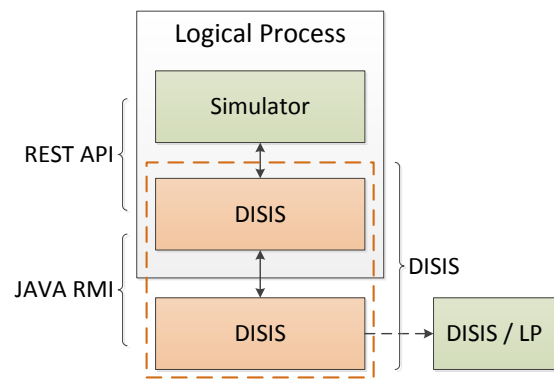


Figure 1: Simulator and DISIS

Every simulator has to communicate with other simulator only through the DISIS service. From the perspective of infrastructure architecture is DISIS a decentralized distributed system that can run on more computing stations. This fact brings us benefits for instance in situation where we have logical processes at geographically distant places (in each location may be one DISIS service) or in the case where we want to control or optimize the message flow.

One case of the use of DISIS infrastructure is a situation where each simulator has its own DISIS service through which it communicates with others (Figure 2).

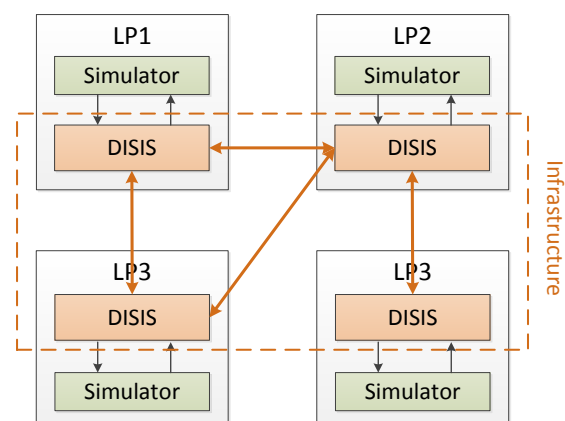


Figure 2: DISIS infrastructure

Figure 2 implies two kinds of messages exchanged in the DISIS network: (i) simulators communicate with DISIS services and (ii) the DISIS services communicate with each other. The simulator has no information about internal messages and cannot interfere with their flow. Internal messages are used only to transfer information between DISIS services. The message contains this set of information:

- sender (DISIS service),
- recipient (DISIS service),
- message timestamp,
- data.



Attribute data usually contains a message sent by logical process during simulation. However DISIS services can exchange messages that are not directly related to the simulation – such messages are called service messages. Amount of service messages in the overall communication is minimal due to optimization. Service messages are sent especially before running the simulation itself because of the initial initialization and exchange information on the locations of all simulators connected to DISIS network.

The DISIS layer brings even more advantages compared to direct interconnection of logical processes. Thanks to the fact that all communications are made via DISIS, it is possible to control the flow of exchanged messages. There is a potential to optimize the communication for maximum efficiency in the large networks (similarly as in computer networks). An example might be two simulators where we know that they will have higher communication overhead during the simulation process. Then it is better if they share one DISIS service (Figure 3).

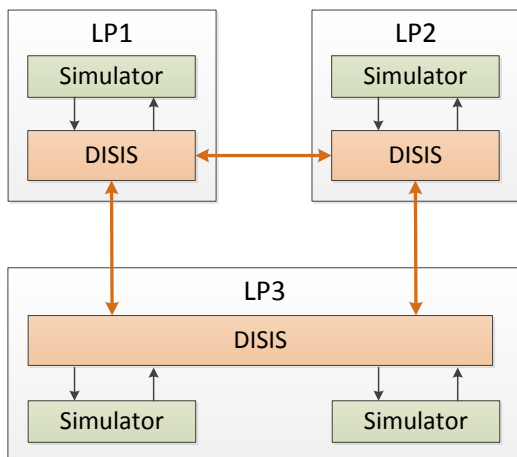


Figure 3: Two simulators sharing one DISIS service

In case of simulation model where we don't expect too high interaction between logical processes or if they are all on local network, we can use a centralized architecture with only one DISIS service (Figure 4).

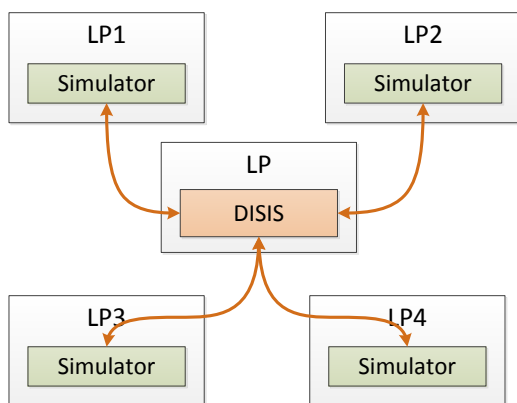


Figure 4: Centralized DISIS network

The DISIS infrastructure architecture is designed by using configuration files. Each DISIS service has its own configuration. It is necessary to have the configuration ready before running the infrastructure. The configuration contains basic information about the configured DISIS service and information about remote DISIS service it will directly communicate with.

The start of DISIS is a very important part. Simulators cannot connect to the DISIS network before all DISIS services are ready. The ready-state in our terminology means the moment when all DISIS services are connected and they established communication with their immediate surroundings. Until this moment, the network is not available for simulators and the simulation cannot be started. The start-up process is realized by these steps performed by each DISIS service:

1. load configuration,
2. initialize local DISIS service,
3. search for remote DISIS services,
4. connect to remote DISIS services,
5. send ready message,
6. wait for simulators.

The connection to a remote DISIS service is successful when the ready message is received. If there are all DISIS services ready, the simulators can connect.

Let's look closer to the realization of distributed simulation using DISIS. Consider the basic simulation model, which consists of two simulators that are supposed to communicate with each other and so create a distributed simulation model (Figure 4).

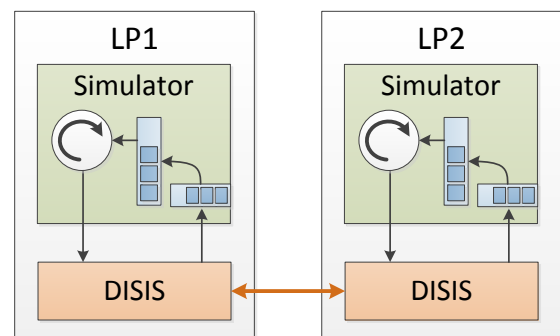


Figure 5: Two simulators

Each simulator communicates through its own DISIS service. DISIS supports only discrete simulation with conservative synchronization method and the actual simulation works on the principle of simulating a life cycle of an entity. In other words, except the timestamp, simulators exchange also an entity and information about what will happen with the entity. In DISIS there is used an algorithm of sending null-messages on request to guarantee the causality. The responsibility of sending null-message requests is taken by DISIS. The simulator doesn't need to worry about

sending null-message requests. The simulator works in a standard way – it contains a local event calendar and input queue for every remote simulator it wants communicate to (Figure 5). All messages from simulators go through DISIS, and therefore it is possible to detect the right moment when to send a null-message request (Figure 6). There is one condition for the simulator in order to have everything working all right. It has to inform the DISIS about every change of its local virtual time. If this condition is met, the DISIS knows the exact moment when to send the null-message request.

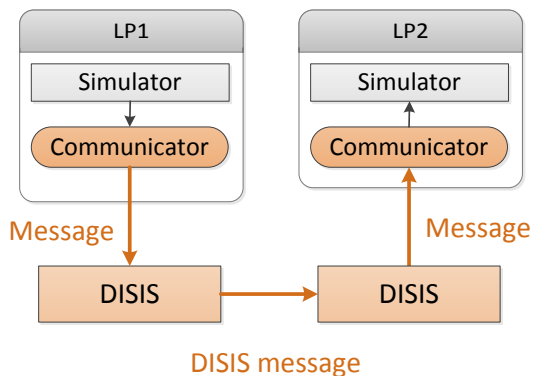


Figure 6: Message transferred from LP1 to LP2

The DISIS infrastructure allows to connect an “observer module”. This module can be connected to a DISIS like standard simulator with the exception that it only consumes data (Figure 7).

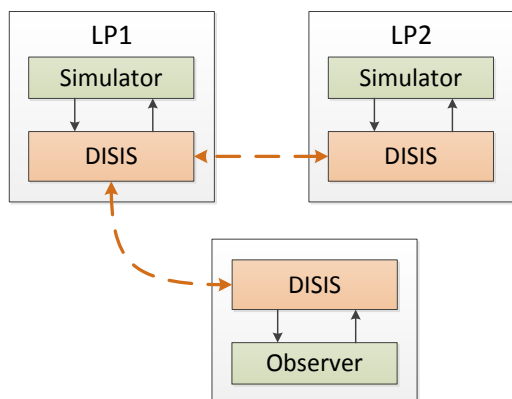


Figure 7: Connected analytic module

This module can be located anywhere. It will be receiving messages from DISIS so it can for example do some analysis of current state of simulation or act like an animation engine.

#### 4. IMPLEMENTATION

The main pillar of DISIS is DISIS-API (Distributed Simulation Infrastructure – Application Programming Interface), which has every DISIS service and through which the simulator communicates with the service. This interface declares four functions:

- connect(RestSimulatorInfo),
- send-message(RestMessage),
- register-observer(ObserverInfo, Observables[]),
- update-simulation-timestamp(RestSimulatorInfo, double).

The connect function has to be called by the simulator as first, because it’s used to connect to the DISIS network. Until this function is called, other functions are unavailable. The simulator sends basic information about itself:

- title (name in human-readable form),
- remote-name (unique network name),
- description (short description, not required),
- end-point-address (network address of the simulator),
- surrounding-simulators (simulators with whom it wants to communicate).

The names of surrounding simulators are necessary for monitoring incoming messages. Since we know the timestamps of all messages from remote simulators and the local virtual time of given simulator in DISIS service, we can detect a situation when it is necessary to send null-message request. Thus, the user is partly shielded from the synchronization algorithm.

As its name implies, the function send-message is used to send messages. The message has to contain the following information:

- from (sender – simulator),
- to (recipient – simulator),
- timestamp,
- message (message content).

The function register-observer mediates the connection of observer module. Its parameters are information about the module (like in the case of connect function) and list of message types that it wants to observe.

The last operation update-simulation-timestamp is used to update local virtual time. It is necessary that every simulator calls this function after every change of its local virtual time.

The communication between simulator and DISIS service is implemented using REST-API, since REST is available across many platforms. The disadvantage of this approach is the HTTP on background. Because it is request-response protocol, the HTTP server has to run on both sides – DISIS service and simulator. The simulator has its own REST-API, which supports these four operations:

- start-simulation,
- stop-simulation,
- process-message,
- null-message-request.

The *start-simulation* function is called at the start of simulation calculation – when all simulators are properly connected to DISIS. On the other hand, the second operation *stop-simulation* is used for an explicit stopping of the simulation (such as instruction from some other simulator).

The function *process-message* is intended for receiving messages. DISIS uses this function for passing messages to simulator.

The last function *null-message-request* is used for sending null-message request.

The previous text describes the establishing communication with DISIS and following messaging, which is done through DISIS-API (Figure 8). It should be recalled that the communication between DISIS and simulator is available only when the entire DISIS infrastructure is ready.

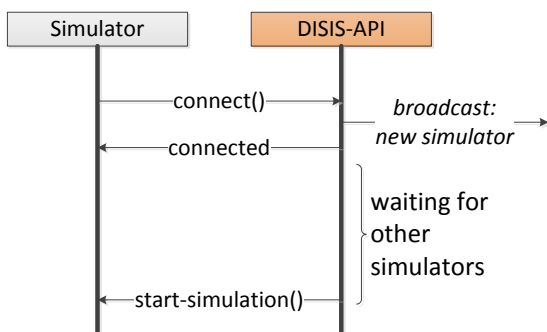


Figure 8: Connection establishing diagram

For the functioning of simulator in DISIS network it is necessary that the simulator meets following requirements:

- It has to send every change of local virtual time to the DISIS service.
- It has to implement REST-API for receiving messages from DISIS service.

DISIS controls the initial impulse that starts the simulation calculation at individual simulators. After running the entire DISIS infrastructure all DISIS services wait for connection of simulators. The DISIS service has no information where the simulators (doesn't know their network location) or what simulators will be connected to it. Information about the connection of each simulator is distributed to all DISIS services. Each service maintains information about which DISIS service is each simulator connected to and the path to given service in the infrastructure (Figure 9). This register is created before starting the simulation.



Figure 9: Indirect communication between two simulators

Once all simulators, which given simulator needs for its running, are connected, the DISIS sends its command to start the simulation.

The simulator starts to execute scheduled events and DISIS service updates its information about the simulator's local virtual time and sends null-message requests.

DISIS uses Java RMI (Remote Method Invocation) for internal communication and it is necessary to transform every incoming message from simulator to DISIS message. Thus, DISIS receives message from simulator, transforms it to internal message and looks for service which the target simulator is connected to. The target DISIS service is found in the register mentioned above. Then it sends the message to the target service. The target service transforms the message back (extracts the encapsulated message for simulator) and sends it to the target simulator.

## 5. CASE STUDY

This chapter will target on a case study, which will reflect a simplified traffic system of a limited scope. There will be demonstrated a process of construction of a simple generic distributed simulation model, that means inputs or outputs of simulation experiments are not closely monitored.

### 5.1. Crossroad simulation

Let's have a restricted transport segment – crossroad, to which traffic flows from four directions (Figure 10).

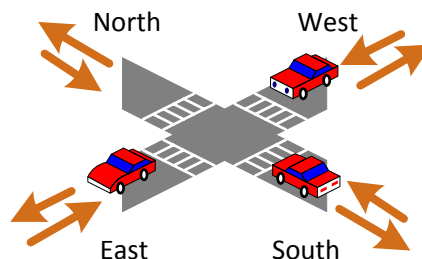


Figure 10: Considered transport segment

The intended case study assumes four traffic input generators and a model of crossroad. One of the possibilities to design an appropriate distributed simulation model is its division into three separated logical parts:

- crossroad,
- traffic – north and south,
- traffic – east and west.

Let's have two simulators that generate, or receive flow of vehicles (in two directions) and a simulator that reflects a model of crossroad. There are several possibilities how to design the communication infrastructure. In this case we used two DISIS services

(Figure 11). One service only handles the crossroad. The second one handles the traffic.

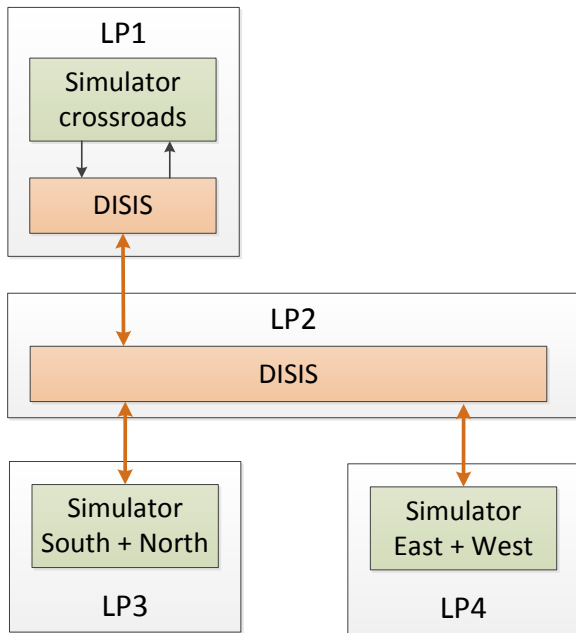


Figure 11: DISIS infrastructure

During the configuration it is necessary to configure properly all simulators and DISIS services. In the intended simulation model we try to find out if it was better to use a crossroad with traffic lights or a traffic roundabout. Replacing the simulator of crossroad with roundabout simulator is very easy. It is possible to use the *observer module* for mining data for analysis, or we can collect the data directly from the simulators.

The described generic simulation model was run in a laboratory of the University of Pardubice – Faculty of Electrical Engineering and Informatics. The computing nodes (on which the logical processes were allocated) were realized by three normal desktop personal computers connected to computer network.

## 6. CONSLUSION

The paper introduces the problems of communication between simulators across different platforms – especially finding a suitable solution with minimal impact to the original simulator in terms of implementation. The main concept of the HLA standard, which solves this problem to some extent, was also introduced.

Distributed simulation has a number of benefits, e.g., the simulation calculation can be easily scaled or, in some cases, more cost-effective resources can be used. There is a problem if we want to implement the distributed simulation in different programming languages / platforms. The HLA standard could be a possible solution of this problem. However, this standard is very complex and quite difficult to implement. The complexity may not be an advantage when we already have implemented simulators and only

want to integrate them to a distributed environment (Fujimoto 2000).

The main topic of the paper is our proposed solution DISIS (Distributed Simulation Infrastructure). The basic concept, main advantages and disadvantages are introduced to the reader.

The issue of different platform is solved by universal API, through which the simulator can communicate with DISIS. All communication between simulators is done through the DISIS infrastructure. This can be used in the simulation process itself – for example to collect data for analysis.

For implementation of DISIS, the Java platform was chosen and the communication interface is designed as REST-API with only small number of necessary operations that are required to run the simulation. It is easy to communicate on almost all commonly used platforms through REST-API. The DISIS requires very little modification of already implemented simulator. Therefore, the integration of DISIS to current solution is relatively simple and fast.

## REFERENCES

- Fujimoto, R.M, 2000, *Parallel and distributed simulation systems*. New York Wiley-Interscience.
- Kuhl, F., Dahmann, J., Weatherly, R., 2000, *Creating computer simulation systems : an introduction to the high level architecture*. Upper Saddle River, NJ Prentice Hall PTR.
- Topper, C., 2002, *Parallel and distributed discrete event simulation*. New York Nova Science.
- McLeod Institute of Simulation Sciences, 2001. *HLA Module 1. Basic Concepts of the High Level Architecture (HLA)*. California State University HLA Courses. Available from: <http://www.ecst.csuchico.edu/~hla/courses.htm> [accessed 15 July 2014].
- Joshi, R., Castellote, G., 2006, *A Comparison and Mapping of Data Distribution Service and High-Level Architecture*. Santa Clara Real-Time Innovations, Inc.

# PREDICTING EQUIPMENT AVAILABILITY USING A HIGH LEVEL ARCHITECTURE FRAMEWORK

Estacio Siemann Santos Pereira<sup>(a)</sup>, Yasser Mohamed<sup>(b)</sup>, Simaan AbouRizk<sup>(c)</sup>

<sup>(a)</sup>University of Alberta

<sup>(b)</sup>University of Alberta

<sup>(c)</sup>University of Alberta

<sup>(a)</sup>[estacio@ualberta.ca](mailto:estacio@ualberta.ca), <sup>(b)</sup>[yasser.mohamed@ualberta.ca](mailto:yasser.mohamed@ualberta.ca), <sup>(c)</sup>[abourizk@ualberta.ca](mailto:abourizk@ualberta.ca)

## ABSTRACT

Breakdown and maintenance can affect equipment availability, generating delays and increasing the cost of a project. The purpose of this paper is to develop a simulation model to predict equipment availability in earthmoving operations using a High Level Architecture framework. Based on user information, the model can verify the influence of the temperature, quantity of crews, location of breakdown and overheating of the tires on the equipment availability. The user can also define different distributions for maintenance and breakdown intervals, according to the equipment type and model. Different scenarios were defined to verify the model, and the results demonstrated that it can be useful to help inexperienced engineers to verify the influence of several parameters on the equipment availability and provide accurate information to the decision maker.

Keywords: Equipment Availability, High Level Architecture, Earthmoving

## 1. INTRODUCTION

Earthmoving operations, which are used in projects such as dams, highways and mines, utilize heavy equipment to increase productivity and decrease costs. Earthmoving operations typically include: excavating, loading, hauling, unloading, compacting and grading.

Inter-dependent activities require continuous work of the equipment during the entire project, without interruption, for better progress, productivity, and profits (Lee 2010). The equipment should be designed to ensure its successful operation though the anticipated service life, but deterioration starts as soon as the project begins (Muchiri et al. 2013). A major hindrance to achieving optimum production rates in earthmoving operations is equipment unavailability caused by maintenance and breakdowns.

There is a conflict of interest between the productivity rate and the maintenance and breakdown of equipment. Managers usually use the optimum use of equipment to create the project schedule. However, this approach does not consider the maintenance time. Lack of maintenance can increase the quantity of breakdowns. Companies should be able to develop maintenance management strategies to decrease the

costs associated with unexpected problems with the equipment, define the optimum quantity of crews to fix broken equipment and define strategies to optimize the maintenance and breakdown intervals (Mohideen and Ramachandran 2014). An optimum scenario between maintenance, breakdown and equipment availability should be achieved in the project to avoid delays.

However, the equipment availability in these operations is generally affected by conditions that may give rise to uncertainty. Contractors should be able to predict these uncertainties to estimate the time and cost of projects, monitor the equipment availability, and make adjustments in response to the actual production performance to achieve a successful project (Mohideen et al. 2011).

According to AbouRizk (2010), when a problem is characterized by uncertainties, “simulation is the most suitable analytical tool to model and analyze the problems at hand.” In earthmoving operations, simulation models can be used to predict equipment availability, developing scenarios that help the decision maker improve the process (Moselhi and Alshibani 2010). However, due to the complexity of the operation, High Level Architecture (HLA) should be used to increase the accuracy of the simulation.

The purpose of this paper is to propose a simulation tool to allow testing different scenarios to better predict equipment availability in an earthmoving operation using an HLA framework.

## 2. EQUIPMENT MAINTENANCE AND BREAKDOWN

Maintenance is defined as a combination of technical and administrative activities required to keep equipment operating in its desired condition (BSI 1984). The main purpose of maintenance is to restore or improve equipment condition to keep it productive, in safe condition, and prolong its working life (Edwards and Holt 2009). According to Muchiri et al. (2013), there are three basic types of maintenance: failure-based maintenance—if the maintenance occurs just when the equipment fails; time-based maintenance—or periodic maintenance (PM), which is carried out at specified time intervals, but, failure can still occur between the intervals of PM; condition-based maintenance (CBM)—done when the PM is carried out whenever a given



system parameter or condition reaches a predetermined value.

If well-established maintenance management is achieved, the company will have higher accuracy in predicting equipment breakdown, avoiding extra costs and being able to produce more reliable project schedules (Mohideen and Ramachandran 2014). Muchiri et al. (2013) concluded that different maintenance policies can lead to different equipment performances and that CBM policies provide better results when the interval time between the PMs is small. According to Edwards et al. (2000), the maintenance costs of an excavator in an opencast mine can represent up to 40% of production costs. However, lack of maintenance can result in project delays, loss of client goodwill, and reduced profit margins (Edwards and Holt 2009; Mohideen et al. 2011).

Another aspect to consider during maintenance is the crews' ability to take efficient action to fix the equipment. According to Muchiri et al. (2013), there are two extreme assumptions on the post-maintenance state of equipment: as good as new (AGAN) or as bad as old (ABAO). Periodic maintenance can reduce the equipment interval failure, but may not leave the equipment as good as new. Also, the equipment type and age affect the cost and time interval between the PMs. Competent plant operators can decrease the maintenance costs, because they can improve health and safety performance on the site and protect the equipment life-time by following the procedures recommended by the manufacturers (Edwards and Holt 2009).

Despite all the advances in maintenance management, equipment breakdown is inevitable. Breakdown can be defined as "the deterioration beyond the threshold level, decreasing the equipment performance in a critical level or loss of function of system performance" (Muchiri et al. 2013). The breakdown of equipment occurs due to the unpredictable failure of components and due to gradual wear and tear of the parts (Mohideen and Ramachandran 2014). Planners and estimators should account for potential lost time during the earthmoving operation, according to Harris and Olomolaiye (1993). Bernold (1989) mentions that breakdown is a well-known phenomenon and the random occurrence of such an event and duration to repair the equipment makes simulation a very efficient tool to analyze its impact. Some examples of random occurrences that affect breakdown are: equipment failure, tire problems (Mohideen et al. 2011), unexpected site conditions (Marzouk and Moselhi 2004), design limits, operational errors (Muchiri et al. 2013), lack (or quality) of maintenance (Snaddon 1988), type of truck, and hours of work (Wakefield and Sears 1997). Mohideen et al. (2011) identify the main causes for earthmoving equipment breakdown as mechanical failure, hydraulic failure and tire burst.

Zakeri et al. (1996) ranked equipment breakdown as the third most common cause of loss of productivity

in construction. According to Nepal and Park (2004), breakdown represents an average of 6% of planned working time for equipment (specific percentages: crushers (19.3%), motor graders (9.9%), and trucks (8.6%)). Breakdown can affect not only the cost and production, but also the safety, behavior of the workers and team environment (Mohideen et al. 2011).

Some researchers have tried to establish connections between the factors that cause breakdown. Harris and Olomolaiye (1993) developed charts that relate the age of the equipment with the probability of breakdown; Elazouni and Basha (1996) correlated the usage and downtime of a piece of equipment due to breakdown; Wakefield and Sears (1997) developed a theoretical simulation model where the probability of breakdown is related to the equipment model.

Based on the literature review, an influence diagram was built to verify the main factors that can contribute to equipment breakdown (Figure 1). The four main factors that influence breakdown are maintenance, work hours, operational errors and equipment type. Although the main factors that contribute to breakdowns are known, there is need for a more detailed observation of the relationships between them. Due to this, some assumptions are made when forecasting the equipment breakdown because of the difficulties to measure factors such as operator skill and the influence of project characteristics on the equipment. The main interest in modeling breakdown occurrences is to analyze their effects on equipment availability and performance. In this research, we assumed that equipment deterioration and failure are factors of equipment utilization, and thus related to operation time.

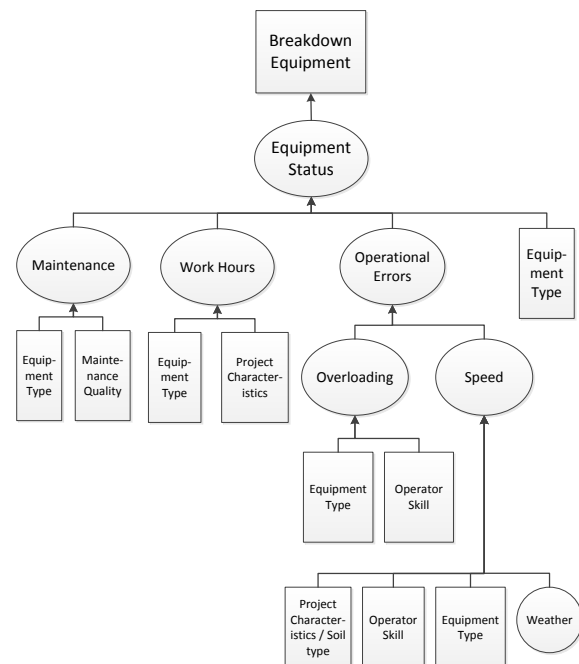


Figure 1: Influence Diagram Equipment Breakdown



### 3. METHODOLOGY

As there are many factors that can affect earthmoving operations, High Level Architecture (HLA) was used to simulate the maintenance and breakdown of equipment. High Level Architecture supports building complex virtual environments (called federations) using distributed computer simulation systems, to create a collaborative research project. In addition, it provides standards (composed by three main components: Rules (IEEE 1516), interface specifications (IEEE 1516.1) and the Object Model Template (IEEE 1516.2), that allow different developers to build their individual components (called federates) and exchange information between them (AbouRizk et al. 2009).

As the HLA is not software, the use of a run-time infrastructure (RTI) software is required to support the operations of a federation execution. The RTI provides services such as synchronization, communication, and data exchange between federates to support an HLA-compliant simulation. For this research, COSYE (AbouRizk et al. 2008) was utilized.

The breakdown and maintenance equipment federate (BMEF) is one of the current federates in this collaborative environment (federation) that receives and sends information through COSYE. To receive updates of an attribute, the federate should subscribe for this attribute, and to send an update, the federate should be able to publish to this attribute. In addition, another four federates also participate in the federation (Controller, Weather, Mover and Dumper). Figure 2 shows a schema of the federates' interaction with the federation. The simulation time is controlled by the RTI and the time step should be the same for all the federates.

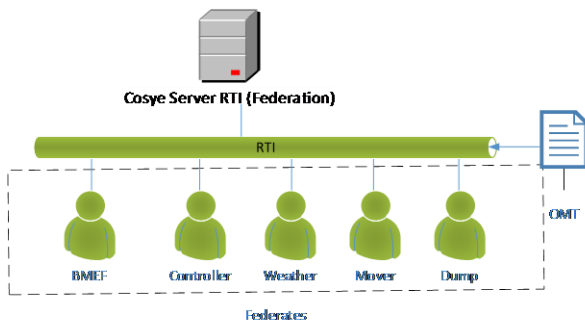


Figure 2: COSYE with Modeling Federates

The Object Model Template (OMT) contains the Federation Object Model (FOM), which contains all the objects, attributes and interactions for the federation, and the Simulation Object Model (SOM), which describes the object, attributes and interactions that a federate will use. If a federate wants to publish or subscribe to any attributes, it should be in the FOM and SOM. Figure 3 presents a conceptual model for the whole federation.

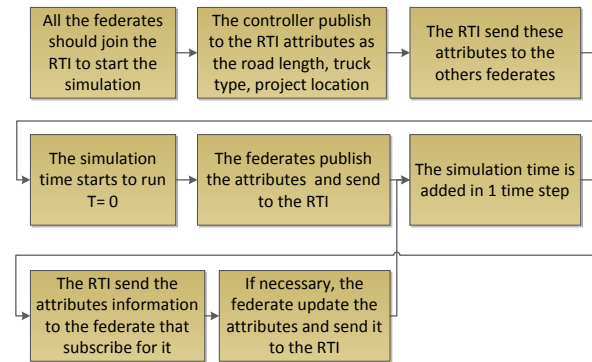


Figure 3: RTI Conceptual Model Process

At the beginning, all federates should join the RTI. The controller federate publishes the project location and start date, the road segments, the quantity of trucks and their characteristics, and the quantity of excavators and their characteristics. The other federates receive the information through the RTI and the controller federate allows the time to advance to start the simulation. In each time step, the RTI will send the attributes that were published from each federate to the federate that subscribes to that specific attribute. Also, the federates can publish new attribute values and send them to the RTI.

Each federate's function and interaction with the BMEF is described below:

- **Controller Federate:** The Controller Federate is an interface that regulates the simulation process and allows the user to update the project characteristics. The BMEF subscribes to the following attributes from this federate: truck model, quantity of each truck model, truck weight when empty, tire's Tons Kilometers Per Hour (TKPH) limit, excavator model and quantity of excavators. The Controller Federate subscribes to three attributes that the BMEF publishes: available equipment time, unavailable equipment time, equipment status and number of times that a truck reached the TKPH limit. With these attributes, the controller federate outputs graphs to the user that allow for easy understanding of the equipment availability. Figure 4 shows the interface where the user can update the project attributes and the truck availability outputs.

The Controller Federate is also responsible for providing a graphic visualization of the truck location during the simulation run. If the truck is working, it is represented by the color green. If the truck needs service, the truck is represented by the color yellow and if it is down, it is represented by the color red (Figure 5).

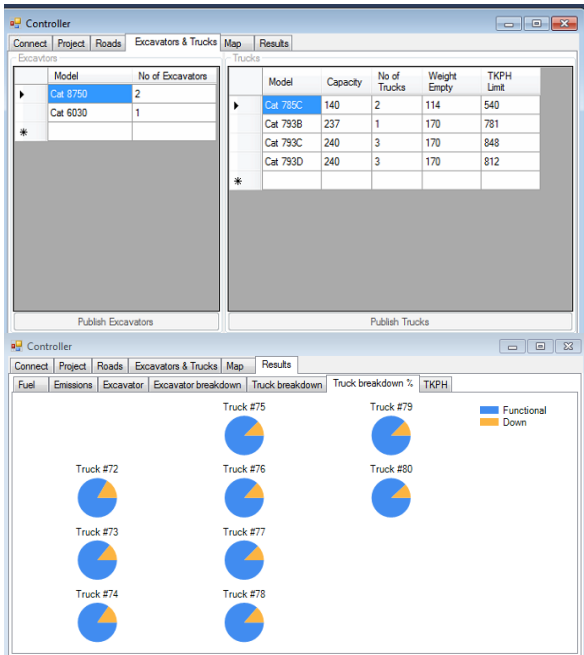


Figure 4: Controller Interface to Update the Truck and Excavator Initial Attributes and the Truck Availability Outputs

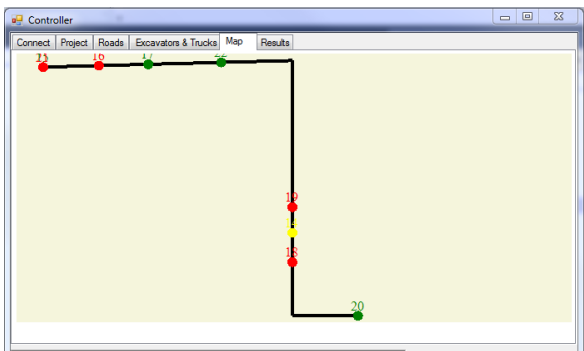


Figure 5: Controller Interface Showing the Truck Location and Status during the Simulation

- Weather Federate: This federate is responsible for publishing the weather conditions for the project location. The BMEF subscribes to all the parameters from this interaction class. Figure 6 shows the BMEF receiving an interaction with the weather parameters (89—Snow Fall; 90—Precipitation; 91—Visibility; 92—Wind Speed; 93—Temperature) during the simulation.

```
TKPH company 848.0
Requesting Time Advance to: 61.0
interaction received
[<89: 0.0, 90: 0.0, 91: 16.0, 92: 3.6, 93: 12.0>]
```

Figure 6: BMEF Receiving

- Mover Federate: This federate is responsible for moving the trucks in the project between the loading and unloading points. The BMEF subscribes to the Truck Location attribute and the Mover Federate subscribes to the truck

Equipment Status. Figure 7 shows the BMEF receiving the Truck Location attribute during the simulation.

```
C:\Python33\python.exe
Truck 18
At time: 359.0 Attribute updated is received for object #19
Attr:21 = <'Z': 0.0, 'Y': 0.8104446507469563, 'X': 1.95>
Truck 19
At time: 359.0 Attribute updated is received for object #20
Attr:21 = <'Z': 0.0, 'Y': 1.225, 'X': 2.2>
Truck 20
At time: 359.0 Attribute updated is received for object #21
Attr:21 = <'Z': 0.0, 'Y': 0.275, 'X': 1.0>
Truck 21
At time: 359.0 Attribute updated is received for object #22
Attr:21 = <'Z': 0.0, 'Y': 0.25714095250822794, 'X': 1.678643>
Truck 22
At time: 359.0 Attribute updated is received for object #14
Attr:21 = <'Z': 0.0, 'Y': 0.9080921047590471, 'X': 1.95>
Truck 14
At time: 359.0 Attribute updated is received for object #15
Attr:21 = <'Z': 0.0, 'Y': 0.275, 'X': 1.0>
Truck 15
```

Figure 7: BMEF Subscribes to the Truck Location Attributes from the Mover Federate

- Dumper Federate: This federate is responsible for loading and unloading the trucks. The BMEF subscribes to the Truck Load State and Payload and the Dumper Federate subscribes to the Excavator Equipment Status. Figure 8 shows the BMEF receiving the Truck Load State attribute (Attr:35) from the Dumper Federate through the RTI.

```
24
Total distance 12.372013117
tkph_per_tyres 595.624084198
TKPH company 812.0
Requesting Time Advance to: 357.0
At time: 356.4318066221646 Attribute updated is received for object #22
e stamp 356.4318066221646
Attr:33 = 240.0
Attr:35 = LoadStates.Loaded
At time: 357.0 Attribute updated is received for object #17 with time st
```

Figure 8: BMEF Subscribes to the Truck Load State Attributes from the Dumper Federate

Other interactions happen between the other federates, but they do not affect the BMEF and are not part of this research's scope.

### 3.1. BMEF Algorithm and Characteristics

The BMEF is responsible for calculating the equipment availability and publishing the equipment states as Functional, Down or Need Service. The federate works in a condition-based maintenance way. Aside from the information provided by the Controller Federate, the user should fill the data\_breakdown.csv file (Figure 9) with the information requested. As the main data required are distributions, the federate supports the following types: Normal, Uniform, Triangular, Exponential, Constant and Beta. The file also accepts values between 0 and 1. In this case, the user should update the distribution type with the word "Float." The parameter columns (P1 to P5) are the parameters necessary for each type of distribution. As an example, a uniform distribution with a low value of 5 and maximum of 10 should be updated in the model as follows: Distribution: uniform; P1 = 5; and P2 = 10.

| Description                            | Distribution | P1 | P2 | P3 | P4 | P5 |
|----------------------------------------|--------------|----|----|----|----|----|
| Cat 785C breakdown                     |              |    |    |    |    |    |
| Cat 793B breakdown                     |              |    |    |    |    |    |
| Cat 793C breakdown                     |              |    |    |    |    |    |
| Cat 793D breakdown                     |              |    |    |    |    |    |
| Time to repair                         |              |    |    |    |    |    |
| Increase time field                    |              |    |    |    |    |    |
| Cat 785C maintenance                   |              |    |    |    |    |    |
| Cat 793B maintenance                   |              |    |    |    |    |    |
| Cat 793C maintenance                   |              |    |    |    |    |    |
| Cat 793D maintenance                   |              |    |    |    |    |    |
| Time to maintenance                    |              |    |    |    |    |    |
| Probability truck break on Field       |              |    |    |    |    |    |
| Increase time repair temperature < -30 |              |    |    |    |    |    |
| Crews available                        |              |    |    |    |    |    |
| Cat 8750 breakdown                     |              |    |    |    |    |    |
| Cat 6030 breakdown                     |              |    |    |    |    |    |
| Time to repair Excavator               |              |    |    |    |    |    |
| Cat 8750 maintenance                   |              |    |    |    |    |    |
| Cat 6030 maintenance                   |              |    |    |    |    |    |
| Time to maintenance Excavator          |              |    |    |    |    |    |

Figure 9: CSV File for Input Data

In Figure 4 it can be observed that the model lets the user define more than one model of truck and excavator. The *Time to repair* is the total time for the truck to be repaired after breakdown. In this federate, two types of truck breakdowns are considered: in the field or in the shop. The user should define the probability (value between 0% and 100%) of a truck breaking down in the field in the row *Probability truck breakdown in the field*. If the truck breaks down in the field, the user can increase the repair time in the row *Increase time field* (between 0% and 100%). The repair time can also be increased (between 0% and 100%) if the temperature is less than  $-30^{\circ}\text{C}$  in the row *Increase time to repair temperature is less than  $-30^{\circ}\text{C}$* . The *Time to Maintenance* will be the same for all truck models.

The excavators will always break down in the field, and the *Time to repair* excavator will also be increased if the temperature is less than  $-30^{\circ}\text{C}$ . The *Maintenance Time* is the same for all the excavators. The user can also specify the quantity of crews available for equipment repair and maintenance. Figure 10 demonstrates the algorithm used in the BMEF to calculate the breakdown and maintenance intervals for the truck equipment.

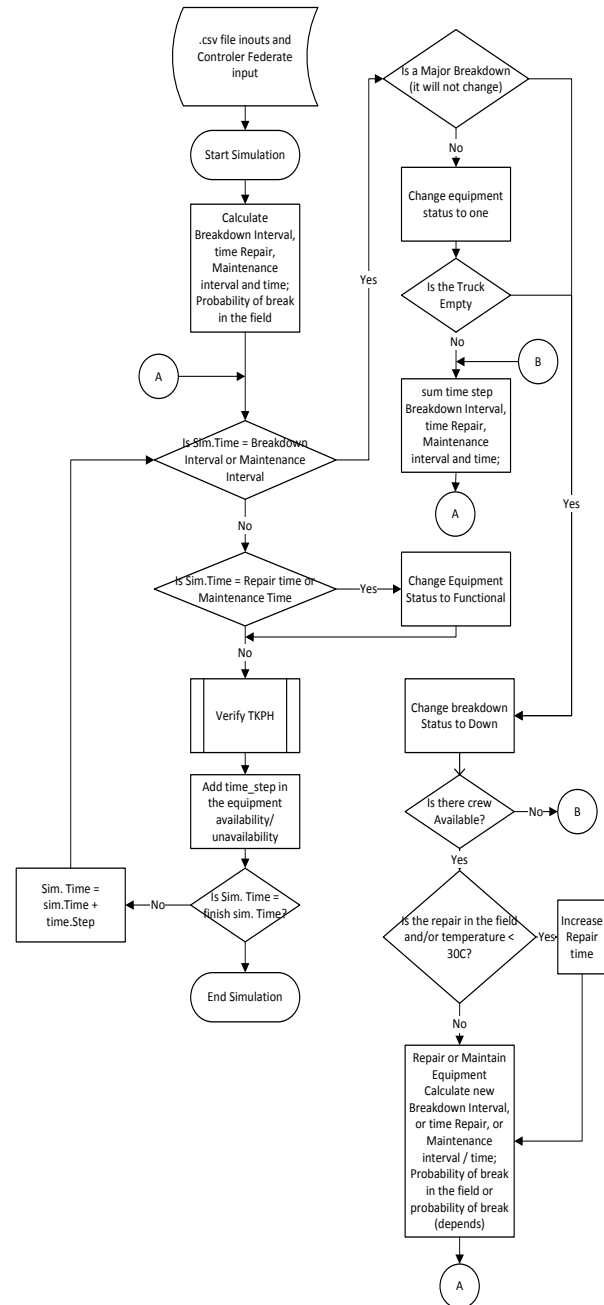


Figure 10: Algorithm Used in the BMEF for the Truck

With the users' inputs from the Controller Federate and the breakdown.csv file, the simulation starts. When the simulation time = 1, the truck (object) should define the following values: the next *Breakdown and Maintenance Interval*; *Time to repair* the equipment; *Time to maintain* the equipment; and a random number between 0 and 1 is generated to compare with the *Probability of breaking in the field*. The BMEF should receive the parameters that it subscribes to in every time step. The simulation time should be increased by one time step. If a truck breaks and it is a minor breakdown or if it is maintenance time, it should be emptied to be repaired or maintained. If the truck is empty or the breakdown is a major breakdown, the breakdown state will change to down and this information is published. If there is a crew available, the truck can be fixed. After

repairing or maintaining the truck, a new interval for breakdown or maintenance should be calculated. Other characteristics of the BMEF are described below:

- If a truck needs to wait to be emptied for repair or maintenance, or if equipment needs to wait for a crew, the *Time to repair* and *Maintenance interval* should be increased by the total waiting time for these resources to avoid the two events from happening at the same time.
- If a piece of equipment is in maintenance, the *Breakdown Interval* should be added with this time as the distributions for breakdown and interval only consider working hours. The same procedure should be performed when the truck is repaired. In this case, the *Maintenance Interval* should be increased by the repair time.
- For each piece of equipment in every time step, the federate verifies the equipment status. If the *Breakdown State* is *Functional* or *Need Service*, the *Available Time* should be increased by the time step, and if the *Breakdown State* is *Down*, the *Unavailable Time* is increased by the time step.

The BMEF also verifies if a truck is exceeding the TKPH limit provided by the user. The TKPH is a measurement of the work load of a tire and is based on the weight and distance (or speed) that the tire can travel without overheating. If the TKPH maximum value is achieved, the truck should stop for 60 minutes to decrease the tire temperature. It is assumed that all the truck tires have the same workload. The TKPH is calculated using Equation 1:

$$TKPH = \sum_{i=1}^{60} \left( \frac{(TEW+P_{(i)})}{6} * (TTD_{(i)} - TTD_{(i-1)}) \right) \quad (1)$$

Where:

- TEW: Truck Empty Weight (tons)
- P: Payload—the amount of material carried by the truck at a specific time (tons)
- TTD - Total Truck Distance—distance traveled by the truck in 1 minute (km)

Figure 11 shows the algorithm used to verify if the truck achieved the TKPH limit defined by the user. Figure 12 shows the calculation executed by the BMEF for each truck (19,20,21), comparing the TKPH calculated and the TKPH provided by the user in the Controller Federate.

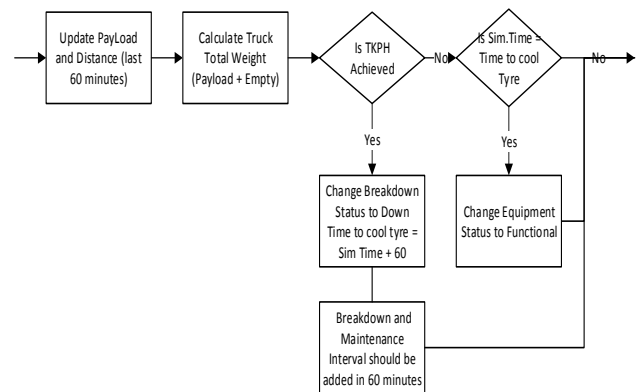


Figure 11: TKPH Algorithm

```

19
Total distance 8.2472682151
tkph_per_tyres 405.002329458
TKPH company 848.0
20
Total distance 11.8426195899
tkph_per_tyres 571.426710237
TKPH company 812.0
21
Total distance 11.5987323874
tkph_per_tyres 555.047531581
TKPH company 812.0

```

Figure 12: TKPH Calculated per Tire in the BMEF

#### 4. SCENARIOS AND RESULTS

To verify the BMEF behavior, two scenarios were tested. In the first scenario, the TKPH limit was not considered, and the parameters used in the simulations were retrieved from empirical data. They were used as input data for the simulation. The weather parameters used in scenario 1 were defined by the Weather Federate and they are from Fort McMurray (Alberta, Canada). The parameters used are described in Table 1. Several combinations were developed to demonstrate the options that the user can test with the BMEF. Only the Weather, Controller and BMEF federates were used in this scenario, and the time step is defined in hours.

Seven combinations were developed with the parameters defined in Table 1. After testing each combination, the BMEF provides the availability for each piece of equipment (Figure 13). The combinations vary in the number of crews, probability of breakdown in the field and weather parameters (Table 2). The total simulation time is 8760 hours (one year) and the equipment runs in 24/7 shifts.

Table 1: Parameters Used in Scenario 1

| Parameters                                                   | Duration (h) / Specification |
|--------------------------------------------------------------|------------------------------|
| Truck Breakdown Interval (same for all truck models)         | Exponential (200)            |
| Truck Maintenance Interval (same for all the trucks)         | Constant (400)               |
| Time to Repair Truck                                         | Uniform (20, 24)             |
| Time to Maintenance Truck                                    | Triangular (18,21,24)        |
| Increase Time to Repair in the Filed                         | 10%                          |
| Increase Time Repair Temperature is less than -30°C          | 20%                          |
| Excavator Breakdown Interval (same for all excavator models) | Exponential (250)            |
| Excavator Maintenance Interval (same for all the excavators) | Constant (350)               |
| Time to Repair Excavator                                     | Uniform (20, 24)             |
| Time to Maintenance Excavator                                | Triangular (18,22,25)        |
| Quantity of Trucks                                           | 9                            |
| Quantity of Excavators                                       | 3                            |

\* Combination 6 used *Constant Distribution* (200h) for *Truck and Excavator Breakdown Interval*.

It is possible to see that between combination 1 and 2, the truck and excavator unavailable times were decreased by 29% (comparing the mean) just by changing the quantity of crews. The unavailability of the trucks between combinations 2 and 3 increased 400% with a change in the probability breakdown in the field. This was expected, because the quantity of crews (4) is not enough if the repair time increases. In the same way, if the crew number increases to 6 (combination 5), the trucks' unavailability time is almost the same, when compared with combinations 2 and 4.

Combination 6 has constant distribution for breakdown interval (200h), but the equipment unavailable time was not increased. This result can be explained because the crew availability will be limited for the first breakdown, but not for the subsequent events, as the following breakdowns will happen at different times. However, the effect of the constant interval time on the cycle time should be investigated.

Combination 7 was the only one to consider the temperature to fix the equipment in the field, and the difference between this combination and combination 2 was not significant. The BMEF will be more influenced by the temperature if the probability of breakdown in the field is also increased. Moreover, the weather will not always affect equipment repair time, since the temperature changes every hour, and the mean temperature in Fort McMurray is higher than -30 °C.

In the second scenario, the TKPH limit was considered. The truck speed limit (50 km/h) and tire TKPH is defined by the user in the Mover and Controller federates, respectively, and the BMEF has no control on these parameters, thus limiting the combinations used in this scenario. The parameters used by the Controller Federate are described in Figure 14. In this scenario, only the Weather Federate was not used and the time step is defined in minutes.

```
total payload 0.0
Requesting Time Advance to: 8761.0
Federation Synchronized at point: ReadyToTerminate
Equipment.Excavator # 69 Available Time 7189 Unavailable Time 1571
Equipment.Excavator # 70 Available Time 6929 Unavailable Time 1831
Equipment.Excavator # 71 Available Time 7158 Unavailable Time 1602
Equipment.Truck # 72 Available Time 6443 Unavailable Time 2316
Equipment.Truck # 73 Available Time 6924 Unavailable Time 1835
Equipment.Truck # 74 Available Time 6724 Unavailable Time 2035
Equipment.Truck # 75 Available Time 7155 Unavailable Time 1604
Equipment.Truck # 76 Available Time 6905 Unavailable Time 1854
Equipment.Truck # 77 Available Time 7122 Unavailable Time 1637
Equipment.Truck # 78 Available Time 6973 Unavailable Time 1786
Equipment.Truck # 79 Available Time 7004 Unavailable Time 1675
Equipment.Truck # 80 Available Time 7269 Unavailable Time 1490
Resigned Federation execution, divest ownership and ready to terminate
Attempt to destroy federation failed
```

Figure 13: Output Availability/Unavailability of Each Piece of Equipment in the BMEF

Table 2: Different Combinations in Scenario 1

| Com-bo | Crew | Prob. Break-down in field | Weather parameters | Trucks unavail-able (%) (min, mean, max) | Excav-ators unava-ailable (%) (min, mean, max) |
|--------|------|---------------------------|--------------------|------------------------------------------|------------------------------------------------|
| 1      | 2    | 50%                       | N                  | (17, 21, 26)                             | (18, 19, 21)                                   |
| 2      | 4    | 50%                       | N                  | (13, 15, 16)                             | (13, 14, 15)                                   |
| 3      | 4    | 100%                      | N                  | (73, 77, 78)                             | (76, 77, 77)                                   |
| 4      | 4    | 75%                       | N                  | (13, 14, 15)                             | (14, 15, 17)                                   |
| 5      | 6    | 100%                      | N                  | (12, 14, 15)                             | (14, 14, 15)                                   |
| 6*     | 4    | 50%                       | N                  | (14, 15, 15)                             | (13, 14, 15)                                   |
| 7      | 4    | 50%                       | Y                  | (12, 14, 15)                             | (13, 16, 18)                                   |

| Model    | Capacity | No of Trucks | Weight Empty | TKPH Limit |
|----------|----------|--------------|--------------|------------|
| Cat 785C | 140      | 2            | 114          | 540        |
| Cat 793B | 237      | 1            | 170          | 781        |
| Cat 793C | 240      | 3            | 170          | 848        |
| Cat 793D | 240      | 3            | 170          | 812        |

Figure 14: Trucks Parameters to Verify the TKPH Limit (Defined in the Controller Federate)

The distribution for the breakdown in scenario 2 was changed to verify the influence on the truck unavailability. The breakdown interval and time to repair are the same for all the truck models and it was based on the assumptions made by Chang et al. (2013)



(Table 3). The simulation total time is 10,080 minutes (one week).

Table 3: Parameters Used in Scenario 2

| Parameters                                          | Duration (min)/ Specification |
|-----------------------------------------------------|-------------------------------|
| Truck Breakdown Interval (same for all truck Types) | Triangular (300, 360, 480)    |
| Time to Repair Truck                                | Triangular (60,80, 150)       |
| Increase Time to repair in the filed                | 10%                           |
| Quantity of Trucks                                  | 9                             |
| Maintenance Interval                                | Constant (1000)               |
| Maintenance Repair                                  | Triangular (60,80,150)        |
| Probability of Breakdown in the Field               | 50%                           |
| Quantity of crews                                   | 4                             |

The results show that the trucks achieved the TKPH limit between 50 and 42 times during the simulation and the mean unavailable time for the trucks is 57% (Figure 15). It is possible to observe that the BMEF was able to verify the TKPH limit and this value cannot be disregarded during the planning phase. Suggestions to increase the available time are: increase the tire TKPH, decrease the truck speed limit, or if possible, decrease the cycle time.

```

Equipment.Truck # 14 Available Time 4414 Unavailable Time 5665
Equipment.Truck # 15 Available Time 4598 Unavailable Time 5481
Equipment.Truck # 16 Available Time 4264 Unavailable Time 5815
Equipment.Truck # 17 Available Time 4760 Unavailable Time 5319
Equipment.Truck # 18 Available Time 4214 Unavailable Time 5865
Equipment.Truck # 19 Available Time 4580 Unavailable Time 5499
Equipment.Truck # 20 Available Time 4187 Unavailable Time 5892
Equipment.Truck # 21 Available Time 4904 Unavailable Time 5175
Equipment.Truck # 22 Available Time 4417 Unavailable Time 5662

```

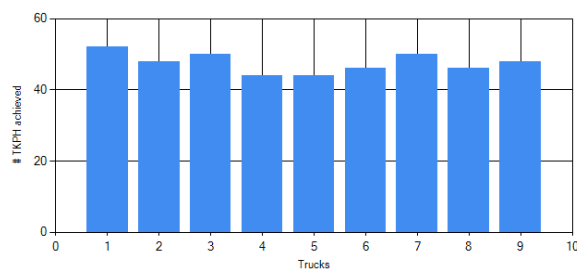


Figure 15: Truck Available and Unavailable Time (Provided by the BMEF) and Quantity that Each Truck Achieved the TKPH Limit (Provided by the Controller Federate)

## 5. VERIFICATION AND VALIDATION

The model output was examined under a variety of settings to perform model verification. Based on the scenarios and combinations developed, the BMEF demonstrated logic, and it is possible to verify that the model was implemented properly. The development of the simulation model made it possible to verify the influence of different parameters on equipment availability (as shown in the scenarios 1 and 2 in

Section 4). The BMEF demonstrated the effect of the quantity of crews, weather parameters, TKPH limits and the probability of the equipment breakdown in the field, as well as the type of distribution for maintenance and breakdown (Figure 13 and 15). The model behavior was also verified through the Controller Federate showing the equipment breakdown status. Also, the outputs provided by the BMEF and the Controller Federate are coherent.

However, as the data used to verify the BMEF was empirical, further research should be performed to validate the model. It is suggested that real data be used to verify the BMEF behavior, and necessary adjustments be done to increase the simulation model accuracy. Another possible way to validate the federate is through a face validation.

## 6. CONCLUSION

The research objective was achieved. The simulation model developed was able to verify the influence of different parameters on the equipment availability in an earthmoving operation and help the decision makers with maintenance management. The results can be used to improve the project schedule and engineers with limited experience can use the simulation model to verify the influence of the parameters on the equipment availability. The BMEF was able to receive data from other federates through the RTI and increase the model accuracy results.

To improve the model results, as indicated, actual data should be used in order to produce real results. The increment of time repair as a consequence of the temperature and localization (shop or field) needs further research. The relationship between maintenance and breakdown intervals should also be studied in greater depth, as it is agreed that shorter maintenance intervals can decrease breakdown occurrences. As an improvement, we should consider adding to the model the tire usage cycle, and the maintenance and breakdown intervals can be divided in different causes to provide a more realistic environment.

## ACKNOWLEDGEMENT

This research was made possible in part by Coordination for the Improvement of Higher Education Personnel (CAPES) Ministry of Education of Brazil (Proc. n° 0393-12-6).

## REFERENCES

- Abourizk, S. (2010). Role of Simulation in Construction Engineering and Management. *Journal of Computing in Civil Engineering*, 136, 1140–1153.
- AbouRizk, S. M., Hague, S., and Moghani, E. (2009). Developing a Bidding Game Using High Level Architecture. In *Computing in Civil Engineering (2009)* (513–522). Reston, VA: American Society of Civil Engineers. doi:10.1061/41052(346)51
- AbouRizk S. M., Robinson Fayek, A. (2008). “Synthetic Environments for Construction Planning



- and Control.” NSERC IRC Renewal application 2007-2011, pp. 11-15
- Bernold, L. (1989). Simulation of Nonsteady Construction Process. *Journal of Construction Engineering and Management*, 115(2), 163–178.
- BSI. 1984. Glossary of Maintenance Terms in Terotechnology. London: British Standard Institution (BSI), BS 3811
- Chi, S., and Caldas, C. H. (2009). Development of an Automated Safety Assessment Framework for Construction Activities. In IAARC (Ed.), *Proceedings of 26th International Symposium on Automation and Robotics in Construction (ISARC)* (pp. 38–46). Austin, TX.
- Edwards, D. J., and Holt, G. D. (2009). Construction plant and equipment management research: thematic review. *Journal of Engineering, Design and Technology*, 7(2), 186–206. doi:10.1108/17260530910974989
- Elazouni, A. M., and Basha, I. M. (1996). Evaluating the Performance of Construction Equipment Operators in Egypt. *Journal of Construction Engineering and Management*, 122, 109–114.
- Harris, F. C., and Olomolaiye, P. O. (1993). A theoretical model for determining equipment service life. *Building Research and Information*, 21(4), 243–245. doi:10.1080/09613219308727305
- Lee, T.-Y. (2010). Development and Validation of Rolling Resistance-Based Haul Road Management For. Purdue University.
- Marzouk, M., and Moselhi, O. (2004). Multiobjective Optimization of Earthmoving Operations. *Journal of Construction Engineering and Management*, 130, 105–113.
- Mccann, M. (2006). Heavy equipment and truck-related deaths on excavation work sites. *Journal of Safety Research*, 37, 511–517. doi:10.1016/j.jsr.2006.08.005
- Mohideen, P. B. A., and Ramachandran, M. (2014). Strategic approach to breakdown maintenance on construction plant – UAE perspective. *Benchmarking: An International Journal*, 21(2), 226–252. doi:10.1108/BIJ-05-2012-0030
- Mohideen, P. B. A., Ramachandran, M., and Narasimmalu, R. R. (2011). Construction plant breakdown criticality analysis – part 1: UAE perspective. *Benchmarking: An International Journal*, 18(4), 472–489. doi:10.1108/14635771111147597
- Moselhi, O., and Alshibani, A. (2010). Optimization of Earthmoving Operations in Heavy Civil Engineering Projects. *Journal of Construction Engineering and Management*, 135(10), 948–954.
- Muchiri, P. N., Pintelon, L., Martin, H., and Chemweno, P. (2013). Modelling maintenance effects on manufacturing equipment performance: results from simulation analysis. *International Journal of Production Research*, 1–16. doi:10.1080/00207543.2013.870673
- Nepal, M. P., and Park, M. (2004). Downtime model development for construction equipment management. *Engineering, Construction and Architectural Management*, 11(3), 199–210. doi:10.1108/09699980410535804
- Snaddon, D. (1988). Breakdown Theories and Complex Units A Critique. *Omega International Journal Of Management*, 16(2), 135–143.
- Wakefield, R. R., and Sears, G. A. (1997). PETRI Nets for Simulation and Modeling of Construction Systems. *Journal of Construction Engineering and Management*, 123, 105–112.
- Zakeri, M., Olomolaiye, P., Holt, G., and Harris, F. (1996). A survey of constraints on Iranian construction operatives’ productivity. *Construction Management and Economics*, 14(5), 417–425.

# DESIGN AND DEVELOPMENT OF A DISTRIBUTED EARTHMOVING SIMULATION

Duanshun Li<sup>(a)</sup>, Cristian Petre<sup>(b)</sup>, Cayce Kerr<sup>(c)</sup>, Tim Joseph<sup>(d)</sup>, Simaan AbouRizk<sup>(e)</sup>, and Yasser Mohamed<sup>(f)</sup>

<sup>(a)</sup>University of Alberta

<sup>(b)</sup>University of Alberta

<sup>(c)</sup>University of Alberta

<sup>(d)</sup>University of Alberta

<sup>(e)</sup>University of Alberta

<sup>(f)</sup>University of Alberta

<sup>(a)</sup>[duanshun@ualberta.ca](mailto:duanshun@ualberta.ca), <sup>(b)</sup>[petre@ualberta.ca](mailto:petre@ualberta.ca), <sup>(c)</sup>[cayce@ualberta.ca](mailto:cayce@ualberta.ca), <sup>(d)</sup>[tim.joseph@ualberta.ca](mailto:tim.joseph@ualberta.ca), <sup>(e)</sup>[abourizk@ualberta.ca](mailto:abourizk@ualberta.ca),  
<sup>(f)</sup>[yaly@ualberta.ca](mailto:yaly@ualberta.ca)

## ABSTRACT

Earthmoving operations are critical components of many mining and construction projects. These operations utilize heavy equipment to execute excavation, loading, hauling, dumping, grading, and compacting tasks in a repetitive and almost continuous manner. The efficiency, environmental impact and cost of these operations are affected by many interdependent factors. These include road layout, road material, road maintenance frequency, trucks and excavator characteristics, and weather conditions. This paper describes the design and development of a comprehensive simulation of earthmoving operations. The simulation presented herein is developed with a distributed approach using High Level Architecture (HLA) Standards to enable inclusion of different models and behaviors that affect these operations. The paper presents the components of the simulation and models implemented in each, with more focus on the transportation and hauling component. It also describes the distributed development process and showcases the implementation outcomes and sample use and verification of the simulation.

Keywords: Simulation, High Level Architecture, Earthmoving

## 1. INTRODUCTION

Despite the considerable complexity and risk involved in construction projects, computer based simulation provides a powerful tool to predict the potential outcome of certain construction plans including time, duration, cash flow, risk, etc. Earthwork operations have been a common application field for simulation since it was introduced to the construction industry in the early 1970s. After the innovation of a graphic based simulation software, CYCLONE (Halpin 1977), application of simulation in construction increased. Earlier research on earthwork simulation mainly focused on estimation (Willenbrock 1972, Clemmens and Willenbrock 1978). The focus then transformed to

productivity prediction and fleet optimization (Karshenas 1989, Smith and et al. 1995, Hegazy and Kassab 2003, Marzouk and Moselhi 2003, Marzouk and Moselhi 2004, Moselhi 2009). Hajjar and AbouRizk (1996) proposed a special purpose earthmoving simulation template based on Symphony (AbouRizk and Mohamed 2000). The special purpose template presents the user with graphical, high-level details through object oriented simulation, so the user does not require simulation knowledge, and thus, it enhances the speed of model construction. However, modifications or adjustment to the template could take considerable time for specific projects due to the uniqueness of each construction project.

One drawback shared by all of these models is that most of them are established, at the lowest level, on 4 major components: loading, hauling, dumping and returning, which ignore the existence of detailed models of individual components. However, considerable research has been done on the modeling of those tasks, especially loading and hauling/returning. A number of models (Filla 2005, Coetzee et al. 2007, Nezami et al. 2007, Wang and Yang 2007, Bošnjak et al. 2008, Coetzee and Els 2009, Schmidt et al. 2010) of the loading process have been proposed for excavator simulation based on bucket-soil interaction. While others (Tam et al. 2002, Maciejewski et al. 2003) try to model the productivity of the excavators. The vehicle operation cost (VOC) has been thoroughly investigated. Tan et al. (2012) provide a thorough review of existing VOC models and divide the evaluation of those models into 4 phases: pre-1970s broad level correlation studies, 1970s-1980s regression models, post-1980s mechanistic models and current research on updated vehicle technology and changing vehicle fleet. In this research, 4 major components are identified: fuel consumption, tire use, repair/maintenance and lubricating oil. As the dominating cost, fuel consumption highly depends on the speed and mass of the vehicle, the engine efficiency, the layout of the road, such as gradient and curvature, and the roughness of the road, and the weather

conditions, such as air speed. More information on widely used fuel consumption models NIMPAC and HDM-types can be found in Thoresen and Roper (1996). The drawback of those models is that the models neglect the evolution of roughness during the operation of the roads. As stated in Thompson et al. (2003), the roughness is not only related to the wearing course of the road, but also the usage and maintenance frequency. Based on experience in South Africa, Thompson and Visser (2003) also proposed a model aiming to derive the fuel consumption based on rolling resistance estimated along the operation time.

Despite of the availability of these models, few of them are applied in earthmoving simulation tools. One of the major barriers is the diversity of the models. Users in different regions have different soil or weather conditions; they may also have different data sources, which make inputs quite different. Considering roughness of the road, for example, it can be measured as international roughness index (IRI) or NRM count per kilometer; however, in a large amount of research, it is modeled as rolling resistance. To make full use of these models, a simulation platform that allows users to switch between models and easily make additions or adjustments is necessary. As an international standard for distributed simulation, High Level Architecture (HLA) is an optimal choice. The standard allows the user to switch between different models regardless of the location, programming language or inputs. HLA enables the integration of numerous existing models based on different operation systems or programming languages for loading, hauling and other components, and thus, form a complete comprehensive earthmoving simulation model.

Based on an HLA earthwork simulation platform, this paper describes the implementation of a Mover federate aiming to provide the location and the rolling resistance for each truck during travel between loading and dumping sites to facilitate visualization and fuel consumption, CO<sub>2</sub> emission, and cost estimation done by other federates. The main contribution of this work is that: 1) it derives the location and the rolling resistance according to the input information, such as breakdown state and loading state of the truck provided from other federates based on the model introduced by Thompson and Visser (2003); 2) it accounts for several scenarios of the truck limiting speed strategies and maintenance strategies that affect operation performance. For example, one performance measure that is usually overlooked in earthmoving simulations is the tonnage moved per hour and its effect on tire deterioration or ton kilometers per hour (TKPH). There are limited ratings with regards to this measure that a tire can handle without overheating and causing premature deterioration for each type of tire (Mining 2014). The federate models these effects and allows creation of scenarios where speed is governed by either user input, sampling from fitted distribution or TKPH limits.

## 2. SIMULATION STRUCTURE

The earthmoving operation is modeled as an HLA federation that has five main federates that interact during runtime to simulate different behaviors. These federates are Controller federate, Loader federate, Mover federate, Weather federate and Breakdown federate. Figure 1 illustrates the federation and the objectives of each federate. The federates are time regulated and the time step used is 1 minute.

Each federate in the federation is implemented by a separate development team/individual. The federation uses an HLA run time infrastructure developed at the University of Alberta (AbouRizk et al. 2009) with API available for .Net languages, Java, and Python. All federates were developed in Python except for the Controller federate, which was developed using C#.

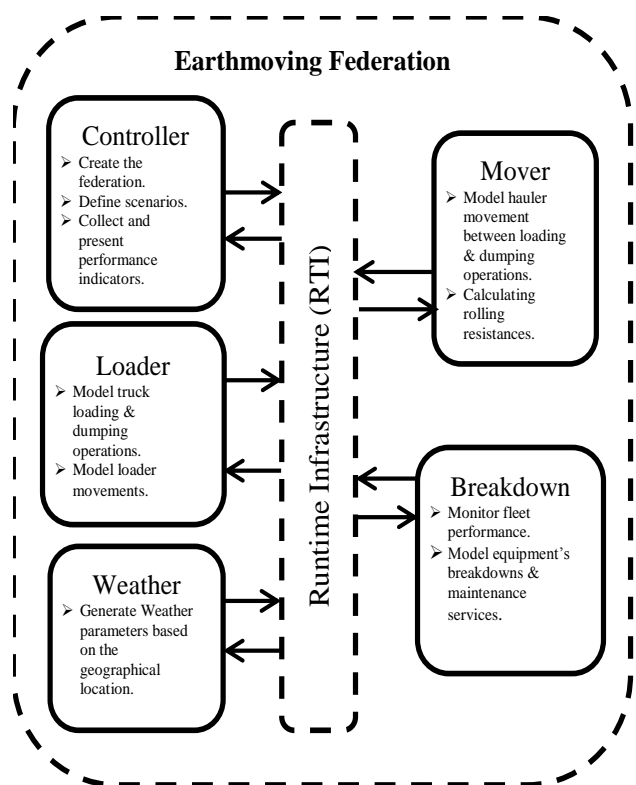


Figure 1: Federation Structure

The Controller federate is responsible for starting the HLA service and creating the earthmoving federation. The federate also controls the starting point of the simulation and initializes some of the essential attributes for the project object classes, such as the fleet size and combination, and the road sections for hauling and returning operations. At the same time, the federate receives results from other federates and presents them in a graphical and statistical fashion.

A Loader federate is established to simulate the excavating operation. The main job of this federate is to model the excavating process, which includes updating the loading status and the gross weight of the vehicle; meanwhile, it is also responsible for the dumping process related to the loading status.

To make the model more complete and realistic, modeling other critical factors, such as breakdown events and the weather condition, is essential. In the federation, the weather condition will be forecasted hourly based on a realistic historical data set by the Weather federate, while the Breakdown federate is responsible for the breakdown events for each piece of equipment.

Most of the time and cost spent on vehicles is during traveling. As the most critical part of the simulation, the Mover federate changes the moving state of the trucks based on the road layout, loading states and the breakdown states received from other federates. The position and rolling resistance will be derived based on the course wear of the road and the speed sampled from fitted distributions. Meanwhile, the results will be published to other federates for further calculation. Focusing on the comprehensive modeling of the traveling process, the mover federate will be detailed in the following sections.

Each federate uses its own model to update object attributes under its jurisdiction separately, while the communication between federates is established through the shared federation object model (FOM). The FOM describes the types of objects and attributes that are exchanged and the parameters that should be transferred for interactions during the simulation. The major objects the earthmoving operation is dealing with are equipment and the road network; thus, they are recognized as the major part of the object classes in our implementation as illustrated in Table 1. The publish/subscribe (P/S) properties for each attribute is also given in the table, where C, W, M, L and B indicate Controller, Weather, Mover, Loader and Breakdown separately. The abbreviation will be applied hereafter for convenience in this paper. In our current implementation, weather updating is treated as the sole interaction class, and the structure of it is presented in Table 2.

### 3. MOVER FEDERATE

The mover federate aims to provide the location and the rolling resistance for each truck during travel between loading and dumping sites to facilitate visualization and fuel consumption, CO2 emission, and cost estimation simulated by other federates. In addition, statistics regarding the cycle time for the trucks are collected and published to the Controller federate for visualization and performance analyzing. The federate requires certain inputs from other federates as well. The key inputs include road layout and road wearing course material, truck model, breakdown state and loading state. Generally, the federate execution can be divided into three stages: Initialization, Instance Creation and Updating. The order of initialization and instance creation is given in Figure 2.

Table 1: Object Class Structure Table

| Attribute            | Type         | C      | W | M      | L | B |
|----------------------|--------------|--------|---|--------|---|---|
| <b>Equipment</b>     |              |        |   |        |   |   |
| Location             | Point3D      | S      |   | P<br>S | S | S |
| Model                | String       | P<br>S |   | S      | S | S |
| BreakdownState       | Enum         | S      |   | S      | S | P |
| Capacity             | Float        | P      |   | S      | S |   |
| EquipmentAvailable   | Float        | S      |   |        |   | P |
| EquipmentUnavailable | Float        | S      |   |        |   | P |
| <b>Truck</b>         |              |        |   |        |   |   |
| LocationType         | Enum         |        |   | P      | S |   |
| LoadState            | Enum         |        |   | S      | P | S |
| MovingState          | Enum         | S      |   | P      | S |   |
| PayLoad              | Float        | S      |   | S      | P |   |
| RollingResistance    | Float        | S      |   | P      |   |   |
| RoadSegmentID        | Integer      | S      |   | P      |   |   |
| WeightEmpty          | Float        | P      |   | S      |   | S |
| CycleTime            | Float        | S      |   | P      |   |   |
| HaulTime             | Float        | S      |   | P      |   |   |
| ReturnTime           | Float        | S      |   | P      |   |   |
| TKPHLimit            | Float        | P      |   | S      |   | S |
| TireTKPHAchieved     | Int64        | S      |   |        |   | P |
| <b>Excavator</b>     |              |        |   |        |   |   |
| ProductionRate       | Float        |        |   |        | P |   |
| UtilityRate          | Float        | S      |   |        | P |   |
| <b>RoadSegment</b>   |              |        |   |        |   |   |
| Id                   | Integer      | P      |   | S      |   |   |
| Node1                | Point3D      | P      |   | S      |   |   |
| Node2                | Point3D      | P      |   | S      |   |   |
| Material             | String       | P      |   | S      |   |   |
| DailyKTONnage        | Float        | S      |   | P      |   |   |
| MaintenanceDate      | Date         |        |   |        |   |   |
| <b>Road</b>          |              |        |   |        |   |   |
| Sections             | IntegerArray | P      |   | S      |   |   |
| <b>Project</b>       |              |        |   |        |   |   |
| Location             | String       | P      | S |        |   |   |
| StartDay             | Date         | P      | S | S      |   |   |

Table 2: Interaction Class Structure Table

| Parameter      | Type  | C | W | M | L | B |
|----------------|-------|---|---|---|---|---|
| CurrentWeather |       | S | P | S | S | S |
| Temperature    | Float | S | P | S | S | S |
| WindSpeed      | Float | S | P | S | S | S |
| Visibility     | Float | S | P | S | S | S |
| SnowFall       | Float | S | P | S | S | S |
| SnowDepth      | Float | S | P | S | S | S |
| Precipitation  | Float | S | P | S | S | S |



Figure 2: Initialization and Instance Creation

To evaluate the performance of the fleet under different speed control strategies, switches and other parameters are provided to allow the user to simulate different scenarios. The lowest speed and highest speed are provided as the hard limit of the maximum and minimum speed of each moving truck. The maintenance interval of the road, which will influence the rolling resistance and furthermore the cost, is also provided and can be modified using one of the switches. The interval can be specified either as a constant or as a fitted distribution. Considering that exceeding the TKPH limit of the tires may cause them to breakdown prematurely and finally influence the production of the system, two Boolean parameters are specified as switches to control the speed of the trucks. The first switch allows the user to choose between sampling the speed of the trucks from past data or setting it as a constant value derived from the TKPH limit and the gross vehicle weight. The other one is related to the speed sampling scenario and allows the user to specify whether to apply a speed limit derived from the TKPH limit as an additional maximum speed limit for each truck, or not. However, whether the TKPH limit is used to derive the speed of the truck or as an additional limit, the truck will never achieve this value as long as there are stops during any given hour due to loading or dumping. The switches and parameters are organized in an initialization file which is easy to modify, while the configurations will be read from the file and applied to the simulation at the beginning of the federate execution.

In the Instance Creation stage, the federate listens to the instance creation messages and creates local records of the instances. The road layout is the combination of a group of road segments and one or several lists of ordered road segment IDs called road sections, indicating different hauling roads. For each truck there must be a hauling road assigned to it, which means the truck should not be created before the creation of the road section; the order of initialization and instance creation is given in Figure 2. Apart from the coordinates, the material of the road should also be provided from the Controller. This is needed in order to derive the rolling resistance. To simplify the interface of the Controller, a local record of properties of the wearing course material is kept by the Mover federate.

In the most important stage of the federation execution, the updating stage, the federate will listen to the messages from other federates related to the breakdown state and loading state in order to derive the location and rolling resistance for each truck before sending the updated values to the RTI. Certain statistics, such as the cycle time, hauling trip duration, return trip duration for each truck and daily tonnage hauled over each road segment, are also collected for visualization and evaluation.

The updating procedure is illustrated in Figure 3. To make the illustration clearer, vertical lines and horizontal lines are used to indicate the time line and the order of the procedures separately. Each updating cycle starts from the end of time  $i$  which is achieved when all federates ask for time advance request to time  $i+1$ . Once it is achieved, the Mover federate starts to receive messages from the RTI. The Mover will calculate the location and rolling resistance of each truck according to that information and the local record and store it in a local updates list. Once the time reaches  $i+1$  and all of the messages have been processed, it will publish all of the updates to other federates and ask for time advance request to time  $i+2$ .

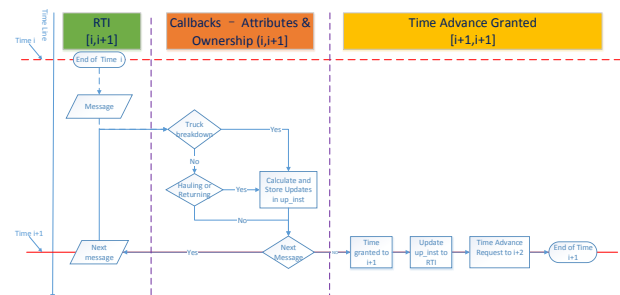


Figure 3: Flow Chart of the Mover Federate Updating Process

### 3.1. Location Updating

The location of the vehicles is updated every minute and published to other federates. The method used to update the location is quite straight forward. The travelling distance is calculated based on kinematics theory using the derived speed. The speed of the vehicles is controlled by the switches and distributions specified in the initialization file, as stated previously. Combined with specified road layout as constraints, the 3D location can be easily derived.

### 3.2. Rolling Resistance Updating

Resistance is one of the most significant factors that will influence the productivity of the road network. The total resistance (TR) is the combination of the grade resistance (GR) and the rolling resistance (RR). It is usually specified as effective grade (EG) in the form of percentage of gross vehicle mass (GVM). It can be written as:

$$TR = RR + GR$$

$$EG(\%) = RR(\%) + GR(\%) \quad (1)$$

The GR can be easily derived from the geometry layout of the road segments and will not be the major concern in this paper. On the contrary, as a factor highly related to the roughness of the road, rolling resistance is usually recognized as an indicator of the roughness, similar to the international roughness index (IRI) or NRM counts per kilometer. Increased rolling resistance reduces fleet productivity, and leads to increases in

operating costs. It is mainly composed of five major components:

- internal power train friction,
- tire flexing under load,
- tire penetration,
- road deflection, and
- air resistance.

Though a rule-of-thumb formula for RR estimation exists and is widely used in the heavy equipment manufacturing industry, it is supposed to be underestimated on an unpaved surface (Karafath 1988, Tannant and Regensburg 2001). To address the problem, a number of models based on tire-soil interaction and surface roughness are proposed. Karafath (1988) argues that the general model underestimates the rolling resistance of non-driven tires and proposes a method to calculate the rolling resistance and the speed attainable during hauling by taking the torque into consideration. However, this model includes too much trivial detail of the tire soil interaction, which makes it too sensitive for simulation. The other disadvantage of this model is that it does not quantify the properties of the wearing course of the road and does not take them into consideration. Another mechanistic-empirical model proposed by Paterson (1987) quantifies the deterioration of the road as roughness which is the combination of structural deformation, surface defects and age and environmental factors. Harvey (2012) revised this formula based on this method which can be illustrated as:

$$R(t) = [R_0 + 134SNCK^{-5}NE(t)]e^{mt} \quad (2)$$

where  $m$  is the environmental coefficient,  $t$  is the pavement age,  $SNCK$  is the modified structural number adjusted for the effect of cracking, and  $\Delta NE$  is millions of equivalent standard axle (ESA) loads per lane during the period, and  $NE(t)$  is cumulative ESAs until time ten millions per lane. The problem with this model is that although the roughness is quantified, the transformation from roughness to rolling resistance is still missing. The problem is addressed by Thompson and Visser (2003), in which a formula for rolling resistance estimation, while taking deterioration of the road and the material of the wearing course into consideration, is provided. In this method, the wearing course is quantified by California Bearing Ration (CBR), Grading Coefficient (GC), Shrinkage Product (SP), Plastic Index (PI) and other characteristics. The roughness is derived from the wearing course material, daily tonnage hauled on the road and the number of days since the last road maintenance.

Based on the survey and evaluation of existing models, the model provided in Thompson and Visser (2003) is applied in the Mover federate for rolling resistance estimation. In this model, roughness defect score (RDS) is defined to quantify the effect of the road wearing course material and its interaction with the tire

in the context of the traffic speed and total traffic volume on a given haul road.

This model considers the rolling resistance as a function of the vehicle speed and aforementioned RDS, which incorporates the wearing course parameters, as well as the number of days passed since the road was last maintained, together with the total volume of traffic within that period. Assume the daily tonnage hauled on the road segment is  $KT$ , the RDS of the specified road segment is provided in Equation 3 given the parameters of the road wearing course. The model requires information with regards to the 100% Mod. California Bearing Ratio (CBR), the Plasticity Index (PI),  $SP$  and  $GC$ . The last two parameters required can be defined either independently or derived from the passing rate through gradation test (Thompson 2003).

$$RDS = RDSMIN + \left[ \frac{RDSMAX - RDSMIN}{1 + e^{RDSI}} \right] \quad (3)$$

where

$$\begin{aligned} RDSMIN &= 31.1919 - 0.05354SP - 0.0152CBR \\ RDSMAX &= 7.6415 + 0.4214KT + 0.3133GC + 0.4952RDSMIN \\ RDSI &= 1.768 + 0.001D(2.69KT - 72.75PI - 2.59CBR - 9.35GC + 1.67SP) \end{aligned}$$

With speed  $V$ , the rolling resistance of the vehicle on this road segment is given in Equation 4.

$$RR = RRMIN + RDS \cdot e^{RRI} \quad (4)$$

where

$$\begin{aligned} RRMIN &= e^{-1.8166 + 0.0028V} \\ RRI &= 6.068 - 0.00385RDS + 0.0061V \end{aligned}$$

### 3.3. Tire Capacity

The travel conditions such as speed, payload, road conditions and cycle time influence the performance of a truck and its breakdown likelihood directly (Zhou 2006). One measure of the work a tire has to perform in a hauling operation is the ton kilometer per hour (TKPH). There are limiting ratings with regards to this measure that a tire can handle without overheating and causing premature deterioration for each type of tire (Ming 2014).

Although dynamic conditions govern the load distribution between tires depending also on payload and ground conditions (Joseph 2003), an empirical method of deriving the tire load when exact data is not available is to divide the truck load by the number of tires used. The Mover federate implements this concept and allows its use either as a limiting factor or as a means of deriving the hauling speeds. For either case, although it is recognized that a truck will not spend its entire time travelling, and will be allowed to cool down during loading and dumping, it was decided that for a safe implementation in the case when the TKPH limits are used to derive hauling speeds, these cooling times will be ignored as a means of implementing a factor of



safety in the operation. However, as will be described in more detail in Section 4, the user will have the ability to choose if they wish to consider this metric in the simulation or not.

For a truck with 6 tires, the speed limit of the truck will be:

$$V_{max} = \frac{6 \cdot TKPH}{GVW} \quad (5)$$

where *GVW* is the total mass of both the vehicle and the payload.

#### 4. SUPPORTED SIMULATION SCENARIOS

The federate user can change types and parameters of the distributions used to sample the travel speed from, as well as select speed limits that are to be applied. The tire speed limitations can also be derived from their TKPH limits received from the controller federate; this option can be switched on and off at the user's choice. The default data used for the speed distributions allows for four different types of trucks. In addition, there is another set accepted in case the truck types received from the controller do not coincide with those in the data input file; for the case study this is an average distribution of the available data sets.

Soil parameters can be modified according to the specifications of a particular application. Currently, the federate supports two sets of material types, each with up to three materials. The first type is specified in terms of the sieve percentages, whereas the second accepts readily defined parameters. Once the parameters of these materials are specified, the user can select the material applicable to each road segment from the Controller federate by specifying the material name.

Specifying soil parameters and speed distribution parameters can be achieved by changing the values in the data initialization .csv files. Once all the required information is introduced in the speed and soil .csv files, the user can specify the speed limits and choose which scenario or combination to run from the Init.csv initialization file. This file allows a user to specify the minimum and maximum speed limits for the operation, as well as a maintenance interval for the roads. By setting the Maximum Tire parameter to 0, the model will use the distributions to sample a speed for a truck each time travel starts or the segment changes. Setting this parameter to 1 will bypass the distribution sampling and use a model that computes the speed based on the TKPH limit. This scenario is one where the trucks tires shall never fail, since their haul speeds are derived from their TKPH limit (Michelin 2004), and they are also allowed to cool during loading and dumping.

The last scenario option available is the "TKPHASLIMIT", which expects either a 0 or a 1 for the value as a Boolean switch. Setting it to 1 applies a speed limit based on the TKPH speed derivation in the case where the federate samples truck haul speeds from the given distributions. This check is performed in addition to checking the Highest Speed value, and in

case the sampled speed is larger than either of them, the smallest value will be used instead.

## 5. EXPERIMENT AND RESULTS

### 5.1. Data Resources

The haul road in the experiments is provided as three road segments with a total length of 2.18 km. Based on a realistic data set collected at an Alberta mining operations site, two distributions are derived to fit the hauling trip speed and the returning trip speed respectively for each truck model. According to the result, the best fits of the hauling trip speed are normal distributions and the best fits of returning trip are triangular distributions based on an evaluation which combines the result from criteria provided by @Risk. The hauling speed distribution parameters for each type of trucks are given in Table 3, where "other" will be applied if the model specified by the user cannot be found in the local record. This table also includes information about the TKPH limit for the tires used and the capacity of the trucks.

Table 3: Hauling trip distributions and TKPH tire limits

| Truck    | Hauling (Normal)<br>(km/hr) |                       | TKPH<br>Limit for<br>Each<br>Tire | Capacity<br>(tons) |
|----------|-----------------------------|-----------------------|-----------------------------------|--------------------|
|          | Mean                        | Standard<br>Deviation |                                   |                    |
| CAT 785C | 20.0146                     | 5.0947                | 540                               | 140                |
| CAT 793B | 20.5110                     | 5.1015                | 781                               | 237                |
| CAT 793C | 20.6625                     | 5.1017                | 848                               | 240                |
| CAT 793D | 21.3953                     | 5.0717                | 812                               | 240                |
| Other    | 20.5613                     | 5.1152                | -                                 |                    |

Three typical road wearing course materials are used in the experiment: Stone, Slag and Dolerite. The characteristics of the materials are described as the following table.

Table 4: Material characteristics

| Material | CBR | LS  | PI | P425 | P2 | P475 | P265 |
|----------|-----|-----|----|------|----|------|------|
| Stone    | 69  | 0.5 | 0  | 27   | 32 | 37   | 60   |
| Slag     | 140 | 0.5 | 0  | 17   | 37 | 61   | 100  |
| Dolerite | 26  | 3.5 | 7  | 34   | 39 | 45   | 70   |

### 5.2. Scenarios and Results

Several experiments are constructed based on the initialization files for scenario configuration. Both the lowest and highest speed limits can be specified in the configuration file. Additionally, another speed limit derived from the TKPH limit can be applied through a switch. If the switch is on, the federate will check if sampled speed satisfies all of the limits or not, if not satisfied the THKPH derived speed limit will be adopted.

The federation execution has been run for 10 days in order to produce an initial set of results. In this

experiment, four types of trucks listed in Table 3 are used; these are: 785C, 793B, 793C and 793D. The number of the trucks used of each type is: 2, 1, 3 and 3 respectively.

Totally three scenarios are simulated:

- A. Sampled speed according to fitted distribution without speed limits derived from TKPH limits;
- B. Sampled speed according to fitted distribution with speed limits derived from TKPH limits;
- C. Fixed speed derived from TKPH limits.

In Scenario A - Sampled Speed without TKPH Limits, the speed is sampled from the distribution fitted, and the speed limit derived from the TKPH limit will not be applied. The lowest and highest values allowed for speed are 1 and 50 km/hr respectively. Because the speed limit corresponding to TKPH limit is not adopted in this scenario, there will be no truck breakdowns due to exceeding of the TKPH limit.

Due to the TKPH derived speed limit being applied in scenario B, there will be no truck breakdown due to exceeding of the TKPH limits in this scenario. What should be noted is that, in scenario C, the potential of the tire capacity is maximized compared to the other two scenarios, by setting the speed according to this limiting factor.

The statistics of the production and the cycle times for each scenario are presented in Table 5, where KT is the average daily tonnage hauled on the road, a measure of the production rate.

Table 5: General statistics of the scenarios for the cycles without breakdowns

| Scenario | KT  | Average Cycle Time (min) |      |      |      |      |
|----------|-----|--------------------------|------|------|------|------|
|          |     | 785C                     | 793B | 793C | 793D | ALL  |
| A        | 369 | 19.5                     | 19.9 | 20.1 | 19.8 | 19.8 |
| B        | 330 | 21.4                     | 22.8 | 22.5 | 22.5 | 22.3 |
| C        | 398 | 17.4                     | 19.1 | 18.2 | 19.1 | 18.5 |

In order to compare the percentages of breakdown cycles between the scenarios, the results have been aggregated for all trucks with the breakdown cycles separated, as summarized in Table 6 below.

Table 6: Cycle Time Summary for all Scenarios

| Cycle Time Summary      |       |       |       |
|-------------------------|-------|-------|-------|
| Scenario                | A     | B     | C     |
| <b>All Cycles</b>       |       |       |       |
| Count                   | 4635  | 4047  | 5059  |
| Average (min)           | 27.9  | 32.0  | 25.6  |
| Minimum (min)           | 13.0  | 16.0  | 16.0  |
| <b>Breakdown Cycles</b> |       |       |       |
| Count                   | 353   | 359   | 343   |
| Average (min)           | 125.9 | 131.4 | 124.1 |
| Percentage              | 7.62% | 8.87% | 6.78% |

### 5.2.1. Scenarios Interpretation

From these results, we can conclude that the maximum production rate is achieved while the potential of the tire is fully utilized (Scenario C). This can be said by looking at the daily average tonnage hauled on the road with all the trucks (KT values in Table 5) or at the total number of cycles achieved in 10 days of simulated operations for each scenario (Table 6).

Little variation exists between types of trucks and the cycle time when distribution sampling alone (Scenario A) is employed - this is because the input distributions for the different types of trucks are very similar (Table 3). However, one should also consider that these trucks have different capacities and therefore their loading and dumping time would vary in a fashion not analyzed within this study. When considering capacity, which is very similar between trucks 793 A, B and C, we can observe that the average cycle time (presented in Table 5) varies in a consistent fashion to the mean hauling speed (presented in Table 3). This means, for example, that the truck with the highest mean speed (CAT 793D) is also the one experiencing the lowest average cycle times in the case of Scenario A, although by a slim margin.

It is expected that the trucks would break down more frequently when the TKPH-derived speed limit is not enforced (Scenario A) due to the addition of the tire cool down periods enforced by the Breakdown Federate. However, the results do not show a decrease in the percentage of breakdown cycles between Scenario A and B. This can most likely be attributed to the random nature of the mechanical breakdowns simulated outside the scope this federate, which lead to the inability to analyze solely the TKPH breakdowns.

By analyzing the cycle times between the two cases where speed was sampled (A and B), it can be observed that applying a speed limit (in scenario B) has a negative impact on the average cycle time; this is valid whether all cycles are included (top part of Table 6) or only the ones without breakdowns (Table 5). When the speeds are derived from the TKPH limits (Scenario C), the average cycle times will be reduced back to similar values of Scenario A since the simulation will not be using lower speeds than those derived. Moreover, the return times decrease significantly due to the high allowable speed limit for the empty trucks; this behavior is induced by the reduced value for the denominator in Equation 4. This leads to the most efficient operation, where the productivity (in terms of either KT or cycle times) is maximized, while experiencing the minimum percentage of breakdowns.

## 6. CONCLUSION

A distributed simulation platform based on HLA standard is established to model earthmoving operation. The platform enables the possibility of utilization of existing comprehensive models of the loading and hauling process. As an example of one comprehensive

model, the Mover federate is able to successfully update the location of trucks during the execution of an HLA compliant distributed simulation of an earth moving operation, while responding and reacting to updates from the Loader and Breakdown federates. The primary experimentation scenarios supported are related to the way the speed of the trucks is derived, which can be either from past data or can be derived in order to comply with the TKPH limit of the tires. The rolling resistance calculations are performed using a realistic and comprehensive model that accounts for the road wearing course deterioration. The experiment shows that the federate is able to reflect the influence of the fleet composition, the truck type, TKPH limit and the speed limits. Thus, the federate can be used to evaluate the effectiveness of the tire, the speed limit settings and fleet composition while planning the project.

## REFERENCES

- AbouRizk, S., and Mohamed, Y., 2000. Symphony: an integrated environment for construction simulation. *In Proceedings of the 32nd conference on Winter simulation*, pp.907-1914.
- AbouRizk, S. M., Hague, S., & Moghani, E., 2009. Developing a bidding game using high level architecture. *In Proc., 2009 ASCE Int. Workshop on Computing in Civil Engineeringpp*, pp. 513-522). Reston, VA: ASCE.
- Bošnjak, S., Zrnić, N., Petković, Z., 2006. Bucket wheel excavators and trenchers - Computer added calculation of loads caused by resistance to excavation. u: Kuzmanović S. (ur.) *Machine Design*, University of Novi Sad, pp.121-128
- Clemmens, J. P., and Willenbrock, J. H., 1978. The SCRAPESIM computer simulation. *Journal of the Construction Division*, 104(4), 419-435.
- Coetzee, C. J., Basson, A. H., and Vermeer, P. A., 2007. Discrete and continuum modelling of excavator bucket filling. *Journal of terramechanics*, 44(2), 177-186.
- Coetzee, C. J., and Els, D. N. J., 2009. The numerical modelling of excavator bucket filling using DEM. *Journal of Terramechanics*, 46(5), 217-227.
- Filla, R., 2005. An event-driven operator model for dynamic simulation of construction machinery. arXiv preprint cs/0506033.
- Hajjar, D., AbouRizk, S., 1996. Building a Special Purpose Simulation Tool for Earth Moving Operations. *Proceedings of the 1996 Winter Simulation Conference*, pp.1313-1320.
- Halpin, D. W., 1977. CYCLONE: Method for Modeling of Job Site Processes, *Journal of the Construction Division*, ASCE, 103(3), 489-499.
- Harvey M.O., 2012. Optimising road maintenance. *International transport forum discussion papers*. OECD publishing.
- Hegazy, T. and Kassab, M., 2003. Resource Optimization Using Combined Simulation and Genetic Algorithms. *Journal of construction engineering and management*, 129(6), 698–705.
- Joseph, T. G., 2003. Large mobile mining equipment operating on soft ground. *In Proceedings of the 18th International Mining Congress and Exhibition of Turkey-IMCET*, pp. 10-13.
- Karafath, L. L., 1988. Rolling resistance of off-road vehicles. *Journal of construction engineering and management*, 114(3), 458-471.
- Karshenas, S. (1989). Truck Capacity Selection for Earthmoving. *Journal of construction engineering and management*, 115(2), 212–227.
- Maciejewski, J., Jarzębowski, A., & Trampczyński, W., 2003. Study on the efficiency of the digging process using the model of excavator bucket. *Journal of terramechanics*, 40(4), 221-233.
- Marzouk, M. and Moselhi, O., 2003. Object-oriented Simulation Model for Earthmoving Operations. *Journal of construction engineering and management*, 129(2), 173–181.
- Marzouk, M. and Moselhi, O., 2004. Multiobjective Optimization of Earthmoving Operations. *Journal of construction engineering and management*, 130(1), 105–113.
- Michelin., 2004. *EM & Industrial Data Book*. Michelin.
- Mining, P., 2014. *TKPH*, Precision Mining Software. Available from: <http://support.precisionmining.com/spry-help/building-a-spry-model/scenarios/trucks/tkph/> [accessed 14 Apr 2014]
- Moselhi, O. and Alshibani, A., 2009. Optimization of Earthmoving Operations in Heavy Civil Engineering Projects. *Journal of construction engineering and management*, 135(10), 948–954.
- Nezami, E. G., Hashash, Y., Zhao, D., and Ghaboussi, J., 2007. Simulation of front end loader bucket–soil interaction using discrete element method. *International journal for numerical and analytical methods in geomechanics*, 31(9), 1147-1162.
- Paterson WD., 1987. The highway design and maintenance standard series – road deterioration and maintenance effects. *Baltimore, MD: The Johns Hopkins University Press*.
- Schmidt, D., Proetzsch, M., and Berns, K., 2010. Simulation and control of an autonomous bucket excavator for landscaping tasks. *2010 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5108-5113.
- Smith, S., Osborne, J., and Forde, M., 1995. Analysis of Earth-Moving Systems Using Discrete-Event Simulation. *Journal of construction engineering and management*, 121(4), 388–396.
- Tam, C. M., Tong, T. K., and Tse, S. L., 2002. Artificial neural networks model for predicting excavator productivity. *Engineering Construction and Architectural Management*, 9(5 - 6), 446-452.
- Tan, F., Thoresen, T., & Evans, C. (2012, September). Review of Vehicle Operating Costs and Road Roughness: Past, Current and Future. *In 25th ARRB Conference*, pp.1-24.
- Tannant, D. D., and Regensburg, B., 2001. Guidelines for mine haul road design. School of Mining and

- Petroleum Engineering, *Department of Civil and Environmental Engineering*, University of Alberta.
- Thompson, R., & Visser, A. (2003). Mine haul road maintenance management systems. *Journal of the South African Institute of Mining and Metallurgy*, 103(5), 303-312.
- Thoresen, T. and Roper, R., 1996, Review and enhancement of vehicle operating cost models: assessment of non-urban evaluation models, *ARR 279*, ARRB Transport Research Ltd, Vermont South, Vic.
- Wang, J., and Yang, H., 2007. Model tests and numerical simulation on bearing capacity of bucket foundations in soft clay under vertical static and cyclic loads. *International Offshore and Polar Engineering Conference*, pp.1547-1552.
- Willenbrock, J. H., 1972. Estimating costs of earthwork via simulation. *Journal of the Construction Division*, 98(1), 49-60.
- Zhou J., Hall R.A., and Fowler G., and Huntingford K., 2006. Applications of Engineering Analysis to Improve Tire Management. *Proceedings of CIM*.

# WEBRTC TECHNOLOGY AS A SOLUTION FOR A WEB-BASED DISTRIBUTED SIMULATION

Stepan Kartak<sup>(a)</sup>, Antonin Kavicka<sup>(b)</sup>

<sup>(a)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice

<sup>(b)</sup> Faculty of Electrical Engineering and Informatics, University of Pardubice

<sup>(a)</sup> [stepan.kartak@student.upce.cz](mailto:stepan.kartak@student.upce.cz), <sup>(b)</sup> [antonin.kavicka@upce.cz](mailto:antonin.kavicka@upce.cz)

## ABSTRACT

The modern web browser is a runtime environment which aspires to replace original (native or desktop) applications. This article describes a new HTML5 technology called WebRTC, which enables a direct connection between browsers and which enables to perform a peer-to-peer network connection, which is suitable for the creation of a distributed simulation model. We compare this new technology with original web-based distributed simulation solutions – implemented using applets – and present one of the possible approaches to distributed simulation model creation.

Keywords: web-based simulation, WebRTC, discrete-event simulation, HTML5

## 1. INTRODUCTION

Web browsers are part of daily life today. During the past few years, web browsers have grown enormously; standards have been unified – especially JavaScript – and the present-day web browser succeeds as a platform for a wide range of applications which used to be implemented as so-called desktop applications, which are bound to the operation system, processor architecture, etc. The web browser overcomes these dependencies and represents an ideal multi-platform runtime environment, which, together with extended HTML5 support, represents minimum restrictions to the deployment of applications which could only be implemented as desktop ones in the past.

Since the second half of 2013, major web browsers have supported WebRTC (included in HTML5), which enables to initiate a peer-to-peer network connection between browsers (clients), and thus to perform a smooth distributed and decentralized simulation. It is this type of simulation that we elaborate on.

The article also covers available competing technologies, the technology and use of WebRTC, and describes a practical implementation of distributed simulation models running in the web browser.

The aim of the solution is not to compete with the existing HLA solutions, but to present new possibilities opened up by modern web browsers, and point out their advantages (as well as disadvantages). This article is a follow-up to a previous work (Kartak 2013), where a

web-based simulation was implemented using Java Applets (which was the only possible solution at that time). The article concludes with a simple comparison of the two solutions.

## 2. AVAILABLE TECHNOLOGIES

The aim of a web-based simulation is to create a browser ecosystem where the deployment of external elements (applets, extensions, etc.) is reduced to minimum.

Distributed compute nodes require a bi-directional communication between individual logical processes in the network. This purpose is best served using the peer-to-peer architecture (figure 1).

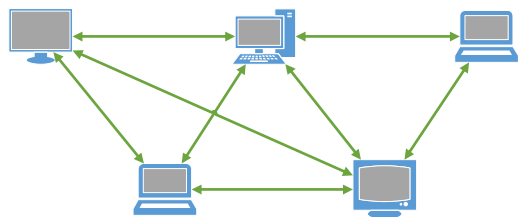


Figure 1: Peer-to-peer Network Topology

### 2.1. HTML Before Introducing HTML5

As mentioned in the previous paragraph, bi-directional communication is necessary. In the past, it was the absence of this functionality that did not allow for this type of connection among devices (typically, browser-browser or browser-server).

There were several solutions dealing with this issue. The simplest one, exclusively at the level of browsers, was accumulating client requests in a queue on a server, which was accessible to all clients involved in the communication, where the requests were waiting for the retrieval by the original target client (browser), see Figure 2.

This way, *de facto*, we pass from the peer-to-peer communication to the client-server one. It follows that this solution can be considered as an “emergency” one and thus not suitable in general. The use of a server within a distributed simulation may have its purpose; however, it is only complementary to the planned simulation topology. Using a server is purposeful for on-line gathering of information regarding the system behavior – statistics, animation output, etc.

Although it works, this solution brings several disadvantages (the major ones are listed below):

- It requires an efficient server (a network element which is virtually missing in the peer-to-peer network architecture).
- As a rule, it requires a low response rate of the client connection. In this case, requests are repeatedly sent to the server (requests repeated after several milliseconds). Most of the requests sent to the central server element are useless, yet they are necessary for the required low response rate of the whole system (in our case, a simulation). This considerably increases the communication traffic in the network.

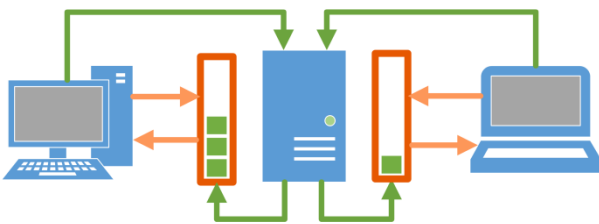


Figure 2: Message Queue For Clients On Server

## 2.2. The Use Of Applets

In the past, the only possibility to deal with the above-mentioned issue was the use of applets. In web browsers, it was the use of an external application (e.g. Java code running in Java Virtual Machine) in the context of a web page. (Byrne, Heavey and Byrne 2010)

All commonly available applets (Java Applet, Adobe Flash Player, Microsoft Silverlight) include features to perform the peer-to-peer connection; however, most of them (with the exception of Adobe Flash Player) are “process virtual machines”, which usually have direct access to the host computer. This is very dangerous, as applets can be loaded from any web page, and thus allow an unauthorized user to access the computer. These applets deal with this issue by using so-called policy files – files which contain a security policy definition (i.e. a definition of enabled and disabled features or operations and network access). In these files, it is necessary to explicitly allow client call from a specific server/client, or more generally defined groups of servers/clients often called “domains”. In addition, the above-mentioned must be allowed by the user. The communication is often blocked by a firewall or another security feature in the target computer or in the network.

The use of applets is easy due to the well-known languages (typically Java and Java Applet); however, in general it is not suitable because of a non-trivial use and network communication safety issues.

The most applicable out of the above-mentioned applet types is Java Applet; its safety policies enable a smooth bi-directional client-server communication (with the server from where the applet was loaded). This, of course, contradicts the notion of the peer-to-

peer connection, and does not bring any vital benefit when compared to the above-mentioned solution based solely on HTML.

All these applet-based solutions had to deal with connection safety issues (connection blocked by firewalls, etc.) and incompatibility among browsers, platforms, or operating systems. (Martin, Rajagopalan and Rubin 2013)

## 2.3. New HTML(5) Possibilities

Together with new technologies (often called HTML5), HTML allows all necessary communication in the network:

- Download and upload data (standard browser features even before HTML5).
- The WebSocket technology allows for performing real bi-directional client-server communication.
- WebRTC technology allows bi-directional communication directly between browsers – real peer-to-peer communication.

This solves the issues related to network communication for the needs of distributed simulation (see Table 1 for the overview of the availability in web browsers). For a detailed description of HTML5 network technologies, refer to Chapter 3.

Table 1: The HTML5 Network Technology Availability In Major Web Browsers

| Web Browser       | WebSocket | WebRTC    |
|-------------------|-----------|-----------|
| Chrome            | 14        | 23        |
| Firefox           | 6         | 22        |
| Internet Explorer | 10        | -         |
| Opera             | 12.10     | 22        |
| Safari            | 6         | -         |
| Available since   | May 2012  | July 2013 |

HTML5 brings further useful technologies which find their use in a web-based simulation:

- *Canvas*: Allows for (mainly vector) 2D drawing using JavaScript. A crucial disadvantage of Canvas is the necessity to always redraw the whole scene (there are some mostly “caching” techniques to minimize the problem to a certain degree).
- *SVG*: Allows for drawing (and animation) using a declarative HTML-like approach.

## 3. HTML5 NETWORK TECHNOLOGIES

### 3.1. WebSocket

WebSocket technology enables us to perform a network connection corresponding to the standard behavior of desktop applications, familiar to us for years. Network sockets establish a bi-directional connection between two applications (client 1 can contact client 2 and vice



versa). In terms of web applications, it means that a web page (client 1) opens a connection with a server (responding to client 2) and this connection is bi-directional – there is no longer a queue (see Figure 2), the server sends messages directly to the client via the opened connection.

This solution, of course, requires a server. The server side does not have to meet any special requirements; it only needs to follow the WebSocket protocol. We also have at our disposal a wide range of ready-made open-source tools, from PHP (e.g. the Ratchet library), via Python (e.g. the Tornado framework) to Java (e.g. TooTallNate) or C# (the Alchemy WebSockets library).

### 3.2. WebRTC

At first it is crucial to mention that the development of this technology is still in progress. According to W3C, WebRTC is in the “Working Draft” phase (as of September 10, 2013, see References), and the behavior of some features in different browsers may not be 100% correct, or different browsers have implemented these features in different ways. Most of these issues can be solved using a JavaScript solution, which overcomes the existing browser-specific differences (at present only minor differences among browsers).

This technology crucially enhances web browser capabilities in terms of network communication. It is possible to perform the peer-to-peer connection without a server (a server is only required to initiate the connection, which is, of course, standard for peer-to-peer communication). To solve routing issues when communicating with a client in a local network using NAT, WebRTC implements directly the use of the ICE (Interactive Connectivity Establishment) protocol. More information is provided in the following chapter 4.2.

WebRTC transfers data in two ways:

- **MediaStream** – used for audio and video streaming.
- **DataChannel** – used for text message transfer – it is this type that we used for the communication in the simulation.

DataChannel (uses the SCTP protocol for the communication between clients; this protocol allows for optional reliability settings). The reliable variant ensures that the message is delivered to the addressed clients (reliability corresponds to TCP); in the opposite case, the delivery is not guaranteed (or, only a limited number of callbacks is guaranteed; this type is similar to the UDP protocol).

## 4. NETWORK COMMUNICATION ISSUES IN A PUBLIC NETWORK

In this section, we elaborate on the two major and restrictive issues which need to be considered when communicating in the network (not only in the web browser).

### 4.1. Same-Origin Policy

Web browsers require following the same-origin policy. In practice, a problem arises when JavaScript needs to call the source – usually to download data – from a server which is located in a different domain than the one from where the page was loaded.

This way, browsers prevent the cross-site request forgery (CSRF or XSRF) attack, where the local script (loaded on a currently viewed web page) might call a script (or data containing executable code) which is not under the control of the application author (is located in another domain), and thus may present a potential threat.

We are likely to encounter this issue if we want to incorporate a logical process (performed using a web page) into a web-based distributed simulation, where the simulation is located in a different domain than the domain where e.g. the initialization server is located.

There are several solutions which can perform this type of communication. The simplest and most dangerous solution is an explicit disabling of the cross-origin request blocked (CORB) in a web browser. The best and most straightforward solution is the use of a server script which runs in a domain from where the web page was loaded. The web page then calls the script with the respective request for loading the source from another domain. On request, the script loads the requested source and sends it to the web page.

### 4.2. Peer-to-peer Communication via NAT

In the peer-to-peer network architecture, a problem in the connection between clients may arise when the clients are located behind NAT (Network Address Translation) routers. In this case, clients are in a local network and communicate with a public network via NAT routers, which serve as a public network gateway. What causes a problem in this case is the addressing of the client behind the NAT router, because from the public network perspective, the client is not visible (only the NAT router is visible), and is thus unreachable. In this case, direct initiation of a peer-to-peer connection is not possible.

The client connection problem can be solved using the ICE protocol (RFC 5245, see References).

The solution requires the use of an initialization server, which provides the connection.

The following example of the ICE protocol function has been simplified to demonstrate the scope of the problem (the performing of this connection algorithm requires 2 computers (clients) A and B which are, from the public network perspective, located behind a single NAT router at maximum (not port-forwarding) – see Figure 3):

1. Client A contacts a STUN server via port X. STUN sends back the number of port Y, from where it was contacted by client A. Based on the response (the number of the communication port of the client STUN request and the number of the port in the

STUN server response is the same), if the client finds out that the computer is accessible from the public network (has a public IP address – in this case, port  $X = Y$ ), there is no problem in the communication – the client can perform a connection and ICE protocol has finished.

2. If the client finds out, according to the STUN server response, that it is not publicly accessible (the port numbers of the request and of the response are not the same, i.e.  $X \neq Y$ ), client A sends the public port number Y via the initialization server (the port which the STUN server was able to contact) to client B, with which it can then communicate.
3. Steps 1 and 2 are also performed by client B.
4. Clients A and B initiate a connection via public IPs and ports, which they exchanged via the initialization server.
5. If direct communication between clients A and B is not possible (is blocked by the NAT router, firewall, etc.), it is the TURN server which can function as a mediating element between the two clients. The TURN server exchanges messages from one client to another. This does not correspond to the peer-to-peer architecture but to the client-server-client network topology. The use of the TURN server is not an ideal solution as there is high latency and server load; however, in this case, it is the only solution to connect clients A and B. This is, however, a rather exceptional state.

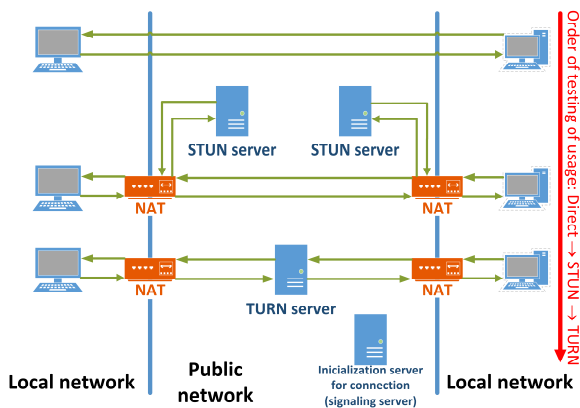


Figure 3: Visualisation Of ICE Protocol

From the above-mentioned algorithm it follows that the procedure is not rival. Importantly, both the STUN and the TURN servers are available as open-source, and there are libraries available for the common programming languages (Java, C#, C++, etc.) which support the ICE protocol. WebRTC is no exception, and when the STUN (eventually TURN) server address is provided, the web-browser itself handles the operation without the need of human interference. Some publicly available STUN and TURN servers also exist and can be used, making the facilitation of direct communication a very easy operation.

## 5. WEBRTC PRACTICAL APPLICATIONS

To be used in practice, WebRTC has to meet yet another requirement. Each peer-to-peer connection requires an initialization server, which serves as an “intermediary” while initiating a connection between clients. WebRTC is no exception; in the context of WebRTC, the initialization server is called a signaling server. In practice, the situation is more complex; another (by no means less important) role is the role of the client, which performs a connection using the SDP and ICE protocol.

Signaling server is a server via which clients exchange SDP (Session Description Protocol) messages, where they provide the information about the network connection via which they are to communicate. (It follows that a signaling server is not required if the SDP message exchange can be facilitated using another option). The communication via a signaling server can be facilitated using a wide range of options:

- The simplest solution is sending requests to a server, or eventually sending the required SDP information (see Chapter 2.1.).
- A more efficient option is the usage of WebSockets.
- A wide range of other options.

If the initialization process is successful, a bi-directional communication channel between clients (web browsers) is created, with the help of which we can exchange messages (or other data types supported by WebRTC – especially audio and video streaming).

We also have at our disposal a number of open-source signaling servers, which we can use and only deal with the logical process modeling.

## 6. IMPLEMENTATION

For testing and application purposes, a collection of programming tools and protocols was introduced to create a simple ecosystem with focus on reusability, simplicity and development speed. We elaborate on the implementation in this chapter.

### 6.1. Concept

The central idea was to create a set of (relatively general) logical processes, which can be integrated into the distributed model. The sample implementation concerned traffic, where logical processes represented:

- A road,
- A road with a turnoff,
- A road with a crossroads.

The three logical processes can be used multiple times in a single distributed model in arbitrary combinations and individual instance configurations of the used logical processes. The three logical processes are sufficient for the purpose of simulations of e.g. traffic infrastructure of a small town.

We elaborate on a web-based simulation, i.e. individual logical processes were implemented as web pages programmed in JavaScript. The communication among the logical processes was performed using the peer-to-peer connection via the HTML5 WebRTC technology.

## 6.2. Used Synchronization Algorithm

To synchronize logical processes, we used a conservative synchronization technique of sending null messages with a lookahead (*Chandy-Misra-Bryant Distributed Discrete-Event Simulation Algorithm*, Fujimoto 2000). The algorithm was modified to a version where null messages are only sent on request. The lookahead ensures that the simulation calculation proceeds forward in time – there is no risk of a deadlock. In our logical process creation concept, the logical process lookahead is not difficult to define. The simplest method to define the lookahead limit is registering the most short-time activities which occur in the logical process.

The algorithm can be described in the following way (the example of simulators SC1 and SC2, where SC1 receives messages from SC2; LVT = Local Virtual Time, Calendar = the event queue).

1. SC1 has no planned activities from SC2.
2. SC1 sends the LBTS activity of the “Request” type, which is labelled as a “service”; it also sends its own SC1.LVT.
3. SC1 is waiting.
4. SC2 receives the LBTS event; as the LBTS activity is labelled as a “service”, SC2 does not queue it, but executes it immediately after the current activity has been finished.
5. SC2 performs the LBTS activity: it calculates  $SC2.LBTS = SC2.LVT + SC2.Lookahead$ .
6. If  $SC1.LVT > SC2.LBTS$ , SC1 does not accept such LBTS answer and the sending on the SC2.LBTS answer is scheduled in SC2.Calendar to the SC1.LVT time.
7. Otherwise, the LBTS event of the “Response” type with the SC2.LBTS time is sent to SC1 (it is no longer labelled as a “service”).
8. SC2 continues its activities.
9. SC1 queues the LBTS event which was sent by SC2.
10. By this time, SC1 knows the lower limit of the SC2 time and can execute the planned events before the received LBTS message.

## 6.3. Administration interface

To facilitate the administration of the distributed models, an administration interface was created which runs as a common web application. It allows for a simple addition of individual logical processes and their integration into the simulation.

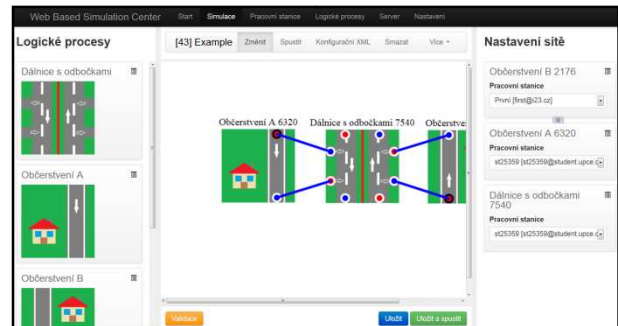


Figure 4: Administration Example

The simulation is created using the Drag & Drop method in a visual editor directly in an HTML page (implemented using the HTML canvas) and brings the editor a whole range of functions (basic features):

- A multiple use of individual logical processes (multiple instances of a single logical process type),
- An intuitive connection of logical process instances – logical processes have distinct input and output connection points. The connection distinguishes various types of received entities by logical processes.
- Individual logical process instances are configurable.
- Global configuration of the whole simulation is available.

## 6.4. A Software Library For Logical Process Implementation

To simplify the development, a software library (a set of classes and functions in JavaScript) was introduced which allows for a simple creation of logical processes and an implicit realization of logical process synchronization.

The library includes the implementation of:

- A simulation kernel,
- An activity prototype (a class from which activities can be created easily).
- A network connection using WebRTC,
- A synchronization mechanism (described in Chapter 6.2).
- Animation.

All was implemented using primarily the CoffeeScript language, which subsequently compiled to JavaScript.

CoffeeScript was used for faster and more transparent implementation than a comparable JavaScript solution. Both types of source files are at the user's disposal. The whole solution is opened; the encapsulation of functions and classes has been reduced to minimum to allow for prototyping (and especially inheritance). This does not represent a typical OOP approach; however, in its very nature, JavaScript was

designed differently. Although this solution may potentially be dangerous (the user can “rewrite” the code while the program is running), as a result it brings increased flexibility and a possibility to expand classes (especially as far as inheritance is concerned, some parts of the solutions are dependent on it).

As a result, the programmer is only required to implement activities which run on the basis of a logical process (eventually to create appropriate animation output). The rest is ready to use without any modifications or alterations.

### 6.5. Logical Process Communication Methods And Other Suggested Standards

If the logical process creator works with a library which has been designed specifically for this purpose (see the previous chapter), they do not have to solve the implementation of logical process communication or loading of configuration files at all.

It is of course possible to create a logical process quite independently of the above-mentioned software library. In this case, we have at our disposal a description of the communication between logical processes (*defacto* an internal communication protocol) and a description of configuration XML files using an XML schema.

The communication implements very easily. JavaScript implicitly solves a correct language coding of messages. Information is transferred in the JSON data format (a common means of data transfer via the Internet actively supported by JavaScript).

The configuration of the whole simulation is available in two XML files (can be processed using JavaScript):

- The logical process configuration, which describes especially the possibilities to connect with other logical processes and the required instance configuration. This file uses especially the administration for corresponding visual editor behavior.
- Configuration of the whole distributed simulation setup. The file contains a global simulation configuration and a configuration of all logical process instances and their interconnection.

### 6.6. A Sample Solution And Testing Logical Processes

The whole solution was tested on 3 logical processes representing:

- Highways with turnoffs,
- Two logical process types representing (highway) fastfood facilities.

The logical processes are accompanied with an animation, which provides visual information about the events within the logical process.

See an an example of an administration interface (Fig. 4) and used logical processes (Fig. 5).

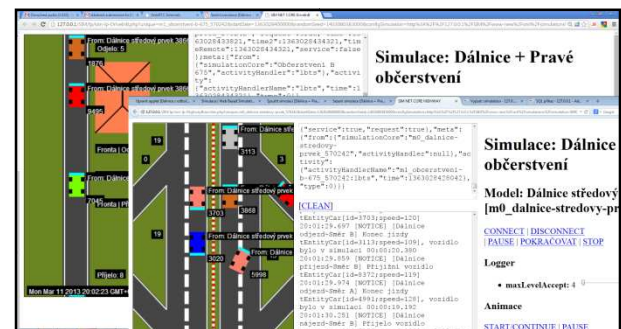


Figure 5: An Example Of Logical Processes Running In A Web Browser

As mentioned above, logical processes can be freely concatenated and allow for multiple usage. In the context of testing, the three logical processes were used in a distributed simulation which consisted of 20 logical processes. All worked smoothly and without issues.

### 6.7. Comparison Of Java Applet And JavaScript Solutions

As mentioned at the beginning, this work is a follow-up to a previous solution, where logical processes were also running in the web browser, yet using Java Applet. This allows us to compare the effort spent on both solutions as well as their results. The aim of this chapter is not to compare which of the two solutions, Java or JavaScript, is better (especially due to the fact that the two languages are quite different one from another); it is rather to point out which kinds of comparison the two solutions inspire.

As expected, the results for the same simulation scenarios in both realizations were equal.

A much more interesting comparison is that of the implementation of the same problem. Both source codes were written by the same author; we can thus claim that the style and the algorithm solution would be similar. The size of the JavaScript source code (realized solely by web technologies) represents approximately 70% of the size of the code in Java (used in the Java Applet). Provided that we do not take into account the used frameworks for the DOM manipulation and for the facilitation of canvas drawing (to a certain extent, both include the basic programming tool collection which is also provided by Java SDK), we realize that the JavaScript code is less than half the size of the Java code.

In slightly exaggerated terms, we can claim that a program which is half the size takes half the time to create and contains half of issues. Shorter code is faster to read and understand. Even from this perspective, the transfer from applets to web technologies is worth to consider, provided that it is feasible. And this article demonstrates that the transfer certainly is feasible.

The simulation speed was not measured, because in both cases, the animation was running in real time,

which itself decreased the speed so that the simulation could be observed by the user. When compared to the program itself, there was a significant time lag in the network communication. In the end, the speed of the code execution is not essential.

It has to be mentioned that JavaScript has numerous implementation drawbacks (as any other programming language, in fact); however, their description is beyond the scope of this article.

## 7. SUMMARY

The article introduced the reader into a web-based simulation and mainly into new HTML5 possibilities of web browsers, which provide opportunities for a web-based simulation using only a web browser, i.e. without complementary third-party software (applets, etc.). A web-based simulation provides the opportunity to create logical processes within the public network irrespective of the platform or processor architecture. A software platform, which is typically realized as a desktop application requiring installation, is to be replaced by the web browser, which is today widely available free of charge, and for almost any computer. The creation of a web-based distributed simulation faces only a limited number of issues (they usually concern network traffic safety policy); as opposed to desktop applications, the web browser brings a variety of ready-made features (in our case, the most important one is WebRTC), whose realization in common native applications is not trivial.

Using the web browser as a runtime environment, we can focus on the main aim, i.e. on the creation of distributed simulation models running in a web browser on any device connected to the network, from computers via cellular phones to e.g. smart TVs. This opens up an opportunity for a user-friendly, interactive simulation available for anybody at any time.

## REFERENCES

- Fujimoto, R.M., 2000. *Parallel and distribution simulation systems*. New York: Wiley.
- Kuhl, F., Dahmann, J., Weatherly, R., 2000. *Creating Computer Simulation Systems: An Introduction to the High Level Architecture*. Upper Saddle River, NJ: Prentice Hall.
- Tropper, C., 2002. *Parallel and distributed discrete event simulation*. New York: Nova Science Pub Inc.
- 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]
- Martin, D., Rajagopalan, S., Rubin, A., 2013. *Blocking Java Applets at the Firewall*. Available from: <http://avirubin.com/block.java.pdf> [accessed date June 2013]
- Hridel, J., Kartak, S., 2013. Web-based simulation in teaching. *Proceedings of The European Simulation and Modelling Conference 2013*, pp. 109–113. September 23–25, Lancaster, UK.
- Byrne, J., Heavey, C., Byrne, P.J., 2010. A review of Web-based simulation and supporting tools. *Simulation Modelling Practice and Theory* 18: 253–276
- Javor, A., Fur, A., 2012. Simulation on the Web with distributed models and intelligent agents. Available from: <http://sim.sagepub.com/content/88/9/1080> [accessed date June 2013]
- Kartak, S., 2013. *The software tool for configuring distributed simulation model using a web simulation*. Thesis. University of Pardubice.



# COMBINING DEVS AND MODEL-CHECKING: USING SYSTEMS MORPHISMS FOR INTEGRATING SIMULATION AND ANALYSIS IN MODEL ENGINEERING

Bernard P. Zeigler<sup>(a)</sup>, James J. Nutaro<sup>(b)</sup>

<sup>(a)</sup>RTSync Corp. and Arizona Center for Integrative Modeling and Simulation, AZ

<sup>(b)</sup> OakRidge National Laboratory, TN.

<sup>(a)</sup>[zeigler@rtsync.com](mailto:zeigler@rtsync.com), <sup>(b)</sup>[nutarojj@ornl.gov](mailto:nutarojj@ornl.gov).

## ABSTRACT

Our objectives here are to discuss the development of a formal framework that exploits the advantages of the Discrete Event System Specification (DEVS) formalism and builds upon recent extensive work on verification combining DEVS and model checking for hybrid systems. The mathematical concepts within the DEVS formalism encompass a broad class of systems that includes multi-agent discrete event components combined with continuous components such as timed automata, hybrid automata, and systems described by constrained differential equations. Moreover, DEVS offers the ability, via mathematical transformations called system morphisms, to map a system expressed in a formalism suitable for analysis (e.g., timed automata or hybrid automata) into the DEVS formalism for the purpose of simulation. Conversely, it is also possible to go from DEVS to formalism suitable for analysis for the purposes of model checking, symbolic extraction of test cases, reachability, among other analysis tasks. We give an example of application of these concepts and discuss the open opportunities for research in model engineering in this direction.

## 1. INTRODUCTION

Model checking, a well-known formal verification method, systematically explores the state space of a system model to check that states satisfy specified behavioral properties (Baier, and Joost-Pieter, 2008). Model checking methods encounter state space explosion in analyzing *autonomous* systems that require complex logical processes to perform complex decision making tasks. Moreover, because they are limited in their expressive capability to restricted logics, such methods must typically make stringent assumptions about physical components and environments. These assumptions and idealizations greatly reduce the methods' applicability to *cyber-physical systems* where the interplay of physical and computational elements is paramount. Finally, *cooperative* multi-agent systems raise the state space explosion exponentially through the cross product of their individual state spaces. In the absence of workable simulation approaches to enable

virtual testing, the only recourse for V&V of Cyber-physical Autonomous Cooperative (CACSoS) is to brute-force methods which are severely limited in the range of conditions they can test.

A key root cause of limitations in current V&V approaches to CACSoS is that they are not based on a general dynamic systems modeling and simulation framework. Such a framework should be capable of expressing the interaction of decision logic, discrete events, and continuous dynamics that are the hallmarks of such systems. We therefore propose that the Discrete Event System Specification (DEVS) formalism, as the computational basis for a general dynamic systems theory (Zeigler, Praehofer, and Kim 2000), provides a sound and practical foundation for enhancing existing V&V methods to address their limitations in addressing in CACSoS.

The value of Modeling and Simulation in defense and other applications is well-known (Shaffer 2012). A DEVS model is a system-theoretic concept specifying inputs, states, outputs, similar to a state machine, (Mittal and Risco 2012). Critically different however, is that it includes a time-advance function that enables it to represent discrete event systems, as well as hybrids with continuous components (Nutaro 2011) in a straightforward platform-neutral manner (Zeigler and Sarjoughian 2012). A recent thesis (Denil 2013) presents a multi-paradigm model-driven approach to design, verification and deployment of software intensive systems, another formulation of cyber-physical systems. It shows that DEVS provides excellent features for modeling such systems. The thesis provides a list of properties of DEVS and their mapping to properties of automotive software and systems – here viewed as instances of CACSoS:

- *Concurrency*: Multiple processors and communication links are concurrent in a CACSoS system. The semantics of DEVS coupled models supports concurrency by appropriate interleaving of the discrete-event behavior of individual sub-models.
- *Time*: Real-time performance is a crucial property of CACSoS embedded software. End-to-end latencies are part of the requirements for these



applications. The time advance function of an atomic DEVS model can be used to model latency.

- **Events:** Event-triggered and time-triggered architectures use triggers in the form of either external events or timing events to start certain pieces of functionality. DEVS implements reaction to events using the external transition functions.
- **Priorities:** Some real-time communication channels use priority-based and other mechanisms for arbitration. DEVS supports such arbitration by means of explicit specification of executable events from the set of simultaneous events.
- **Simulation of the physical parts of the system:** DEVS is a very general formalism and is able to include different other formalisms. This generality stems from the infinite possible states that DEVS allows to model and the (continuous) time elapse between the different state transitions. The hierarchical coupling techniques are used to integrate the different formalisms using DEVS as a common denominator.

## 2. BACKGROUND

The *Modeling and Simulation Framework* (MSF) (Zeigler 1976) presents entities and relationships of a model and its simulation as background for the

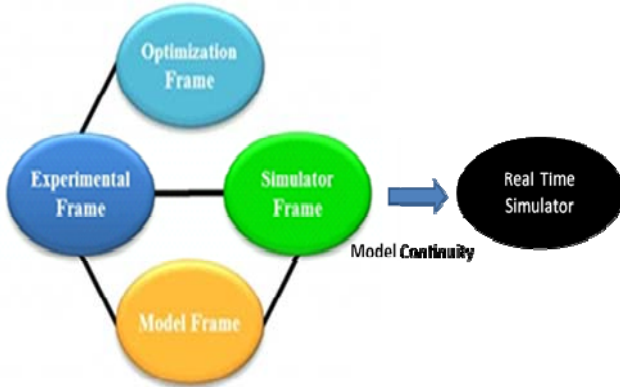


Figure 1. Model and Simulation Framework showing Model Continuity for SoS V&V

proposed work (Figure 1.)

The MSF separates models from simulators as entities that can be conceptually manipulated independently and then combined in a relation which defines correct simulation. The Experimental Frame defines a particular experimentation process, e.g., Latin hypercube sampling for yielding model outcome measurements in accordance with specific analysis objectives. Figure 1 depicts the notion of an *Optimization Frame* to supplement the MSF Experimental Frame, where the Optimization Frame directs search among the possible models for one or more that satisfy design space criteria, including those that minimize uncertainty about how well the implemented design will work. Figure 1 also

emphasizes the ability enabled by model continuity to transfer a simulation model from a logical-time simulator to a real-time simulator. In particular, DEVS models for autonomous control in CACSoS can be shifted without alteration (avoiding error-prone and tedious reprogramming) to interact with real environments in cooperative configurations after being verified in virtual environments (Zeigler and Sarjoughian 2012, Hu and Zeigler 2008.) The MSF underlies the DEVS Simulation Protocol which provides provably correct simulation execution of DEVS models thereby obviating time and state conflicts arising in simulation of multi-formalism models. There are numerous implementations of DEVS simulators (Nutaro 2011, Mittal and Risco 2012).

## 3. LIMITATIONS OF V&V METHODS APPLIED TO CACSoS

Linear Temporal Logic (LTL) and Computation Tree Logic (CTL) which are used for expressing desired behavior and model checking have been applied to the development of vehicle routing and road monitoring in multi-UAV systems (Karaman and Frazzoli, 2008, Sirigineedi, et al/, 2010). Humphrey (2013) explored the use of LTL, the SPIN model checker, and the modeling language PROMELA (Gerth 1997, Baier, and Joost-Pieter, 2008, Holzmann 2004), for high-level design and verification in UAV related applications, reporting some success while suggesting limitations and needed extensions. Table 1 shows three UAV related cases she discussed.

Table 1. Example Applications of model checking to CACSoS

|                                                                                                                                                      |
|------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Model</b> A centralized UAV cooperator controller that coordinates the actions of multiple UAVs performing a monitoring task                      |
| <b>Focus of Model Checking:</b> Assuring that All sensors are eventually visited.                                                                    |
| <b>Sample Simplifying Assumptions</b> Communication between UAVs and sensors can only occur when in the same location and is error free.             |
| <b>Model</b> A leader election protocol for a decentralized system of unattended ground sensors sending estimates of an intruder's position to a UAV |
| <b>Focus of Model Checking</b> At least one leader exists at every time step.                                                                        |
| <b>Sample Simplifying Assumptions</b> The sensors all use sampling epochs of the same length enabling a single time step for time advance.           |
| <b>Model</b> Verification of high level UAV mission plans for a scenario in which multiple UAVs must be used to safely escort an asset across a      |

|                                                                                                                                       |
|---------------------------------------------------------------------------------------------------------------------------------------|
| road network.                                                                                                                         |
| <b>Focus of Model Checking</b> The path travelled by the asset is safe, i.e., all road segments in the path have been scanned by UAV. |
| <b>Sample Simplifying Assumptions</b> UAVs and VIP were assumed to travel at the same speed.                                          |

In each case, the focus of model is shown along with a simplifying assumption. Because they are oriented to verification, model checking tools tend to lack many functions that exist in DEVS environments and require abstractions that fit the tools' operation. This forces an abstraction of the real system that on the one hand enables the modeler to better understand the model, and on the other hand entails numerous assumptions to enable the model checker to verify the focal requirement. Despite these drastic simplifications, state space explosion prevents employing more than a handful of UAVs and sensors.

Zervoudakis et al (2013) write that "Research in model checking has focused on enhancing its efficiency and scalability thereby enabling model builders to verify larger, more elaborate models. Popular model checkers tend to support low-level modeling languages that require intricate models to represent even the simplest systems. For example, PROMELA, the language of the model checker SPIN, is essentially a dialect of the low-level programming language C. Another example is the modeling language used by the probabilistic model checker PRISM, whose lack of control structures forces model builders to pollute model components with counter variables that explicitly encode the components' state transitions." These authors show that mapping of domain knowledge, assuming it exists in the right form, can be used to reduce the manual and error-prone encoding of state transitions at relatively low levels of abstraction. Here we propose to develop such mappings using the domain knowledge contained within simulation models and their ontological representations within the DEVS-based Modeling and Simulation Framework.

#### 4. DEVS SUPPORT FOR CACSoS

Cyber-physical systems are real-time hybrid systems, i.e., include both discrete and continuous dynamics, which, as earlier indicated, are well represented within the DEVS-based MSF. Typically such a system is described by a state consisting of both discrete control phases and continuous variables (Nutaro 2008., Saadawi, Wainer, and Moallemi 2012). developed a methodology that combines DEVS and Timed Automata (TA) (Bengtsson and Yi 2004. Henzinger., 1997, Alur 1995, Courcoubetis et al. 1995) to allow the designer to model, simulate, verify, and deploy real-time hybrid systems. This is achieved by guaranteeing

the correctness of the model with a methodology that verifies DEVS models with TA model-checking techniques and tools. Under model continuity (Figure 1) the verified DEVS models are then made executable on the target platform, thus eliminating the risk of introducing errors in the final system implementation. TA provides a solid theory and algorithms for model checking, and many existing tools implement these algorithms (e.g., UPPAAL). The combined DEVS/TA methodology deals with RTA-DEVS - a restriction of DEVS to Rational (a subset of real) Time Advance values. For this subclass the methodology provides automated mappings to TA's abstract formal system specification that is verifiable by *decidable* model checking. In this methodology, if UPPAAL (or other model checker) faces a problem of state explosion, and no answers can be obtained in reasonable time, the user can use model checking on an abstraction of the system while employing DEVS-based simulations to empirically check out properties not included in the formal analysis. In particular, the methodology applies to hybrid systems whose continuous components are expressed using differential equations solved using Quantized State System (QSS) integration. Concurrently, a burgeoning literature is developing on the use of QSS, a class of DEVS models, to efficiently and accurately model such systems, for example using multicore processors.

While multi-agent-based simulation (ABS) is well established using DEVS (e.g. (Perez et al. 2010), time-step scheduling is still used in classic ABS models (e.g., Repast. 2009) However, Zhang et. al. (2014) developed a DEVS simulation model which is significantly more efficient than the Repast ABS model (350 times faster for 10000 agents) while keeping high model spatial fidelity and the same agent cognitive capability, collision avoidance, and low agent-to-agent communication cost.

Work on non-DEVS model checking for hybrid systems includes that based on timed automata (Henzinger 1997, Alur 1995, Courcoubetis et al. 1995), abstraction and simplification of systems (Chauhan et al. 2002, Clarke, et al. 2003, Long 1993), and statistical model verification (Younes et al. 2006).

#### 5. INTEGRATING DEVS AND NON-DEVS VERIFICATION METHODS

As discussed earlier, several DEVS methodologies have been developed which incorporate non-DEVS verification methods (Zeigler and Sarjoughian 2012, Hu X. and Zeigler 2008.) These methodologies attempt to employ DEVS to enable loosening the simplifying assumptions typically made by non-simulation models. For example, the Sample Simplifying Assumptions in Table 1 suggest that such simplifications include those concerning perfect communication among component systems occurring when they are in exactly specified locations, that time advances for all components using the same fixed time step, and that speeds of vehicles are

constant or change instantaneously neglecting realistic accelerations. However, so far, such methodologies have not provided a general approach to combining simulation and verification using available system theoretic concepts.

Our objective here is to develop and employ system morphisms and model transformations to integrate the various types of models to be included in a general DEVS-based framework for verification in model engineering. Model transformations are a key means of converting between different model types and must preserve desired aspects of structure and behavior to qualify as system morphisms. Our approach is to define system morphisms for such transformations and to prove these morphisms are mathematically correct using existing theory of system morphisms (Zeigler, B. P., 1976). This will enable us eventually to automate verification of properties for complex simulation models as opposed to simplified models developed specifically for model checking. Then we will explore algorithmic approaches to automate the construction of these types of systems mappings. Such automation is necessary to create a practical tool for engineering systems of systems.

Figure 2 compares the methodology of (Saadawi, Wainer, Moallemi 2012.) which provides automated transformations from RTA-DEVS to TA enabling tractable model-checking using UPPAAL (Behrman 2004) with the approach we discuss here. The RTA-DEVS approach seeks to verify DEVS models by transforming them into a subset of TA that can be verified using UPPAAL. To do this, it must

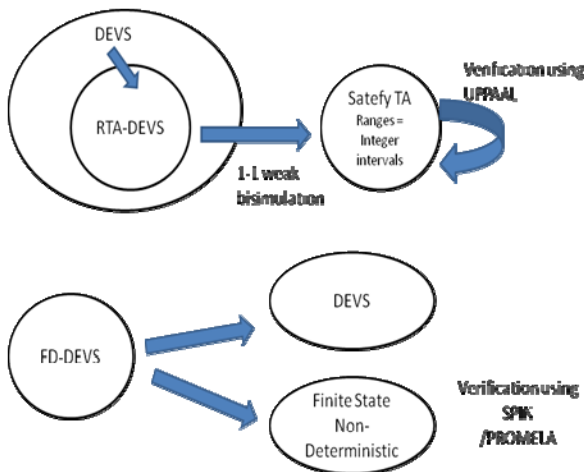


Figure 2 Comparing DEVS-based Verification Approaches

appropriately limit the class of DEVS models to a subclass that can be mapped to the input class of TA for UPPAAL via 1-1 weak bi-simulation with the Safety TA subclass of TA. Weak bi-simulation will be shown to be a system morphism. The mappings are not automated.

In contrast, as shown in Figure 2, our approach is to start with the FD-DEVS (Zeigler and Sarjoughian 2012) subclass of DEVS models having finite sets of states, inputs, and outputs and whose construction is supported by MS4 Me. Then two mappings are defined:

1. Map FD-DEVS models into non-deterministic automata that are subject to model checking using SPIN/PROMELA.
2. Elaborate FD-DEVS models into full-fledged DEVS models that can be simulated to obtain behavior of interest and possibly to discover unexpected or emergent behaviors via simulation.

To illustrate the general idea, we will construct and illustrative, mathematically verifiable transformations from DEVS models to model checking models and vice versa.

### 5.1. Example of System Morphism between PROMELA and DEVS

For an example of this approach, consider the alternating bit protocol introduced by (Bartlett, Scantlebury and Wilkinson, 1969) for implementing full-duplex communications over half-duplex communication lines. This protocol is illustrated in Figure 3. It has been used to illustrate fundamental elements and analysis capabilities of the PROMELA language by proving that the protocol operates correctly (see the SPIN manual, Gerth 1997); that is, that "Every message fetched by A is received error-free at least once and accepted at most once by B". It is apparent from the figure that this PROMELA model is a finite state automaton, and all finite state automata are instances of DEVS models that have a fixed time advance (see (Zeigler, Praehofer, & Kim. 2000)). Hence,

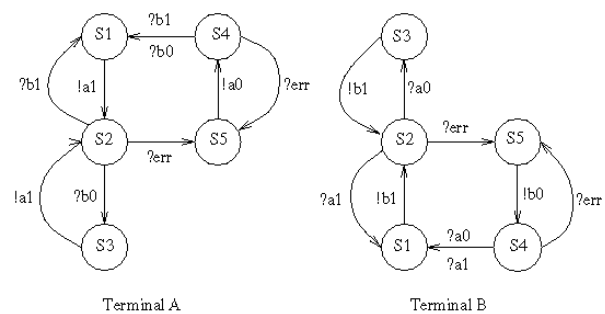


Figure 3. Model for automatic verification of the

alternating bit protocol can be built in a simple way. We do this by adding two pieces of information to the PROMELA description: the time  $T$  to transmit a bit and the probability  $p$  of an error. One important use of this DEVS model is to answer questions about performance of the protocol. Conversely, we may map any instance of this DEVS model with parameters  $p$  and  $T$  onto a PROMELA model by abstracting away the specific probability distribution and stating only that a bit may arrive or not arrive.

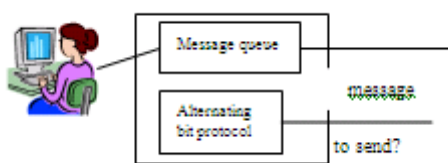
Table 2 shows the results of a simulation study, using the DEVS model sketched above, to discover how the bit rate of the protocol varies as a function of  $T$  and  $p$ . This table was constructed with the DEVS simulation model by sweeping over a range of values for  $T$  and  $p$ , and for each combination recording the bits per second that could be exchanged by the system assuming 100% utilization of the communication channel. These simulations provide important, systems level performance metrics that cannot be obtained via a query of the PROMELLA model. At the same time, we may be certain that the simulation model preserves the formal properties that we have proven about the protocol by use of the PROMELLA model. The capability to construct performance studies, like this bit rate study, using a simulation model derived directly from the formal verification model illustrates the power of the proposed approach for engineering complex systems.

**Table 2. Performance study of the alternating bit protocol**

| $T$<br>(seconds) | $p$  | Bits per<br>second | $T$<br>(seconds) | $p$  | Bits per<br>second |
|------------------|------|--------------------|------------------|------|--------------------|
| 1E-9             | 0    | 5.0E8              | 1E-6             | 0    | 5.0E5              |
|                  | 0.25 | 3.8E8              |                  | 0.25 | 3.8E5              |
|                  | 0.5  | 2.5E8              |                  | 0.5  | 2.5E5              |

## 5.2. Extending the Range of Verification Models with Simulation Models

Transformations between simulation models and verification models can also facilitate the reuse of model components throughout the lifecycle of an M&S



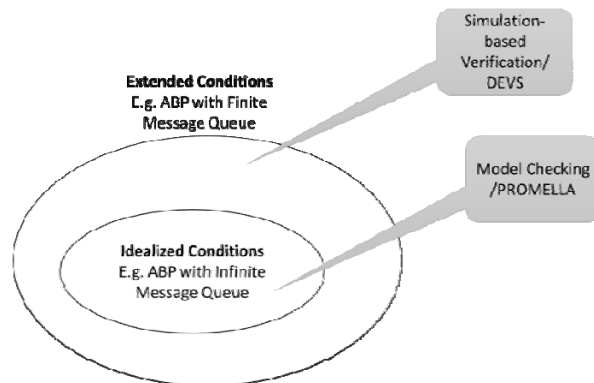
*Figure 4 Message Layer addition to ABP*

system, and this reuse can have the important effect of revealing implicit assumptions in proofs constructed for the verification model. For instance, it is natural to use the alternating bit protocol as a media access control layer within a more comprehensive network simulation. This more comprehensive model could have a sender and receiver each with two components. This is illustrated in the Figure 4. The first component is our DEVS model of the alternating bit protocol. The second component sends and receives messages, rather than just bits, and queues messages that are pending transmission. This second component appends to each message the bit that it receives from the MAC layer, and sends to the MAC layer the first bit in each message received from the network or an error indicator, as

appropriate. Transmissions in the upper layer occur at the instant that a bit is received from the MAC layer below. Central to understanding the behavior of this model is its queue capacity, rate of requests to send messages, and bit rate.

### 5.2.1. Alternating Bit Protocol with Infinite Queue

Let us first consider a combination of these that may be reduced to a slightly more complex version of the verification model for the alternating bit protocol. To obtain this model we reduce the queue to two states: empty and occupied. The former indicates no messages waiting for transmission and the latter indicates a message waiting to be transmitted. Adding these states to the transmitter increases the size of the verification model from 25 states to 50 states, and we may prove for this larger model that every message transmitted is received once and only once. This proof is significant because it reflects the intended, but idealized, behavior



*Figure 5 Relation between Verification and Simulation*

of the system. If, for instance, this assertion could not be proved then the design of the system should be reconsidered before moving to other forms of testing.

### 5.2.2. Alternating Bit Protocol with Finite Queue

However, this proof does not indicate how the system will behave over its entire range of structural variations and realizable behaviors. In particular, we may encounter a case where the queue's capacity is finite. In this case, the bit rate and rate of requests to send messages may be such that at certain points in time the queue's capacity is exceeded as can be predicted by simple queuing theory (Jain 1991). The consequence of this will be messages that are lost: a clear violation of the above proof! In this simple example, the cause of this violation is obvious. The verification model assumes that the queue never reaches its capacity, and so the proof implicitly assumes a restricted range of values for the queue capacity, rate of transmission requests, and bit rate.

## 6. COMBINING SIMULATION AND FORMAL VERIFICATION

These types of implicit assumptions can be difficult to identify in a large model, and simulation offers an



opportunity to explore the system's parameter space and identify boundaries beyond which any particular proof fails to hold.

As illustrated in Figure 5, the combination of simulation and formal verification gives a much more powerful capability to test designs than can be achieved with either alone. In a design process that incorporates both types of analysis, verification models can be used to obtain absolute answers concerning system behavior under idealized conditions. Failures in this verification stage clearly indicate a need to find and correct fundamental flaws in the system design. On the other hand, a successfully verified model can be formally extended into a simulation model for which the verification model is a homomorphic simplification. Hence, the simulation model retains the properties that were verified with the simpler model, and then can be used to explore scenarios that are necessarily outside the scope of formal verification. In some cases, other simplifications of a full-fledged simulation model can be applied (for example, as mentioned above, queueing theory can predict the probability of a finite capacity queue being exceeded.) However, in general, simulation models can incorporate the complexity needed to deal with real systems using the base model concept. The traceability between these two types of models, which is obtained by the application for formal system morphisms, is central to the success of this two tiered approach to testing.

## 7. FUTURE RESEARCH

Future research must develop robust transformation techniques that can be shown to be systems morphisms of suitable types. This will ensure consistency between the results of verification of emergent foreseen behaviors using analytical techniques and the discovery of emergent unforeseen behaviors through dynamic simulations. This will require formulating a meta-modeling approach that will lead to a multi-step verification process that can handle the dynamical complexity of CACSoS models. We recognize that the most expeditious way to develop an inclusive framework is to build on existing methods and software tools to the extent possible. The multi-step verification process will help to manage and cross-check results obtained from the various analytical and simulation methods, as well as from integrated methods that are supported by the proposed framework. This approach helps to deal with the complexities of CACSoS by enabling more robust designs and a more thorough and organized simulation and verification process. These complexities can be addressed with the help of the theory of M&S (Zeigler 1976) to develop methods for hierarchical decomposition and simplification as essential tools within the emerging field of model engineering.

## REFERENCES

- Alur, R., Courcoubetis, C., Halbwachs, N., Henzinger, T.A., Ho, P.H., Nicollin, X., Olivero, A., Sifakis, and J., Yovine, S. 1995, The algorithmic analysis of hybrid systems. *Theoretical Computer Science* 138(1), 3–34
- Baier, C. and Joost-Pieter, K. 2008, *Principles of Model Checking*. The 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, 260-265,
- Behrmann, G., David A., Larsen K. 2004. "A Tutorial on UPPAAL". *Proceedings of the 4th International School on Formal Methods for the Design of Computer, Communication, and Software Systems*. LNCS 3185.
- Bengtsson, J. and Yi W.. 2004. "Timed Automata: Semantics, Algorithms and Tools". *Lectures on Concurrency and Petri Nets*, 3098.
- Chauhan, P., Clarke, E., Kukula, J., Sapra, S., Veith, H., and Wang, D., 2002, Automated abstraction refinement for model checking large state spaces using SAT based conflict analysis. In: Aagaard, M.D., O'Leary, J.W. (eds.) *FMCAD 2002*. LNCS, vol. 2517, pp. 33–51. Springer, Heidelberg
- Clarke, E., Grumberg, O., Jha, S., Lu, Y., and Veith, H., 2003, Counterexample-guided abstraction refinement for symbolic model checking. *Journal of the ACM (JACM)* 50(5), 752–794()
- Denil, J, 2013 *Design, Verification and Deployment of Software Intensive Systems: A Multi-Paradigm Modelling Approach*, Ph. D. Dissertation, University of Antwerp.
- Gerth, R., 1997, Concise PROMELA reference (), <http://SPINroot.com/SPIN/Man/Quick.html>
- Henzinger, T.A., Ho, P.H., and Wong-Toi, H. 1997, HyTech: A model checker for hybrid systems. *International Journal on Software Tools for Technology Transfer (STTT)* 1(1), 110–122 ()
- Holzmann, G.J., 2004, *The SPIN Model Checker: Primer and Reference Manual*. Addison Wesley Publishing Company ()
- Hu X. and Zeigler B. P., 2005. A Simulation-Based Virtual Environment to Study Cooperative Robotic Systems, *Integrated Computer-Aided Engineering (ICAE)*, 12:4, pp. 353-367,
- Humphrey, L. R. 2013, *Model Checking for Verification in UAV Cooperative Control Applications Recent Advances in Research on Unmanned Aerial Vehicles*, *Lecture Notes in Control and Information Sciences* Volume 444, 2013, pp 69-117
- Jain R. K. 1991, *The Art of Computer Systems Performance Analysis: Techniques For Experimental Design, Measurement, Simulation, And Modeling* Hardcover: 685 pages Wiley.
- Karaman, S. and Frazzoli, E., 2008 ,Vehicle routing with linear temporal logic specifications: Applicationsto multi-UAV mission planning. In:

Proceedings of the AIAA Conference on Guidance, Navigation, and Control

- Long, D.E. 1993, Model Checking, Abstraction, and Compositional Verification. PhD thesis, Carnegie Mellon University
- Mital, S., Risco Martin José L 2012, Netcentric System of Systems Engineering with DEVS Unified Process, CRC Press; 1 edition
- Nutaro J. 2008. On constructing optimistic simulation algorithms for the discrete event system specification, ACM Transactions on Modeling and Computer Simulation, 19(1), pp. 1-21.
- Nutaro J., 2011. Building Software for Simulation: Theory and Algorithms with applications in C++. Wiley.
- Perez E, Ntairo L, Bailey C and McCormack P , 2010, Modeling and simulation of nuclear medicine patient service management in DEVS. Simulation-Transactions of the Society for Modeling and Simulation International 86(8-9): 481-501.
- Repast, 2009. Repast home page. Available at <http://repast.sourceforge.net>
- Saadawi, H, Wainer, G, Moallemi, M.. 2012. "Principles of Models Verification for Real-Time embedded Applications". *Real-Time Simulation Technologies: Principles, Methodologies, and Applications*. K. Popovici, P. Mosterman Eds. Taylor and Francis. CRC Press. 2012.
- Shaffer, Alan R 2012 The Value of Modeling and Simulation for the Department of Defense M&S Journal pp 2-3
- Sirigineedi, G., Tsourdos, A., White, B., Zbikowski, and R. Kripke, 2010, Modelling and model checking of a multiple UAV system monitoring road network. In: Proceedings of the AIAA Guidance, Navigation, and Control Conference.
- Younes, H.L.S., Kwiatkowska, M., Norman, G., and Parker, D. 2006, Numerical vs. statistical probabilistic model checking. International Journal on Software Tools for Technology Transfer (STTT) 8(3), 216-228
- Zeigler, B. P. and Sarjoughian, H. S. 2012, Guide to Modeling and Simulation of Systems of Systems Springer; pp. 393
- Zeigler, B. P., Praehofer, H., & Kim, T. G., 2000, *Theory of Modeling and Simulation* (2nd ed.). Academic Press.
- Zeigler, B. P., 1976. *Theory of Modeling and Simulation* (1<sup>st</sup> ed.). Academic Press.
- Zervoudakis, F., Rosenblumy D. S., Elbaumz, S., and Finkelstein, A, 2013 Cascading Verification: An Integrated Method for Domain-Specific Model Checking, ESEC/FSE 2013, Saint Petersburg, Russia
- Zhang, B, Chan, W K V, and S V Ukkusuri, 2014, On the modelling of transportation evacuation: an agent-based discrete-event hybrid-space approach *Journal of Simulation* advance online publication 21 February 2014

## AUTHORS BIOGRAPHY

**Bernard P Zeigler**, Chief Scientist of RTSync Corp., and Emeritus Professor from Arizona Center for Integrative Modeling and Simulation is internationally known for his seminal contributions in M&S theory, and has published several books including *Theory of Modeling and Simulation*, IEEE named him a Fellow of the IEEE for his invention of the Discrete Event System Specification (DEVS).

**James Nutaro** is Senior Research Staff at Oak Ridge National Laboratory. He has extensive experience in M&S and systems modeling in both defense and commercial domains. He has applied M&S techniques for design, analysis, and testing in diverse enterprises including missile systems, space systems, communications, uranium processing, electrical power, disease processes, and high performance computing technology. He is also the author of *Building Software for Simulation: Theory and Algorithms, with Applications in C++*.



# TWO STAGE SIMULATION USE IN PROJECT VERIFICATION AND VALIDATION

Egils Ginters<sup>(a)</sup>, Artis Aizstrauts<sup>(b)</sup>, Miquel-Angel Piera Eroles<sup>(c)</sup>, Roman Buil<sup>(d)</sup>

<sup>(a,b)</sup>Sociotechnical Systems Engineering Institute of Vidzeme University of Applied Sciences, Cesu street 4, Valmiera LV-4201, Latvia

<sup>(c,d)</sup>Dept. de Telecomunicació i Enginyeria de Sistemes, Universitat Autònoma de Barcelona, Carrer dels Emprius, 2 08202 Sabadell-Barcelona, Spain

<sup>(a)</sup>[egils.ginters@va.lv](mailto:egils.ginters@va.lv), <sup>(b)</sup>[artis.aizstrauts@va.lv](mailto:artis.aizstrauts@va.lv), <sup>(c)</sup>[miquelangel.piera@uab.cat](mailto:miquelangel.piera@uab.cat), <sup>(d)</sup>[roman.buil@uab.cat](mailto:roman.buil@uab.cat)

## ABSTRACT

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

If the developed policy decision-making support system does not correspond to the requirements of the decision maker, funding will have been spent pointlessly. The basic task of the sociotechnical approach is to bring together the wishes of the decision maker with the technical possibilities of the developer, thereby promoting project clearness and giving the beneficiary an opportunity to validate the offered solutions and establish conceptual mistakes in the early stages of the software designing.

The article dealt with two step simulation use based on Zagreb Open Green Park application development.

Keywords: simulation, simulators architecture, policy modeling, agent-based models

## 1. INTRODUCTION

The development of suitable support tools for the previously mentioned requirements is rather complicated because a fundamentally important factor is product functionality and algorithm conformity to objective requirements e.g. political decision quality assurance. Political decision-making usually is implemented in the framework of a classic sociotechnical system where respect has to be shown for both technical and social aspects.

Each project has to be transparent otherwise the user cannot take part in the design, and the expected result mostly will not be acceptable to the user. Result visualization also has to correspond to guidelines accepted within the user's professional field. The simulator has to maintain a beneficiary's alphabet, which is a set of terms that the beneficiary encounters during his/her daily professional activities, to ensure user-friendly visualization. The simulator has to be open so it often can be integrated into goal system software. Policy decision makers i.e. policy, tourism, economy

and other specialists could create simulation models, but then the design syntax have to be understandable by people without specific programming knowledge.

To enhance the use of simulation models by politicians for decision-making process such as policy design, it is essential to verify and validate (V&V) the simulation model (Sargent 1992, Burton 2003). While verification concerns whether the model is working as the modeller team expects it to, validation concerns whether the simulation is a good model of the target. Roughly speaking, a model is considered valid for a particular goal, if the results that can be obtained from the simulation have the same statistical properties and patterns as those obtained from the real system. However, in the area of sociotechnical system there are several aspects to be considered when comparing simulation results with a sample of the real system:

- Most social systems are characterized by its stochastic behaviour. Individual decisions strongly depend on human behaviour which is characterized by a diversity of options in front a particular choice problem, selecting a different choice even under similar circumstances. Thus, the simulation model considers also random factors to represent the stochastic process, which somehow justify a lack of rigorous correspondence between the results of a simulation run with respect to a sample. Comparisons between real data and simulation data usually is carried by statistical methods such as hypothesis test, which is a quite difficult problem to avoid accepting a false hypothesis or rejecting a true one, based always on modeller subjective considerations.
- Some simulation models are quite sensitive to the initial conditions dealing with drastic different results. Meanwhile the real system behaves in the same way, so the dynamics are highly dependent on the environmental conditions, the sensitivity of the model to the initial conditions is not a problem for explanatory targets validation, but it constraints its use for predictability purposes.

- Some social dynamics are quite dependent on time varying scenarios such as for example those urban policies which depend on seasonal weather conditions, or economic cycles. Consider the planning activities in a public green park for integration of autistic people with the neighbourhoods: affluence to the park and duration of the stance depends considerably on the weather conditions. Thus, even if the dynamics formalized is correct for certain boundaries conditions, they can lead to wrong results when the model is used for predictive purposes if the real scenario doesn't fulfil some of the boundary conditions.

To create trust and increase the credibility of the model and the simulation results delivered, it is essential to deal with a validation approach in which non-simulation trained end-users (ie. practitioners) could feel comfortable with the computer experimentation technique and trust the simulation model.

In this paper it is presented a two stage simulation framework, based on the model plausibility notion, also known as “conceptual validity” or “face validity”, which expresses the conformity of the model with a priori knowledge about the process.

## 2. CONCEPT

The conception about simulator or goal system functionality will always differ between the policy decision maker and the developer because the decision maker is guided by specific functionality requirements, nuances and particular conditions only known to him, whereas the developer places importance on technical solutions, development time and potential costs. The complexity of simulators to be developed increases not only from a technical (territorially distributed systems, cloud computing, Future Internet architectures, semantic search etc.), but also from a functional standpoint (decision-making algorithms; fusion, mining and useable visualisation of complex data). This significantly complicates the development of suitable simulation software.

To achieve successful designing results it is reasonable to split goal system design in at least two steps: conceptual modelling and implementation. In each step simulation can be used as an active verification and validation method (see Figure 1) of provided ideas.

Assessment of model plausibility is tightly related to expert judgment of whether the model is good or not. The level of plausibility, or better said the expert opinion about it, is basically related to two features of the model:

- The first one considers the question whether the model “looks logical”. This question concerns characteristics of the concept simulator.

- The second one is related to the question whether the model “behaves logically”. This part concerns assessment of the reaction of the model outputs (dynamics, shape, etc.) to typical events (scenarios) on the inputs. If the model in different situation reacts in accordance with expectations of the experts, then the confidence about its validity is increased.

Concept simulator (V&V-1) is aimed to provided concept verification and validation to compare different potential possibilities of bottlenecks, conflicts and to determine border conditions. This first preliminary step is carried out to understand the trend instead of precise quantitative solution, because it is necessary to decide is it reasonable to continue the project if the forecasts are unpromising.

The V&V-1 simulation results are input for the second implementation stage. However also designing results are verified and validated on Implementation simulator (V&V-2). These simulations are oriented to detail testing of some preliminary results obtained, before launching design of real more expensive prototypes.

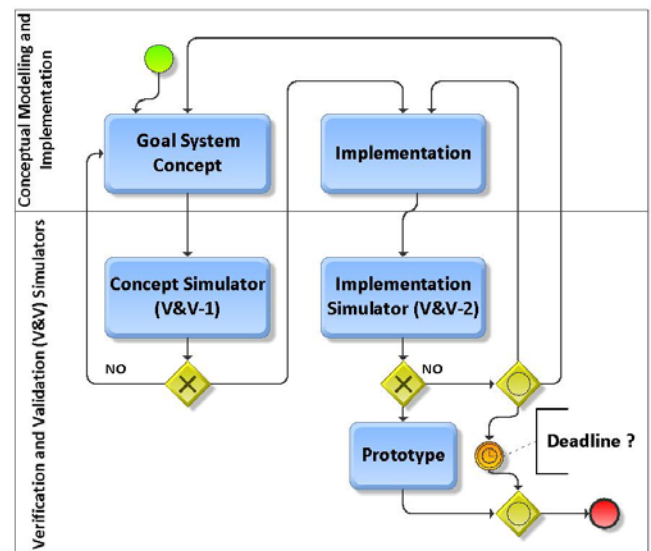


Figure 1: Two Stage Simulation BPMN2 Diagram of Project Verification and Validation

## 3. ZAGREB OPEN GREEN PARK SIMULATOR

One distinctive application of the above-mentioned approach is the FP7 FUPOL project No. 287119 “Future Policy Modeling” case related with designing of the Zagreb Open Green Park simulator (Ginters and Aizstraus et.al. 2013; Piera and Buil et. al. 2013).

The territory of Zagreb Open Green Park is approximately 20,000 square meters. Autism Centre patients and visitors from other neighbourhoods are expected to visit the park. Zagreb municipality has statistical forecasts about the possible number of visitors on weekdays and weekends.

Zagreb Open Green Park simulator consists of two parts and is based on multi-agent simulation (MAS) designed in Repast Symphony and GUI implemented in Java environment (see Figure 2).

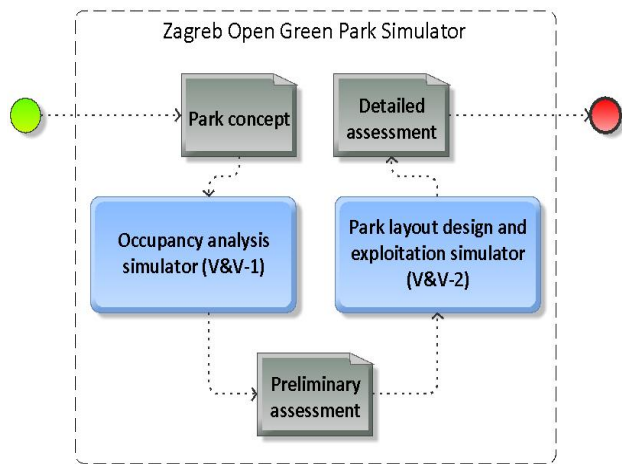


Figure 2: Zagreb Open Green Park Simulator Architecture

Despite, validation of multi-agent models can quickly become intractable (Batty et al., 2003), one of

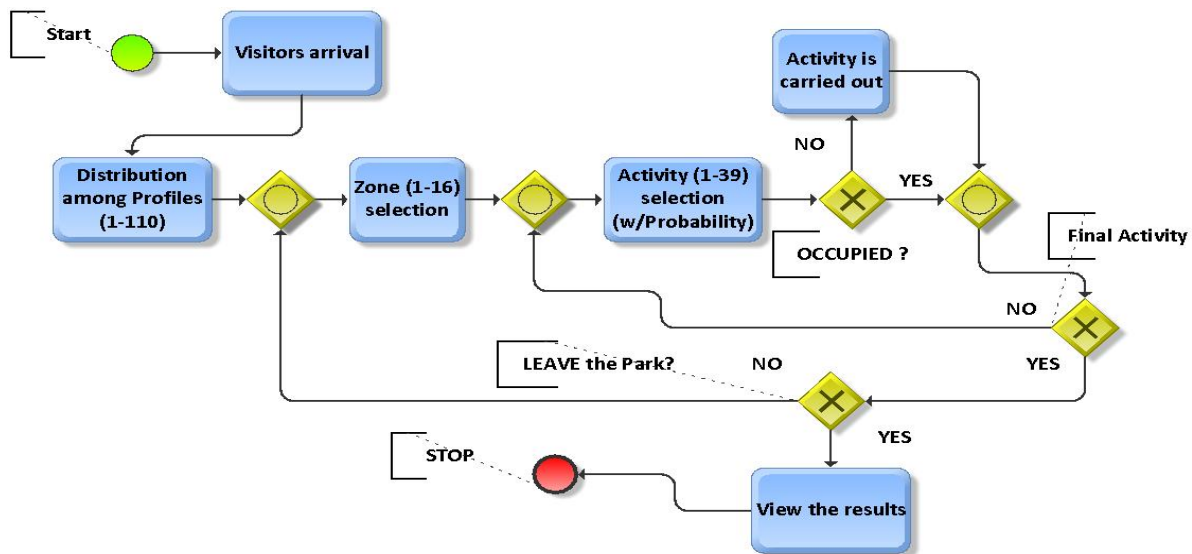


Figure 3: Zagreb Open Green Park Occupancy Analysis Simulation

Each profile has been appointed several most typical activities. These priority activities have been determined with a certain probability for each profile. The most likely time slot for visiting the park is provided for each profile. Simulation length determines result credibility because it anticipates the distribution of simulation results to conform to the normal distribution of probabilities. Result levelling can be observed visually but in keeping with statistical methods, it can be assumed that a 95% credible result can be achieved if the number of simulation cycles

the advantages of MAS as a modelling technique is that it allows achieving an ontological correspondence between artificial and real agents. Thus, pre proposed two stage simulation framework proposed for V&V provides an innovative qualitative validation approach.

### 3.1. Concept Simulation

The aim of the simulator is the realisation of two related tasks: area breakdown general planning through potential occupancy analysis (stage 1) and park layout design, and exploitation (stage 2) (see Figure 1).

The Occupancy analysis part and the Layout design and exploitation part can be used together or separately. The Occupancy analysis simulation model is based on statistical data (typical visitor and accompanying person characteristics, time of arrival, list of desirable activities and probability of their choice), profile distribution on weekdays and weekends, as well as the minimal determined visitor comfort area. For each park visitor a corresponding profile is chosen according to recommendations of Zagreb municipality and Autism centre, and other statistical data. The set of activities have been determined to be implementable by visitors (see Figure 3).

allow non-experts check if their previous knowledge about the socio-technical system is properly understood by the modeller.

Figure 4 illustrates by means of a flowchart the sequence of decisions implemented in the citizen agent model, each time a person with a parameterized profile enters in the park.

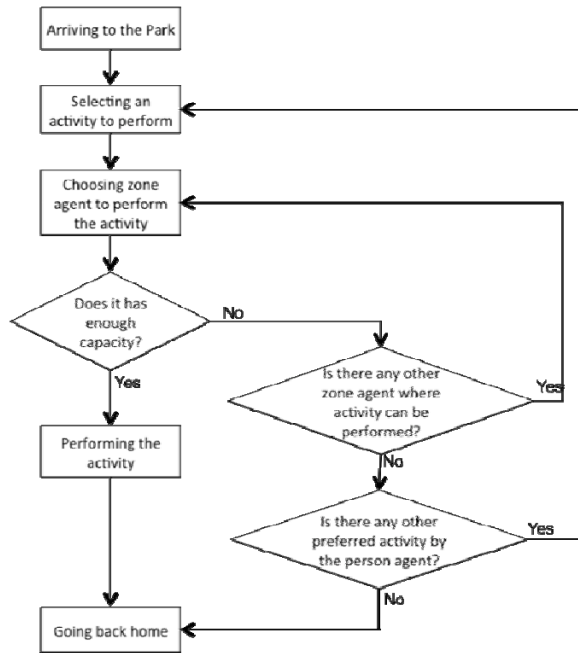


Figure 4: Citizen Behavior Arriving to the Park

In the Zagreb Open Green Park model, to succeed with the Concept Simulator (V&V-1), it has been necessary to formalize users affinity dynamics. It is well known that the original preferences of users are changed in order to satisfy different criteria, such as the possibility to meet some residents (ie. families with kids) while avoiding some other scenarios, such as overcapacity in certain zones. To deal with this important dynamic which is the result of several interdependencies between the different park users, it has been proposed a mechanism to evaluate the satisfaction of a park user considering the active agents in the zone.

Figure 5, summarizes using a flowchart the decision if the time scheduling preferences of a user of the park are updated to have the possibility to match users with similar affinities in the next visit to the park. The right parameters used to obtain a stable behaviour are tuned at the implementation model.

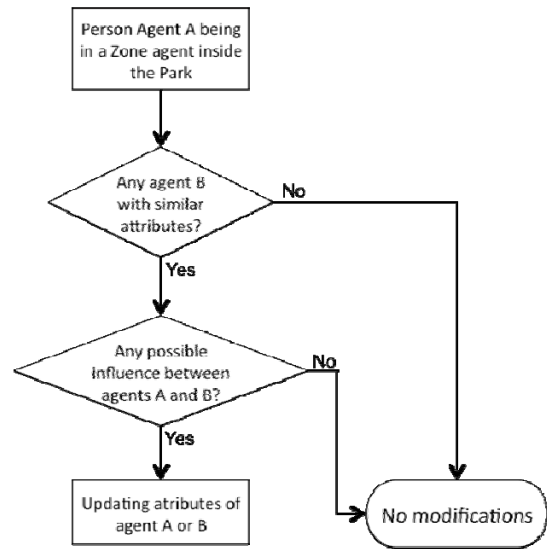


Figure 5: Citizen Affinity Flowchart

The simulation allows creating a rough and hypothetical breakdown of park zoning. It is possible to ascertain the number of visitors in each zone in the simulated time slot, as well as determine the number of rejected visitors if the particular zone/activity was fully occupied.

However, the above-mentioned project assessment stage does not take into account the locations and square size of activities/facilities and reasonable distances between intermediary activities.

Such preliminary park concept evaluation would be recommended before implementation phase.

Important question would be how to recognize that simulation model of the concept is right and valid. Because the park does not exist it is possible to identify the trend only. Therefore social networks are used comparing discussions and opinions with concept simulation results.

### 3.2. Implementation

The Zagreb Open Green Park Layout design and exploitation part can use the operational data from a preliminary park zone area conceptual assessment provided by first simulation stage and allows park layout interactive design using spatial graphical information (see Figure 6).



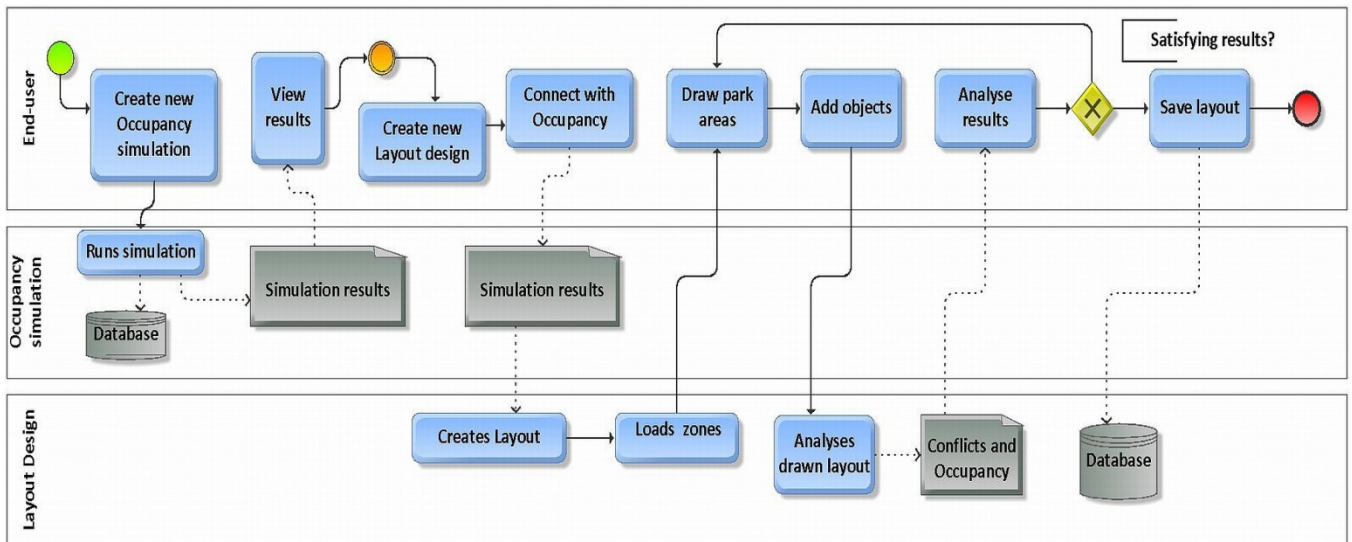


Figure 6: Zagreb Open Green Park Layout Design Diagram

Diagram shows processes of connecting Occupancy Analysis simulation results with Layout Design and modifying a Layout.

The process is distributed among three pools:

- End-user – this pool represents the user of a system, who is working with Zagreb-Open Green park simulation software and wants to receive the results of a simulation and Layout Design;
- Occupancy Analysis – first part of Zagreb-Open Green park simulation software that runs simulation and provides results regarding park zone sizes and overall visitor attendance;
- Layout Design – second part of Zagreb-Open Green park simulation software that can be connected with first part simulation. Layout Design provides options for planning a park layout including zone and park objects' disposition.

To connect an Occupancy simulation with a Layout Design user has to create one. He can do that by choosing new Occupancy simulation at the home page and inputting necessary initial data for simulation including data about the number of total visitors per day, preferred zones, its' sizes and performable activities in the park etc.

Software runs a simulation and provides results for the user. Simulation results show how big should the park zone sizes be, according to the total size and initial values, as well as total park occupancy at each hour point of simulation (that simulates two week period).

After seeing the results the first part is done and intermediate point is reached.

End user returns to home page to create a Layout Design. While creating new Layout project user

specifies Occupancy simulation to connect with the new Layout Design.

Chosen simulation's results are taken from the database to help creating a start for new park layout, loading only those resources (zones and park objects for each zone), that were chosen for the simulation. Additionally, Layout Design shows proposed zone sizes that should be drawn (also taken from simulation results).

Layout Design includes park zone division in areas that can be placed anywhere on the map showing its' real sizes. Each park area can be filled with park objects – activities and facilities (see Figure 7).

After any creation/modification of the layout, software calculates the situation and provides results on the map for the layout including conflicts between park objects and occupancy per hour for each object.

If the user is satisfied with the layout and provided results, he can finish the work by saving the project or continue modifying until decent layout is created. Layout can also be saved at any point of creation and work can be continued from the last saved point.

The second stage provides an opportunity to create a more accurate breakdown and zoning distribution, because it is based on layout visualization providing designers with an extra capacity to deploy the zones considering the maximum number of simultaneous visitors for a certain activity and also respect allowed distances among different activities.

The development of the Zagreb Open Green real planning and realization can be continued if both steps have given promising results.



Figure 7: Zagreb Open Green Park Layout Design Desktop

#### 4. CONCLUSIONS

The designing of support tools for planning and making political decisions is an almost impossible mission because in politics, agriculture and medicine everyone is an expert. Voluntary and unsubstantiated decisions are made as a result and the losses, as usual, are covered by the tax payer. Analytical modelling methods are usually not applicable as they are too complex and cases have a significant number of important stochastic factors.

To ensure at least some possibility to evaluate potential project solutions before investing significant financial resources the authors propose performing an initial conception simulation that would allow determining trends and ascertaining whether further implementation of the project is expedient.

The proposed framework supports an efficient try-and-error approach, which is one of the bases of the simulation techniques, since errors or misleading functionalities can be detected at early stages of the concept simulation. Simulation users with different profiles can be involved at the V&V-1, minimizing the risk of developing a complex simulation model that could not satisfy end-users expectations.

This approach is demonstrated on the basis of the Zagreb Open Green Park designing project.

#### ACKNOWLEDGEMENTS

The article publication is funded by FP7 FUPOL project No. 287119 "Future Policy Modeling".

#### REFERENCES

- Batty, M., Desyllas, J., Duxbury, E., 2003. *Safety in numbers? Modelling crowds and designing control for the Notting Hill Carnival*. *Urban Studies*, 40, 1573–1590.
- Burton R., 2003. Computational Laboratories for Organization Science: Questions, Validity and Docking. *Computational & Mathematical Organization Theory*, 9(2): 91- 108.
- Ginters, E., Aizstrauts, A., et. al., 2013. Deliverable 4.4: FUPOL Simulator Software Prototype (Zagreb Open Green Park), 324.
- Ahmed, A., Greensmith, J., Aickelin, U., 2012. Variance in system dynamics and agent-based modelling using the SIR model of infectious disease. *Proceedings 26th European Conference on Modelling and Simulation*. K. G. Troitzsch, M. Möhring, U. Lotzmann, eds. ISBN: 978-0-9564944-4-3.
- Piera, M-A., Buil, R., Ginters, E., 2013. Validation of Agent-Based Urban Policy Models by Means of State Space Analysis. *Proceedings 8th EUROSIM 2013 Congress on Modelling and Simulation*. In: K. Al-Begain et.al., eds. Wales, United Kingdom, 10-13 September 2013, IEEE, ISBN 978-0-7695-5073-2, 403-409.
- Sargent, R., 1992. Validation and Verification of Simulation Models. In *1992 Winter Simulation Conference*, Piscataway, New Jersey, Institute of Electrical and Electronics Engineers.
- 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.



## **AUTHORS BIOGRAPHY**

**Egils Ginters** is director of Socio-technical Systems Engineering Institute. He is full time Professor of Information Technologies in the Systems Modelling Department at the Vidzeme University of Applied Sciences. He is a Senior member of the Institute of Electrical and Electronics Engineers (IEEE), member of European Social Simulation Association (ESSA) and Latvian Simulation Society. He participated and/or coordinated some of EC funded research and academic projects: FP7 FUPOL project No. 287119 (2011-2014), FP7-ICT-2009-5 CHOREOS project No. 257178 (2010-2013), e-LOGMAR-M No.511285 (2004-2006), SocSimNet LV/B/F/PP-172.000 (2004-2006), LOGIS MOBILE LV/B/F/PP-172.001 (2004-2006), IST BALTPORTS-IT (2000-2003), LOGIS LV-PP-138.003 (2000-2002), European INCO Copernicus DAMAC-HP PL976012 (1998-2000), INCO Copernicus Project AMCAI 0312 (1994-1997). His main field of interests involves: systems simulation technologies, logistics information systems, and technology acceptance and sustainability assessment. He has more than 140 scientific articles related with the research fields.

**Artis Aiztrauts** 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.

**Miquel Angel Piera Eroles** is the delegate for Technical Innovation Cluster, and director of LogiSim, a recognized research group on Modeling and Simulation of Complex Systems. He is Full time Professor in the System Engineering Department at Universitat Autònoma de Barcelona (UAB). Graduated with excellence from UAB in Computer Engineering (1988), Msc from University of Manchester Institute of Science and Technology in Control Engineering (1991), and he got his Phd in 1993. He is member of the Editorial board of 3 international journals and Editor in Chief of IJBRM. Dr Piera has been general chair and invited speaker of many International conferences. He has been nominated as Deputy Director of the UAB Engineering School and Coordinator of the Spanish CEA-IFAC research team on Modeling and Simulation. He has coordinated many research and industrial projects, he has also participated in some EC funded research and academic projects, such as: LOGIS MOBILE LV/B/F/PP-172.001 (2004-2006), Curriculum Development on Logistics and Supply Chain Management. 134522-LLP-1-2007-1-ESERASMUS-ECDSP and FP7 FUPOL No.287119.

**Roman Buil** received the B.S. degree in Mathematics from Universitat Autònoma de Barcelona (UAB) in 2002, the M.S. degree in industrial engineering -

advanced production techniques from UAB in 2004. He is currently working towards the Ph.D. degree in industrial engineering - advanced production techniques at the same University. He is research scientist, assistant teacher and project manager at the Logistics and Aeronautical Unit of the Telecommunications and Systems Engineering Department of UAB. His research interests include modelling and simulation methodologies, optimization techniques, policy modelling, production planning and decision making for production planning and logistics. He is member of LogiSim, a recognized research group on Modelling and Simulation of Complex Systems and he has been involved in industrial projects working as consultant of different companies. He participates in FP7 FUPOL project No.287119.

# MIGRATION AMONG SIMULATION PARADIGMS AND TOOLS

Mairita Zake<sup>(a)</sup>, Egils Ginters<sup>(b)</sup>

<sup>(a,b)</sup>Sociotechnical Systems Engineering Institute of Vidzeme University of Applied Sciences, Cesu street 4, Valmiera LV-4201, Latvia

<sup>(a)</sup>mairita.zake@va.lv, <sup>(b)</sup>egils.ginters@va.lv

## ABSTRACT

The Simulation is used as an active research method when analytical solutions are inconvenient or even impossible. One problem is the incompatibility of created simulations on a tool level, but other - is model incompatibility because of different simulation methods. The article dealt with determination of basic principles for effective model migration and accentuates the use of appropriate simulation methods for the development of specific model groups.

Keywords: simulation, migration, system dynamics, agent-based models

## 1. INTRODUCTION

A typical research object of simulation is sociotechnical systems with a significant number of important parameters that are usually stochastic. Several simulation technologies exist corresponding to the nature and conceptual model of the target system, for example discrete-event systems (DEVS), system dynamics (SD), agent-based simulation (ABM) (Bruzzone, Verbraeck, Ginters et.al. 2002; Ginters et.al. 2011) and others. However some problems remain. First, is the incompatibility of created simulations on a tool level. This is usually addressed by defining common data exchange rules, creating joint communication environments (CORBA, ECE, HLA) (Aizstrauts et.al. 2012; Zhang and Deng 2012) or even shells for data query processing (SimQL/SimAL, MeshSQL) (Wiederhold 2002). The second challenge is model incompatibility because of different simulation methods. As any software can be written in almost any programming language, any simulation model can be developed using any of the before-mentioned simulation methods. However, developer work efficiency and the computing resources required to run the model will differ. Quite often models developed using one simulation method have to be altered to use another to ensure integrity in a unified environment, as well as to avoid using specific communication tools.

The most common model migration type is from SD to ABM and vice versa (Wakeland et. al. 2004; Norling; Scholl 2001; 2007; Macal 2010; Figueredo and Aickelin 2011; Ahmed, Greensmith and Aickelin 2012; Ahmadzadeh, Teose and Gomes 2011).

SD (sometimes called differential equation modeling or dynamical systems modeling) concerns itself with the high-level behavior of a system or macroscopic view. It helps to understand the aggregate operations of system on a macro-scale. It is great for cutting away unnecessary detail and focusing on what is truly important in a model. SD models are constructed from set basic building blocks also known as "primitives". The key primitives are Stocks, Flows, Variables and Links (Insight Maker 2014).

ABM allows to model individual agents within a system. Where in SD might only look at the system as a whole, in ABM is possible to model each individual element and explore the differences and interactions between these elements (NetLogo User Manual 2013).

It would be reasonable to determine of basic principles for model migration and accentuates the use of appropriate simulation methods for the development of specific model groups. Further analysis will be performed based on a practical example: Insight Maker (Insight Maker 2014) SD model migration to an ABM simulation model, as well as to an ABM model in the NetLogo (NetLogo User Manual 2013) environment.

Several comparative studies between ABM and SD have been undertaken (Jaffry and Treur 2008). Some notable discussions in these studies include the issue of computing power and control. In some studies, it also the case that the ABM is computationally expensive compared to the classical mathematical model although this may be overcome in future by highly parallel computing architectures (Tang et al. 2008). Traditional continuous models are generally easier to implement but many aspects of biological systems are intrinsically stochastic in nature (Wilkinson 2009) so the ABM could be viewed as a more „faithful“ interpretation of the processes being modeled (Ahmed et al. 2012).

Ahmed et al. (2012) in their studies shows that ABM is able to capture natural variation without recourse to modification of any parameters for a simulation. The classic SD model has no variation. Therefore an ABM with tool that has built-in randomness is able to capture the natural variation better than a classic SD model simulation. The source of variation for the ABM is the contact between the agents between the different experiments. (Ahmed et al. 2012)

As Figueredo et al. (2011) shows in the results of their experiments that there are simulation cases where SD and ABM derived from the same mathematical model do not have the same output. Therefore, it is not possible to compare which approach would be more suitable for some situations. Similar conclusions draw Macal (2010) in his research for example the equivalency of the model results is not exact in terms of numerical accuracy for the reasons noted. The ABM is able to provide additional information over what the SD model provides due to the explicit stochastic nature of the ABM.

## 2. DEMOGRAPHY MODEL IMPLEMENTATION UNDER DIFFERENT PLATFORMS

### 2.1. System Dynamics (SD) Simulation under Insight Maker Framework

SD modeling essence is based on the data set that have common characteristics, and which can form common types of activities. For this type of models the main concentration is to the stock. There are not separated individuals and different type of activities can be performed for the stock as a whole.

Initial conditions were to make SD demography model to demonstrate demographic changes in a small country like Latvia. The model has to be made under Insight Maker framework. The results of the model have to show demographic changes in the country in the time period of 20 years.

In Figure 1 there is illustrated a simple representation of a Population using SD Stocks and Flows made using Insight Maker tool. There is the single Stock in the middle of the model, “population”, which represents everyone who is alive in this population. There is only one Flow going into population, called births and one Flow out, called deaths. There are also four variables that is neither a Stock nor a Flow. Flow called “birth” that represents “birth per year” is dependent from “birth rate” and “population”. But “birth rate” is dependent from “possibility of birth” that just adds simple random possibility.

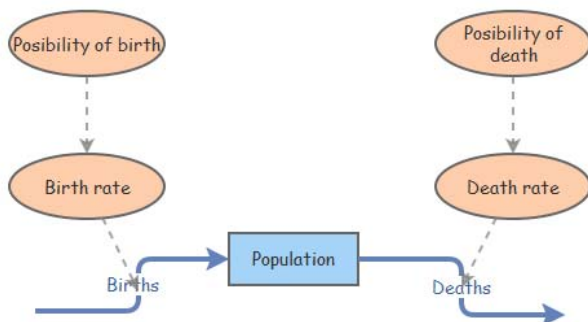


Figure 1: SD model displaying the demographic changes with Insight Maker

Plot graphs in the Figure 2 shows the SD model for simulating demographic changes. Model is very simple and there is only birth rate and death rate that is taken into account just to show a modeling with Insight Maker. Plot graphs shows population decreasing in this specified situation.

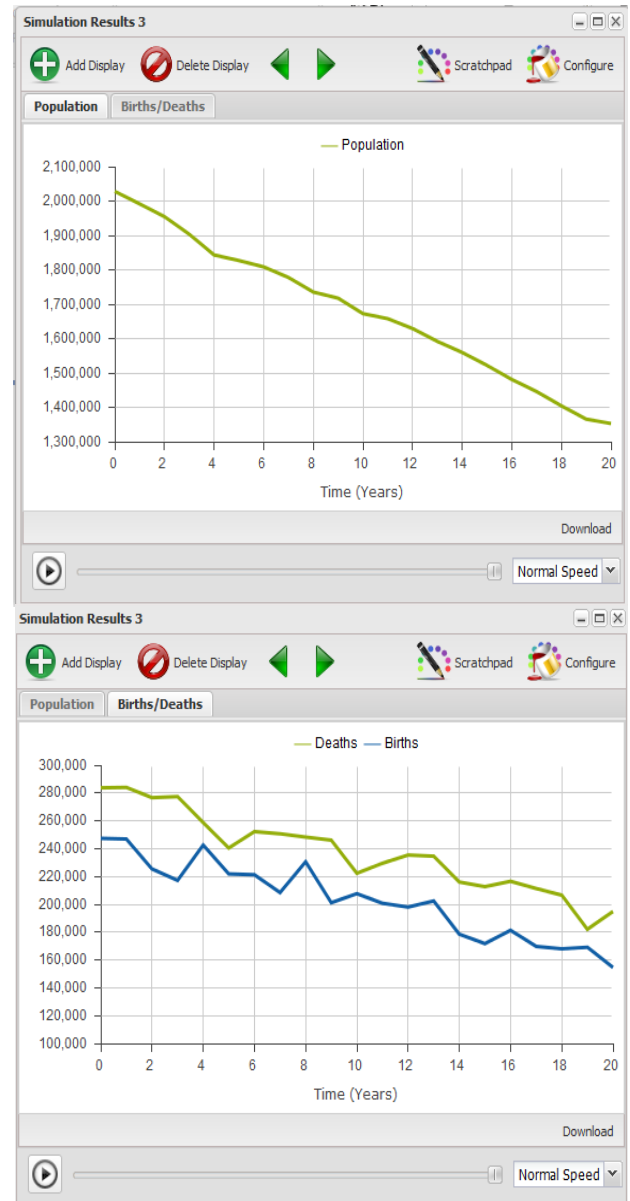


Figure 2: Results of population SD model made with Insight Maker

As it is shown in the Figure 2 in the upper image that with the conditions that have been put in the model, that demography in the time period of 20 years in these conditions is decreasing. The reason for population decreasing is that death level in the population is higher than birth level in the population as it is seen in the lower image of the Figure 2.

The SD approach deals with a population as total rather than with individual humans. The simulation calculates the total population of human at each time step, but each individual is not represented. This makes

it hard to model relationship among the humans. This type of simulation is good for simulating a population in some region or to compare it with another region. SD should be used to simulate a population of human as a whole.

## 2.2. Insight Maker Use for Agent-Based Modeling (ABM)

The ABM main concentration is on each agent separately. Each agent has certain characteristics and with that various types of operations can be performed. In ABM there can be distinguished different types of agents as well.

For modeling demographic situation in ABM style there is the agent - human that has its two main stages – alive or dead. But in SD there is main stock that consists of several agents that has similar stages – alive or dead. In each of these modeling styles these stages could appear different depending on simulation tool specifics.

In this case, the SD model has been modeled the demographic situation for a specific place so in this particular case agent will be a human. Due to the fact that there has been modeled human birth rate and death rate in the system dynamics model, which means that in some way, either as a state or as characteristics of the agent there has to appear the birth and death rates in the agent-based model. Depending on various simulations modeling applications execution of some things can be different but the result stays the same.

Initial conditions for ABM using Insight Maker environment were similar to SD model – to make demography model to demonstrate demographic changes in a small people environment. The results of the model have to show demographic changes in 200 people environment in the time period of 20 years.

Unlike the SD population model where model is based on stocks and flows, the population ABM works with conditions and actions. In order to simulate a population there has to a separate agent for each individual in that population. Each of these agents has a set of attributes that defined their state. States represent the condition someone is in. In this sample use case there is only one state for “human” and that is “alive” as it is shown in Figure 3. There are two actions – “births” and “deaths”. An Action can be used to manipulate a model during a simulation, in this case, to get human to be born or to be dead. When agents in this model die, then they have been deleted from model therefore there is no need for state “dead” for agents.

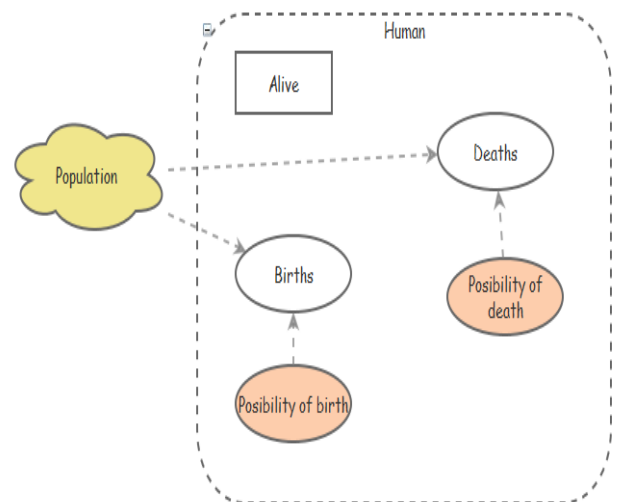


Figure 3: ABM that displays the demographic changes made with Insight Maker

Plot graphs in Figure 4 shows agent based population model with Insight Maker in time of 20 years. Plot graphs for agent based model approach are very similar to SD models.

Since in the demography ABM initial amount of population was 200 people, then the results is hard to compare with previous but as it is seen in Figure 4 that also in this case population curve is decreasing, so comparing these results it can be concluded that results coincide. Exactly like previous model also in this the reason for population decreasing is that death level is higher than birth level.

Insight Maker has lot of convenient menus and graphical model construction, so it is easier to make a System Dynamics model, but for making an Agent Based model these pop-ups and buttons makes it more time-consuming to make an agent based model because it's not as intuitive as it should be for this kind of simulation tool.

Also ABM is not very suitable for modeling conditions where there is a lot of agents with similar conditions especially Insight Maker as an environment is not made for modeling large amount of agents.

It is hard if not impossible to model a simple population model within a region like Latvia not to mention for larger countries using Insight Maker environment.

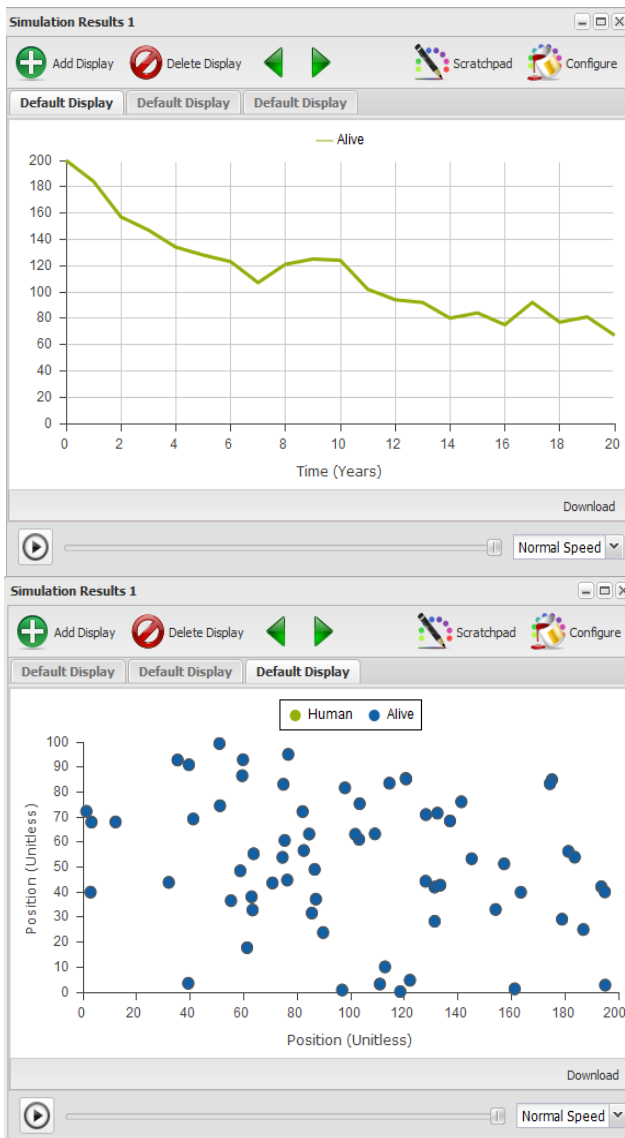


Figure 4: Results of population ABM made with Insight Maker

And also ABM should be used to model tasks where individual approach for modeling agents is needed. Modeling large amount of same type of agents like in population model is not even hard to do in Insight Maker environment but also it is not very practically used recourses.

### 2.3. Multi-Agent Simulation in NetLogo

The ABM will allow the exploration of a range of variations that would be difficult (if not impossible) to encode in the SD model. The SD model has provided some results that are of interest to some fields for example to ecologists; ABM provides a tool for further exploration of the results carried out in SD model, and possibly an understanding at the micro level of why they are so. Furthermore, the ABM provides the opportunity to examine the impact of the assumptions that are encoded in the SD model (Norling 2007).

Initial conditions for ABM using NetLogo were to make demography model to demonstrate demographic

changes in a small region. Conditions for NetLogo environment were a little bit more complicated – there were separated women population and men population separately.

Unlike the agent based model in Insight Maker in order to make an agent based model in NetLogo user has to write almost everything using programming language as it is seen in the Figure 5.

```

breed [women woman]
breed [men man]

women-own [age]
men-own [age]

globals [random-birth]

to birth-women
 set-default-shape women "person"
 create-women initial_number_of_women []
 set color red
 setxy random-xcor random-ycor
 set age random 70
end

to birth-men
 set-default-shape men "person"
 create-men initial_number_of_men []
 set color blue
 setxy random-xcor random-ycor
 set age random 70
end

to Setup
 clear-all
 reset-ticks
 birth-women
 birth-men
end

to Go
 tick
 if not any? turtles [stop]
 ask turtles [
 right random 50
 left random 50
 forward 1
]
end

ask men [
 if age > max-age_of_men
 [die]
 get-birth
 set age (age + 1)
]

ask women [
 if age > max-age_of_women
 [die]
 get-birth
 set age (age + 1)
]

to get-birth
 set random-birth (random 2)
 if random-birth = 0 [
 birth-male
]
 if random-birth = 1 [
 birth-female
]
end

to birth-male
 if random-float 200 < men_birth_rate[
 hatch-men 1 [
 rt random-float 360
 fd 1
 set age 0]
]
end

to birth-female
 if random-float 200 < women_birth_rate[
 hatch-women 1 [
 rt random-float 360
 fd 1
 set age 0]
]
end

```

Figure 5: Source code for ABM displaying the demographic changes made with NetLogo

It makes it difficult for modeling opportunities for end-users in NetLogo environment but for more advanced professionals it creates the conditions which can lead to much more complex models. In order to make agent do anything, there has to be written a procedure. Similar to any other programming language with these written procedures there are a lot of possibilities that can be made in this simulation tool.

In Figure 6 it is shown that in this population simulation model button “setup” calls procedures that create women and men. Button “Go” allows agents to move around and interact with each other. When each and individual agent age has reached maximum then agent dies. Population births and deaths are also affected by birth rate.

In this ABM it is possible for user to interact with model more easier because there has been made slots



for initial number of men and women, maximum age of men and women and women's and men's birth rate.

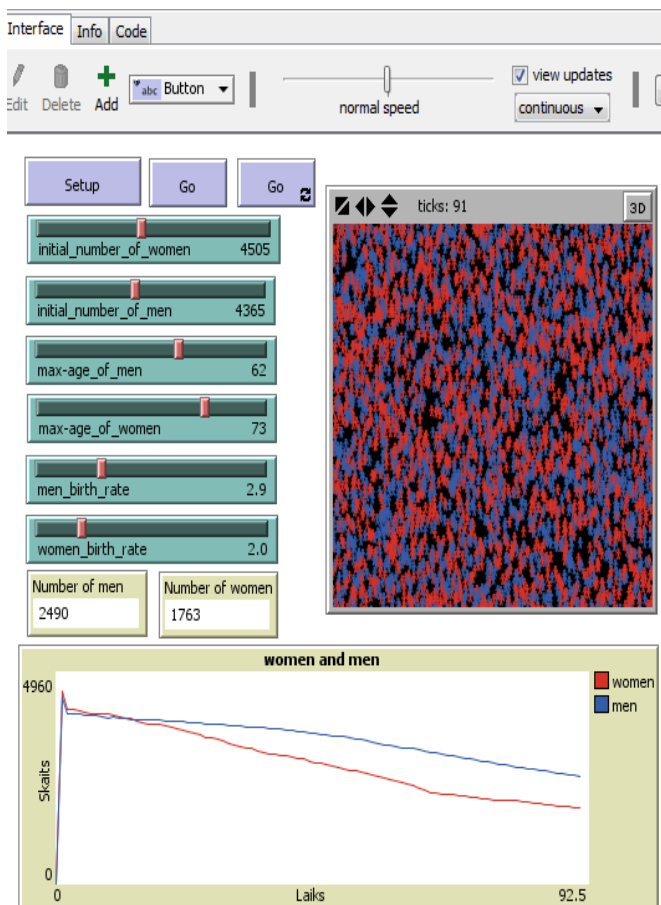


Figure 6: An interface window of the ABM that displays the demographic changes made with NetLogo

The results of the ABM in NetLogo also shows that population with these types of conditions is decreasing. Despite the fact that there has been added some extra functions in the model that wasn't in previous models, simulation results stay the same.

NetLogo as a tool is more convenient to simulate the conditions with large amount of agent because it has more recourse that processes all these agents. But despite that ABM should still be used for more individual tasks that can't be done by SD modeling. Also comparing NetLogo with Insight Maker for modeling in NetLogo environment user needs to be having more programmer skills to make more complicated models. In order to make some model in NetLogo environment user needs to write it programming code.

### 3. STEP BY STEP FROM SD TO ABM - IS IT REASONABLE?

In order to move from SD modeling to ABM there has to be certain steps that need to be followed:

- Step 1 – in the SD model's stock find out what is the agent. Define what the agent is;

- Step 2 – when the agent is founded the next step is to define what kind of characteristics or stages the agent has to have;
- Step 3 – after finding out what the agent is and what kind of characteristics or stages the agent has the next step is to find out what types of activities will be conducted with these agents in order to get the needed results.

Building SD tasks in ABM environment is quite inexpedient because a lot of resources are consumed in the model to simulate each individual agent separately. Tasks that are modeled in the SD environment usually requires large amount of different types of agents with similar types of activities. Migration from SD to ABM for these kind of tasks that requires large amount of agents with similar states of activities are very unreasonable due to the fact that these large amount of agents with similar states doesn't affect each other. Simulation tools needs to simulate each of these agents separately meaning that it takes a lot of resources to simulate that kind of model and if these agents doesn't affect each other at all then migrating it to ABM loses its point.

Tasks that require agents to interact with each other are more useful to migrate from SD to ABM because of the fact that modeling situations where agents interact with each other are more precise in ABM than SD. That means when these kinds of tasks will be migrated from SD modeling to ABM then it will show more precise and more meaningful results. Also it is useful to migrate from SD to ABM assignments where there are different kinds of agents with different kind of states that affect each other.

The Figure 7 shows the basic processes for migration from SD modeling to ABM. The most important part for migration a model is to understand who the agent is. Also the other important par for migration is verification and validation of the model. Every model needs to have verification and validation in order to have successfully made simulation model.

Firstly there has to be made a theoretical model of the example SD model. To develop a theoretical model there should be defined model architecture, the main concept of the model has to be understood. The main part is to define the agent. If that part is not correctly done then the whole model can be not correct. Usually agent can be founded more easily if it is possible to find out what represents stocks.

Very often stocks are filled with agents, there just has to be founded out what is this agent. If the agent is founded next part is to define states of agent. In order to better understand agent's behavior, agent has to have states. If the concept of the agent has been defined then there has to be determine what agent's environment is.

Conceptual model has to be so clear that it is easy to make a simulation model after that and in the conceptual model it should be clear what the agent is and all other details concerning it. Before making the simulation model there has to be done verification and



validation in order to check if developed conceptual model is consistent with the SD model that has been

done before.

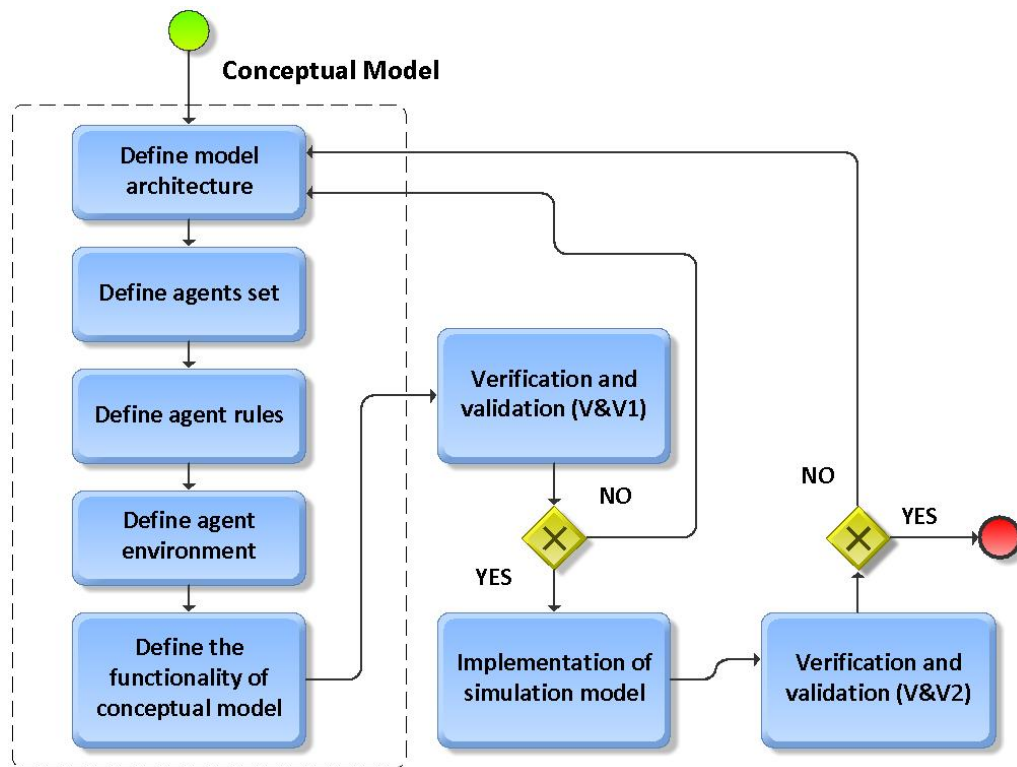


Figure 7: Way of migration from SD to ABM

If conceptual model has done clearly, it can be verified and validated, but after the simulation model can be implemented. In the end there also needs to be another validation to check if developed ABM works exactly as SD model, as well as, is ABM consistent with the SD model.

Verification and validation (V&V1) was done to check if developed conceptual model is consistent with the SD model. Conceptual model was closely compared with the SD model to test if all of the important key points from SD model are added in the conceptual model of ABM and to verify if all of the functionality of SD model is added in the conceptual model.

Verification and validation (V&V2) was done to check if developed simulation model is consistent with the SD model as well as the conceptual model. Developed ABM firstly was compared with the conceptual model. Model output was also closely examined under a variety of settings of the input parameters as well as the model input-output transformations was compared to input-output transformations for the real system to check if it corresponds. Verifying that developed simulation model is consistent, it may be inferred that the model is verified and validated.

#### 4. CONCLUSIONS

If there is needed migrating from SD to ABM then firstly it should be really considered if migration to

other modeling approach is even recommended and good for the model. There are some situations when migration from SD to ABM is very time consuming and takes too much resources. The main important thing in migrating model is to define who the agent in the model is. It is the most important thing in model and if that is not correctly defined then whole model could be done wrong. If the agent is correctly defined then rest of the things in model will be easier to figure out.

Each of the simulation approach is the better option for different kind of situations. SD model is more useful in order to make simple model that asks for large amount of data. In these types of tasks using stock as a representative of a whole of something is more useful than making it as each agent separately. ABM will be more useful for modeling more individual tasks. The ABM is able to provide additional and more detailed information over what the SD model can provide.

#### REFERENCES

- Ahmadizadeh, K., Teose, M., Gomes, C., 2011. Embedding System Dynamics in Agent Based Models for Complex Adaptive Systems. *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, Barcelona, Catalonia, Spain, July 16-22, 2011
- Ahmed, A., Greensmith, J., Aickelin, U., 2012. Variance in system dynamics and agent-based

- modelling using the SIR model of infectious disease. *Proceedings 26th European Conference on Modelling and Simulation*. K. G. Troitzsch, M. Möhring, U. Lotzmann, eds. ISBN: 978-0-9564944-4-3.
- Aizstrauts, A., Ginters, E., Aizstrauta, D., Sonntagbauer, P., 2012. Easy Communication Environment on the Cloud as Distributed Simulation Infrastructure. *Proceedings of the 5th WSEAS World Congress on Applied Computing Conference (ACC '12)*. Recent Advances in Computing Engineering Series 2, ISBN: 978-1-61804-089-3, ISSN: ISSN: 1790-5109, 2-4 May, 2012, Faro, Portugal, pp. 173-179.
- Bruzzone, A., Verbraeck, A., Ginters, E. et al., 2002. *Logistics Information Systems*. Part 1-2. ISBN 9984-30-021-8, 682 p. Riga: Jumi.
- Figueredo, G.P, Aickelin, U., 2011. Comparing System Dynamics and Agent-Based Simulation for Tumour Growth and its Interactions with Effector Cells. *Proceedings of the International Summer Computer Simulation Conference 2011*, pp. 15-22.
- Ginters, E. et al., 2011. Simulation Highway – Direct Access Intelligent Cloud Simulator. *Proceedings of 23th European Modelling & Simulation Symposium (EMSS 2011)*, ISBN 978-88-903724-4-5, 12-14 September, 2011, Rome, Italy, pp. 62-72.
- Insight Maker, 2014. *Types of modeling*. Available from: <http://insightmaker.com/modeling> [accessed 11 May 2014]
- Jaffry, S.W. and Treur, J., 2008. Agent-Based and Population-Based Simulation: A Comparative Case Study For Epidemics. *Proceedings of the 22nd European Conference on Modelling and Simulation 2008*, 123-130.
- Macal, C.M., 2010. To agent-based simulation from system dynamics. *Proceedings of the 2010 Winter Simulation Conference*, IEEE, pp. 371-382.
- NetLogo User Manual, 2013. *What is NetLogo*. Available from: <http://ccl.northwestern.edu/netlogo/docs/> [accessed 11 May 2014]
- Norling, E., 2007. Contrasting a System Dynamics Model and an Agent-Based Model of Food Web Evolution. *Lecture Notes in Computer Science*. Volume 4442, pp. 57-68.
- Scholl, H.J., 2001. Agent-based and System Dynamics Modeling: A Call for Cross Study and Joint Research. *HICSS '01 Proceedings of the 34th Annual Hawaii International Conference on System Sciences ( HICSS-34)*. Volume 3. ISBN:0-7695-0981-9, pp. 3003. Washington: IEEE Computer Society.
- Tang, W., Bennett D.A., and Wang, S., 2008. A Parallel Agent-Based Model of Land Use Opinions. *Land Use Science*, 6(2), 121-135.
- Wakeland, W.W. et. al., 2004. A Comparison of System Dynamics and Agent-Based Simulation Applied to the Study of Cellular Receptor Dynamics. *Proceedings of the 37th Hawaii International Conference on System Sciences – 2004*, IEEE, pp. 1-10.
- Wiederhold, G., 2002. Information Systems That Also Project into the Future. *Databases Networked Information Systems*, 2544. pp. 1–14. doi:10.1007/3-540-36233-9\_1.
- Wilkinson, D.J., 2009. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, 10, 122-133.
- Zhang, J., Deng, H., 2012. Research and Design on High Level Architecture General Gateway. Recent Advances in Computer Science and Information Engineering. *Lecture Notes in Electrical Engineering*, Volume 126, 2012, pp. 295-300.

#### ACKNOWLEDGEMENTS

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

#### AUTHORS BIOGRAPHY

**Mairita Zake** is an associate at Socio-technical Systems Engineering Institute and technical assistant in the faculty of Engineering of the Vidzeme University of Applied Sciences. Her research interests are simulation methods and models migration.

**Egils Ginters** is director of Socio-technical Systems Engineering Institute. He is full time Professor of Information Technologies in the Systems Modelling Department at the Vidzeme University of Applied Sciences. He is a Senior member of the Institute of Electrical and Electronics Engineers (IEEE), member of European Social Simulation Association (ESSA) and Latvian Simulation Society. He participated and/or coordinated some of EC funded research and academic projects: FP7 FUPOL project No. 287119 (2011-2014), FP7-ICT-2009-5 CHOREOS project No. 257178 (2010-2013), e-LOGMAR-M No.511285 (2004-2006), SocSimNet LV/B/F/PP-172.000 (2004-2006), LOGIS MOBILE LV/B/F/PP-172.001 (2004-2006), IST BALTPORTS-IT (2000-2003), LOGIS LV-PP-138.003 (2000-2002), European INCO Copernicus DAMAC-HP PL976012 (1998-2000), INCO Copernicus Project AMCAI 0312 (1994-1997). His main field of interests involves: systems simulation technologies, logistics information systems, and technology acceptance and sustainability assessment. He has more than 140 scientific articles related with the research fields.

# UTILIZATION OF COMPUTER SIMULATION FOR DETECTION NON-STANDARD SITUATIONS WITHIN THE NEW DATA LAYER OF RAILWAY NETWORK MODEL

Jan Fikejz<sup>(a)</sup>, Emil Řezanina<sup>(b)</sup>

<sup>(a)</sup> Department of Software Technologies, FEEL, University of Pardubice, Pardubice, Czech Republic

<sup>(b)</sup> Department of Information Technologies, FEEL, University of Pardubice, Pardubice, Czech Republic

<sup>(a)</sup>[Jan.Fikejz@upce.cz](mailto:Jan.Fikejz@upce.cz), <sup>(b)</sup>[Emil.Rezanina@upce.cz](mailto:Emil.Rezanina@upce.cz)

## ABSTRACT

The article deals with further development of a railway network design and its subsequent use within the simulation of rolling stock operation. The article primarily focuses on the options of rolling stock localization using the GNSS and secondly, on the description of the existing railway network model using a two-layer data model. Attention is further paid to the design of another data layer of the model reflecting a more realistic view of the rail infrastructures between the stations on regional routes. Consequently, simulation of the rolling stock traffic and induction of different non-standard situations is performed within the new model layer, whose detection could be used as an additional support to the dispatcher control system.

Keywords: Railway infrastructure models, train positioning, simulation of railway traffic, agent-based simulation

## 1. INTRODUCTION

Rolling stock localization has been a largely discussed issue involving a number of subjects. The problem can be divided into two main areas of interest: localization for the needs (i) of safety technology and localization for the needs (ii) of information and telematics systems. The earlier mentioned underlines the safety and reliability, however, these systems usually call for higher implementation costs because they require the complementing of railway infrastructure by additional communication or identification elements/ equipment.

In the later case, certain scope of inaccuracy or reduced reliability can be accepted, which often results in a significantly lower implementation of such solutions. Rolling stock localization has recently been connected with the use of satellite navigation system (GNSS – Global Navigation Satellite System).

## 2. STATE OF THE ART TECHNOLOGY IN ROLLING STOCK LOCALIZATION

Rolling stock localization can be divided into the three crucial parts:

- Localization without GNSS,

- GNSS using localization,
- GNSS based, involving further support systems.

### 2.1. Localization without GNSS

This localization type often requires the rail network infrastructure by additional construction elements, which represents higher costs of implementation as such. On the other hand, this localization type shows a high accuracy and reliability and is often used in safety technology. The system in question is the following:

- ETCS (Ghazel, 2014),
- Automatic train control (Lieskovský and Myslivec, 2010),
- Track circuit (Dorazil, 2008),
- RFID.

### 2.2. GNSS using localization

GNSS system used for different application levels requires a thorough consideration of the indicated position error usually based on the satellite navigation character. Systems that operate with the information about the position only on informative level can tolerate a certain error; however, such inaccuracy is unacceptable in safety technology. In such situations, various additional systems can be implemented to eliminate the error in whole or at least partially, thus increasing the tracked object's position. The following systems can be listed in this group:

- EGNOS (Senesi, 2012),
- Differential GPS (O'connor1997; Winter and Xiaogang 2005).

### 2.3. GNSS based localization involving additional support systems

As it was indicated above, precise localization of rolling stock using GNSS, especially for the needs of signalling technology, is a priori impossible. Nevertheless, position of a rail vehicle can be put much more precisely with the use of additional systems. This concerns especially the solutions using inertial systems

(Stadlmann, 2006), and also the less known GNSS using systems based on the GNSS and the contact free eddy current measurement (Becker and Poliak, 2008).

### 3. LOCALIZATION ON REGIONAL TRACKS

GNSS-only based rolling stock localization can be used e.g. for identification of certain position within single-track regional lines in the Czech Republic that, unlike the main corridors, do not dispose of such a high-level technical equipment as the modern ETCS safety systems.

At regional lines it is often only possible to record a train entering/leaving a station, however, a rail car localization on a wide line without complementing the infrastructure by identification elements is much more complicated. As it was said above, GNSS-based localization is always loaded with an error resulting from the satellite navigation character. Despite this, the GNSS using localization can be applied, for example within additional support of dispatcher's control or in various information systems working with a rail vehicle position. Among the essential methods of a rail car localization is the option to use a communication terminal involving the GNSS system, that selected rail cars are equipped with (Fikejz and Kavička 2012). Selected rail cars in the Czech Republic are currently equipped with the following communication terminals:

- Telerail TLR-ZJ (producer Unicontrols, a.s.),
- Radiostation VS67 (producer T-CZ, a.s.).

These remotely configurable communication terminals periodically send defined messages including information about a rail car position. Data messages are transferred to the dispatching centre by means of the UDP protocol. GSM-Ra (Global System for Mobile Communications – Railway) transmission network is primarily used for data transfer. In the case of unavailability hereof, the data is transferred by means of classical GSM network, see Figure 1.

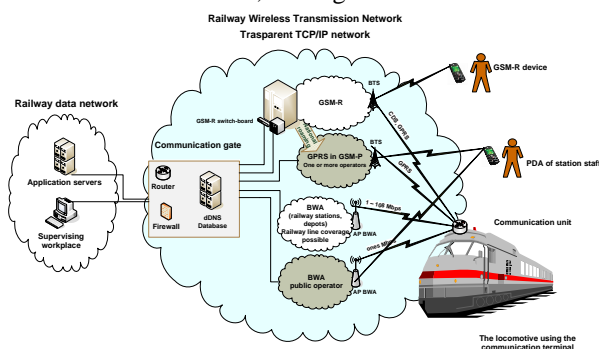


Figure 1: Scheme of communication between trains and dispatching centres

### 4. RAILWAY NETWORK MODEL

One of the key problems in rolling stock localization is the identification of a rail car position in relation to a

railway network infrastructure. (Fikejz and Kavička, 2012b). Experiments have been performed showing that a railway network model can be based on the mileposts (hecto-meters), where every stationing disposes, among others, with a GPS coordinate.

The data does not precisely describe a complete railway network infrastructure (mainly at the stations), nevertheless, it can be used for the design of a railway network model (Fikejz, 2012). Further, based on the data provided by the SŽDC-TUDC (Railway Infrastructure Administration), it was determined that the four following data tables are sufficient for a railway network representation:

- Mileposts / hectometres,
- Super-routes table,
- Railway stations table and,
- Definition supra-sections table.

where the key connector between individual tables is always the line definition section (LDS). Basic aspects of a railway network description are shown in Figure 2.

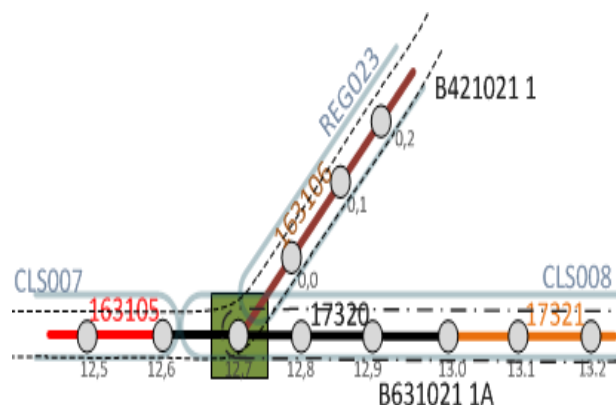


Figure 2: Railway network description methodology

The designed algorithms enabled the authors (Fikejz and Kavička, 2011) to design a railway network (RN) infrastructure reflecting the data structure non-oriented graph in two data layers (micro/macro layer) and three visualization layers (micro/mezo/macrolayer).

This model include data layer:

- **Data-micro**, consisting of vertices and edges,
- **Data-macro** containing super-vertices and super-edges.

The visualization layer then consists of the following layers:

- **Visual-micro**, containing the stations, mileposts representing the vertices and Data-micro layer edges,
- **Visual-mezo**, containing the stations and Data-micro layer edges,

**Visual-macro**, containing the super-peaks and super-edges from the abstract Data-macro layer, see Figure 3.

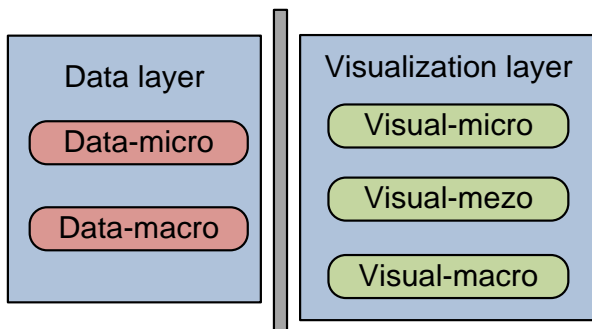


Figure 2: Rail network model concept

The data structure non-oriented graph was 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 network representation, involving also the object scheme and the communication interface API.

The objects scheme includes metadata and network tables. The interface contains on the server side PL/SQL API (an SDO\_NET packet) for the creation, control and analysis of the database network, and a middle layer Java API (on client's side) for the network analysis. The actual network is then defined by means of two compulsory tables:

- Node table,
- Link table.

Concept of the technology is described in Figure 4.

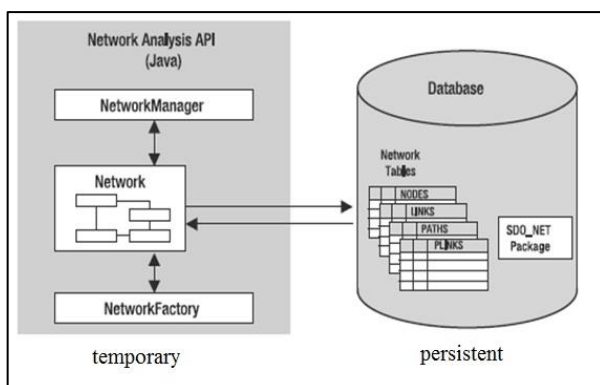


Figure 3: Oracle Network Data Model

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.

#### 4.1. New layer of the model

The edge length in the first data layer (Data-micro) is approximately 100m, while in the second data layer with a higher abstraction scope (Data-macro), the edge length (a so called super-edge) is as many as several tens of kilometres, regardless the occurrence of a rail station on the edge.

The third data layer, the **Data-mezo** attempts to remove this disadvantage by means of the super-edge decomposition. The original two layers used an equal database for vertices, hectometres table, while other vertex types representing individual stations entered the new mezo layer. The formation of a new Data-mezo layer has three parts:

1. Preparation of new input data and layer formation,
2. Loading the layer with data,
3. Generating additional data.

The designed algorithm was implemented directly on the database level by means of the PL/SQL language. However, the algorithm had to be adjusted several times and to be generalized because of the occurrence of various non-standard situations, such as mileposts jumps (non-linear growth of the kilometric sequence between the mileposts) or a change in the growing kilometric sequence to the declining type or vice versa.

##### 4.1.1. Preparation of input data and Data-mezo layer formation

Prior to the launch of the main algorithm, it was essential to prepare the input data and to create individual Data-mezo layer. The input data of the algorithm was obtained from the detail (Data-micro) and the abstract (Data-macro) layers. Within this step, a path in the Data-micro is formed for each super-edge. It is required to consequently find out the initial and ending vertex at each super-edge, located in the Data-micro layer. Dijkstra algorithm of the shortest ways was used in the next step to find the partial edges sequence for each super-edge in the Data-micro layer. One auxiliary SEQ\_TUDU\_ABS\_LINKS\_KM view and two auxiliary coupling tables SUPERNODE\_BINDING and STATION\_BINDING were formed in the data preparatory stage.

##### 4.1.2. Loading the new data layer network

An algorithm for building a new Data-mezo layer in the preparatory stage was implemented. This algorithm for each found way will look up the railway stations and super-vertices located on them. The found vertices are inserted in the network including the edges between them. The main algorithm then performs the following actions:

1. **Initial vertex processing** – the initial vertex is tested for its repeated occurrence in the network. If the vertex is already found in the network, then it is inserted as a new one and recorded in the coupling table (table of edges).
2. **Processing of track route sections** – this step presents a crucial part of the algorithm serving for the railway stations inclusion as vertices into the network and forming the relevant links/edges.



- a) **Check up of the vertices linking** – the check up of the correct sequence of the first vertex. In the case that the whole section is not found in the same sequence, the initial and end vertex of the first partial edge must be swapped for the correct operation of the algorithm.
- b) **Finding out the route direction**– route direction detection.
- c) **Finding out the route sections** – finding out all the sections located on the route in the order from the route beginning to the end. An identification number is recorded at each route section found this way, as well as the initial and final kilometre of the section.
- d) **Finding out the railway stations** –for each route section in a given kilometric interval are determined the relevant railway stations. The stations are returned in the ascending or declining order according to the route direction found out in **step b**.
- e) **Processing of the railway stations**– each railway station is tested for its occurrence in the railway network. In the case of its existence, a new vertex is formed and included in the network and in the coupling table (edges table).
- f) **Processing of a new edge**– a new edge is formed in this step. The previously inserted/found vertex is selected as an initial one, and as a final vertex, it is the newly inserted/found one.

### 3. Processing the final route vertex–

The final vertex is tested accordingly to the initial route one. Providing that the vertex is not located in the network, it is inserted as a new one and recorded in the coupling table. The last step is the formation of a new edge between the penultimate inserted/found vertex and the final route vertex; consequently, the new vertex is inserted in the generated network.

#### 4.1.3. Generating additional data

After creating and filling new Data-mezo layer by the algorithm which is described above it was important to complete the network by the correct shapes (the corresponding data type form SDO\_GEOMETRY) for the required data visualization. Also the relations between the initial and final vertices needed to be solved. To set up the correct geometry, an algorithm was used going through individual edges of the Data-mezo layer calculating the geometry from the GPS coordinates of the initial and final vertex of the relevant edge.

In the processing of the initial and final route vertex not yet inserted in the network, it is necessary to additionally solve the situation when a vertex not only is a super-vertex but also a railway station. In such case, it was needed to include additional information in the coupling table (edges table).

Figure 5 shows a new concept of a railway network model containing three data and four visualization layers.

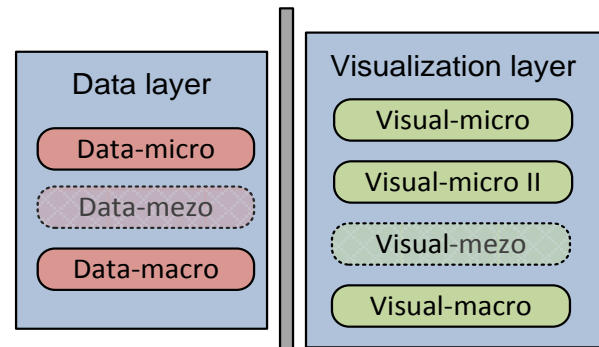


Figure 4: New concept of railway network model

## 5. VISUALIZATION

Taking in consideration a railway infrastructure model visualization, we have to solve the way of multi-dimensional data storage. In the case of using an ORACLE database with a supra-structure Spatial, a visualization tool MapViewer can be applied (Murray et al. 2010) developed in Java language. MapViewer is a J2EE service for drawing map-type documents based on spatial data (e.g. an object data type SDO\_GEOMETRY) administered by means of the ORACLE Spatial. This technology helps to establish a scalable map layers with different detail level of the described information (Fikejz and Kavička, 2011). The basic concept of the MapViewer architecture is demonstrated in Figure 6.

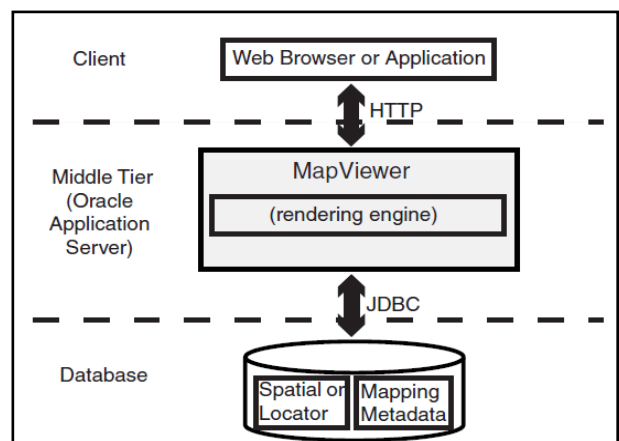


Figure 5: Basic concept of MapViewer service

Figure 7 shows all the three data layers. The black line presents the highest abstraction level layer Data-macro (the vertex is defined only in the rail embranchment nodes, i.e. the super-vertex).



The blue line shows the Data-micro layer (the peak of this layer is each milepost, i.e. the hectometer). The red line (the sickest line) shows the Data-mezo layer. This layer's vertices are formed by the super-vertices and by the stations. The new Data-mezo layer also represents a good compromise between the actual railway infrastructure to be reflected in the model and the number of elements by which is the layer represented. Overall data reduction is almost 28 times higher compared with the elementary Data-micro layer. Such a data reduction than has a positive impact on the total time required for the visualization.



Figure6: Visualization of all data layers

## 6. OPERATION OVER SPATIAL DATA

It is possible to use different operators and functions over the ORACLE database objects with the Spatial supra-structure. One of them is operator SDO\_NN (Near Neighbor) enabling its users to determine the nearest geometry (i.e. a neighbour). In this case it is the

nearest vertex or an edge of a non-oriented graph. If we dispose of a GPS-based information about the real rolling stock position, this operator can be applied for the location of the nearest vertex/edge and then perform e.g. visualization of the position into the map built by means of the MapViewer technology. Besides the rolling stock localization, we can obtain a number of other information closely relating to the rolling stock position and to the railway network infrastructure, especially:

- The node on which the rolling stock is located,
- Kilometric position according to the mileposts,
- Train occurrence in the station area,
- Direction of the rolling stock motion by means of azimuth (to/from a station),
- Distance to the nearest station,
- Name and number of the route according to the regular civil schedule,
- Relevance of the real GPS position.

## 7. SIMULATION OF RAILWAY TRAFFIC

As it was indicated above, the selected driving cars are equipped with the communication terminals transmitting the data that include the real GPS coordinates of the rolling stock. In the event that a car is in motion, the communication terminal sends the information containing the position data every 30 seconds.

Table 1 contains the data obtained from the communication terminals.

The designed simulation model contains the core of a discrete simulation using a standard calendar of events performed within the simulation according to their time stamp, see Figure 8.

The simulation model was consequently implemented in a demonstration application InfraRail designed for additional support to the dispatching control.

Simulation of the rolling stock traffic can be divided into two parts on the basis of:

- The real historical data (traffic emulation)
- The generated data.

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 |

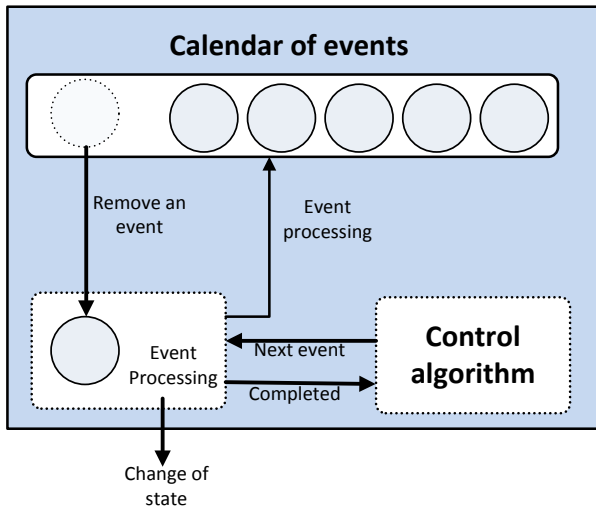


Figure 7: Simulation /core

Simulation based on the generated data enables the user to induce different non-standard situations. For example, a motion of the rolling stock moving against each other can be simulated on a single track without rail embranchment. A new data model provides for a more precise detection of such situations because railway stations can be located on single-track routes without embranchment (super-edges – macro-layer); the rolling stock can be shunted to a rail siding at such a station. Figure 9 shows a running demonstration application in a situation where two cars are moving against each other on a single-track route.

## 8. CONSLUION

The article deals with the design of a new layer of the Data-mezo railway network. The purpose is to complete the previous two-layer model built on the basis of the

SŽDC-TUDC data analysis. The newly designed model thus contains a significantly reduced base of data elements than the detailed Data-micro layer. Despite this, the model is rather realistic in the reflection of physical railway network infrastructure. The new data model was built by means of the designed algorithm, which is described in the article in detail. The newly designed model was then implemented into the demo application InfraRail, and using a discrete simulation, the rolling stock traffic on a rail network was simulated and various non-standard situations on single-track regional routes lacking the track embranchment. The system was able to correctly detect such non-standard situations and can be used as an additional tool of dispatching control of the railway traffic.

## ACKNOWLEDGMENTS

This work has been supported by the project “SGFEI04/2013 Models of infrastructure and land transport and service systems operation” (financed by the University of Pardubice).

## REFERENCES

- Becker, U., Poliak, J. DEMOORT repositions trains with satellite. In: EURAILmag Business & Technology. 18. vyd. BLUE LINE & Bro, France, s. 216-219.
- Chudaček, V., Lochman, L. Vlakový zabezpečovací systém ERTMS/ETCS in *Vědeckotechnický sborník CD*, č. 5/1998
- Dorazil, P. Základní vlastnosti kolejových obvodů bez izolovaných styků. Pardubice, 2008. Dostupné z: <http://dspace.upce.cz/bitstream/10195/28903/1/text.pdf>. Bachelor thesis. Univerzita Pardubice. Milan Kunhart.
- Fikejz, J. Lokalizace polohy kolejových vozidel v rámci modelu železniční sítě s využitím technologie ORACLE Network Data Model. *Perner's Contacts* [online]. 2012, roč. 7, III, s. 33-49 [cit. 2013-02-28].

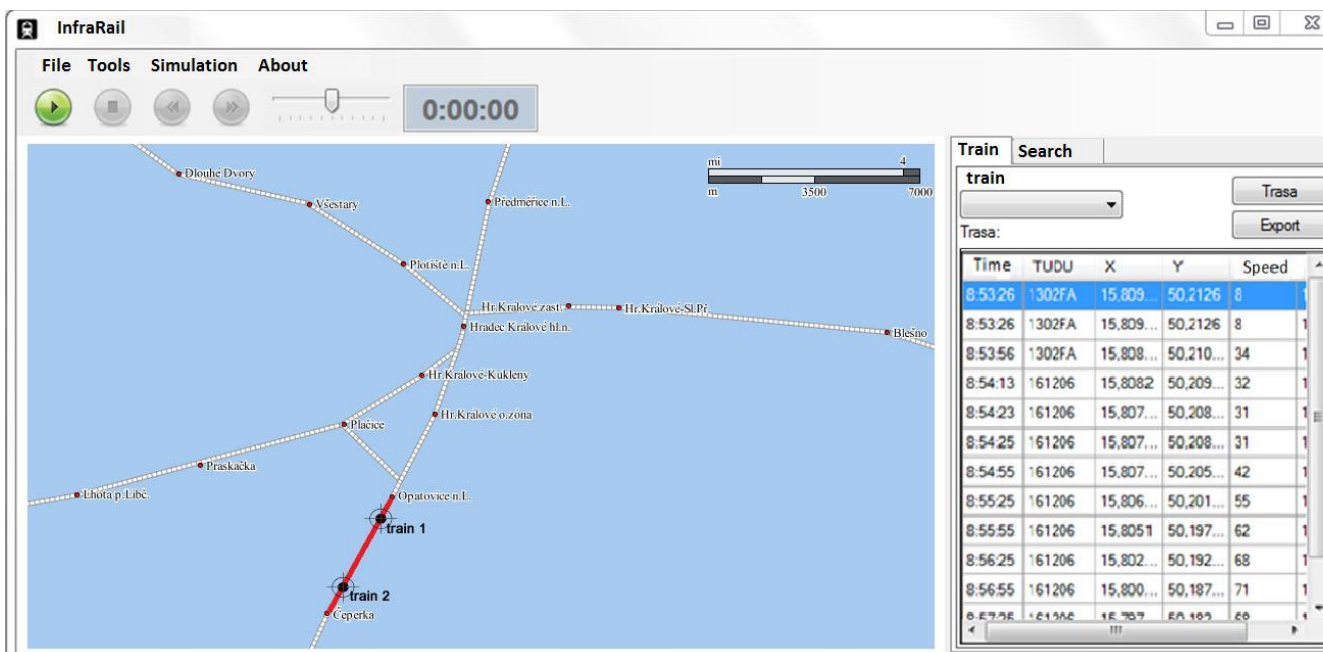


Figure 8: Traffic simulation

- ISSN 1801-674x. Available from: WWW:  
[http://pernerscontacts.upce.cz/27\\_2012/Fikejz.pdf](http://pernerscontacts.upce.cz/27_2012/Fikejz.pdf)
- Fikejz, J., Kavička, A. Modelling and simulation of train positioning within the railway network. In: KLUMPP, Matthias. *ESM'2012. The European simulation and modelling conference*. Ostende: EUROSIS - ETI, 2012a, s. 366 -376. ISBN 978-9077381-73-1.
- Fikejz, J.; Kavička, A. Rolling Stock Localization Within the Model of Railway Infrastructure. In: *Proceedings of Third International Conference on Computer Modelling and Simulation*, Brno, CZ, FIT VUT, 2012b, s. 80-85, ISBN 978-80-214-4576-5
- Fikejz, J.; Kavička, A. Utilisation of computer simulation for testing additional support for dispatching rail traffic. *ESM'2011: European Simulation and Modelling Conference 2011*. Belgium : Reproduct NV, 2011. p. 225-231. ISBN 978-90-77381-66-3.
- Ghazel, M., Formalizing a subset of ERTMS/ETCS specifications for verification purposes. *Transportation Research Part C: Emerging Technologies*, Elsevier Limited, 2014, pp. 60-75 ISSN:0968-090X
- Kothuri, R., Godfrind, A., and Beinat, E., *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. Automatické vedení vlaků Českých drah. Automatizace. Praha: Automatizace, 2004, roč. 10. ISSN 0005-125x.
- Lieskovský, A., Myslivec, I. ETCS a AVV - poprvé společně. EuroZel, Žilina, 2010
- Lieskovský, A., Myslivec, I., Špaek, P. ETCS a AVV - bezpečně a hospodárně. *Moderní technologie a diagnostika v železniční telekomunikační a zabezpečovací technice*, České Budějovice, 2007
- O'connor, M. L. Carrier-phase differential gps for automatic control of land vehicles, December 1997, STANFORD UNIVERSITY
- Senesi, F., Satellite application for train control systems: The Test Site in Sardinia, *Journal of Rail Transport Planning and Manageme*, Elsevier BV, 2012, pp. 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 17-20, 2006.
- Winter, J., Xiaogang Gu. GNSS-supported train location for safety-relevant applications on branch lines, in *Proc. of 12th International Symposium ZEL 2005*, Žilina, pp. 171-178

# FEEDBACK DTC-SVM BASED A FRACTIONAL PI CONTROLLER: APPLIED TO ROTOR'S SPEED OF INDUCTION MOTOR

Zennir Youcef<sup>(a)</sup>, Bouras Lakhdar<sup>(b)</sup>

<sup>(a)</sup> Automatic Laboratory of Skikda, Route El-Hadeaik, BP26. 21000 Skikda, Algeria

<sup>(b)</sup> Electrical Engineering Laboratory of Skikda, Route El-Hadeaik, BP26. 21000 Skikda, Algeria

<sup>(a)</sup>[youcef.zennir@univ-skikda.dz](mailto:youcef.zennir@univ-skikda.dz), <sup>(b)</sup>[bouras\\_nour@yahoo.fr](mailto:bouras_nour@yahoo.fr)

## ABSTRACT

In this paper we present a new DTC approach for induction motor control, the approach is based on the use of both classical PI and fractional PI controllers, classical PI is used to control the magnetic flux and the torque whereas Fractional PI controller is used for rotor's speed control. In the first part we pointed out some structures of DTC control along with mathematical modeling for induction motor, in the second part a description of software simulation is given and finally we end up with discussion of obtained results and prospective for future works.

Keywords: workstation design, work measurement, ergonomics, decision support system

## 1. INTRODUCTION

The induction motors are widely used in industrial drives because they are rugged, reliable and economical (Casadei,1996), (Bin-jun 2012). The induction motors are increasingly being used with variable-frequency drives in variable-speed services. In the other hand Squirrel cage induction motors are used in both fixed-speed and VFD applications. The rotor circuit for induction motor consists of conducting bars regularly distributed between two metal crowns forming the ends. Since the induction motor have no field winding, then it is necessary to provide magnetic energy for the motor. The power supply for the motor can be provided using two ways: from the electrical network or a renewable energy sources. In the first case and in the upstream of the motor generally there is a rectifier and an inverter circuits to convert the continuous feed to alternative. The speed of the motor depends on the frequency, the number of pole's pairs, as well as the torque opposed by its load. If the motor is supplied from renewable energy source we only find an inverter, in the motor upstream (Jadhav 2011). In these two configurations the control of inverter frequency is needed ensure a good performance of the motor. The use of voltage inverter was the subject of many researches and it is necessary in this case to mention that the appearance of thyristor controller GTO and, thereafter, transistors IGBT help in the development of powerful, reliable MLI and with low costs voltage inverters (Jadhav 2011; Depenbrock

1988), which solve the problem associated to motor's feed, knowing that the control of the induction motor can be applied in steady state conditions. In the literature, there exist several works concerning the development new strategies that allow a control uncoupled from the motor with induction. Among these methods there are the field-oriented controls (FOC) which, ensure of the dynamic performances equivalent to those obtained by the motor with D.C. current (Bettou 2008). During last years, the development of new signal processing methods allow the realization of more advanced control structures, as example we mention the direct torque control (DTC), the synoptic ones of the order evolved certainly in the direction to improve some aspects like minimization of the influence of the motor parameters without requiring a mechanical sensor speed (Dingy 2006).

## 2. CLASSICAL DIRECT TORQUE CONTROL APPROACH

This control approach is based on the direct control of induction motor's torque. This approach was proposed by I. Takahashi and T. Noguchi (Chikh 2011) and Depenbrock (Mansour 2012). This control approaches the torque by the application of the inverter's with various voltage vectors, which determines its state. The control variables are the magnetic stator flux and the electromagnetic torque which are usually ordered by hysteresis controllers. The problem is how we can maintain the magnitude of stator flux and electromagnetic torque inside these bands of hysteresis controller. The output of these controllers determines the optimal tension vector to apply every moment at commutation. The use of this controller's type supposes the existence of a variable commutation frequency in the converter requiring a very weak step of calculation. The use of voltage inverter makes it possible to reach six distinct positions in the plan from phase, corresponding to the eight tension vector sequences on the outlet inverter side. The application of this approach gives a very fast dynamic response torque's motor. It seeks for obtaining the flux and the stator currents close to the sinusoidal forms. The commutation's frequency inverter depends on the hysteresis magnitude bands. The basic

structure of the direct torque control is presented on the following figure:

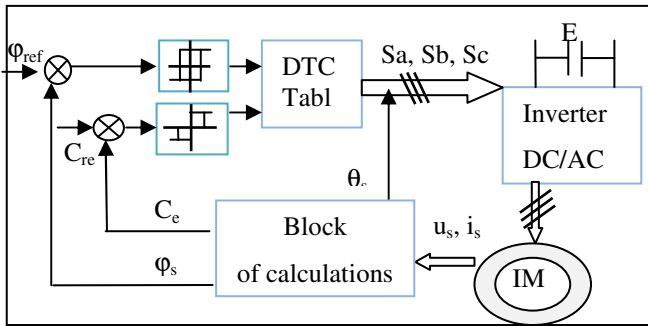


Figure 1: Architecture of classic DTC control.

The basic disadvantages of DTC classic approach using hysteresis controllers are the variable switching frequency, the current and torque ripple. The movement of stator flux vector during the changes of cyclic sectors is responsible for creating notable edge oscillations of electromagnetic torque. We can see also another great issue is the implementation of hysteresis controllers which requires a high sampling frequency. When a hysteresis controller is implemented its digital operation is quite different to the analogue one. In the analogue operation the value of the torque and the magnitude of the flux are limited in the exact desirable hysteresis band. That means, the inverter voltage can change state each time the torque or the flux magnitude are throwing the specified limits. The digital implementation uses specific sample time on which the magnitudes of torque and flux are checked to be in the desirable limits. That means, very often, torque and flux can be out of the desirable limits until the next sampling period. For this reason, an undesirable torque and flux ripple is observed. (Depenbrock 1988). During the last years, a lot of developments and modifications in classic Direct Torque Control approach of control applied of induction motor control (Chikh 2011), have been made. The objective of these modifications was to improve the motor's start up. Also our modifications proposed aimed to reduce the torque and current ripple and to avoid the variable switching frequency. Then we focused our study to develop new control's architecture for switching frequency control and to return it as possible constant in DTC closed loop structure. This new structure of control based on the space vector modulation (SVM) for voltage inverter control and different standard controller (classic and fractional PI) to improve the performances of the motor and to solve the problems of the classic DTC

### 3. FEEDBACK DTC CONTROL WITH PI CONTROLLER

This type of control use space vector modulation (SVM) to inverter control and PI controller for electromagnetic Torque and the magnetic flux control (Ortega 2005; Casadei 2000). With this approach we have constant switch frequency; it improves the dynamic and the static response, and decreases also the undulations of torque and the current. These results are more interesting than obtain with classic DTC, but it's more complicated than

the classic DTC. The algorithm of this new control with rotor's speed control is represented by the following figure.

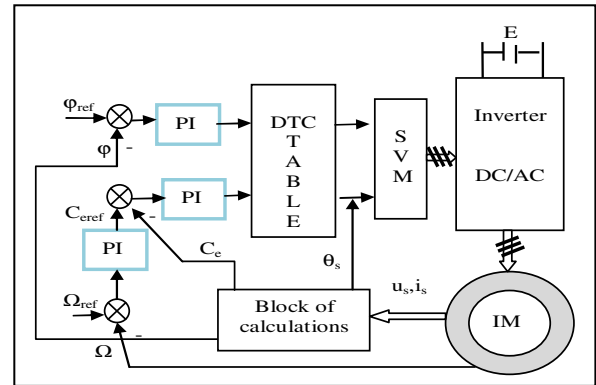


Figure 2: Feedback DTC control with PI controller.

We describing in the following section the synthesis of classic PI controller used to control magnetic flux, the electromagnetic torque and rotor's speed.

#### 3.1. DTC-SVM control with classic PI controller

The mathematical model for induction motor is described by the following equations (Bouras 2005):

- Electric equations (stator's and rotor's voltages):

$$V_{sd} = R_s \cdot I_{sd} + \frac{d}{dt} \varphi_{sd} - \omega_s \cdot \varphi_{sq} \quad (1)$$

$$V_{sq} = R_s \cdot I_{sq} + \frac{d}{dt} \varphi_{sq} + \omega_s \cdot \varphi_{sd} \quad (2)$$

$$V_{rd} = R_r \cdot I_{rd} + \frac{d}{dt} \varphi_{rd} - (\omega_s - \omega_r) \cdot \varphi_{rq} = 0 \quad (3)$$

$$V_{rq} = R_r \cdot I_{rq} + \frac{d}{dt} \varphi_{rq} + (\omega_s - \omega_r) \cdot \varphi_{rd} = 0 \quad (4)$$

- Magnetic flux equations:

$$\varphi_{sd} = L_s \cdot I_{sd} + L_M \cdot I_{rd} \quad \varphi_{sq} = L_s \cdot I_{sq} + L_M \cdot I_{rq} \quad (5)$$

$$\varphi_{rd} = L_r \cdot I_{rd} + L_M \cdot I_{sd} \quad \varphi_{rq} = L_r \cdot I_{rq} + L_M \cdot I_{sq}$$

#### 3.2. Control of stator flux

The induction motor have nonlinear mathematical model represented by the coupling between the torque and magnetic flux. Then for linearization of this model, we decouple the torque and magnetic flux, by the orientation of magnetic flux according to direct "D" component. Which gives \$\varphi\_{sd} = \varphi\_s\$ and \$\varphi\_{sq} = 0\$. And then applying this in the above equations we get the following model:

$$V_{sd} = R_s \cdot I_{sd} + \frac{d}{dt} \varphi_s; \quad V_{sq} = R_s \cdot I_{sq} + \omega_s \cdot \varphi_s \quad (6)$$

$$C_e = \frac{3}{2} \cdot p \cdot \varphi_s \cdot I_{sq} \quad (7)$$

$$I_{rd} = \frac{1}{L_M} \cdot (\varphi_s - L_s \cdot I_{sd}); \quad I_{rq} = -\frac{L_s}{L_M} \cdot I_{sq} \quad (8)$$

$$\varphi_{rd} = \frac{L_r}{L_M} \cdot (\varphi_s - \sigma \cdot L_s \cdot I_{sd}) \varphi_{rq} = -\sigma \cdot \frac{L_s \cdot L_r}{L_M} \cdot I_{sq} \quad (9)$$

$$\text{With: } \sigma = 1 - \frac{L_M^2}{L_s \cdot L_r}$$

With (7) and (8) in (9) we obtain

$$\varphi_s = \frac{L_s \cdot [(1 + \sigma \cdot T_r \cdot S) \cdot I_{sd} + \sigma \cdot T_r \cdot I_{sq} \cdot \omega_{sl}]}{1 + T_r \cdot S} \quad (10)$$

$$I_{sq} = \frac{T_r \cdot \omega_{sl} [\varphi_s - \sigma \cdot L_s \cdot I_{sd}]}{L_s \cdot (1 + \sigma \cdot T_r \cdot S)} \quad (11)$$

$$\text{With: } T_r = \frac{L_r}{R_r}$$

From (10) and (11) we obtain:

$$V_{sd} = \frac{\varphi_s}{G_{\varphi_s}} + E_d \quad (12)$$

$V_{sq} = R_s \cdot I_{sq} + \omega_s \cdot \varphi_s$ . The term  $R_s \cdot I_{sq}$  it's very small then:  $V_{sq} \cong \omega_s \cdot \varphi_s$

$$G_{\varphi_s} = \frac{T_s \cdot (1 + \sigma \cdot T_r \cdot S)}{1 + (T_r + T_s) \cdot S + \sigma \cdot T_r \cdot T_s \cdot S^2} \quad E_d = -\frac{\sigma \cdot R_s \cdot T_r \cdot I_{sq} \cdot \omega_{sl}}{1 + (T_r + T_s) \cdot S} \quad (13)$$

Hence we can control stator's magnetic flux by the component d of the stator's tension. And we have a system of second degree with disturbance  $E_d$  represented by the following block diagram:

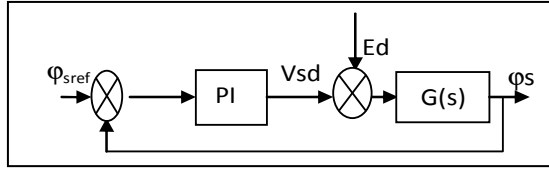


Figure 3: Block diagram of the flux loop

### 3.3. Torque control

To control the torque we have:

$$\varphi_s = \frac{L_s \cdot [(1 + \sigma \cdot T_r \cdot S) \cdot I_{sd} + \sigma \cdot T_r \cdot I_{sq} \cdot \omega_{sl}]}{1 + T_r \cdot S} \quad (15)$$

$$I_{sq} = \frac{T_r \cdot \omega_{sl} [\varphi_s - \sigma \cdot L_s \cdot I_{sd}]}{L_s \cdot (1 + \sigma \cdot T_r \cdot S)} \quad (16)$$

From where:

$$I_{sd} = \frac{(1 + T_r \cdot S) \cdot \varphi_s - \omega_{sl} \cdot \sigma \cdot L_s \cdot T_r \cdot I_{sq}}{L_s \cdot (1 + \sigma \cdot T_r \cdot S)} \quad (17)$$

$$I_{sq} = \frac{T_r \cdot \omega_{sl}}{L_s \cdot (1 + \sigma \cdot T_r \cdot S)} \cdot \left[ \varphi_s - \frac{\sigma \cdot L_s \cdot [(1 + T_r \cdot S) \cdot \varphi_s - \omega_{sl} \cdot \sigma \cdot L_s \cdot T_r \cdot I_{sq}]}{L_s \cdot (1 + \sigma \cdot T_r \cdot S)} \right] \quad (18)$$

$$I_{sq} = \frac{\frac{(1 - \sigma) \cdot T_r}{(L_s)} \cdot \varphi_s^2 \cdot \omega_{sl}}{(1 + \sigma \cdot T_r \cdot S)^2 + (\omega_{sl} \cdot \sigma \cdot T_r)^2} \quad (19)$$

While replacing in the formula of the torque we obtains:

$$C_e = \frac{3}{2} \cdot p \cdot \frac{\frac{(1 - \sigma) \cdot T_r}{(L_s)} \cdot \varphi_s^2 \cdot \omega_{sl}}{(1 + \sigma \cdot T_r \cdot S)^2 + (\omega_{sl} \cdot \sigma \cdot T_r)^2} \quad (20)$$

With:  $\sigma$  it's very small then

$$C_e = \frac{3}{2} \cdot p \cdot \frac{\frac{(1 - \sigma) \cdot T_r}{(L_s)} \cdot \varphi_s^2}{1 + 2 \cdot \sigma \cdot T_r \cdot S} \cdot (\omega_s - \omega_r) \quad (21)$$

$$C_e = G_{ce}(s) \cdot (\omega_s - \omega_r) \quad (22)$$

$$\text{With: } G_{ce}(s) = \frac{3}{2} \cdot p \cdot \frac{\frac{(1 - \sigma) \cdot T_r}{(L_s)}}{1 + 2 \cdot \sigma \cdot T_r \cdot S} \cdot \varphi_s^2 \quad (23)$$

From the above equations we conclude that we can control the electromagnetic torque by the stator's pulsation as it is indicated in the following block diagram:

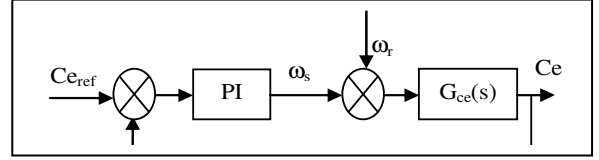


Figure 4: Block diagram of the torque loop

From the above block diagrams the control structure of DTC-SVM control with a classical PI controller is presented, in the next section we present the synthesis of a fractional PI controller for rotor's speed control.

### 3.4. DTC-SVM control with fractional PI controller for rotor's speed

The fractional PI controller has better flexibility in the operating concept because it has another parameter that represents the fractional control of integral action in the PI controller. This parameter can be used to satisfy additional performances in the systems design controlled. For the adjustment fractional parameters controller  $PI^\alpha$  ( $K_c$ ,  $T_i$ ,  $\alpha$ ) we us the following algorithm (Bettou 2008)(Bouras 2013):

**Step1:** calculate the parameters  $\theta_i$  for  $0 \ll i \ll 2$

$$\theta_0 = \frac{1}{2}; \theta_1 = \frac{-m}{4 \cdot \omega_u}; \theta_2 = \frac{m}{4 \cdot \omega_u^2} \quad (24)$$

With:  $\omega_u$ : frequency of the profit unit of the reference model;  $m$ : the derivation fractional order of the reference model;  $\theta_i$ : derived from the transfer function of the reference model  $G_d(p)$

**Step 2:** calculate the parameters  $y_i$  for  $0 \ll i \ll 2$

$$y_0 = \sum_{k=0}^N G_p(kT) \cdot e^{-K \cdot T \cdot \omega_u} \quad (25)$$

$$y_1 = -\sum_{k=0}^N (kT) \cdot G_p(kT) \cdot e^{-K \cdot T \cdot \omega_u} \quad (26)$$

$$y_2 = \sum_{k=0}^N (kT)^2 \cdot G_p(kT) \cdot e^{-K \cdot T \cdot \omega_u} \quad (27)$$

With:  $y_i$ : derived from the transfer function  $G_p(p)$  compared to the variable  $p$  at the point  $\omega_u$ ;  $N$ : many samples.

**Step 3:** calculate the parameters  $x_i$  for  $0 \ll i \ll 2$

$$x_0 = \frac{\theta_0}{y_0 \cdot (1 - \theta_0)} \quad x_1 = \frac{\theta_1}{y_0 \cdot (1 - \theta_0)^2} - \frac{x_0 \cdot y_1}{y_0} \quad (28)$$

$$x_2 = \frac{\theta_2}{y_0 \cdot (1 - \theta_0)^2} + \frac{2 \cdot \theta_1^2}{y_0 \cdot (1 - \theta_0)^3} - \frac{2 \cdot x_1 \cdot y_1 + x_0 \cdot y_2}{y_0} \quad (29)$$

with:  $x_i$ : derived from the transfer function of the controller  $C(p)$

**Step 4:** calculate the parameters  $K_c$ ,  $T_i$ ,  $\alpha$   
 $\alpha = -\frac{\omega_u \cdot x_2}{x_1} - 1$ ;

$$T_i = -\frac{\omega_u^{(1+\alpha)} \cdot x_1}{\alpha}; \quad K_c = x_0 - T_i \cdot \omega_u^{-\alpha} \quad (30)$$

Concerning the approximation of fractional action we used the approximation method developed by Charef (Bettou 2008). The transfer function in closed loop for rotor's speed is regarded as stable. The problem system's design is thus to regulate the three parameters of fractional controller to guarantee that the transfer



function in closed loop behaves the frame of reference which itself answers the specifications of the fractional control system (transfer function in closed loop). The frame of reference fractional model used is:

$$G_d(p) = \frac{1}{1 + (\frac{p}{\omega_u})^m} \text{ avec } 1 < m < 2 \quad (31)$$

#### 4. SIMULATION

Our aim is studying the effectiveness of fractional PI controller and comparing the results with the use of traditional PI controller. We carried out the following simulations: DTC-SVM control with a classic PI controller for the torque, stator's magnetic flux and rotor's speed (in close loop), and DTC-SVM control with a fractional PI controller for the rotor's speed. All simulations developed in the Matlab-Simulink environment and the parameters of the controllers, induction motor and reference model are illustrated in the following table:

Table 1: Simulation's Parameters

|                   | Parameters                                                                   |
|-------------------|------------------------------------------------------------------------------|
| PI                | of magnetic flux : $K_p=20$ ; $K_i=200$ ;<br>of torque: $K_p=5$ ; $K_i=20$ ; |
| PI <sup>α</sup>   | $\alpha=0.5$ ; $K_p=0.46$ ; $K_i=3.2$ ;                                      |
| IM                | Nominal power 1.5k W , $J = 0.013 \text{ Kg.m}^2$                            |
|                   | $R_s=4.75 \Omega$ , $R_r=6.3 \Omega$ ; $L_m=0.612 \text{ H}$                 |
|                   | $L_s=0.6550 \text{ H}$ , $L_r=0.6520 \text{ H}$ ; $P = 4$                    |
|                   | Coefficient of friction $F = 0.002 \text{ Nms}$                              |
| Reference model : | $\omega=10 \text{ rd/s}$ ; $\xi=0.707$                                       |

The transfers function of fractional PI controller:

$$G_{PI^\alpha}(p) = k_p * (1 + k_i * G1(p)) = \frac{\text{num}(p)}{\text{den}(p)} \quad (32)$$

Degrees of the num (p) =26 and den (p) =28 with zeros and poles distinguish. The results obtained are illustrated in the following figures:

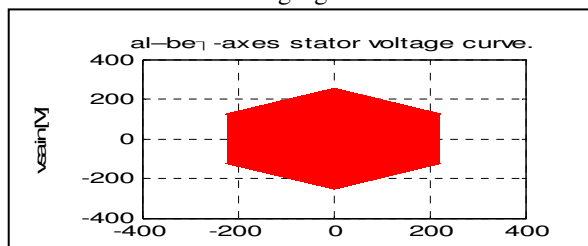


Figure 5: Stator voltage curve with classical PI.

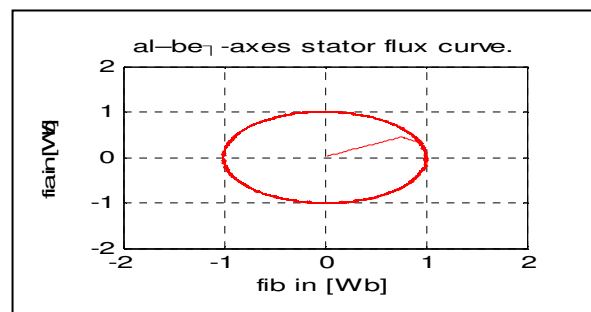


Figure 6: Stator magnetic flux curve with classical PI.

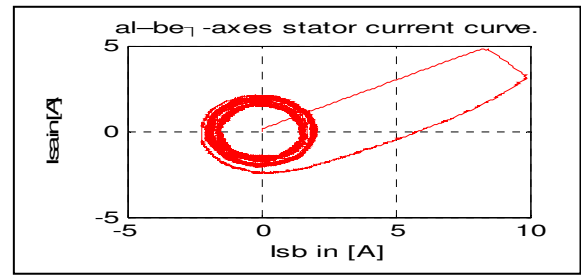


Figure 7: Stator current curve with classical PI.

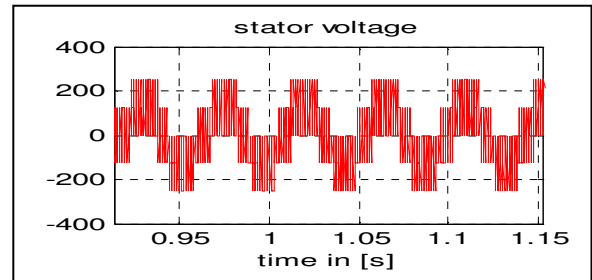


Figure 8: Stator voltage with classical PI.

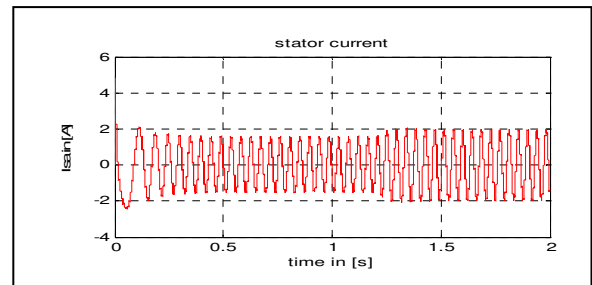


Figure 9: Stator current with classical PI.

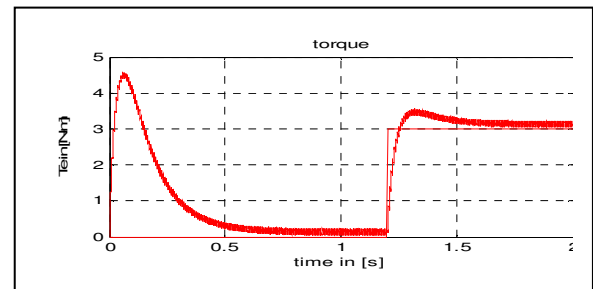


Figure 10: Torque with classical PI.

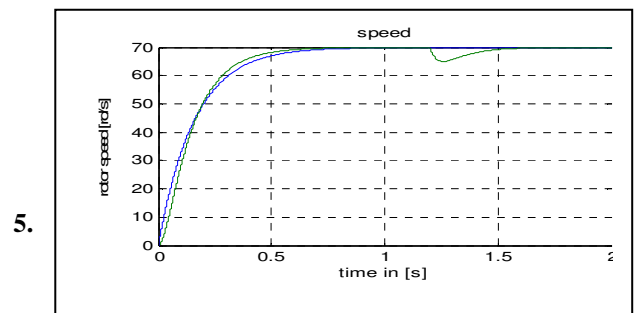


Figure 11: Rotoric Speed with classical PI.

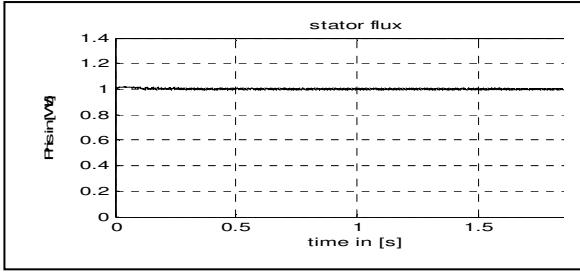


Figure 12: Stator magnetic flux with classical PI.

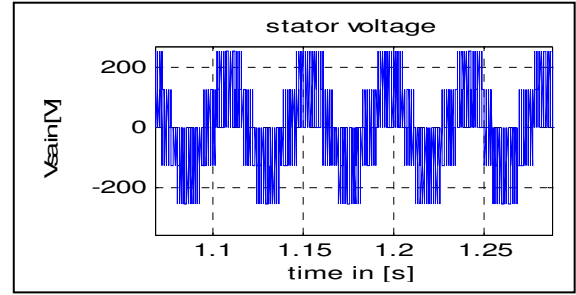


Figure 17: Stator voltage with fractional PI.

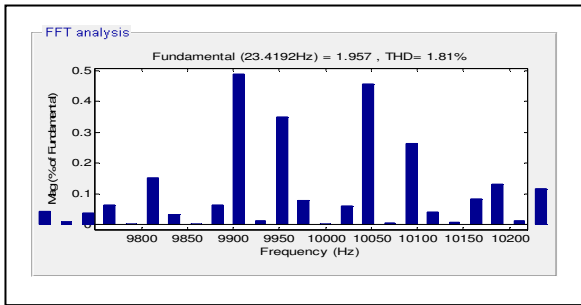


Figure 13: FFT of stator current with classical PI.

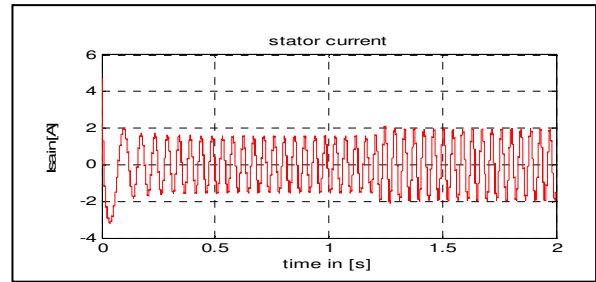


Figure 18: Stator current with fractional PI.

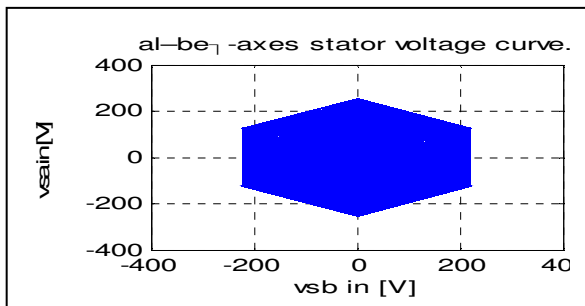


Figure 14: Stator voltage curve with fractional PI.

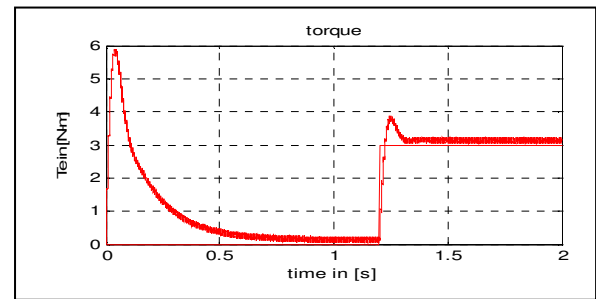


Figure 19: Torque with fractional PI.

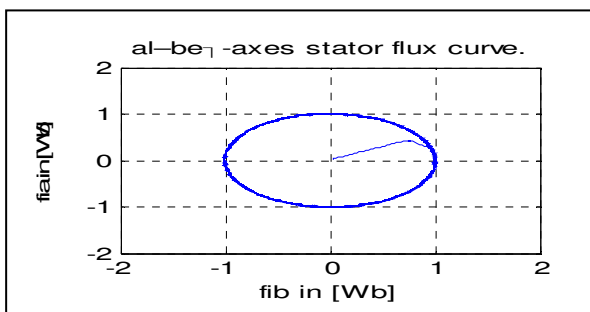


Figure 15: Stator magnetic flux curve with fractional PI.

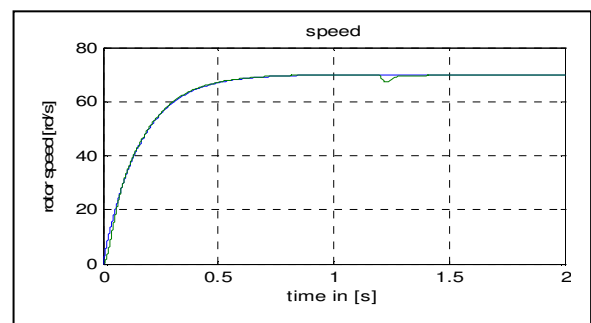


Figure 20: Rotoric Speed with fractional PI.

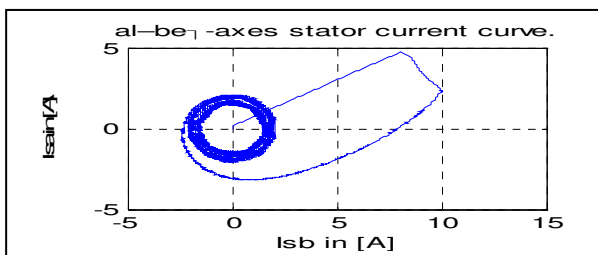


Figure 16: Stator current curve with fractional PI.

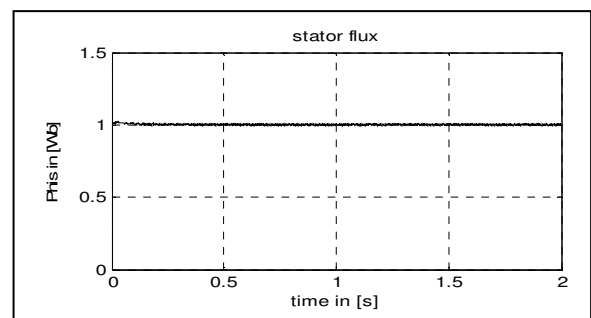


Figure 21: Stator flux with fractional PI.

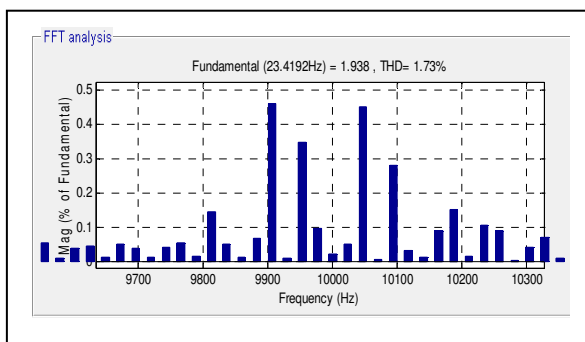


Figure 22: FFT of stator current with fractional PI.

The simulation results obtained starting from the DTC-SVM with classic PI controller applied to electromagnetic torque, flux magnetic and rotor's speed are illustrated in Figure 5 to 13. Figure 14 to 22 show the results obtained with fractional PI controller applied for rotor's speed in DTC-SVM control. We note that the results obtained with classic PI are more interesting than classic DTC (8)(9) and switch frequency is constant  $f=10$  kHz. We also notice that all results obtained with PI fractional applied in rotor's speed is more interesting than obtained with classic PI and the correction is more quickly with fractional PI controller (figure 20). We can also see that fractional PI controller gives good performances compared to classic PI and when the harmonic content (THD) is weaker (compared between figure 13 and figure 22), the form of the stator's current is closer to the form of a sinusoidal signal (figure 18).

### CONCLUSION

We presented in this paper a new approach of DTC control. We use in this approach the space vector modulation applied to the voltage's inverter, PI classic controller of torque and stator's flux and classic or fractional PI controller applied in the rotor's speed. According to the results obtained and the remarks observed before, we can conclude that the DTC-SVM with a fractional PI controller it's more efficient than DTC-SVM used classic PI controller for rotor's speed on the level of precision, response time and the robustness. Nevertheless there are some remaining points that should be considered in the future work for example the choice between the null vectors ( $V_0$  and  $V_7$ ) and to validate this control with PI fractional controller for torque, flux magnetic and rotor's speed.

### REFERENCES

Casadei, D. Serra, G. and Tani, A. 1996. Constant frequency operation of a DTC induction motor drive for electric vehicle. Sept, 10-12, pp.224-229 In proc, ICEM'96, vol.III, Vigo, Spain,.

Bin-jun, C. Xiao hong, N., 2012. A Novel MC-DTC Method for Induction Motor Based on Fuzzy-neural Network Space Vector Modulation. Journal of software, Vol. 7, N° 5, May 2012, pp 966-973.

Jadhav, S. V. Mohan Krishna Reddy, M., and Chaudhari, B. N. 2011. An Improved SVM-DTC

based Induction Motor Drive Scheme using d-SPACE. pp 388-393. IEEE PEDS, Singapore,

Depenbrock, D., 1988. Direct self control of inverter-fed induction motor. IEEE Trans. Power Electron. Vol. (3), pp.420-429.

Takahashi, I. and Noguchi, T., 1986. A new quick-response and high efficiency control strategy of an induction motor. IEEE Trans. Ind. Applicat., Vol. (22). pp.820-827.

Bettou, K. Charef, K.A., 2008. A New design method for fractional PI $\lambda$ D $\mu$  controller, IJSTA, vol.2, N° 1, pp 414-429.

Dingy, X. Chunna, Z., and Yang, Q.C., 2006. A Modified Approximation Method of Fractional Order System. pp.1043- 1048. IEEE International Conference on Mechatronics and Automation, June 25 - 28, 2006, Luoyang, China, ,

Mansour, M. Rachdi, S., Mansouri, M. N., Mimouni, M. F. 2012. Direct Torque Control Strategy of an Induction-Motor-Based Flywheel Energy Storage System Associated to a Variable-Speed Wind Generator. Journal of Energy and Power Engineering, Vol. (4), pp. 255-263.

Chikh, K. Saad, A. Khafallah, M. and Yousfi, D. 2011. A Novel Drive Implementation for PMSM By using Direct Torque Control with Space Vector Modulation. Canadian Journal on Electrical and Electronics Engineering, Vol.(2), pp. 400-408.

L. Bouras, Y. Zennir and F. Bourourou. 2013. Direct torque control with SVM based a fractional controller: Applied to the induction motor. Proceeding of the 3rd international conference on system and control pp.702-707. 978-1-4799-0275-0/13/\$3100 © IEEE, 29-31 October. Algeria.

Ortega, M., & al., 2005. Direct Torque Control of Induction Motors using Fuzzy Logic with current limitation, IEEE Industrial Electronics Society, IECON. 32nd Annual Conference.

Casadei, D. & al. 2000. Implementation of a Direct Torque Control Algorithm for Induction Motors Based on Discrete Space Vector Modulation, IEEE Trans. Ind. Applicat. vol. 15, No.4 pp. 769-777.

L. Bouras, M. Kadjoudj, N. Golea. 2005. Contrôle Direct du Torque Basé sur la Modulation Vectorielle avec Régulateurs à Hystérésis Appliqué à la Motor à Induction. pp 319-324. First International Conference on Electrical Systems, PCSE'05, May 9-11, O. E. Bouaghi Univ. Algeria.

# INFLUENCE OF PARKED CARS ON SMOKE PROPAGATION DURING CAR PARK FIRE

Weisenpacher, P.<sup>(a)</sup>, Halada, L.<sup>(b)</sup>, Glasa, J.<sup>(c)</sup>, Astalos, J.<sup>(d)</sup>

<sup>(a)</sup> Institute of Informatics, Slovak Academy of Sciences, Dubravská cesta 9, 84507 Bratislava, Slovakia

<sup>(b)</sup> Institute of Informatics, Slovak Academy of Sciences, Dubravská cesta 9, 84507 Bratislava, Slovakia

<sup>(c)</sup> Institute of Informatics, Slovak Academy of Sciences, Dubravská cesta 9, 84507 Bratislava, Slovakia

<sup>(d)</sup> Institute of Informatics, Slovak Academy of Sciences, Dubravská cesta 9, 84507 Bratislava, Slovakia

<sup>(a)</sup> [Peter.Weisenpacher@savba.sk](mailto:Peter.Weisenpacher@savba.sk), <sup>(b)</sup> [Ladislav.Halada@savba.sk](mailto:Ladislav.Halada@savba.sk), <sup>(c)</sup> [Jan.Glasa@savba.sk](mailto:Jan.Glasa@savba.sk), <sup>(d)</sup> [Jan.Astalos@savba.sk](mailto:Jan.Astalos@savba.sk)

## ABSTRACT

In this paper, the influence of number and shape of cars parked in underground car park on smoke propagation during the car park fire is examined. A series of high performance parallel computer simulations of fire scenarios is performed on cluster of computers using the NIST FDS system, version 5.5.3. Special emphasis is on evaluation of tenability conditions in the car park under different circumstances. Variable circumstances include different number and size of parked cars, ventilation configuration as well as fire source. Visibility at head level determined by simulation shows a significant impact of the cars number and shape on smoke propagation. The effect is more significant if cars are high enough to disrupt hot layer like vans and less significant for passenger cars. Air flow patterns differences of selected fire scenarios are analysed as well.

Keywords: fire dynamics, fire simulation, smoke propagation

## 1. INTRODUCTION

Computational Fluid Dynamics (CFD) is a powerful tool, which has been recently used for both theoretical research of fire and smoke movement and also in practice for design of fire safety measures. Smoke movement simulation in large, complex compartments with non-trivial geometry like underground car parks is especially important CFD application, in particular if there is a significant threat of fire with probable severe impact on people and property. Because of non-linear nature of fire, proper design of mechanical ventilation in car parks requires careful evaluation of its possible unintended effects. Temperature under ceiling, smoke back-layering and non-trivial flow patterns in case of car park fire is studied in (Deckers, Haga, Sette, and Merci 2013) by full-scale experiments and by CFD simulations in (Deckers, Haga, Tilley, and Merci 2013). Several fire parameters were varied in simulations including presence or non-presence of beam. The impact of parked cars on smoke propagation in the case of tunnel fire is dealt in (Glasa, and Valasek 2014).

However, due to complexity in flow and smoke patterns there is a difference between car parks and tunnels.

In this paper, the influence of parked cars on ventilation efficiency is evaluated and quantified. In order to increase the reliability of the simulation, relatively fine resolution of 6 cm is used, which requires significant computing resources. Therefore, parallel computation on cluster is used. The modelling of fire source is very realistic and based on experimental observation. Two passenger cars with very detailed representation are used in simulation as a fire source. Their representation includes interior equipment geometry and material properties, whose reliability is based on full scale experiments in open air including a fire in automobile interior and its spread onto a near standing vehicle (Polednak 2010; Svetlik 2010). During this experiment, temperatures inside and outside passenger compartment were measured to be used for the calibration of computer simulation. Subsequent simulations confirm credibility of this model (Weisenpacher, Glasa, and Halada 2012; Weisenpacher, Halada, Glasa, and Slizik, 2013) and the values of heat release rate (HRR) obtained by simulation are also in agreement with car fire experiments (Okamoto, Watanabe, Hagimoto, Chigira, Masano, Miura, Ochiai, Satoh, Tamura, Hayano, Maeda, and Suzuki 2009). The model enables to take into account mutual interaction between fire and car park ventilation as well as specific geometry of particular scenario, given by number and size of parked cars. Similar approach to car modelling for simulation purposes is used in (Partanen, and Heinisuo 2013).

## 2. FDS OVERVIEW

The Fire Dynamics Simulator (FDS) is CFD simulation system of fire-driven fluid flow, developed at National Institute of Standards and Technology (NIST), USA (McGratten, Baum, Rehm, Mell, McDermott, Hostikka, and Floyd 2010). FDS solves numerically a form of the Navier-Stokes equations appropriate for the low-speed, thermally-driven flow with an emphasis on the smoke and heat transport from fires. It includes also modules for physical and chemical processes like thermal radiation, pyrolysis, combustion of the pyrolysis

products, conductive heat transfer and fire suppression by sprinklers.

The basic set of equations includes conservation equations for conservation of mass, species, momentum and energy and equation of the state. These equations must be simplified in order to filter out sound waves, which are much faster than the typical flow speed. The final numerical scheme is an explicit predictor-corrector finite difference scheme, which is second order accurate in space and time. The flow variables are updated in time using an explicit second-order Runge-Kutta scheme. This part of the algorithm is highly suitable for parallel computation. Momentum equation is simplified and transformed into Poisson equation, which is solved in every time step.

All input data for a simulation are required in the form of a text file in the prescribed format, which describes the coordinate system, geometry of the domain and its location in the given coordinates, mesh resolution, obstacles, boundary conditions, material properties and different simulation parameters. Important limitations of the program is that the domain and all obstructions representing real objects, which can burn, heat up, conduct heat etc., should be rectangular, conforming with the underlying grid. Boundary conditions are prescribed on the walls and vents.

The strategy of parallel computation consists in decomposition of the computational domain into multiple meshes and computation of the flow in each mesh separately. Each computation is usually performed as an individual MPI process assigned to one CPU core. Values at mesh boundaries must be averaged in every time step in order to maintain solution stability. Therefore, the way of domain decomposition influence the precision of simulation results.

### 3. CAR PARK FIRE SIMULATION

Model of an underground single storey car park with dimensions 23.04 x 38.88 x 3.0 m (the total area of 895.8 m<sup>2</sup>) with parking places for 24 cars is used in simulation. Its geometry includes two concrete beams and eight columns, tube system under the ceiling, car park entrance, lift shaft and other alcoves. Two different metal ductwork systems for mechanical extraction of smoke (D1 and D2) or none ventilation are considered. The first scenario (D1, see Fig. 1) uses 9 inlet fans located in the proximity of the elevator and corridor entrance and 9 outlet fans, more or less regularly distributed within the ductwork system. Both inlet and outlet fans are represented by 96 x 48 cm surfaces with given normal air velocity. The second scenario (D2, see Fig. 2) uses 9 inlet fans (three of them with different location than in D1) and 12 outlet fans. Fans velocities are piecewise-linear functions of time given by Tab. 1.

Two burning cars are located in the middle part of the car park. Two types of cars are considered: passenger cars with maximal dimensions 421 x 180 x 150 cm and vans with maximal dimensions 534 x 192 x 210 cm. Seven scenarios are evaluated: scenario with empty car park (only 2 burning cars are included),

scenarios with half-full car park (10 additional cars are included, of which 0, 4 and 8 are vans) and scenarios with full car park (22 additional cars are included, of which 0, 4 and 8 are vans). As a main quantity influencing tenability conditions, visibility in head level is evaluated. Visibility averaged over the whole car park area is a good, though simplified aggregate measure of untenability evolution. Taking into account a simple shape of this quantity curves, their time averaged value is even simpler scalar quantity illustrating the rate of untenability rise in the car park.

Table 1: Inlet and outlet ventilation velocities [ms<sup>-1</sup>]

| Conf. | V <sub>180s,</sub> | V <sub>180s,</sub> | V <sub>360s,</sub> | V <sub>360s,</sub> | V <sub>600s,</sub> | V <sub>600s,</sub> |
|-------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|       | inlet              | outlet             | inlet              | outlet             | inlet              | outlet             |
| D1    | 1.5                | 2.0                | 1.5                | 3.0                | 2.0                | 4.0                |
| D2    | 1.5                | 2.0                | 1.5                | 2.0                | 2.5                | 3.0                |

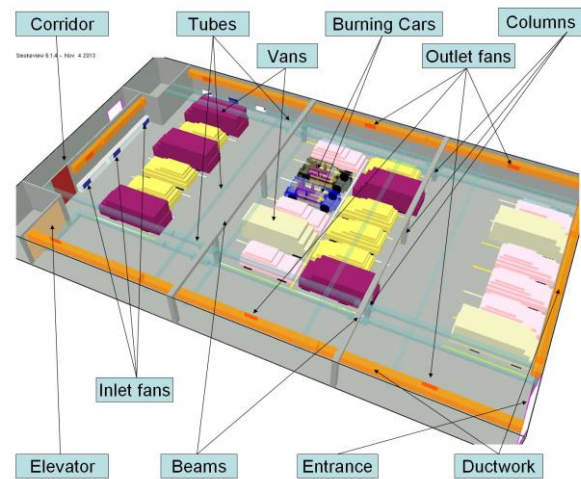


Figure 1: Scheme of car park with full number of parked cars including eight vans.

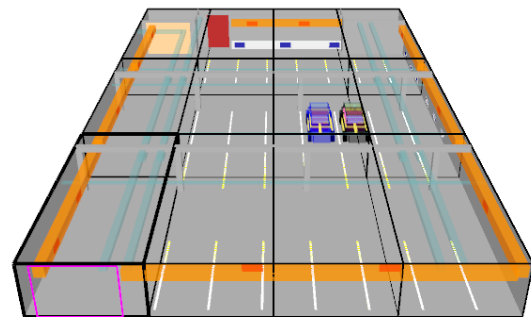


Figure 2: D2 ventilation configuration and the way of decomposition 12M (4-3-1).

In order to use parallel computation, FDS requires dividing the whole computational domain into particular computational meshes, each assigned to one CPU core. For all simulation scenarios, 12 meshes are used (12M, the way of decomposition 4-3-1, see Fig. 2), while in some cases also 6M (2-3-1), 36M (4-9-1) and 144M (8-9-2) simulations are performed to test reliability of parallel computation. The simulation



includes 384 x 648 x 50 cells (the total cells number is 12,441,600) with 6 cm resolution. 600 s of fire is simulated on SIVVP cluster at the Institute of Informatics SAS, Bratislava, Slovakia. It is an IBM dx360 M3 cluster with 54 computational nodes (2x Intel E5645 @ 2.4 GHz CPU, 48 GB RAM). The number of cores is 648. The total computational time is about nine days for 12M simulations.

#### 4. SIMULATION RESULTS

Time averaged values of visibility at head level (see Tab. 1 and Figures 3-9) show significant impact of car number on smoke propagation.

Table 2: Time averaged visibility for particular fire and ventilation scenarios and 12M

|                | $S_{no\ vent}$ [m] | $S_{D1}$ [m] | $S_{D2}$ [m] |
|----------------|--------------------|--------------|--------------|
| Empty          | 10.76              | 18.26        | 19.94        |
| Half-full-0van | 10.69              | 18.36        | 20.75        |
| Half-full-4van | -                  | 18.46        | 20.63        |
| Half-full-8van | -                  | 17.47        | 20.41        |
| Full-0van      | 10.66              | 18.02        | 19.89        |
| Full-4van      | -                  | 17.64        | 20.12        |
| Full-8van      | 10.35              | 16.38        | 19.43        |

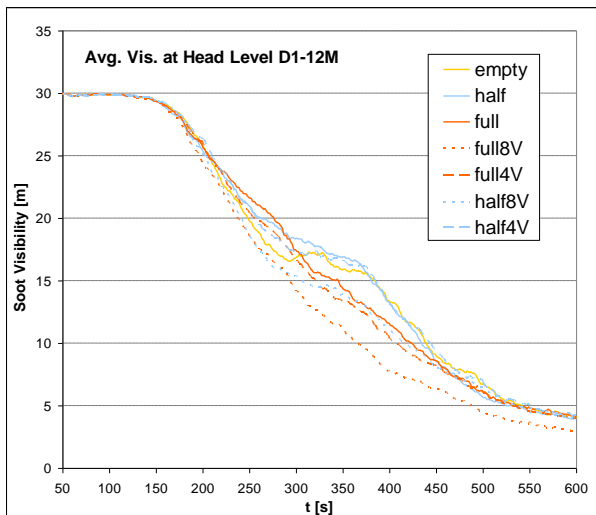


Figure 3: Averaged visibility for particular fire scenarios with D1 ductwork and 12 meshes

In D1 scenarios including only passengers cars, the impact of their number on average visibility decrease is small (maximal difference is 34 cm). Up to 300th s, "full" scenario provides slightly more efficient ventilation, probably due to dissipative effect of parked cars on inlet air flows (see Fig. 6), while in "empty" scenario fast inlet flow disrupt hot layer slightly. After 300th s, the low hot layer is disrupted by car bodyworks in "full" scenario (see smoke fluctuations in Fig. 5) and "empty" scenario increases its relative performance. "Half" scenario is the most efficient, as it combines advantages of other two scenarios. Some differences between smoke patterns can be also observed (see Fig. 5). In "empty" scenario there are two large compact smoke-free areas in the car park, whereas in "full"

scenario visibility fluctuates significantly within almost all car park area after smoke layer is torn.

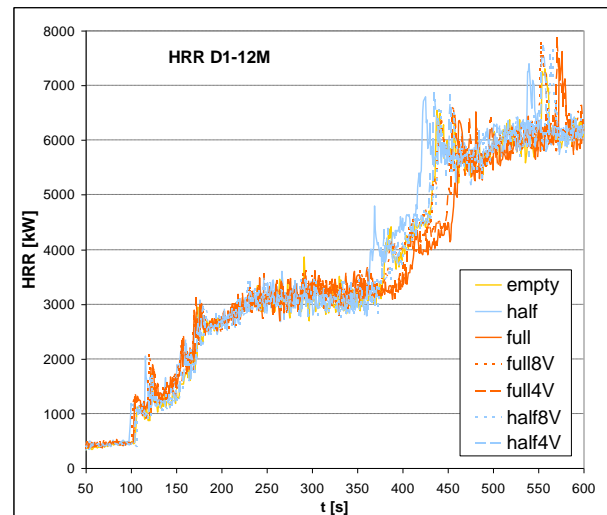


Figure 4: Heat Release Rate for particular fire scenarios with D1 ductwork and 12 meshes.

Inlet fans produce horizontal air flow in the car park slice, which is diverted towards the top of the slice slightly after it interacts with fire in the central part of the car park. Dissipated inlet flow in "half" scenario are diverted more considerably, which increases heat transfer towards the second car, accelerates its ignition (see Fig. 4) and increases smoke production in this scenario. Despite this fact this scenario is more efficient than other two. Its efficiency is even more clearly demonstrated if pool fires with the same HRR are used in all three scenarios and visibility difference of 0.71 m between "half" and "empty" scenario is then achieved. Note that efficiency is not surprising, as the fans velocities are optimised for half-full car park. Optimisation for other scenarios would lead to slightly different fans velocities, nevertheless, typical phenomena causing the differences between particular scenarios are still the same.

A different kind of behaviour can be observed in D2 configurations, as the phenomena observed in D1 are strengthened by specific configuration of D2 (see Figures 7-9). Again, "empty" scenario provides worse results than "full" in the first half of the fire and better results in the second half, though the difference is small. Scenario "half-full-0van" leads to significantly higher averaged visibility value than "empty" scenario. In the former one specific character of air flows created by the presence of the cars enhance the efficiency of the ventilation. In "empty" scenario there is a considerable flow leading from the left part of the car park slice to the central bottom part (see Fig. 9). It disrupt smoke layer and cumulates smoke on the left bottom part of the slice and after interaction with air flow leading from the portal also in the right part of the slice. In the "half" scenario, this flow is hindered and dissipated by cars and directed more horizontally to the central part of the slice, where it enhance the rotation of air mass and thus



also ventilation efficiency. Ignition of the second car occurs more or less at the same time in these two scenarios, however, in "full" scenario it occurs sooner, which exacerbate ventilation performance.

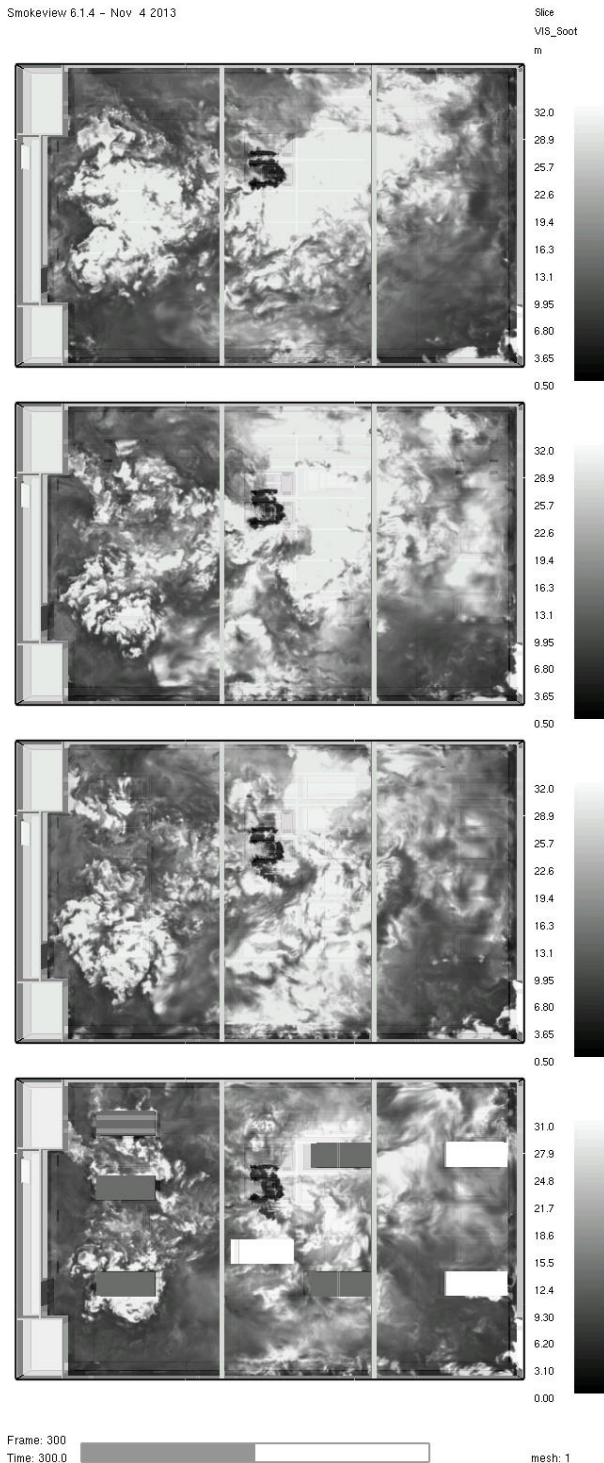


Figure 5: Visibility slices at human head level after 300 s for the fire scenarios "empty", "half", "full" and "full-8van" and D1 configuration.

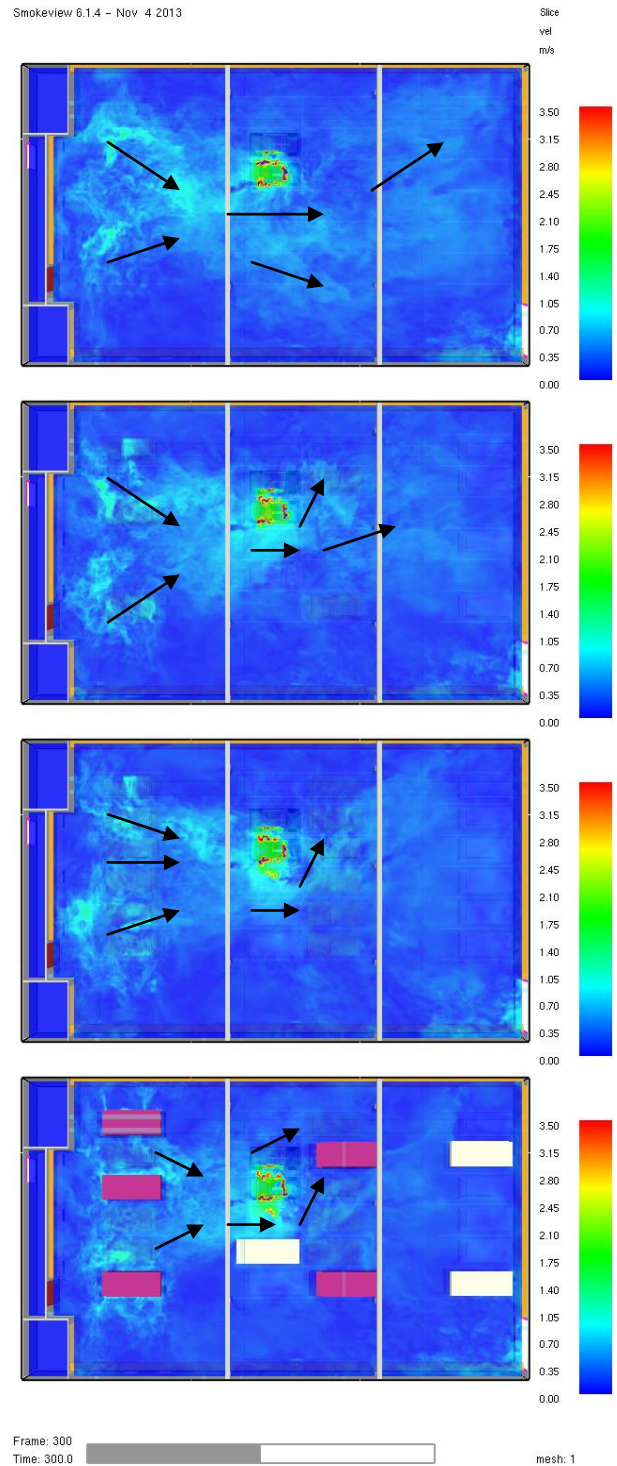


Figure 6: Velocity slices at head level after 300 s for the scenarios "empty", "half", "full" and "full-8van" and D1 configuration.

In scenarios without ventilation the difference between empty and full car park scenario is almost negligible (10 cm for 12M simulations) because of small air velocities. Smoke patterns are very similar as well.

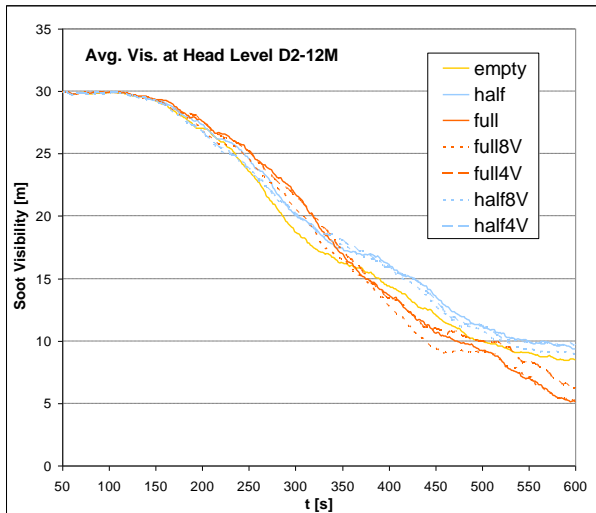


Figure 7: Averaged visibility for particular fire scenarios with D2 ductwork and 12 meshes.

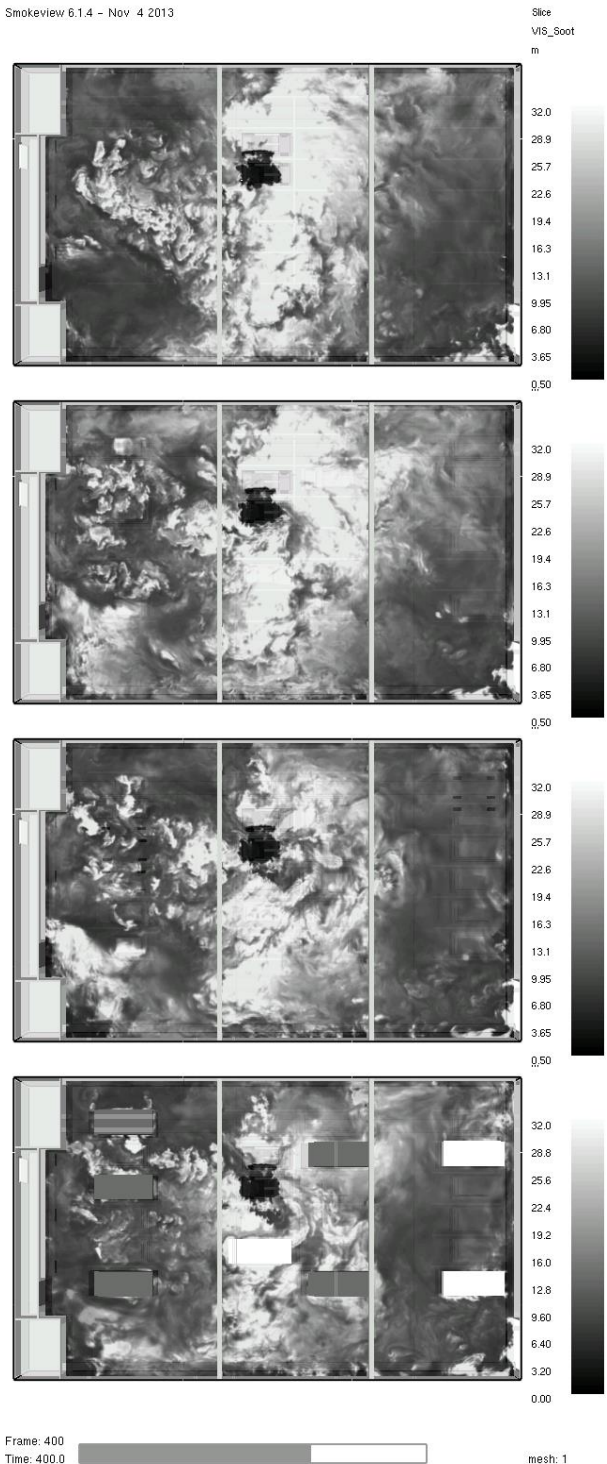
In configurations including also vans, the hot layer is disrupted, smoke descend to head level sooner and the average visibility decreases with increasing cars number. This effect is stronger in D1 than in D2 configuration. Visibility at transversal slice of the rear part of the car park is shown in Fig. 10.

In D1, the difference between "empty" and "full-8van" configuration is about 1.88 m. Average untenable conditions occur about 67 s sooner in latter scenario. The biggest differences between configurations occur between 300<sup>th</sup> and 400<sup>th</sup> s. At  $t = 367$  s "full-8van" conditions become untenable, while "half-0van" visibility is more than 6 m better.

In D2, the difference between "empty" and "full-8van" configuration is smaller, about 0.51 m. Untenable conditions occur about 60 s sooner, although visibility decreases very slowly in both cases in this phase of fire. In scenario "half-full-0van" untenable conditions occur even 50 s later than in "empty" scenario and visibility drops only slightly below 10 m. In general, the differences between scenarios behaviour are relatively small. As the fans velocities are not optimal for "full" scenarios, it is not surprising that the presence of vans can decrease them to more optimal values and thus enhance ventilation efficiency, which moderates general tendency of full scenarios to decrease visibility. In the case of "full-4van" scenario this tendency is even overcome and the averaged visibility is better than in "full-0van" (air flow pattern is similar as in scenario "full-8van", see Fig. 9). The impact of vans on air flow pattern is more significant in D2 scenario, in which vans limits the size of rotating whirl, while in D1 air flows mostly in horizontal direction between vans.

Note that better visibility in D2 scenarios is achieved because of specific configurations of burning cars. Air flows created by D2 avert flames from the second car and thus decrease the fire HRR. In general, D2 is not more efficient than D1.

Smokeview 6.1.4 - Nov 4 2013



Frame: 400  
Time: 400.0

mesh: 1

Figure 8: Visibility slices at human head level after 400 s for the fire scenarios "empty", "half", "full" and "full-8van" and D2 configuration.

Averaged visibility does not reflect the local character of smoke concentration. In the early phase of the fire in scenarios including vans, smoke cumulates not only on the same locations as in "empty" scenarios, but also in the proximity of vans (see Fig. 11). However, the general pattern of smoke cumulation is very similar in all cases, although in "vans" scenarios smoke cumulation is accelerated. The largest difference



in D1 occurs at about 300<sup>th</sup> s, when in "empty" scenario a large smoke free area is created in the central part of the car park, while in "vans" scenario the area is filled with smoke (see Fig. 4). Similar tendencies, although less considerable can be observed also in D2 configuration (see Fig. 8).

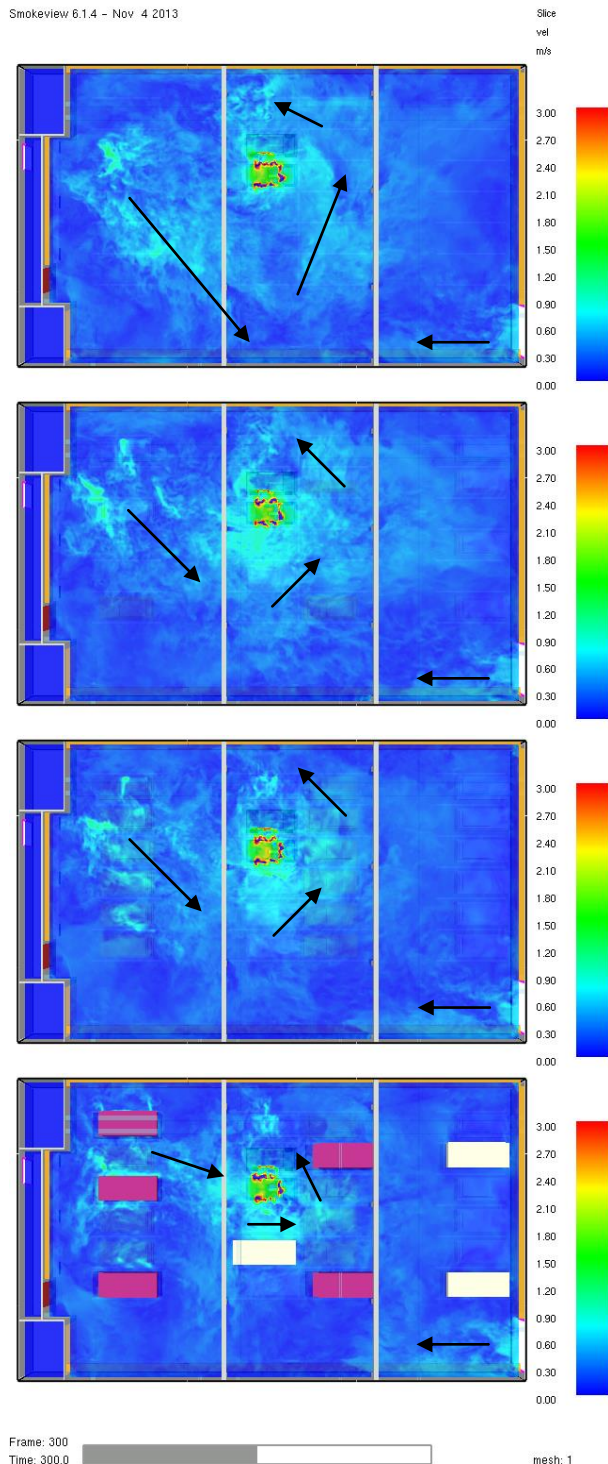


Figure 9: Velocity slices at head level after 300 s for the scenarios "empty", "half", "full-8van" and D2 configuration.

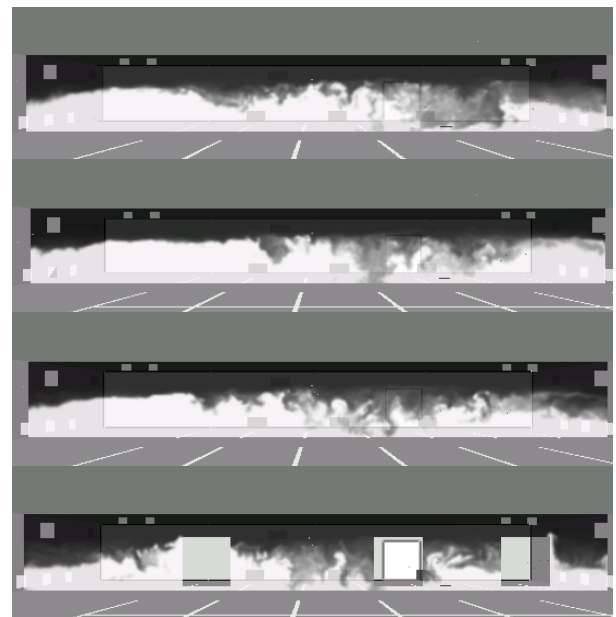


Figure 10: Visibility at transversal slices of the rear part of the car park after 350 s for the fire scenarios "empty", "half", "full" and "full-8van" and D1 configuration.



Figure 11: Fire scenarios "full" and "full-8van" and D1 configuration. Creation of smoke clusters in vans proximity at t = 140s.

## 5. PARALLEL COMPUTATION

Simulations 144M tend to overestimate averaged visibility values as well as differences between empty and full scenarios (see Tables 3 and 4). Fewer mesh number results in better precision as well as smaller

differences between particular scenarios results. However, even simulations 144M provide good visibility estimates which are precise enough to conclude visibility decrease in "full" scenarios correctly. Typical difference between 6M and 12M simulations is only several centimeters for scenarios without ventilation (small air flows, see Table 3) and up to 50 centimeters for D1 scenarios (faster air flows, see Table 4). In Fig. 12 can be seen that the differences between simulations 6M, 12M and 144M increase after 350th s of fire with increasing fans velocity. It suggests that faster flows result in worse simulation precision, if higher number of meshes is used. Typical duration of 12M simulations is about 9 days, up to 1.5 days for 144M simulations and more than 11 days for 6M simulations.

Table 3: Time averaged visibility for particular fire scenarios and mesh decomposition without ventilation.

|           | 6M     | 12M   | 36M   | 144M  |
|-----------|--------|-------|-------|-------|
| Empty     | 10.73* | 10.76 | 10.92 | 11.34 |
| Half-full | 10.66* | 10.69 | 11.04 | 10.78 |
| Full      | 10.71* | 10.66 | 10.90 | 10.77 |
| Full-8van | 10.49  | 10.35 | 10.84 | 10.84 |

Table 4: Time averaged visibility for particular fire scenarios and mesh decomposition and D1 ventilation.

\* - Windows PC computation.

|                | 6M    | 12M   | 144M  |
|----------------|-------|-------|-------|
| Empty          | 18.70 | 18.26 | 19.07 |
| Half-full-0van | 18.81 | 18.36 | 19.51 |
| Half-full-4van | -     | 18.46 | 18.67 |
| Half-full-8van | 17.52 | 17.47 | 17.81 |
| Full-0van      | 18.39 | 18.02 | 18.62 |
| Full-4van      | -     | 17.64 | 18.47 |
| Full-8van      | 16.80 | 16.38 | 17.25 |

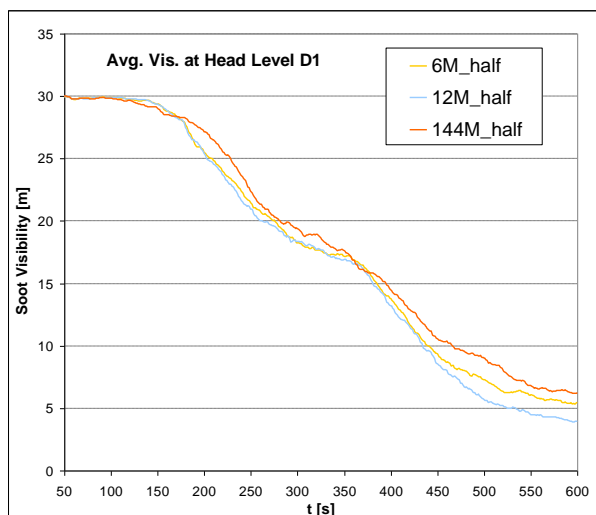


Figure 11: Averaged visibility for "Half-full-0van" scenario with D1 ductwork and 6M, 12M and 144M mesh decomposition .

## CONCLUSION

Series of simulations of car park fire have been performed for two ventilation configurations and seven scenarios with different number and size of parked cars. Simulations have confirmed and quantified significant impact of parked vans on smoke visibility decrease, which can accelerate the rise of untenability by about 60 s for some specific scenarios (averaged visibility differences of up to 2 m) due to smoke layer disruption. If only passenger cars with lower bodyworks are included, differences about dozens of seconds can be expected for scenarios with different car number (averaged visibility differences of up to 1 m). The number and locations of cars influence inlet air flows, which can be hindered, dissipated, strengthened or redirected. Such differences influence the ventilation efficiency, fast flows can even disrupt smoke layer when it is low enough. Tops of car bodyworks tear smoke layer after it descend at that level in later phase of the fire, which results in more significant visibility decrease in "full" scenarios.

## ACKNOWLEDGMENTS

The authors would like to thank to P. Smardz for fruitful discussions, to P. Polednak, J. Svetlik, M. Simonova and J. Flachbart for the organization of automobile fire experiments and to M. Dobrucky and V. Sipkova for technical support during FDS simulations. This paper was supported by VEGA (project VEGA 2/0184/14).

## REFERENCES

- Deckers, X., Haga, S., Sette, B., Merci, B., "Smoke control in case of fire in a large car park: Full-scale experiments", *Fire Safety Journal* 57: 11–21 (2013)
- Deckers, X., Haga, S., Tilley, N., Merci, B., "Smoke control in case of fire in a large car park: CFD simulations of full-scale configurations", *Fire Safety Journal* 57: 22–34 (2013)
- Glasa, J. and Valasek, L., "Study of Applicability of FDS+Evac for Evacuation Modeling in Case of Road Tunnel Fire", *Research Journal of Applied Sciences, Engineering and Technology* 7(17): 3603-3615 (2014).
- McGratten, K., Baum, H., Rehm, R., Mell, W., McDermott, R., Hostikka, S. and Floyd, J.: *Fire Dynamics Simulator (Version 5), Technical Reference Guide*, NIST Special Publication 1018-5, NIST, Gaithersburg, Maryland, USA, 2010.
- Okamoto, K., Watanabe, N., Hagimoto, Y., Chigira, T., Masano, R., Miura, H., Ochiai, S., Satoh, H., Tamura, Y., Hayano, K., Maeda, Y. and Suzuki, J., "Burning Behaviour of Sedan Passenger Cars", *Fire Safety Journal* 44:301-310 (2009)
- Partanen, M. and Heinisuo, M., 2013. Car fires with sprinklers: A study on the eurocode for sprinklers. *Proc International Conference Applications of Structural Fire Engineering*, pp. 23-28. April 19-20, Prague, Czech Republic.

- Polednak, P., 2010. Experimental Verification of Automobile Fires (in Slovak). *Proceedings of the 4th International Conference on Fire Safety and Rescue Services*. June 2-3, Zilina, Slovakia.
- Svetlik, J., 2010. Fire in the passenger car engine compartment (in Slovak). *Proceedings of the 4th International Conference on Fire Safety and Rescue Services*. June 2-3, Zilina, Slovakia.
- Weisenpacher, P., Glasa, J. and Halada, L., 2012. Parallel simulation of automobile interior fire and its spread onto other vehicles. *Proceedings of Fire Computer Modeling: International Congress*, pp. 329-338. October 19, Santander, Spain.
- Weisenpacher, P., Halada, L., Glasa, J. and Slizik, P., 2013. Smoke Propagation In Car Park Fire: A Parallel Study. *Proceedings of Eight Mediterranean Combustion Symposium*. September 8-13, Cesme, Izmir, Turkey.

#### AUTHORS BIOGRAPHY

**Peter Weisenpacher** studied theoretical physics at Comenius University, Faculty of Mathematics and Physics, Department of Theoretical Physics, Bratislava, Slovakia and received PhD. in 2003. He works as research scientists at Slovak Academy of Sciences, Institute of Informatics, Department of Numerical Methods and Algorithms. His current research interests include computational fluid dynamics, fire computer simulation and parallel computing. He participates in various research projects on fire simulation.

**Ladislav Halada** graduated in mathematics in 1971, received the RNDr. degree in mathematics and physics from the Faculty of Mathematics and Physics of Comenius University in Bratislava, and the C. Sc. degree (equivalent to Ph.D.) in mathematics in 1982 also from this university. He is Associate Professor from 1991. During 1972-2001 he worked on different scientific institutions and universities in Slovak Republic and also abroad. He is co-author of scientific books and numerous scientific papers.

**Jan Glasa** graduated in numerical mathematics in 1986, received the RNDr. degree in numerical mathematics and optimization methods and algorithms in 1986 at the Faculty of Mathematics and Physics of Comenius University in Bratislava. He has received the C. Sc. degree (equivalent to Ph.D.) in computer science at the Slovak Academy of Sciences. He is with the Institute of Informatics of Slovak Academy of Sciences in Bratislava as senior scientist. He serves as head of scientific council of the institute. His current research interests include mathematical modelling and computer simulation of fires, theory of curves, efficient algorithms, and parallel computing. Currently he is the leader of computer fire simulation project at the Institute of Informatics of Slovak Academy of Sciences.

**Jan Astalos** studied informatics at Slovak University of Technology, Faculty of Electrical Engineering and Information Technology, Bratislava, Slovakia. He works as research scientists at Slovak Academy of Sciences, Institute of Informatics, Department of Parallel and Distributed Information Processing. He focuses on high-performance computing systems and distributed computing infrastructures.

# ProOpter, production dynamics analysis and optimization tool

Gašper Mušič<sup>(a)</sup>, Miha Glavan<sup>(b)</sup>, Dejan Gradišar<sup>(b)</sup>, Stanko Strmčnik<sup>(b)</sup>

<sup>(a)</sup>University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

<sup>(b)</sup>Jozef Stefan Institute, Ljubljana, Slovenia

<sup>(a)</sup>[gasper.music@fe.uni-lj.si](mailto:gasper.music@fe.uni-lj.si)

## ABSTRACT

The paper deals with analysis and optimization of production performance dynamics expressed by key performance indicators (KPI). The architecture of a supporting software tool that integrates into production information systems is presented. The tool is able to support efficiency indicators composition, determination of performance influential variables and building of prediction model that enables a short-term prediction of the expected performance. The model can be used within predictive control algorithms, composing a production management decision support, which assists production operators at on-line production management. The tool is built modularly and employs industrial standards that facilitate the tool integration into production information systems of different vendors. The functionality of the tool is demonstrated by simulation benchmark as well as by real industrial production data.

Keywords: Production control, production efficiency, key performance indicators, modelling, simulation, optimization

## 1. INTRODUCTION

Manufacturing companies are faced with several challenges in their struggle to gain a competitive advantage. Among others, energy and resource efficiency are becoming increasingly decisive factors for competitiveness, companies are forced to shorten innovation cycles, and markets are becoming more volatile.

These challenges can be addressed by the extensive use of contemporary information technology within all organization levels of a company. A careful orchestration of the related information system improvements throughout the company has to be maintained in order to avoid bottlenecks on the one hand, and oversized capacity on the other. In any case, production itself is a base of any manufacturing company operation. Effective production management is one of the fundamental operational activities that has to be carefully designed and integrated into the overall management structure in order to meet the given requirements.

A number of initiatives emerged with a goal to establish a continual production efficiency improvement framework by the use of the latest technological advancements. In Europe, the German Industry 4.0 initiative (Industrie 4.0

2013) is perhaps the most known. It proposes to employ the Internet of Things paradigm on the factory floor and derive intelligent, intercommunicating autonomously operating production units, co called cyber-physical systems. This should result in flexible and efficient Smart Factory, with consideration of ergonomics and customer needs, and integration of supply-chain partners along the value chain (Brettel et al. 2014).

A similar North American initiative is Smart Manufacturing Leadership Coalition (SMLC 2011), a coalition of companies, manufacturing consortia and consultants working on Smart Manufacturing. This has been defined as the dramatically intensified application of 'manufacturing intelligence' throughout the manufacturing and supply chain enterprise (Davis et al. 2012).

In terms of organizational advancements Japanese manufacturing has long tradition dating back in 1980's with Just-in-time manufacturing and lean production (Holweg 2007). Recently, also other Asian manufacturers are forced to adopt these advancements as they are faced with rising consumer sophistication that makes traditional low cost mass production strategies inadequate.

While these initiatives improve the organization of the production process and establish a decision-support framework with improved insight into the current state of production and its current performance, an obvious further advancement is to employ sophisticated data analysis and system modelling methods in order to predict the effect of production control measures on the future performance. This way a decision-making process at the production control level would be substantially improved by the possibility of beforehand evaluation of the various decisions.

To this end a holistic production control concept is proposed in Zorzut et al. (2009) and further elaborated in Glavan et al. (2013a) and Glavan et al. (2013b). The proposed modelling approach uses historical production performance data jointly with history of most influential decision variables to derive a black-box model by contemporary soft computing techniques, e.g. neural networks. Several additional steps are provided, such as data preparation, influential variable selection, etc. (Glavan et al. 2013a).

This paper describes the design and prototype implementation of a software tool that supports the application of modelling methodology in the context of production performance prediction and optimization. This is one of



the fields where information technology has an immediate and considerable impact on the efficiency and quality of production control and related manufacturing processes. The tool modular architecture is described and the used industrial standards are presented, which facilitate the tool integration into production information systems of different vendors. The functionality of the tool is demonstrated through a selection of results from case studies, which include simulation benchmark as well as historical data from the actual industrial production.

## 2. HOLISTIC PRODUCTION CONTROL

The presented tool is based on the holistic production control concept, which attempts to bring classical feedback control approach, extensively used on the factory floor, also to the higher production control levels.

The concept of holistic production control can be best explained by the scheme depicted in Figure 1. The process we would like to control is indicated by the block *Production process*. Note that this block also covers the low-level process control. Different inputs ( $u$ ) are available to manipulate the production process. These inputs are mostly the reference values for the process control loops and/or other manipulating variables not used within the stabilisation loops. On the other hand, there are many measurable disturbances ( $d$ ) and outputs ( $y$ ), which are used to calculate on-line the key performance indicators ( $K$ ) – *KPI calculation* block. The key performance indicators (KPIs) are the production variables that are used by production manager to determine the appropriate input values ( $u$ ) in order to optimise the production process. The demands from the business control level are given as reference KPIs ( $K^*$ ). The attempt of the concept described here is to help the production manager with the decision-making process, which would close the loop through the introduction of the *Production controller & Optimiser*.

One of the possible solutions with this approach is to apply model-based control concept on the production management level. To enable this, an appropriate model describing the behaviour of the process projected on KPIs is required (the *KPI model M*). The model can be updated online and can provide the controller with the predicted outputs  $K|M$ . The model can also consider the measurable disturbances ( $d$ ). Based on current values  $K$ , the predicted outputs  $K|M$  and the reference values  $K^*$  (i.e., the planned business goals) the *Production controller & Optimiser* determines the appropriate input values  $\hat{u}$  and in this way supports (or substitutes) the production manager (Eq. (1)).

$$\hat{u} = \arg \min_{u \in R} C(K, K^*, K|M) \quad (1)$$

This concept has been introduced in Zorzut et al. (2009), and further elaborated and extended in Glavan et al. (2013a) and Glavan et al. (2013b).

How to derive an appropriate KPI model represents the main challenge of the HPC approach.

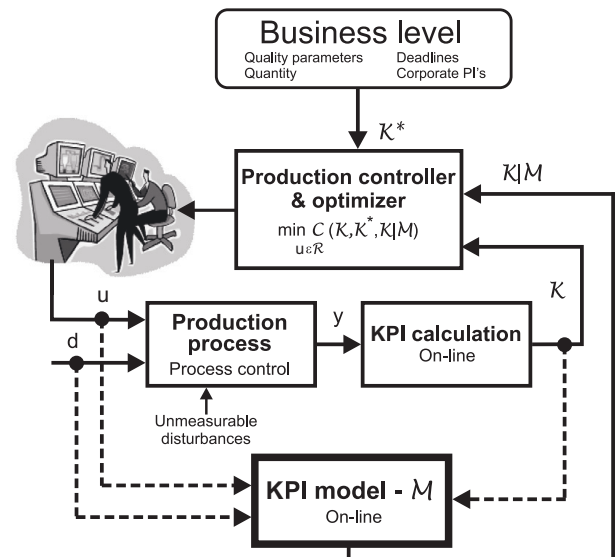


Figure 1: Holistic production control

### 2.1. Production modelling for holistic production control

Modern manufacturing systems are in many cases and for various reasons too complex to be accurately described analytically from first principles. Instead, one can assume that the relationship between the inputs and the outputs can be described by a stochastic, high-dimensional model from a class of generally nonlinear model structures.

The production model has to include enough details of the production process to reflect the dynamics for production control. This model should be relatively simple in comparison to the models used for the process control level, yet because of the overall complexity and the limitations of testing the process, this task is extremely complex. Production control usually requires that the model is easy to adapt online as well. Therefore, the main objective is the development of the concept of identifying a relatively simple input–output model of the production.

The main steps of the production modelling are shown in Figure 2. More detailed discussion about each of these steps and a short overview of the potential methodologies are given in the following subsections.

#### Data preprocessing

Special attention is needed when data from a historical production database are used. From the vast amount of data, the informative portions need to be identified. These data segments should cover the interesting dynamics of the KPIs, for which we would like to determine the future behaviour. Furthermore, any outliers or missing data need to be properly substituted, and to cover all the process operating conditions an uneven data distribution is needed.

Data-cleaning procedures can be applied to detect and remove any outliers present in the data. As pointed out in Pearson (2006), nonlinear data-cleaning procedures are recommended. We can find many filters in the literature proposed for this task: the Martin–Thomson filter, the FIR-median hybrid (FMH) filter, the Hampel filter, etc.



Figure 2: Holistic production control design steps

### Definition of production performance indicators – KPIs

The research field of performance measurement systems (PMSs) is becoming increasingly important for industry and academics (Neely et al. 1995). Performance indicators (PIs) are commonly used by organisations to evaluate their overall economic success and many recommendations have been presented on how to specify such indicators (Folan and Browne 2005). These indicators should cover the relevant business aspects of the specific process, where the status of the process should be evaluated with relatively short-term indicators, since longer-scale business estimations (e.g., an annual profit report) are useless for quick process adjustments (Gerry and Buckbee 2005).

The production objectives are usually aggregated in key production performance indicators (KPIs). The selection of these KPIs should be performed manually, with extensive consideration of the production expert’s knowledge of the process.

### Input variables selection

To simplify the model and to enhance model’s accuracy only the most relevant manipulative variables need to be identified. On the basis of the historical process data, an extensive analysis is needed to evaluate which inputs have the most significant impact on the selected KPIs.

Variable selection is already a widely applied methodology in the field of data mining. But, as noted by some authors, like Smits et al. (2006), in modelling projects it is mostly assumed that true inputs are a-priori known or all the available inputs are used in a model. To avoid the so-called curse-of-dimensionality, which essentially limits the robustness of a data-based model, only the most relevant inputs need to be selected. This represents an especially important step for HPC design, as in real-world production processes many potential variables are available. Furthermore, as aggregated KPIs are connected with many process variables it is often found that some a-priori excluded inputs are later identified as significant, and vice versa.

In the literature, three major principles for variable selection are used (Guyon and Elisseeff 2003):

- feature construction,
- variable ranking,
- variable subset selection.

For detailed discussion and evaluation of the related methods, see Glavan et al. (2013b).

As we are dealing with dynamical systems, the current values of production performance indicators are not dependent only on the current input values, but also on their time-delayed values. The input–selection problem is therefore augmented by the selection of lagged inputs and outputs that are used as regressors.

### Black-box process modelling

The HPC efficiency is closely related to the production model, which should describe the main features of the production process with an acceptable level of approximation. The production process is typically a highly complex process, with nonlinear relationships among the vast quantity of process variables. Since our model should be simple enough and the development time needs to be short, black-box modelling techniques are preferred. Furthermore, the production model should be extracted mainly from the historical process data, since extensive experimentation on the real process is often too expensive or restricted. If the process characteristics were to change during the use of the production model, new process data should be analysed and a better model extracted. The cyclical generation and validation of new models will enable a rather conservative adaptation of the model-in-use to long-term changes in the production.

The main idea of the parametric black-box modelling techniques is to trim some universal input–output functions, with a fixed number of parameters, to represent the true process dynamics (2).

$$y(t) = g(\varphi(t), \vartheta) + e(t) \quad (2)$$

The goal is to minimise the mismatch  $e(t)$  between the true process response  $y(t)$  and the model prediction  $g(\cdot)$ , where the trimming is performed solely on the basis of the process input–output data pairs  $Z^N = \{u(t), y(t)\}_{t=1}^N$

Black-box models can be seen as the connection of two mappings (Sjöberg et al. 1995). The first mapping constructs the regression vector  $\varphi(t)$  from past inputs and outputs, which enables a representation of the dynamical model behaviour. Another mapping predicts the future behaviour of the system  $\hat{y}(t)$ , with the nonlinear mapping from the regressor space to the system output. This nonlinear mapping is found within a family of functions (3), parameterised by the parameter vector  $(\vartheta)$ , where  $g_k$  refers to the basis function, usually derived from a single, mother basis function.

$$\hat{y}(t) = g(\varphi(t), \vartheta) = \sum_{k=1}^N \alpha_k g_k(\varphi(t), \beta_k, \gamma_k) \quad (3)$$

$$\vartheta = [\alpha_1 \dots \alpha_n, \beta_1 \dots \beta_n, \gamma_1 \dots \gamma_n]^T$$

From such a flexible structure, popular nonlinear mappings can be derived, like Neural Networks, Wavelets, Kernel Estimators, Nearest Neighbors, B-splines, Fuzzy models, etc. (Ljung 1999).

### 2.2. Optimization and holistic production control

Optimization is a vital component of holistic production control concept. Once a production performance dynamic model is derived, a natural way of incorporating the model in a production control scheme is to apply the model predictive control (MPC) approach (Maciejowski 2002, Grüne and Pannek 2011).

The MPC approach is based on repetitive solution of an optimal control problem, taking measured system state as the initial state and using system model to evaluate the

effect of possible system input sequences. A discrete time representation of system dynamics is used and only the first sample of the calculated optimal input sequence is applied to the system. In the next sample time the calculation is repeated with the newly acquired system state.

The optimization is performed over a finite moving horizon, which always starts at the current sampling instant. Depending on the method, prediction and control horizons can be of different lengths. The approach can be applied to feedback control of nonlinear systems, the method is then denoted NMPC.

For simplicity, we assume prediction and control horizon of length  $N$ . The determination of the optimal production manipulative values  $\hat{u}$  can therefore be formulated as an optimal control problem

$$\begin{aligned} & \text{minimize} && J_N(x_0, u(\cdot)) := \sum_{k=0}^{N-1} l(x_u(k, x_0), u(k)) \\ & \text{with respect to} && u(\cdot) \in \mathbb{U}^N(x_0) \\ & \text{subject to} && x_u(0, x_0) = x_0, \\ & && x_u(k+1, x_0) = f(x_u(k, x_0), u(k)) \end{aligned} \quad (4)$$

Here  $x_0 = x(n) \in X$  is the current state of the system at the sampling instant  $n = 0, 1, 2, \dots$ ,  $u(\cdot) \in \mathbb{U}^N(x_0)$  is a control sequence where  $\mathbb{U}^N(x_0) \subseteq U^N$  is a set of admissible control sequences over which we optimize,  $l(x, u)$  is a distance (deviation) measure,  $J(x, u)$  the cost function, and  $f(x, u)$  is the nonlinear state transition map representing discrete time dynamics.

When the solution is obtained,  $\hat{u} = u(0)$  is chosen. For a non-constant reference and in the presence of terminal constraints the formulation is slightly modified, but the general idea remains the same.

In any case, the NMPC problem can be reformulated to match the standard problem in nonlinear optimization (NLP) (Grüne and Pannek 2011)

$$\begin{aligned} & \text{minimize} && F(z) \\ & \text{with respect to} && z \in \mathbb{R}^{n_z} \\ & \text{subject to} && G(z) = 0 \quad \text{and} \quad H(z) \geq 0 \end{aligned} \quad (5)$$

with maps  $F : \mathbb{R}^{n_z} \rightarrow \mathbb{R}$ ,  $G : \mathbb{R}^{n_z} \rightarrow \mathbb{R}^{r_g}$  and  $H : \mathbb{R}^{n_z} \rightarrow \mathbb{R}^{r_h}$ . Here  $r_g$  and  $r_h$  denote the resulting number of equality and inequality constraints, respectively.

This reformulation, called discretization, can be done in different ways. Among them, the so-called recursive discretization recursively computes  $x_u(k, x_0)$  from the open-loop system dynamics outside of the optimization problem, thus maintaining low dimension of the optimization variable  $z$  and the number of constraints  $p = r_g + r_h$ . The optimization variable  $z$  reduces to  $z := (u(0)^T, \dots, u(N-1)^T)$ .

Utilizing this form, common solution methods for NLP problems can be applied, such as Sequential Quadratic Programming (SQP) or Interior Point Methods (IPM).

The recursive discretization is optimal regarding the number of optimization variables and constraints. Still,

the method has some drawbacks compared to the full discretization. In particular, it is difficult to use the information on current control sequence solution in search for good initial guess in the next step, and the solution  $x_u(k, x_0)$  may depend very sensitively on the control sequence  $u(\cdot)$ , in particular when  $N$  is large (Grüne and Pannek 2011).

The advantage of the method is that it separates the NLP solver from the solver of the dynamics. This is particularly well suited to models derived by soft-computing techniques, because an arbitrary numeric implementation of the model can be used. In the case of HPC, this enables to use the above mentioned black-box models in the form of nonlinear mappings in a straightforward way.

### 2.3. Software support

To assist the modeller in deriving production performance model for the HPC, and facilitate the application of the NMPC methods within HPC, a user-friendly software tool is needed. Using such a support tool, the system integrator and the production manager would have the possibility to identify a production model based on the historical operational data of the process, and integrate it in a model-based HPC solution.

The main purpose of the developed tool is to facilitate implementation of performance indicators, analyze the main performance influences, automate the model-development procedure and to support the manipulation and maintenance of already existing models, as well as their use in production optimization. As the potential users of such a tool are non-modelling experts (e.g., production managers), the program tends to simplify the model identification and control optimization procedures, where the user would not need to understand a detailed identification and control theory.

## 3. PROOPTER - PRODUCTION DYNAMICS ANALYZER AND OPTIMIZER

ProOpter extends the functionality of classical Manufacturing Execution Systems (MES) with embedded intelligence. It enables the analysis of production dynamics using complex analytical functions and application of advanced production control concepts that are based on embedded models. The Production dynamics analyser and optimizer represents an upgrade of classical MES systems, and thus increases the functionality and efficiency of production information systems.

### 3.1. Architecture

ProOpter enables the analysis of the production dynamics using advanced methods like data mining, data reduction, determination of relevant manipulated variables and production performance indicators model identification, as described in Section 2. The obtained models enable a short-term prediction of production dynamics, which is the basis for production optimization. The general architecture is shown in Figure 3.

ProOpter is composed of several modules, where some of them are used on-line, and others off-line. Connectivity with classical MES systems is enabled by standard IT interfaces, described in the following subsections.

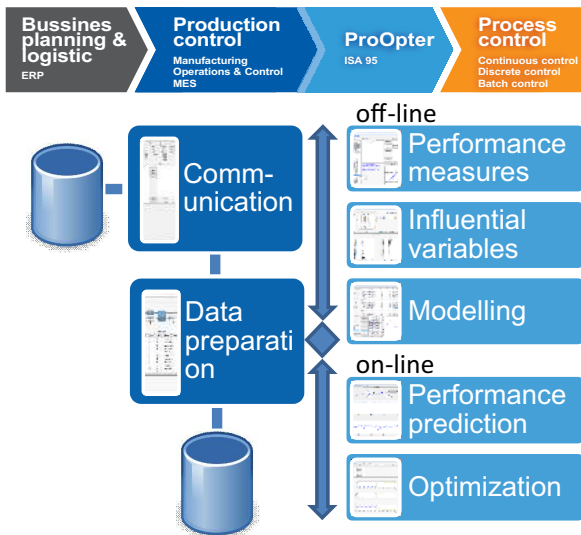


Figure 3: ProOpter architecture

### 3.2. ProOpter workflow

ProOpter workflow is depicted in Figure 4. The main ProOpter output is a set of manipulative variable values that should bring the manufacturing process close to the optimal operating mode.

The performance of the process is estimated through a set of KPI values. The indicators are calculated from the data acquired by the production management information system (MES) and stored in the production database. Various data irregularities typically appear in a real production environment, therefore the data is preprocessed and cleaned. During ProOpter development, a particular attention has been devoted to simplification of the KPI definition, and a corresponding data viewer and KPI formula editor have been implemented. When using on-line, the performance monitoring module supports various views with adjustable level of aggregation. Complex KPIs can be monitored and can also be drilled down to monitor their components.

The determination of the most influential variables is one of the key steps in deriving applicable performance dynamic model. To support this a set of standard input variable selection (IVS) methods can be called from ProOpter IVS module. A user interface has been designed, which enables an aggregate representation and evaluation of results obtained by various methods. Due to predominantly non-linear relations within the production process the IVS analysis results are typically non-uniform, and final decision on the set of variables that will be used in model identification has to be performed by the modeller. The developed user interface assist the modeller by providing graphical representation of the variable impact in various configurations.

A separate module has been developed for model identification support. The modeller can choose the data segments used for model identification and model validation, and adjust various identification parameters. Two soft computing methods are currently supported: neural-network based identification and fuzzy model identification.

The on-line operation of ProOpter provides the insight into production performance as described above, as well as supports the production manager in determining the optimal production settings. The nonlinear model predictive control described in Section 2 is applied here with an important distinction comparing to the classical MPC applications on the process level: the calculated inputs are not automatically applied to the production process, the decision on the application of the calculated manipulative values to the process is left to the production operator instead.

Some of the described ProOpter features are further illustrated by case studies in Section 4.

### 3.3. Data integration and communication standards

ProOpter was designed as an add-on to existing production management information systems. These typically consist of ERP and MES systems on the higher levels, which are connected to SCADA systems, PLCs and other automation equipment on the factory floor.

One of the main requirements for such an add-on is to support established industrial standards to be able to connect and integrate into existing information-communication infrastructure.

This was achieved by supporting standardized data interchange message format, which is typically used for communication among business planning and production control applications. Equally important is to use standardized communication messages among tool modules. This enables simple addition or removal of the functional modules as well as their upgrade without compromising the functionality of the rest of the system.

Therefore the following standards were considered and used in the software tool development:

**IEC/ISO 62264 (ANSI/ISA 95)** standard defines models and terminology that are used for marking information, which has to be exchanged between business level and production control level. The information is structured by UML models and is a basis of production control integration into overall business information systems. The standard is used by IT developers as a guideline for specifying user requirements and as a basis of systems and databases development. The standard can be used in any type of production: continuous, batch or discrete.

Within the ProOpter tool the standard was used for designing internal communication messages. The messages were designed based on XML implementation of the standard that was developed by WBF (The Organization for Production Technology). Each data model of the standard has an XML equivalent within the B2MML (Business to Manufacturing Markup Language), and the model structure is defined by XML schemes (XSD).

Among the standard data models ProOpter tool uses the model that is intended for production response reporting (Production Performance).

**Open Math** is a standard for representation of mathematical objects that has been developed by OpenMath

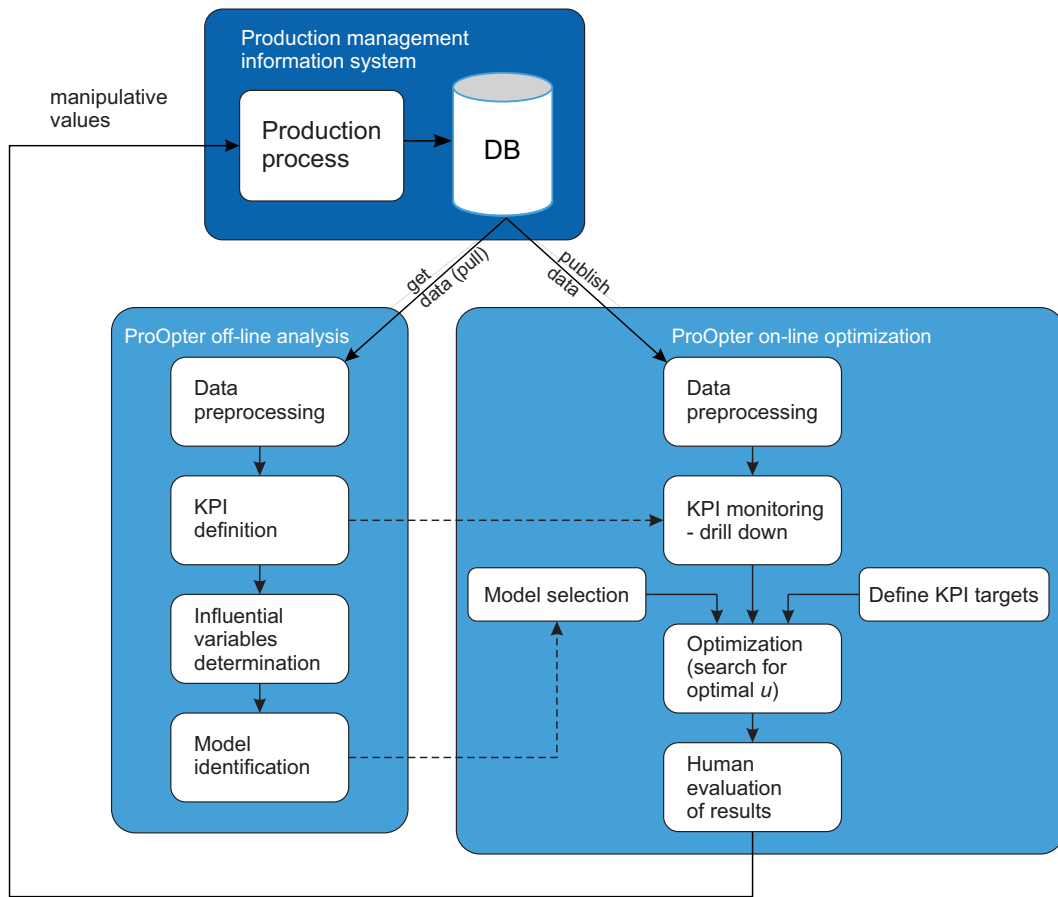


Figure 4: ProOpter workflow

Society. Two equation representations are supported, XML and binary format. The standard is based on the use of extendable content dictionaries, which structurally document mathematical definitions and unambiguously define mathematical symbols. Their openness enables definition of new functions and symbols. Custom content dictionaries can be added to common OpenMath repository and used by all interested parties. The standard does not require that all of the newly developed content dictionaries are public, but have to be accessible to all users of the particular concept.

By using OpenMath the KPIs can be described in a standard way that is recognised by various ProOpter modules, and can be exported to other tools.

**PMML** is used for transfer of identified KPI performance models among the ProOpter modules. PMML (Predictive Model Markup Language) is a standardized XML scheme for model representation in the area of predictive analytics and data mining. The PMML standard has evolved during the years and the recent version includes a rich set of model types and is supported by large number of statistical software tools.

The use of the described standards for inter-module communication within ProOpter is illustrated in Figure 5.

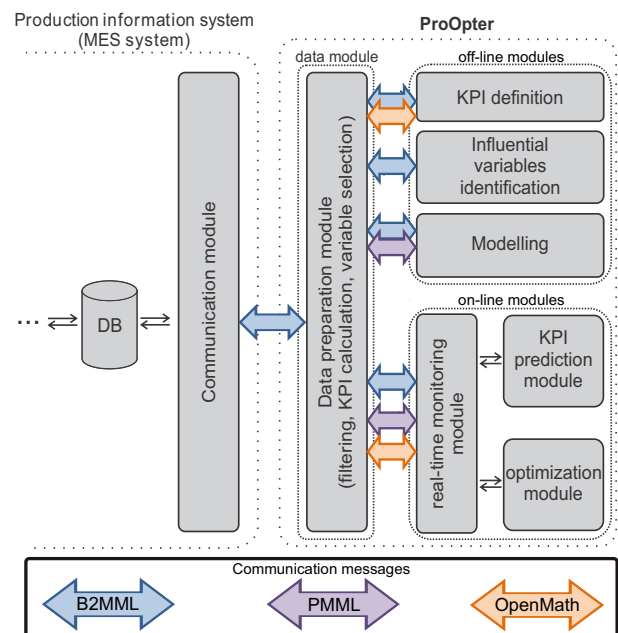


Figure 5: ProOpter standard messages

### 3.4. Communication

The use of unified communication protocols within production information system is of significant importance.



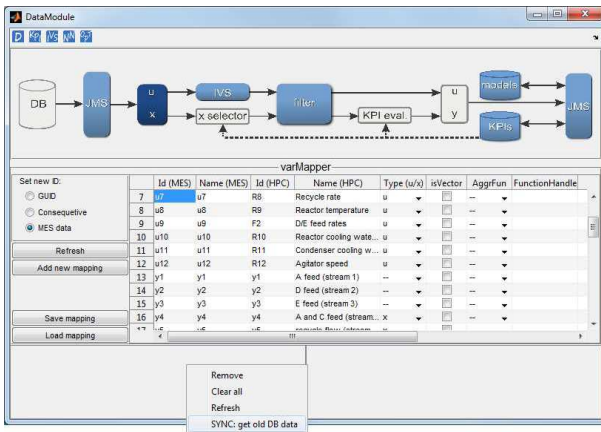


Figure 6: ProOpter data module

Use of communication standards guarantees the long term stability of developed applications as well as easier integration with the existing information infrastructure. In compliance with IEC 62264 two standard ways of data transmission are used within ProOpter:

- Pull model - data user requests the data from the data supplier, which sends the data upon request. This is a point-to-point communication.
- Publish model – data supplier is sending data to the recipients that are subscribed to specific data. The communication is carried on according to publish-subscribe principle.

The use of MOM (Message Oriented Middleware) systems is best suited for such an asynchronous XML message exchange among several clients. It supports both point-to-point and publish-subscribe communication principles. MOM enables distributed communication among loosely coupled clients, meaning that communicating applications do not necessarily all have to be active within the network, nor do they all have to be aware of each other.

Within ProOpter, MOM communication was implemented by JMS (Java Message Service) specification (Figure 6). Even if JMS originates from Java environment it is general and extensible enough to be useful also for clients that are designed within other development environments and platforms.

#### 4. CASE STUDIES

The performance of the tool was tested through a set of production monitoring and control case studies of various production types. The continuous production was covered by a simulation benchmark, while in discrete and batch production the tool was tested on a real production data.

##### 4.1. Batch production

Actual batch production data were considered in connection to performance measurement and influential variable analysis in water based paints and coatings production. The production is a typical batch process consisting of dosing and mixing stage, milling stage, production stage, and

packing stage. In between production and packing stages the quality control is performed, where the acceptability of product for packing is determined. In the case of negative result, the product can be scrapped, or can go to re-work where the quality is improved.

#### Performance indicators

The performance monitoring in the considered batch production was mainly oriented toward determination of suitable batch parameter settings. Therefore batch related indicators were of primary concern. In cooperation with production managerial staff the following relevant indicators were identified:

- Quality - calculated per batch as a result of laboratory analysis. Different products have different measured parameter sets.
- Raw materials consumption ratio - ratio of the actual raw material consumption per work order to the normative consumption.
- Timeliness - difference between the actual finish time and the planned finish time.
- Scrap rate - ration of the actual scrap quantity to the planned produced quantity.

Among the chosen indicators, Quality is the most complex and needs additional explanation. As the measured parameter sets differ among different products they have to be normalized and aggregated to derive indicator values that are comparable among several batches. This is illustrated in Figure 7. For every product (or product family) the relevant set of laboratory measured quality parameters is identified, and these are normalized and aggregated into a standard valued quality indicator (common estimate in Figure 7).

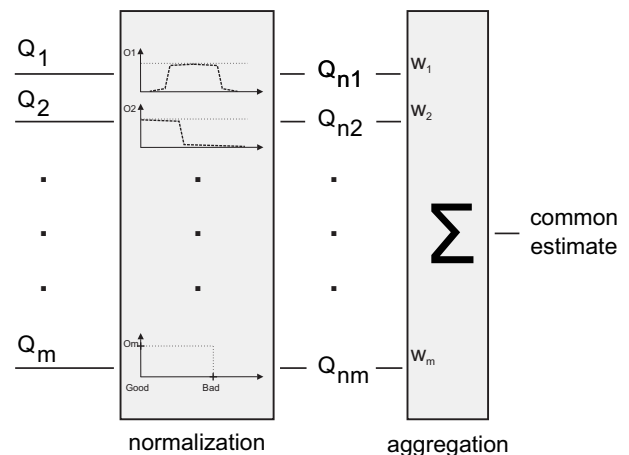


Figure 7: Aggregated KPI composition

The normalization and aggregation is supported by ProOpter KPI definition module. Two of the normalization screens are shown in Figures 8 and 9. For every measurement, the user can chose one of the predefined normalization types and then adjust the relevant parameters. Finally, the normalized measurements are aggregated into a KPI.





Figure 8: ProOpter normalization

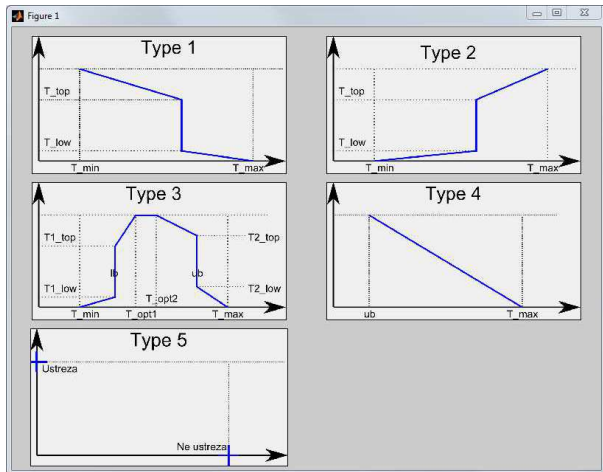


Figure 9: Predefined normalization types

### Influential variables

In the considered case of coatings production the quality indicator is most relevant for the production efficiency. Bad quality products go either to re-work, which decreases productivity, or go to scrap, which is a direct cost. The analysis of influential production variables that determine the quality level is therefore the most important. Unfortunately, the analysis can not be generalized but has to be performed for every product family due to various production recipes.

The ProOpter tool supports the analysis by embedding several standard and advanced variable selection methods that are used in data mining: Linear Correlation, Partial correlation (with forward selection approach), PLS (variable importance in projection – PLS VIP), Non-Negative Garrote, LASSO, DMS search (pareto search of minimum error of linear model as objective function), and others. The problem of variable selection is elaborated in detail in (Glavan et al. 2013b).

The results for one of the product families are shown in figure 10. The aggregated results of a subset of available variable selection methods are shown as a box plot. In the given case, Max. RPMs of the mixer dominate the batch quality, and interestingly also the temperature in the production hall points out to be important as well.

Figure 10 points out another specific of the given production case. As several production variables are continuously sampled we typically have a range of corresponding variable values for a batch. Most of the used variable selec-

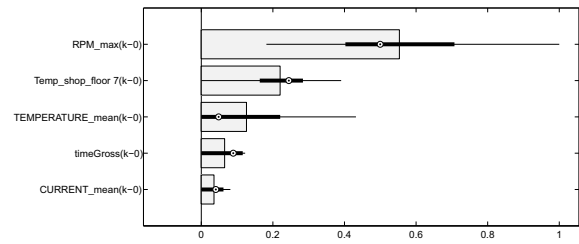


Figure 10: Selection of Quality indicator influential variables for a product family

tion methods require for such values to be aggregated into a single data point when considering the variable influence to the batch quality. For this purpose a set of standard data aggregation functions are provided within the ProOpter Data module, e.g. average, min, max.

### 4.2. Tennessee Eastman benchmark

The Tennessee Eastman (TE) benchmark process was introduced by Downs and Vogel (1993) as a model of a real chemical production process. The model represents a test problem for researchers to experiment with different control-related solutions. The process consists of five major units: a chemical reactor, a product condenser, a vapour-liquid separator, a product stripper and a recycle compressor. Four reactants (A, C, D, E) and an inert component (B) are entering the process, where four exothermic, irreversible reactions result in two products (G, H) and one byproduct (F). The process products leave the process through an output stream, and are later separated in a downstream refining section. The production process has 41 measured variables ( $y$ ) and 12 different manipulative variables ( $u$ ).

A specific combination of the production rate and/or the product mix are usually demanded by the market or some capacity limitations. Therefore, six typical operational modes are defined in the benchmark proposal. The model also provides 20 different process disturbances, which imitate the disturbances typical of real TE production.

As the research literature provides several results on the control of TE process by various methods, the process was used as a primary test bench for verifying the operation of the various ProOpter modules and validation of the obtained results.

### Production modelling

The historical data records needed for production performance evaluation and modelling were generated by simulation. To make the problem more realistic the process measurements have some intentionally added noise, typical for the specific measurement.

Production performance is monitored through three KPIs: Costs, Production, and Quality. The definition of Production KPI is quite straightforward, as the quantity of product leaving the process is directly measured. Directly from the process objectives an indicator for the process

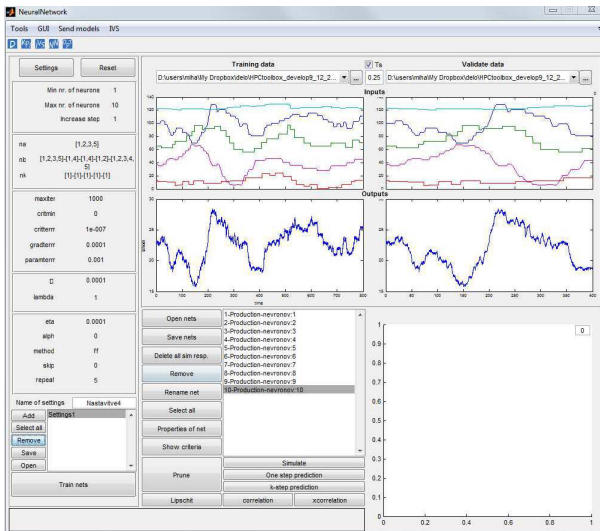


Figure 11: Modelling with ProOpter

quality is also derived, since the product quality can be viewed as a desired mass ratio between the two final products, G and H. For details on the Cost KPI definition, the reader is referred to (Downs and Vogel 1993) or (Glavan et al. 2013a).

Next, the influential variable analysis is performed considering available manipulative variables and the defined KPIs. As described in Glavan et al. (2013b), five manipulative variables are selected, and then a neural-network based model is identified that enables to calculate a short-term prediction of the defined KPIs, provided chosen manipulative values are specified. The modelling is supported by the ProOpter modelling module, shown in Figure 11.

### On-line production efficiency monitoring and optimization

The use of the derived performance models for real-time monitoring and performance prediction is illustrated in Figure 12. Observation variables can be chosen in the upper part of the window, while the chosen variable history can be seen in the lower part. The right sections of the plots indicate the estimated future KPI evolution based on the identified performance model and the manipulative values. A detailed investigation of the variables that contribute to the performance measure as well the insight into the manipulative values is possible by choosing the *drill-down* option.

Additional optimization module supports the adjustment of the manipulative values in a model based predictive manner. The identified performance model is used to determine optimal manipulative variable settings, while the user has to specify appropriate KPI targets. E.g., Figure 13 illustrates how Production and Quality can be maintained at certain levels while the production Cost is minimized.

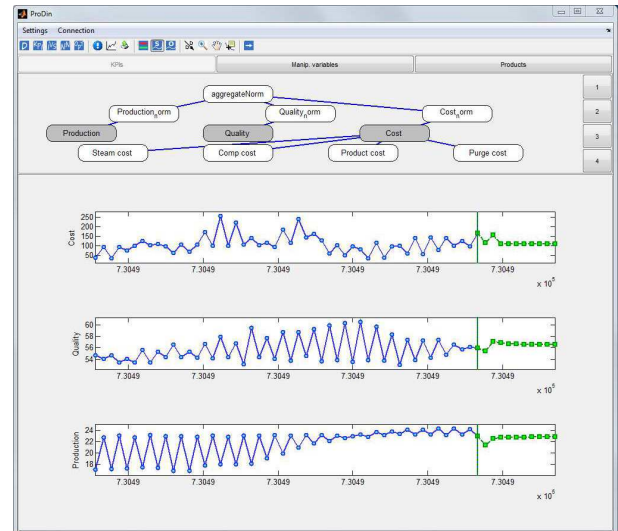


Figure 12: Efficiency monitoring and prediction



Figure 13: Efficiency optimization

## 6. CONCLUSIONS

Production information systems currently used in production companies are very efficient in data collection and data presentation, but are quite limited in the support of decision making and production optimization. The ProOpter - Production dynamics analyser and optimizer enables effective control of various types of manufacturing processes. Its main advantage is in use of simplified dynamic KPI models that are identified from historical data.

ProOpter will unburden the production manager and help him take better decisions in order to improve the production process. With the introduction of Production dynamics analyser and optimizer we can expect savings in various segments in the production through better product quality, increasing the efficiency, reduction of waste, and production cost reduction.

## ACKNOWLEDGMENTS

The presented work has been partially performed within the Competence Centre for Advanced Control Technologies, which is partly financed by the Republic of Slovenia, Ministry of Education, Science and Sport, and European Union, European Regional Development Fund. The support is gratefully acknowledged.

## REFERENCES

- Brettel, M., Friederichsen, N., Keller, M. and Rosenberg, M., 2014. How virtualization, decentralization and network building change the manufacturing landscape: An industry 4.0 perspective, *International Journal of Mechanical, Industrial Science and Engineering*, 8 (1), 37 – 44.
- Davis, J., Edgar, T., Porter, J., Bernaden, J. and Sarli, M., 2012. Smart manufacturing, manufacturing intelligence and demand-dynamic performance, *Computers & Chemical Engineering*, 47 (0), 145 – 156. {FOCAPO} 2012.
- Downs, J. and Vogel, E., 1993. A plant-wide industrial process control problem, *Computers & Chemical Engineering*, 17, 245–255.
- Folan, P. and Browne, J., 2005. A review of performance measurement: towards performance management, *Computers in Industry*, 56, 663–680.
- Gerry, J. and Buckbee, G., 2005. *The link between automation and enterprise KPIs, ExpertTune, Plant Performance Supervision & PID Tuning Software*, White Paper.
- Glavan, M., Gradišar, D., Atanasijević-Kunc, M., Strmčnik, S. and Mušič, G., 2013b. Input variable selection for model-based production control and optimisation, *The International Journal of Advanced Manufacturing Technology*, 68 (9-12), 2743–2759.
- Glavan, M., Gradišar, D., Strmčnik, S. and Mušič, G., 2013a. Production modelling for holistic production control, *Simulation Modelling Practice and Theory*, 30, 1–20.
- Grüne, L. and Pannek, J., 2011. *Nonlinear Model Predictive Control: Theory and Algorithms*, Communications and Control Engineering, 1st edn, Springer.
- Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection, *Journal of Machine Learning Research*, 3, 1157–1182.
- Holweg, M., 2007. The genealogy of lean production, *Journal of Operations Management*, 25 (2), 420 – 437. Special Issue Evolution of the Field of Operations Management SI/ Special Issue Organisation Theory and Supply Chain Management.
- Industrie 4.0, W. G., (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0, *Final report*, Industrie 4.0 Working Group.
- Ljung, L., 1999. *System Identification: Theory for the User*, 2nd edn, Prentice Hall.
- Maciejowski, J., 2002. *Predictive Control with Constraints*, Prentice-Hall.
- Neely, A., Gregory, M. and Platts, K., 1995. Performance measurement system design - a literature review and research agenda, *International Journal of Operations & Production Management*, 15 (4), 80–116.
- Pearson, R., 2006. Nonlinear empirical modeling techniques, *Computers & Chemical Engineering*, 30, 1514–1528.
- Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P.-Y., Hjalmarsson, H. and Juditsky, A., 1995. Nonlinear black-box modeling in system identification: a unified overview, *Automatica*, 31, 1691–1724.
- Smits, G., Kordon, A., Vladislavleva, K., Jordaan, E. and Kotanchek, M., 2006. Variable selection in industrial datasets using pareto genetic programming, in T. Yu, R. Riolo and B. Worzel (eds), *Genetic Programming Theory and Practice III*, Vol. 9 of *Genetic Programming*, Springer US, pp. 79–92.
- SMLC, (2011). Implementing 21st century smart manufacturing, *Workshop report*, Smart Manufacturing Leadership Coalition and U.S. DOE.
- Zorzut, S., Gradišar, D., Jovan, V. and Mušič, G., 2009. Use of a procedural model in the design of production control for a polymerization plant, *The International Journal of Advanced Manufacturing Technology*, 44 (11-12), 1051–1062.

## AUTHORS' BIOGRAPHIES

**GAŠPER MUŠIČ** received B.Sc., M.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana, Slovenia in 1992, 1995, and 1998, respectively. He is Associate Professor at the Faculty of Electrical Engineering, University of Ljubljana. His research interests are in discrete event and hybrid dynamical systems, supervisory control, planning, scheduling, and industrial informatics. His Web page can be found at <http://msc.fe.uni-lj.si/Staff.asp>.

**MIHA GLAVAN** received B.Sc. in electrical engineering from the University of Ljubljana, Slovenia in 2010. He is Ph.D. student at Department of Systems and Control, Jožef Stefan Institute. His research interests are in modelling of dynamical systems, neural networks and model predictive control. His Web page can be found at [http://dsc.ijs.si/en/people/miha\\_glavan/](http://dsc.ijs.si/en/people/miha_glavan/).

**DEJAN GRADIŠAR** received B.Sc. and Ph.D. degrees in electrical engineering from the University of Ljubljana, Slovenia in 2001 and 2006, respectively. He works at the Jožef Stefan institute as a postdoctoral associate. His research interests are in production control, scheduling and control systems. His Web page is at [http://dsc.ijs.si/en/people/dejan\\_gradisar/](http://dsc.ijs.si/en/people/dejan_gradisar/).

**STANKO STRMČNIK** received his B.Sc., M.Sc., and Ph.D. from the Faculty of Electrical Engineering, University of Ljubljana in 1972, 1975 and 1979 respectively. In 1973 he joined J. Stefan Institute in Ljubljana. In years 1986–2011, he has been the head of the Department of Systems and Control at the same institute. His research interests concern mathematical modelling, identification, process control and non-technical aspects of automation. He is also an associated professor at the Faculty of Electrical Engineering, University of Ljubljana and full professor at University of Nova Gorica. His Web page is at [http://dsc.ijs.si/en/people/stanko\\_strmcnik/](http://dsc.ijs.si/en/people/stanko_strmcnik/).

# EVOLUTIONARY ALGORITHMS FOR HYPERPARAMETER TUNING ON NEURAL NETWORKS MODELS

David Orive<sup>(a)</sup>, Gorka Sorrosal<sup>(b)</sup>, Cruz E. Borges<sup>(c)</sup>, Cristina Martin<sup>(d)</sup>, Ainhoa Alonso-Vicario<sup>(e)</sup>

<sup>(a-e)</sup> Deusto Institute of Technology - DeustoTech Energy, University of Deusto, Avda. Universidades 24, Bilbao, Spain

<sup>(a)</sup> [david.orive@opendeusto.es](mailto:david.orive@opendeusto.es), <sup>(b)</sup> [gsorrosal@deusto.es](mailto:gsorrosal@deusto.es), <sup>(c)</sup> [cruz.borges@deusto.es](mailto:cruz.borges@deusto.es), <sup>(d)</sup> [cristina.andonegui@deusto.es](mailto:cristina.andonegui@deusto.es)  
<sup>(e)</sup> [ainhoa.alonso@deusto.es](mailto:ainhoa.alonso@deusto.es)

## ABSTRACT

In this work we present a comparison of several Artificial Neural Networks weights initialization methods based on Evolutionary Algorithms. We have tested these methods on three datasets: KEEL regression problems, random synthetic dataset and a dataset of concentration of different chemical species from the Bioethanol To Olefins process. Results demonstrated that the tuning of neural networks initial weights improves significantly their performance compared to a random initialization. In addition, several crossover algorithms were tested to identify the best one for the present objective. In the post-hoc analysis there were found significant differences between the implemented crossover algorithms when the network has four or more inputs.

Keywords: Evolutionary Algorithm, Artificial Neural Networks, hyperparameter tuning

## 1. INTRODUCTION

The hyperparameter tuning of soft computing modelling techniques, such as Artificial Neural Networks (ANN) or Support Vector Machines (SVM), is an important step in the model development. This task used to be performed manually based on expert experience and using a reduced number of trials because of the low computational power available (Bardinet et al. 2013). Presently, computer clusters and GPU processor are able to run more trials, allowing the use of some methods and algorithms to automate the procedure such as Estimation of Distribution Algorithms (EDA) (Nannen and Eiben 2007), local search based algorithms (Hutter and Leyton-brown 2009) or sequential hyperparameter optimization algorithms (Bergstra et al. 2011).

In other areas of knowledge such as image processing, it was proved that state of the art of image classification could be improved not only by building new algorithms but also by tuning the hyperparameters of existing techniques (Pinto et al. 2009; Bergstra et al. 2011). In the case of ANN, the tuning of some hyperparameters such as layers, number of neurons, activation functions or recurrences has been widely researched (Murata, Yoshizawa, and Amari 1994; Hagiwara, Toda, and Usui 1993; Panchal et al. 2010; Leung et al. 2003). Nonetheless, ANN initial weights are quite often not tuned with the above mentioned

hyperparameters but only initialized by training functions. Frequently used training functions like Levenberg-Marquardt backpropagation (Marquardt 1963; Hagan and Menhaj 1994) or Bayesian Regularization backpropagation (MacKay 1992), use random weight initializations. However, as final results are dependent of the neural network initial weights, optimization of these parameters should be considered as a technique for the improvement of training functions performance. Moreover, optimizing the initial weights reduces both the possibility of stalling at local optima and the number of iterations required to achieve the objective training error criterion (Yam and Chow 2000; Lee, Oh, and Kim 1991).

The simplest techniques are random samples or grid search strategies over the hyperparameters operating ranges. The first case uses pure random sampling while the second option defines a grid of points evenly distributed over the whole operation range. However, as the number of hyperparameters becomes large these techniques are inviable because their computational cost (Do, Foo, and Ng 2008). Evolutionary Algorithms (EA) are, in general, more efficient as intelligent tuning strategies. Based on biological evolution, EA allow a selective exploration of the operation range using fitness functions that determine which is going to be the next point to be sampled (Holland 1975). In hyperparameter tuning problems, EA have been proved to perform better than grid search techniques based on accuracy-speed ratio (Martino et al. 2011; Friedrichs and Igel 2005; Huang and Wang 2006).

In the case of ANN, many weight initialization methods have been proposed (Yam and Chow 2000). Some of these methods are quite simple, like the methodology proposed by Waghmare, consisting on initializing all weight to the same value (usually 0 or 1) instead of random initialization (Waghmare, Bidwai, and Bhogle 2007). Other researchers use statistical techniques such as the Independent Component Analysis (ICA) to extract useful information from training datasets in order to determine the network weights (Y.-F. Yam et al. 2002), or Partial Least-Squares (PLS) algorithms to determine the initial weights and the optimum node number of hidden layers (Ryan Hsiao, Lin, and Chiang 2003).

EA have been used to improve or to develop new training algorithms for ANN. Some researchers



maintain a random initialization and use EAs for the training itself (Montana and Davis 1989), while others use these techniques both for the initialization and training of neural network weights (Al-Shareef and Abbod 2010). Initial weight optimization followed by backpropagation training algorithms has also been used with very good results (Venkatesan, Kannan, and Saravanan 2009; Chang et al. 2012). These works have been implemented only using two-point and four-point crossover operations respectively. Although, the use of EA in the ANN models weight initialization improves the performance compared to a random initialization, the results obtained by the previous algorithms could be further improved by selecting the optimal genetic operators.

This paper demonstrates not only that the tuning of initial weights by EA improves the performance of ANN, but also the existence of significant differences between some genetic operators. The current study allows determining which of the implemented crossover operators for weight initialization provide better results.

This paper is arranged as follows. Section 2 explains the implemented evolutionary algorithm. The datasets used for the experimentation are explained in Section 3. Experimental results are in Section 4. Conclusions and future works are in Section 5.

## 2. EVOLUTIONARY ALGORITHM

EA are optimization algorithms that operate by modifying a set of candidate solutions (population) according to certain rules called *operators*. In this study we will consider EA as the algorithms that follow the pseudocode of Figure 1. One of the main advantages of the EA is their generality, i.e., they can be used in a broad range of conditions due their simplicity and independence of the underlying problem. In this sense, only the codification of the population and the FitnessOperator depends on the specific problem, the rest of the operators are (almost) independent of it.

In this paper, the problem to be resolved is the optimization of the initial weights and biases of an ANN. In this sense, we have codified a solution as two weights matrices and two biases vectors (Figure 2). The first matrix represents the relationship between the input layer and hidden layer and the second one represents the relationship between the hidden layer and output layer. In the case of the biases, each vector sets the biases for the neurons of each layer. Please note that crossovers operators only cross matrices and vectors that have same code position within the individual.

As our main objective is to assess the performance of different crossovers operators, in order to do so we have defined a broad collection of them using strategies of other EA.

- **Arithmetic crossover (A):** Being  $P_1$  and  $P_2$  the individuals to be crossed and  $\lambda$  a random variable uniformly distributed between  $[0,1]$  the two offsprings  $C1$  and  $C2$  are defined as:

```

Using the GenesisOperator, a population of
randomly generated individuals is initialized;
Using the FitnessOperator, evaluate each individual
of the population;
while a criteria is satisfied do
 Create an empty new population;
 Select between crossover and mutation
 GeneticOperators;
repeat
 Use SelectionOperator to select the parents
 from the old population;
 Apply a GeneticOperator to generate
 offsprings;
 Calculate the offsprings fitness values;
 With ReproductionOperator select between
 the parents and the offsprings and add them
 to the new population;
until A new population is filled;
Apply the EliteOperator to maintain the best
individual from previous generation;
Substitute the old population with new
population;
end

```

Figure 1: Pseudocode of the Evolutionary Algorithm

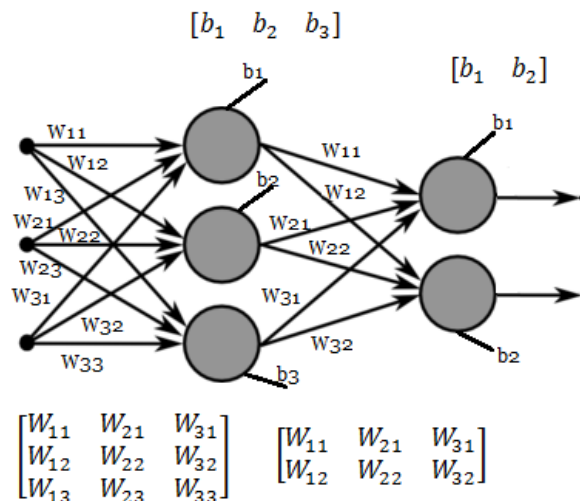


Figure 2: Structure of an individual

$$C_1 = \lambda \cdot P_1 + (1 - \lambda) \cdot P_2 \quad (1)$$

$$C_2 = \lambda \cdot P_2 + (1 - \lambda) \cdot P_1 \quad (2)$$

- **Differential crossover (D):** Being  $P_1$ ,  $P_2$  and  $P_3$  the three individuals to be crossed and  $\lambda$  a random variable uniformly distributed between  $[0,1]$  the six offsprings are:

$$C_x = P_y + \lambda(P_z - P_w), \quad (3)$$

where  $\{y, z, w\}$  are all possible permutations of the set  $\{1,2,3\}$ .

- **Uniform crossover (U):** Given two individuals  $P_1$  and  $P_2$  with  $L$  attributes,  $L$  binomial random variables  $\lambda_i$  with probability 0.5 are flip.  $C_1$  is made

with the features from  $P_1$  such that  $\lambda_i$  are 0 and the features from  $P_2$  such that  $\lambda_i$  are 1.  $C_2$  is made switching  $P_1$  with  $P_2$ .

- **Swarm crossover (S):** For each position  $i$  in the population we denote by  $P_i^e$  as the best qualification obtained by the elements at that position. Moreover,  $P^e$  will represent the best individual in the population. If the element  $P$  to be crossed is at position  $i$  in the population, then the crossover could be defined as:

$$C = \lambda_1 P_i + \lambda_2 (P_i^e - P_i) + \lambda_3 (P^e - P_i), \quad (4)$$

where  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are three random real numbers uniformly distributed over the range  $[0, 1]$ .

- **2 Points crossover (2P):** Given two individuals  $P_1$  and  $P_2$  as a matrix of size  $m \times n$  as attribute and two randomly chosen cross points  $t_1, t_2 \in (1, m)$  we define the two point crossover as:

$$C_1 = [P_{1(1,t_1) \times (1,n)}, P_{2(t_1+1,t_2) \times (1,n)}, P_{1(t_2+1,m) \times (1,n)}], \quad (5)$$

$$C_2 = [P_{2(1,t_1) \times (1,n)}, P_{1(t_1+1,t_2) \times (1,n)}, P_{2(t_2+1,m) \times (1,n)}], \quad (6)$$

where  $C_1$  and  $C_2$  are the offsprings generated by the operator.

The rest of operators are as follow:

- **Mutation Operator (Non Uniform):** If  $P$  is the individual to be mutated and  $min_i, max_i$  the inferior and superior limits of its  $i$  attribute,  $P_i$ , the two offsprings are defined as:

$$C_1 = P_i + \Delta(t, max_i - P_i), \quad (7)$$

$$C_2 = P_i - \Delta(t; P_i - min_i), \quad (8)$$

where  $\Delta(t, y) = yr(1-t/T)$ , being  $r$  a random number uniformly distributed between  $[0,1]$  and  $T$  the maximum number of generations.

- **Genesis Operator:** The values of the three matrices and vectors are initialized using uniform random variables between  $[0,1]$ .
- **Fitness Operator:** In order to avoid overfitting (Hawkins 2004) we have used as fitness function the validation error of the ANN over the dataset. The datasets have been divided in two parts: the training set and the validation set. In order to determine the performance of the developed models, k-cross validation has been used. This technique uses an iterative procedure in which the validation data section is changed at each simulation. Training each time with the training dataset section and validating the models with the portion of data that have been excluded from the training. In our case the validation error is based in the Mean Absolute Percentage Error (MAPE) defined by Equation 9:

$$MAPE = 100 \cdot \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \bar{x}_i}{x_i} \right|, \quad (9)$$

where  $n$  is the number of samples in the dataset,  $x_i$  is the real value and  $\bar{x}_i$  the output of the trained neural network. The validation strategy will be

different depending of the dataset so we will discuss this topic further in the next section.

- **Selection Operator (Round Robin):** Being  $P$  the population and  $S$  the size of the population then the selected individual for  $k + 1$  iteration would be  $P_{\phi(k+1)}$  where  $\phi(k+1) = \text{mod}(\phi(k), S) + 1$ . In the first iteration  $\phi(1)$  is assigned a random value between 1 and  $S$ .
- **Reproduction Operator (N-Different):** The best  $N$  different individuals are selected where  $N$  is the number of individuals that will survive.
- **Elite Operator (best selection):** Selects the individual with the lowest prediction error. Please note that the last individual selected by this operator will be the solution.

### 3. EXPERIMENTAL SETUP

The proposed method has been validated on three different datasets: KEEL dataset, a synthetic dataset built by random sampling the space of neural networks and a dataset containing the concentrations over the time of several chemical species in a catalytic reaction. The KEEL dataset repository was created with the aim of providing the researchers with a set of benchmarks to analyze the behavior of their learning algorithms. There are benchmarks for classification, regression and unsupervised learning (Alcalá-Fdez et al. 2011). The FitnessOperator used in this experiment is the mean of the MAPE errors obtained applying 10-fold cross-validation by initializing the ANN with the weights of an individual.

The synthetic datasets were created by randomly creating feedforward Neural Networks (rNN). A rNN is defined as a neural network with the number of inputs, outputs, layers and neurons selected randomly within some range values. Several random inputs are given to the rNN and outputs are recorded as a dataset. Using this technique infinite datasets can be obtained representing each rNN a different process to model. In order to create realistic conditions, normally distributed noise is added to the outputs of the training set. It was also taken into account the curse of dimensionality phenomenon (Bellman 1961) by increasing the amount of data according to the number of inputs, that is, the more number of inputs the bigger the training dataset has to be. For this reason, the size of the synthetic dataset has been set to depend on the number of inputs following the relation  $m = 100n$  where  $m$  is the number of instances in the training data and  $n$  the number of inputs. Rigorously, the relationship should be  $m = 100^n$ , however this would yield an exponential grow in the training set that would be impossible even to store. In this dataset, the FitnessOperator used is the same as the previous one.

The last used dataset contains the concentration of different chemical species of a Bioethanol To Olefins (BTO) process. This dataset comes from experimental results in a laboratory scale reactor (Gayubo et al. 2010). The BTO process consists on the catalytic transformation of bioethanol, as a substitute of oil, into



olefins, which are commodities for the petrochemical industry. The modelling objectives are the olefins production rate curves at different operational conditions. The activity of the catalyst is one of the most important variables to take into account because of its influence in the final production of olefins. The FitnessOperator used in this problem is the results of applying Leave-One-Out Cross Validation to the 23 experiments available. The fitness value is the maximum absolute error obtained from the 23 iterations.

Regarding to evolutionary algorithms, 75 generations with 50 individuals in the population for each experiment are done. In all the experiments the evolutionary algorithm is set with non-uniform mutation in company with a crossover. ANN with only one hidden layer are used with the number of neurons determined by the relation  $h = 10n$  where  $h$  is the number of neurons in the hidden layer and  $n$  represents the number of inputs. Sigmoid function is used as activation function and Levenberg-Marquardt backpropagation is used as a training algorithm with initial weight given by our algorithm.

We have used Nguyen-Widrow (NW) random weight initialization as contrast method. This method generate weights that follow a normal distribution (Nguyen and Widrow 1990). In order to be fair with the rest of the proposed methods,  $75 \cdot 50 = 3750$  repetitions are made. Please note that in all the proposed methods we have selected the neural network with best validation error as the solution of the problem.

#### 4. EXPERIMENTAL RESULTS

In order to assess if the differences between proposed methods are significant, we have executed a Friedman test with Hochberg's and Bergmann's post hoc analyses (Demšar 2006; Derrac et al. 2011; Bergmann and Hommel 1988). First, a Friedman test was carried over the results of the different crossovers. The null hypothesis of this test states that the outcomes of each algorithm have the same probability distribution. I.e. this means that all algorithms show the same performance. In the case of positive results, i.e. the p-value is less than 0.05, this test only states that there are differences, but not between which pair of methods. In this case a post hoc analysis is needed. We have used Hochberg's post hoc to compare all methods against NW (as the statistical power of this method is low but it is computationally simple) and Bergmann's method to compare the crossover operators between them (as this method gives us a better statistical power but has exponential complexity).

Table 1 shows the results of the different crossover operators over every dataset of the KEEL regression database. These results show that ANN without hyperparameter tuning (NW) tends to be the poorest (around 7 points worse than the others) while there are little differences between the rest of used techniques. The p-value of the Friedman test carried on over the KEEL regression is  $4.5 \cdot 10^{-9}$  way below the significance

level so we can assure without any doubt that there are significant differences between at least one of the methods. Table 2 shows the results of Hochberg's post hoc analysis over the KEEL database. Please remember that this method compares all crossovers against NW. The adjusted p-values are way below the significance level given by the Hochberg's post hoc analysis (the second column is way below the third one) so we can conclude that the NW is worse than the rest of proposed methods. Finally, Bergmann's procedure has been carried out in order to test if there are significant differences between any pair of crossovers. In this case the procedure did not found any significant differences between the crossovers.

Table 3 shows the results of the different crossover operators over the 100 synthetic datasets. We only show the mean result for space purposes. Again, the NW seems to be the poorest but in this case the differences are smaller. However, the p-value of the Friedman test carried on over this dataset is  $5.1 \cdot 10^{-11}$ , again way below the significance level so we can assure without any doubt that there are significant differences between at least one of the methods. Table 2 shows the results of Hochberg's post hoc analysis over this dataset. Once again, the adjusted p-values below the significance level so we can conclude that the NW is worse than the rest of proposed methods. Next, we use the Bergmann's post hoc to study the synthetic datasets. In this case we have divided the samples in two groups: datasets with 3 or less inputs (group A) and the rest (group B). In group A Bergmann's procedure only found significant differences between NW and the rest of crossovers. However, in the case of group B we have found three groups. The first one consists on NW. The second group is composed by U, S, A, and 2P crossovers; between them the Bergmann's procedure has not found any significant difference. The final group consist only on the D crossover that has a result significantly better. Figure 3 shows this information graphically. Please note that algorithms that are linked are not statistically different according to Bergmann's procedure.

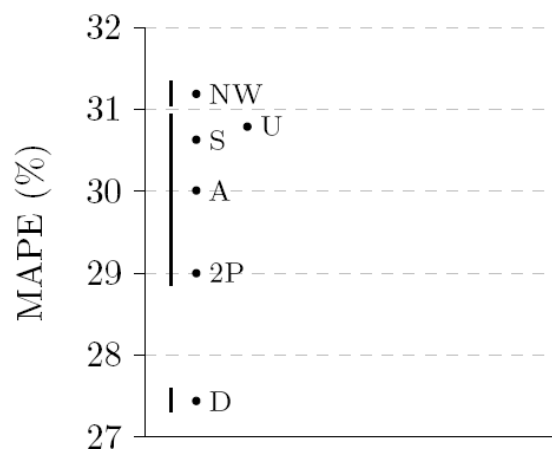


Figure 3: Groups of crossover operators that are statistically different for examples of the synthetic dataset with more than 3 inputs (group B)

Table 1: KEEL Dataset Results with the Different Crossover Algorithms. MAPE (%)

| Dataset     | A      | D      | U      | 2P     | S      | NW Random |
|-------------|--------|--------|--------|--------|--------|-----------|
| ANACALT     | 1,99   | 1,93   | 2,04   | 2,01   | 1,94   | 2,09      |
| abalone     | 10,66  | 10,75  | 10,46  | 10,60  | 10,76  | 13,27     |
| aileron     | 8,32   | 8,45   | 8,71   | 8,51   | 8,33   | 9,93      |
| autoMPG6    | 24,71  | 24,53  | 23,91  | 24,68  | 24,73  | 29,37     |
| autoMPG8    | 10,43  | 9,32   | 10,41  | 10,35  | 10,18  | 10,66     |
| california  | 134,72 | 152,16 | 133,92 | 148,02 | 152,72 | 264,34    |
| compactiv   | 14,27  | 15,17  | 14,33  | 14,56  | 14,57  | 15,90     |
| concrete    | 50,77  | 50,47  | 50,55  | 49,81  | 51,07  | 51,84     |
| dee         | 13,29  | 13,23  | 13,37  | 13,48  | 13,3   | 15,6      |
| diabetes    | 36,33  | 37,54  | 33,34  | 36,28  | 33,64  | 53,29     |
| ele-1       | 15,11  | 15,37  | 14,92  | 15,24  | 14,92  | 15,55     |
| ele-2       | 1,13   | 1,14   | 1,12   | 1,14   | 1,15   | 1,16      |
| elevators   | 92,64  | 81,65  | 92,43  | 91,57  | 93,42  | 94,36     |
| forestFires | 7,90   | 8,09   | 8,04   | 8,12   | 8,08   | 8,20      |
| friedman    | 24,74  | 24,77  | 24,74  | 24,68  | 25,05  | 25,03     |
| house       | 2,40   | 2,38   | 2,41   | 2,41   | 2,40   | 2,41      |
| laser       | 2,44   | 2,47   | 2,50   | 2,46   | 2,64   | 2,84      |
| machineCPU  | 85,43  | 79,72  | 87,42  | 87,33  | 81,27  | 82,81     |
| mortgage    | 2,07   | 2,01   | 2,26   | 2,26   | 1,82   | 2,26      |
| mv          | 2,48   | 2,37   | 2,38   | 2,44   | 2,29   | 2,47      |
| plastic     | 1,57   | 1,53   | 1,59   | 1,64   | 1,51   | 1,67      |
| quake       | 8,53   | 8,79   | 8,39   | 8,40   | 8,48   | 10,36     |
| stock       | 8,81   | 8,67   | 8,42   | 8,30   | 8,53   | 11,22     |
| tic         | 50,26  | 54,16  | 52,68  | 52,01  | 51,67  | 76,19     |
| treasury    | 25,76  | 25,91  | 26,08  | 25,96  | 25,34  | 26,06     |
| wankara     | 2,62   | 2,60   | 2,65   | 2,62   | 2,60   | 2,66      |
| wizmir      | 20,52  | 21,04  | 20,01  | 20,74  | 20,09  | 22,44     |
| MEAN        | 24,44  | 24,67  | 24,41  | 25,02  | 24,91  | 31,63     |

Table 2: Post Hoc Analysis done for the Different Crossover Algorithms of KEEL datasets and Synthetic datasets

| KEEL datasets   |                      |                    | Synthetic datasets |                       |                    |
|-----------------|----------------------|--------------------|--------------------|-----------------------|--------------------|
| Null hypothesis | p-value              | Significance level | Null hypothesis    | p-value               | Significance level |
| A = NW?         | $2.47 \cdot 10^{-9}$ | 0.006              | D = NW?            | $2.15 \cdot 10^{-22}$ | 0.006              |
| S = NW?         | $6.10 \cdot 10^{-9}$ | 0.007              | S = NW?            | $5.14 \cdot 10^{-15}$ | 0.007              |
| U = NW?         | $1.10 \cdot 10^{-8}$ | 0.008              | 2P = NW?           | $4.52 \cdot 10^{-12}$ | 0.008              |
| D = NW?         | $6.08 \cdot 10^{-8}$ | 0.010              | A = NW?            | $1.33 \cdot 10^{-11}$ | 0.010              |
| 2P = NW?        | $1.43 \cdot 10^{-6}$ | 0.012              | U = NW?            | $1.81 \cdot 10^{-8}$  | 0.050              |

Table 3: Synthetic Dataset Results with the Different Crossover Algorithms. MAPE (%)

| Group | A     | D     | U     | 2P    | S     | NW Random |
|-------|-------|-------|-------|-------|-------|-----------|
| A     | 2,91  | 2,54  | 2,88  | 2,79  | 2,68  | 3,41      |
| B     | 30,01 | 27,44 | 30,79 | 29,00 | 30,63 | 31,19     |
| Total | 16,46 | 15,00 | 16,84 | 15,90 | 16,66 | 17,30     |

Finally, the results of the BTO process modelling using swarm crossover are shown in Table 4. Taking into account that this experiment is very time consuming (more than one week for of computation for each crossover), we can only afford executing this simulation with one of the crossovers. We have selected the swarm crossover as it seems the most promising given the previous results. Fortunately, using this method we have been able to reduce the error of the model in 4 points when using EA for ANN

hyperparameter optimization. This represents a significant improvement in this particular dataset.

Table 4: BTO Process Modelling Dataset Results

| Error    | S     | NW    |
|----------|-------|-------|
| MAPE (%) | 18.09 | 22.04 |

## 5. CONCLUSIONS AND FUTURE WORKS

In this paper it is statistically demonstrated that the tuning of artificial neural network initial weights

improves significantly the performance compared to Nguyen-Widrow random initialization. Moreover, it is also demonstrated that when the number of inputs is four or more, there are significant differences between the results shows by the different crossover operators used internally in the evolutionary algorithm. For synthetic data the differential crossover present significantly better results than other crossover operator types.

As future work it is necessary to improve the weight initialization techniques by customizing the existing crossover algorithms in order to make an intelligent selection of parents attributes taking into account the neural nodes specialization.

#### ACKNOWLEDGMENTS

This work has been carried out with the financial support of the Basque Government via the PI2013-40 project and a researchers training grant (COL-2013-1-135).

#### REFERENCES

- Alcalá-Fdez, J., Fernandez, A., Luengo, J., Derrac, J., Garcia, S., Sánchez, L., Herrera, F., 2011. KEEL Data-Mining Software Tool : Data Set Repository, Integration of Algorithms and Experimental Analysis Framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17 (2-3), 255–287.
- Al-Shareef, A.J., Abbod, M.F., 2010. Neural Networks Initial Weights Optimisation. *IEEE 12th International Conference on Computer Modelling and Simulation*, pp. 57–61. March 24-26, Cambridge (UK).
- Bardenet, R., Brendel, M., Kégl, B., Sebag, M., 2013. Collaborative Hyperparameter Tuning. *30<sup>th</sup> International Conference on Machine Learning*. Vol. 28. June 16-21, Atlanta (USA).
- Bellman, R., 1961. *Adaptive Control Processes - A Guided Tour*. Princeton: Princeton University Press.
- Bergmann, B., Hommel, G., 1988. Improvements of General Multiple Test Procedures for Redundant Systems of Hypotheses. *Medizinische Informatik Und Statistik*, 70, 100–115.
- Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for Hyper-Parameter Optimization. *Advances in Neural Information Processing Systems (NIPS)*, 24, 1–9.
- Chang, Y., Lin, J., Shieh J., Abbod, M.F., 2012. Optimization the Initial Weights of Artificial Neural Networks via Genetic Algorithm Applied to Hip Bone Fracture Prediction. *Advances in Fuzzy Systems, 2012*, 1–9.
- Demšar, J., 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. *Journal of Machine Learning Research*, 7, 1–30.
- Derrac, J., García, S., Molina, D., Herrera, F., 2011. A Practical Tutorial on the Use of Nonparametric Statistical Tests as a Methodology for Comparing Evolutionary and Swarm Intelligence Algorithms. *Swarm and Evolutionary Computation*, 1 (1), 3–18.
- Do, C.B., Foo, C.S., Ng, A.Y., 2008. Efficient Multiple Hyperparameter Learning for Log-Linear Models. In: J.C. Platt and D. Koller and Y. Singer and S.T. Roweis ed. *Advances in Neural Information Processing Systems*, 20, 377–384.
- Friedrichs, F., Igel C.. 2005. Evolutionary Tuning of Multiple SVM Parameters. *Neurocomputing*, 64, 107–117.
- Gayubo, A.G., Alonso, A., Valle, B., Aguayo, A.T., Bilbao, J., 2010. Selective Production of Olefins from Bioethanol on HZSM-5 Zeolite Catalysts Treated with NaOH. *Applied Catalysis B: Environmental*, 97 (1-2), 299–306.
- Hagan, M.T., Menhaj, M.B., 1994. Training Feedforward Networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks*, 5 (6), 989–993.
- Hagiwara, K., Toda, N., Usui, S., 1993. On the Problem of Applying AIC to Determine the Structure of a Layered Feed-Forward Neural Network. *International Joint Conference on Neural Networks*, pp. 2263–2266. October 25-29, Nagoya (Japan).
- Hawkins, D.M., 2004. The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, 44 (1), 1–12.
- Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems*. Michigan: University of Michigan Press.
- Huang, C.L., Wang, C.J., 2006. A GA-Based Feature Selection and Parameters Optimization for Support Vector Machines. *Expert Systems with Applications*, 31 (2), 231–240.
- Hutter, F., Leyton-brown, K., 2009. ParamILS: An Automatic Algorithm Configuration Framework. *Journal of Artificial Intelligence Research*, 36, 267–306.
- Lee, Y., Oh, S.H., Kim, M.W., 1991. The Effect of Initial Weights on Premature Saturation in Back-Propagation Learning. *International Joint Conference on Neural Networks*, 1 pp. 765–770. July 8-12, Seattle (USA).
- Leung, F.H.K., Lam, H.K., Ling, S.H., Tam, P.S., 2003. Tuning of the Structure and Parameters of a Neural Network Using an Improved Genetic Algorithm. *IEEE Transactions on Neural Networks*, 14 (1), 79–88.
- MacKay, D.J.C., 1992. Bayesian Interpolation. *Neural Computation*, 4 (3), 415–447.
- Marquardt, D.W., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11 (2), 431–441.
- Martino, S.D., Ferrucci, F., Gravino, C., Sarro, F., 2011. A Genetic Algorithm to Configure Support Vector Machines for Predicting Fault-Prone Components. *Lecture Notes in Computer Science*, 247–261.

- Montana, D.J., and Davis, L., 1989. Training Feedforward Neural Networks Using Genetic Algorithms. *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pp. 762–767. August 20-25, Detroit (Michigan, USA).
- Murata, N., Yoshizawa, S., Amari, S., 1994. Network Information Criterion-Determining the Number of Hidden Units for an Artificial Neural Network Model. *IEEE Transactions on Neural Networks*, 5 (6), 865–872.
- Nannen, V., Eiben, A.E., 2007. Relevance Estimation and Value Calibration of Evolutionary Algorithm Parameters. *International Joint Conference on Artificial Intelligence*, pp. 975–980. January 6-12, Hyderabad (India).
- Nguyen, D., Widrow, B., 1990. Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights. *International Joint Conference of Neural Networks*, pp. 21–26. June 17-21, San Diego (California, USA).
- Panchal, G., Ganatra, A., Kosta, Y.P., Panchal, D., 2010. Searching Most Efficient Neural Network Architecture Using Akaike 's Information Criterion (AIC). *International Journal of Computer Applications*, 1 (5), 41–44.
- Pinto, N., Doukhan, D., DiCarlo, J.J., Cox, D.D., 2009. A High-Throughput Screening Approach to Discovering Good Forms of Biologically Inspired Visual Representation. *PLoS Computational Biology*, 5 (11), 1–12.
- Ryan Hsiao, T.C., Lin, C.W., Chiang, H.K., 2003. Partial Least-Squares Algorithm for Weights Initialization of Backpropagation Network. *Neurocomputing*, 50, 237–247.
- Venkatesan, D., Kannan, K., Saravanan, R. 2009. A Genetic Algorithm-Based Artificial Neural Network Model for the Optimization of Machining Processes. *Neural Computing and Applications*, 18 (2) (January 15), 135–140.
- Waghmare, L.M., Nikhil, N.B., Bhogle, P.P. 2007. Neural Network Weight Initialization. *2007 International Conference on Mechatronics and Automation*, pp. 679–681. August 5-8, Harbin (China).
- Yam, J.Y.F., Chow, T.W.S., 2000. A Weight Initialization Method for Improving Training Speed in Feedforward Neural Network. *Neurocomputing* 30 (1-4) (January), 219–232.
- Yam, Y.F., Leung, C.T., Tam, P.K.S., Siu, W.C., 2002. An Independent Component Analysis Based Weight Initialization Method for Multilayer Perceptrons. *Neurocomputing*, 48 (1-4), 807–818.

## AUTHORS BIOGRAPHY



optimization problems.

**DAVID ORIVE** finished his studies in computer engineering at University of Deusto in 2014. Currently he is a research assistant at DeustoTech Institute of Technology working on evolutionary algorithms applied to



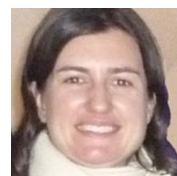
collaborating with DeustoTech Energy in projects related to computer vision and processes optimization. Currently, he is developing his Ph.D. Thesis in optimization and control of the BTO process.

**GORKA SORROSAL**, graduated in Automatic and Industrial Electronic engineering at University of Deusto in 2011, and Master in Control, Automatic and Robotic Engineering at University of the Basque Country in 2013. Ever since 2009, he has been



working in prediction models of power demand and energy consumptions as well as the introduction of Automatic Modelling in the catalyst design processes in the Energy Unit of DeustoTech.

**CRUZ E. BORGES** In 2005 he received his diploma degree in Mathematics from University of La Laguna (Spain). Then he joined Ph.D program “Mathematics and its applications” in University of



techniques for the uncertainty analysis. Cristina worked for eight years in the CEIT (San Sebastián) where she participated in projects of water quality mathematical modelling, development of systems engineering techniques, Bayesian optimization algorithms, etc. Her career continued with a postdoc at Laval University (Canada), where she participated in various research projects for the design and optimisation of wastewater treatment plants. She is currently researcher at the Department of Environment and Energy of DeustoTech (University of Deusto).

**CRISTINA MARTIN**, PhD in Industrial Engineering at the University of Navarra. Her research work has been related to the development of mathematical models, systems analysis and numerical



**AINHOA ALONSO-VICARIO**,  
PhD in Chemical Engineering from  
the University of the Basque  
Country and Head of the Energy  
Unit of DeustoTech. After his PhD,  
he joined the R&D Division of  
Sustainable Energy Production and

Leia CDT Foundation where he worked for two years.  
In November 2009 he joined as a researcher at the  
Energy Unit of DeustoTech, and since September 2011  
he is Head of the Unit. His research interests focus on:  
the integration of distributed energy sources and CO<sub>2</sub>  
capture processes, and improving the energy efficiency  
and environmental processes using artificial  
intelligence.



# MODELING THE MULTI-COMPARTMENT VEHICLE ROUTING PROBLEM WITH STOCHASTIC DEMANDS

Jan Melechovský<sup>(a)</sup>

<sup>(a)</sup>Department of Econometrics, University of Economics, náměstí W. Churchilla 4, 130 67 Prague 3, Czech Republic

<sup>(a)</sup>[jan.melechovsky@vse.cz](mailto:jan.melechovsky@vse.cz)

## ABSTRACT

Vehicle routing problems represent a class of combinatorial optimization problems widely studied in the literature. Practical applications often involve a level of uncertainty, which complicates the decision making process. This paper presents two models of the multi-compartment vehicle routing problem with stochastic demands (MCVRPSD). The problem consists in finding a minimum cost set of vehicle routes serving a set of customers. Each customer can require the delivery of  $m$  products. The products must be transported in different compartments due to their physical incompatibility. The problem is modeled as a stochastic program with recourse. The recourse action consists in a return to the depot vertex whenever the servicing vehicle cannot satisfy the customer's demand for a particular product. In an alternative model the failure can also occur due to time constraints. A hybridized evolutionary algorithm is presented to address the problem.

Keywords: vehicle routing, stochastic programming, metaheuristic, multiple compartments

## 1. INTRODUCTION

The multi-compartment vehicle routing problem (MCVRP) extends the classical formulation of the vehicle routing problem (VRP). In the VRP, a set of customer nodes is given. Each customer requires the delivery of a nonnegative quantity of a product. A vehicle fleet situated at a depot node is available to service the customers. In the classical formulation of the VRP, the vehicle fleet is unlimited and homogeneous, i.e. all vehicles have equal capacity. The solution of the VRP consists in determining a minimum cost set of vehicle routes satisfying the total demand of each customer. Each route is originating and ending at the depot node and respects the capacity of a vehicle. The VRP was first formulated by Dantzig and Ramser (1959). A survey on VRP applications and solution approaches can be found in e.g. Toth and Vigo (2002) or Golden et al. (2008). The VRP is a NP-hard problem since it generalizes the well-known travelling salesman problem. Exact algorithms are limited to instances of a moderate size. Hence metaheuristic algorithms have been proposed to address VRPs of a larger size.

In the MCVRP, each customer can require the delivery of multiple products. The products have specific characteristics and cannot be transported together in one room. Each vehicle is therefore equipped with multiple separated compartments, each dedicated to transport a particular product. In the basic formulation of the MCVRP the vehicle fleet is again assumed to be unlimited and homogeneous. The objective is to satisfy all customer demands at minimum cost. Each product must be transported in the right vehicle compartment and the capacity of each compartment must be respected. Each vehicle route starts and ends at the depot. In the MCVRP, the delivery of two different products can be done in two separate routes. This feature is called partial split of the delivery.

A typical application of the MCVRP is e.g. the delivery of petroleum products to petrol stations using tank trucks with several compartments (see e.g. Avella et al. (2004), Brown and Graves (1981), Brown et al. (1987). In this application, the demand of a customer for a particular petrol type is quite large and often saturates the whole compartment capacity. The main problem then consists of determining the assignment of requests to tank trucks, such that the available capacity is used optimally. The routing part of the problem is then relatively easy since a TSP with a small number of stops is solved for each truck separately. Another practical application is the delivery of groceries to convenience stores Chajakis and Guignard (2003). Each product requires different temperature, e.g. low and ambient temperature compartments must be used. Customers have limited inventory capacities and small but frequent orders are therefore preferred. Animal food distribution to farms is an example of the MCVRP addressed by El Fallahi et al. (2008). The authors modeled the case when the food must be transported separately for certain types of animals due to sanitary reasons. The MCVRP can also model systems of waste co-collection when different types of waste can be collected using vehicles with multiple compartments Muyldermans and Pang (2010).

Practical applications of routing problems often involve a level of uncertainty in the input data. Such a situation arises when the demand of customer is rather a random variable than a deterministic value. The

MCVRP with stochastic demands (MCVRPSD) models this issue. For each customer, the demand for a given product is a random variable with a known probability distribution characterized with a mean and a deviation. The real demand is always nonnegative and is less than or equal the vehicle capacity  $Q_p$ . However, the exact demand is known only upon vehicle's arrival to the customer.

The basic modeling approaches to stochastic routing problems can be classified into static or dynamic models. Routing decisions in dynamic models are performed in multiple stages and depend on prior demands realizations. A re-optimization action is taken at each stage depending on the remaining vehicle capacity and set of unserved customers. An example of a dynamic model is the Markov Decision Process proposed by Secomandi and Margot (2009).

In the static model, routing decisions remain unchanged. This approach is more convenient for difficult-to-solve problems since the problem is solved only once. A typical static stochastic routing model consists of two stages. In the first stage an *a priori* solution is calculated while in the stage the routes are executed. In case of a route failure, e.g. the vehicle does not have enough quantity of a particular product to satisfy the total demand of a customer, a recourse action is performed. The recourse action may e.g. consist in a return trip to the depot, refill of the vehicle compartments up to their capacity, and a resume trip back to the customer node where the failure occurred. A two stage stochastic programming model for the MCVRPSD was proposed by Mendoza et al. (2010).

This paper presents a two-stage stochastic programming model for the MCVRPSD. A solution based on the expected demand quantities is determined in the first stage. The presented model of the problem considers two particular situations. In the first case the stochastic demand affects only the available quantity of a particular product to be delivered. The solution must be adapted so that all customers receive their demanded quantities. In the alternative model, the factor of time is additionally taken into account. The total time duration of a route is limited to  $T_{max}$  while the unloading time  $s_i$  at customer  $i$  is a function of the unloaded quantity.

## 2. PROBLEM MODEL

The problem is defined on undirected graph  $G(V,E)$ , where  $V=\{0,1,2,\dots,n-1\}$  denotes the set of  $n$  nodes and  $E$  denotes the set of edges. The subset  $N = V \setminus \{0\}$  represents the customers indexed from 1 to  $n-1$  while the index 0 is reserved for the depot node. Each customer  $i \in N$  can require the delivery of a nonnegative quantity  $q_{ip}$  of product  $p \in P$ . If  $q_{ip} = 0$  the product is not ordered by the customer. The number of products is denoted  $m$  in the sequel. In the stochastic model, the demand of customer  $i$  for product  $p$  is a random variable  $\lambda_{i,p}$  with a known probabilistic distribution with mean  $\mu_{i,p}$  and standard deviation  $\sigma_{i,p}$ . The exact value of  $q_{ip}$  is therefore known only upon vehicle's arrival to  $i$ . A service time  $s_{ip} = \alpha \cdot q_{ip}$  is associated with the delivery of

product  $p$  to customer  $i$ .  $\alpha$  is a constant value representing the time needed to unload one unit of a product. A fleet  $K$  of identical vehicles based in the depot is available to service the customers. Each vehicle is equipped with  $m$  compartments. Each compartment has a fixed capacity  $Q_p$  to accommodate product  $p$ . Each edge  $e(i,j) \in E$  is associated with travel cost  $c_{ij}$  and travel time  $t_{ij}$ . Both values are assumed to be nonnegative and both sets of values satisfy the triangle inequality. The solution of the MCVRP consists in determining a set of vehicle routes, each route starting and ending at the depot, such that the demands of each customer are totally satisfied, the capacities of each vehicle are respected and the total travel cost is minimized. The problem is NP-hard since it extends the classical VRP.

The two stage stochastic programming model for the MCVRPDS was proposed in Mendoza et al. (2010). Let  $S$  be the planned (a priori) solution consisting of  $n_s$  routes. Each route  $r \in S$  starts and ends at the depot and visits a sequence of customers  $(v_1, v_2, \dots, v_l)$ . The total cost of the solution in the second stage is given by the equation:

$$C(S) = \sum_{r \in S} C_r + \sum_{r \in S} F_r(\vec{\lambda}) \quad (1)$$

where  $C_r$  is the planned cost of route  $r$  and  $F_r(\vec{\lambda})$  is the cost of route failure (recourse).

Then the problem to solve in the first stage is to determine the set of routes  $S$  minimizing the expected cost:

$$E[C(S)] = \sum_{r \in S} C_r + \sum_{r \in S} E[F_r(\vec{\lambda})] \quad (2)$$

The maximum route duration constraints can be expressed as follows:

$$T_r + E[G_r(\vec{\lambda})] + E[H_r(\vec{\lambda})] \leq T_{max} \quad \forall r \in S \quad (3)$$

where  $T_r$  denotes the travel time of the planned route  $r$ ,  $G_r(\vec{\lambda})$  the time of the recourse trips and  $H_r(\vec{\lambda})$  the expected unloading time that is counted for  $r$ .

The cost of the recourse depends on the failure probability  $P_r(v_i)$  at customer  $v_i$  and the cost  $c_{0i}$ :

$$E[F_r(\vec{\lambda})] = \sum_{i=1}^l 2c_{0v_i} \times P_r(v_i) \quad (4)$$

Analogically, the time of the recourse can be calculated as:

$$E[G_r(\vec{\lambda})] = \sum_{i=1}^l 2t_{0v_i} \times P_r(v_i) \quad (5)$$

For the calculation of the failure probability  $P_r(v_i)$  at node  $v_i$ , Mendoza et al. (2010) proposed the following approximation scheme:

$$P_r(v_i) = \sum_{j=0}^{i-1} \left[ \prod_{p=1}^m \Phi(v_{j+1}, v_{i-1}, p) - \prod_{p=1}^m \Phi(v_{j+1}, v_i, p) \right] \cdot P_r(v_j) \quad (6)$$

The term  $\Phi(v_{j+1}, v_{i-1}, p)$  expresses the cumulative probability that the demand for product  $p$  between nodes at positions  $j+1$  and  $i-1$  does not exceed the capacity  $Q_p$  of the corresponding vehicle compartment. Hence the first product approximates the probability that the routes remains feasible after servicing customers  $v_{j+1}, \dots, v_{i-1}$ . Similarly, the second product approximates the probability that the route remains feasible after servicing customer  $v_i$ .

Finally, the expected unloading time depends on the unloaded quantity at each node:

$$E[H_r(\vec{\lambda})] = \sum_{i=1}^l \sum_{p=1}^m \alpha \lambda_{v_i, p} \quad (7)$$

### 3. THE MULTI-START EVOLUTIONARY ALGORITHM

The evolutionary local search method (ELS) was originally proposed by Merz and Wolf (2007) for a peer-to-peer problem in telecommunications. It extends the classical iterated local search (ILS). A detailed description of the ILS can be found in Lourenço et al. (2003). The ILS, given an incumbent solution  $s$ , successively generates child solutions using a perturbation mechanism and a local search. At each iteration,  $s$  is updated with a child solution only if some acceptance criterion is met. The perturbation mechanism serves as a diversification tool while the local search intensifies the search in the current solution neighborhood. The ELS additionally generates multiple child solutions at one iteration. Only the best child solution is kept and it becomes the incumbent solution  $s$  only if it fulfills the acceptance criterion.

The multi-start feature extends the capability of ELS. It can be viewed as the greedy randomized adaptive search procedure (GRASP) in which the local search is implemented as the ELS. In the MS-ELS the search is restarted with a randomized initial solution

each time when the algorithm is being trapped in the local optimum. The original idea of multi-start ELS (MS-ELS) is due to Prins (2009).

```

1: $tc \leftarrow 0$
2: $ni \leftarrow 0$
3: repeat
4: if $tc = 0$ then $\rho \leftarrow 1$
5: else $\rho \leftarrow \rho_{max}$
6: $s \leftarrow \text{Randomized_Initial_Heuristic}(\rho)$
7: if $tc = 0$ then $s^* \leftarrow s$
8: $\pi \leftarrow \pi_{min}$
9: for N_e iterations do
10: $cs^* \leftarrow \emptyset$
11: $Cost(cs^*) \leftarrow \infty$
12: for N_c child solution iterations do
13: $cs \leftarrow s$
14: $cs' \leftarrow \text{Perturbation}(cs, \pi)$
15: $cs'' \leftarrow \text{Local_Search}(cs')$
16: if $Cost(cs'') < Cost(cs^*)$ then
17: $cs^* \leftarrow cs''$
18: endif
19: $tc \leftarrow tc + 1$
20: if $Cost(cs^*) < Cost(s^*)$ then
21: $ni \leftarrow 0$
22: $s^* \leftarrow cs^*$
23: else $ni \leftarrow ni + 1$
24: if $tc \geq N_t$ or $ni \geq NI_t$ goto line 31
25: endfor
26: if $Cost(cs^*) < Cost(s)$ then
27: $\pi \leftarrow \pi_{min}$
28: $s \leftarrow cs^*$
29: else $\pi \leftarrow \min(\pi + 1, \pi_{max})$
30: endifor
31: until $tc \geq N_t$ or $ni \geq NI_t$

```

Algorithm 1. The Multi-Start Evolutionary Local Search

The MS-ELS is used to solve the first stage problem minimizing the expected cost  $E[C(S)]$ . The algorithm implementation is shown in Algorithm 1. First the total iterations counter  $tc$  and the non-improving iterations counter  $ni$  are initialized with 0. Multiple restarts of the algorithm with a different incumbent solution  $s$  are ensured by the main loop between lines 3 – 31. The solution diversity within each restart is ensured with a randomized initial heuristic using a randomization parameter  $\rho$  denoting the number of admissible extensions of a partial solution at an iteration of the initial heuristic. The first initial solution is obtained with no randomization ( $\rho = 1$ ). The best solution  $s^*$  is initialized at the beginning of the first iteration with the initial solution. The perturbation parameter  $\pi$  of the ELS is set to its minimum value  $\pi_{min}$ . This parameter controls the perturbation strength and it is dynamically updated during the run. The ELS (lines 9 – 30) performs at most  $N_e$  iterations. At every ELS iteration,  $N_c$  child solutions are generated using the

perturbation mechanism and the local search. Only the best child solution  $cs^*$  among all child solutions generated is kept in the memory. In case it improves also the global best solution  $s^*$  the latter is updated with  $cs^*$  and  $ni$  is set to 0. The algorithm can be terminated prematurely if one of the iterations counter attains its maximum limit. If  $cs^*$  improves the incumbent solution  $s$ , it is updated with  $cs^*$  and  $\pi$  is reset to  $\pi_{min}$ . Otherwise  $\pi$  is incremented. The algorithm stops when either  $tc$  reaches the number of maximum total iterations  $N_t$  or  $ni$  reaches the number of maximum non-improving iterations  $NI_t$ .

The key components of the MS-ELS are the randomized initial heuristic, perturbation mechanism and a local search mechanism. The following subsections provide implementation details of each component.

### 3.1. Initial heuristic

The initial solution is obtained with a simple construction heuristic. The solution is initialized with an empty route. Then all customers with some unsatisfied demand are scanned and the best insertion of a customer into a route is determined. If the customer cannot be inserted into any already existing route, the insertion into a new route is considered instead. The criterion for the insertion is the least increase of travel costs. The procedure terminates when the demand of all customers is fully satisfied. The method is randomized with a parameter  $\rho$  denoting the number of best insertions (in terms of insertion cost) determined in each iteration which are recorded into a list. The insertion which is finally performed is selected randomly from that list.

### 3.2. Perturbation

The perturbation mechanism is one of the two routines of ELS that modify the incumbent solution. It plays the role of a diversification tool in the general ELS framework since it performs several random operations on a solution. Thus it can be interpreted as a mutation operator used in genetic algorithms. The procedure is controlled with a parameter  $\pi$  denoting the number of operations to be performed. In our implementation the operation is a removal and relocation of a customer. Given a route  $r$  in the solution,  $\pi$  customer nodes are randomly selected and removed from  $r$ .

Each of the  $\pi$  removed customers is then tested for a feasible insertion into some of the remaining routes of the solution. If such feasible insertion is detected, the customer is relocated to its new position. If not, new route visiting only this single customer is added to the solution.

### 3.3. Local search

Local search is applied to the solution modified by the perturbation mechanism. Its purpose is to improve the solution using a set of operators. It intensifies the search and the improved solution represents a local optimum within the given solution neighborhood. Together with the perturbation mechanism it enables the algorithm to

explore effectively the solution space and find potentially good solutions.

The implemented local search procedure relies on three operators:

- 2-opt* – replaces two arcs and reorders nodes,
- Path Exchange* – interchanges two sub-paths between two routes,
- Relocate* – relocates a sequence of nodes within the same route or between two routes,
- Swap* – swaps two sequences of nodes between two routes.

Moreover, *Relocate* and *Swap* can operate with sequences of  $k$  consecutive nodes in a route instead of one single customer. The parameter  $k$  defines the size of the neighborhood to be explored. Starting with  $k = 1$ , the size is dynamically increased up to a maximum possible value  $k_{max}$  if no improvement was found with the current value of  $k$ .

The operators are illustrated in Figures 1 – 6. Each operator replaces a certain number of arcs in the solution. The cancelled arcs are depicted with dashed lines. *2-opt* in Figure 1 operates on a single route. It replaces a pair of arcs and reorders the intermediate sequence of customers. *Path Exchange* in Figure 2 is a variation of *2-opt* but it operates on a pair of routes. It replaces one arc in each route and the order of customers remains preserved. *Relocate* can operate on a single route (Figure 3) or on a pair of routes (Figure 4). This operator replaces three arcs. Finally, *Swap* replaces four arcs either within a single route or a pair of routes. The operator is depicted in Figure 5 and Figure 6.

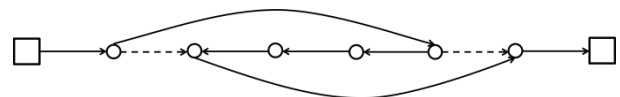


Figure 1. 2-opt exchange.

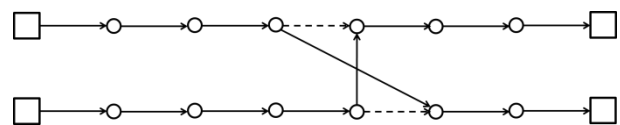


Figure 2. Path exchange.

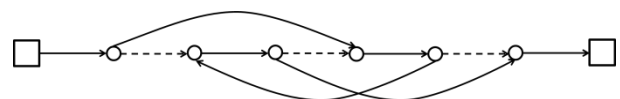


Figure 3. Relocate on a single route.

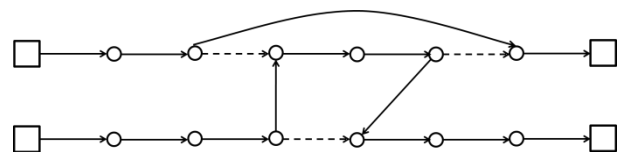


Figure 4. Relocate on a pair of routes.

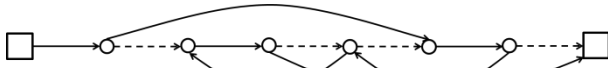


Figure 5. Swap on a single route.

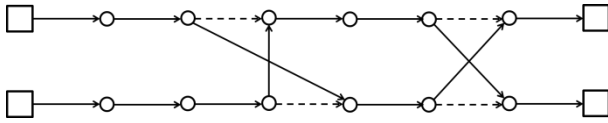


Figure 6. Swap on a pair of routes.

The operators are applied sequentially in the above given order with the first improvement strategy – i.e. the first detected improving move is performed. The search stops when no improving move can be found by any operator. Each move must be checked for the feasibility. This involves the time windows as well as the capacity. The time windows feasibility can be checked in  $O(1)$  if an information of the maximum feasible shift of each node visit is kept in the memory. However, the update of this information must be done for each node and requires  $O(n)$ . The capacity must be checked for all vehicle compartments if the move involves two routes. It can be implemented in  $O(1)$ .

Every insertion or removal of a customer requires the evaluation of the cost difference generated by the operation. The cost difference is composed of the part given by the deterministic travel cost and the part expressing the impact of the operation on the failure probability. The second part requires the recalculation for each customer subsequent the operation. Even the approximation scheme for the evaluation of the failure probability given by Equation 6 is computationally expensive. As shown in Mendoza et al. (2010), the recalculation of the failure probability requires  $O(mn^2)$ . For this reason the authors suggested to evaluate the recourse cost only in some cases. In the presented algorithm, the recourse cost is evaluated only when selecting the best insertions within the initial heuristic and the perturbation. The local search procedure relies only on the deterministic travel cost.

## 4. COMPUTATIONAL EXPERIMENTS

### 4.1. Testing environment

Two sets of testing instances for deterministic MCVRP were proposed by El Fallahi et al. (2008). The authors derived two data sets from standard VRP instances available at Beasley's *OR Library* (<http://people.brunel.ac.uk/~mastjjb/jeb/>). Each dataset is based on 20 instances. The first 14 instances (vrp1 – vrp14) contain 50 – 199 customers. The last six instances contain 76 – 484 customers. In both datasets the number of products is 2. The first dataset (*Set 1*) is derived in a straightforward way: the capacity of compartments and customer demands is divided into two equal parts. In the second dataset (*Set 2*), the demand  $q_{i1}$  for product 1 is generated randomly for each  $i$ . The demand for the second product is defined as  $q_{i2} = q_i - q_{i1}$ , where  $q_i$  is the original demand of customer  $i$ .

The capacity  $Q_1$  of the first compartment is determined as follows from the average demand  $\bar{Q}_1$  for the first product, the average demand  $\bar{Q}_2$  for the second product, and the vehicle capacity  $Q$  in the original VRP:  $Q_1 = (Q \times \bar{Q}_1) / (\bar{Q}_1 + \bar{Q}_2)$ . The capacity of the second compartment is given by  $Q_2 = (Q \times \bar{Q}_2) / (\bar{Q}_1 + \bar{Q}_2)$ .

Mendoza et al. (2010) adapted the instances for the stochastic MCVRP. The demands follow normal distribution with  $\mu_{i,p} = d_{i,p}$  for each  $i \in N$  and  $p \in P$ . The standard deviation is calculated as  $\sigma_{i,p} = 0.3 \times \mu_{i,p}$ . Some instances involve the distance upper limit implicitly. For the others the limit is  $L = 4 \times \max_{i \in N} c_{0i}$ . These instances do not involve the service time. For the purpose of preliminary tests, *Set 3* has been derived from *Set 2*. The service time is defined as  $s_{ip} = \alpha \cdot q_{ip}$ , where  $\alpha = 1$ . The travel time is set equal to the travel cost ( $t_{ij} = c_{ij} \forall e(i,j) \in E$ ) and the maximum time of a route is defined as  $T_{max} = L + \alpha \cdot n_r \cdot (\bar{Q}_1 + \bar{Q}_2)$ .  $n_r$  denotes the maximum number of customers serviced in one route in the best solution of the original instance in *Set 2*.

### 4.2. Parameters setting

The proposed algorithm requires eight parameters to be set up. The parameters are summarized in Table 1. The determination of an appropriate set of parameters is not trivial and the task constitutes a particular part of computational experiments. Several tests of the MS-ELS have been therefore carried out on the 14 MCVRP instances from *Set 1*. In these preliminary experiments the demand was considered to be deterministic. Each instance was solved with 41 different parameter configurations and each test was executed five times with various rand seed settings.

Table 1: List of algorithm parameters

| <i>Notation</i> | <i>Description</i>                                                            |
|-----------------|-------------------------------------------------------------------------------|
| $N_t$           | Maximum number of total iterations                                            |
| $NI_t$          | Maximum number of non-improving iterations                                    |
| $N_e$           | Number of ELS iterations                                                      |
| $N_c$           | Number of child solutions                                                     |
| $\pi_{min}$     | Minimum value of perturbation parameter                                       |
| $\pi_{max}$     | Maximum value of perturbation parameter                                       |
| $\rho_{max}$    | Randomization parameter                                                       |
| $k_{max}$       | Maximum size of a sequence in local search ( <i>Relocate</i> or <i>Swap</i> ) |

The results are shown in Table 2. Columns 2 – 9 present the configuration values. The last two columns

present the average gap to the best known solutions (BKS) of the concerned instances and the average running time in seconds respectively. The results are sorted in non-decreasing order of the average gap (ranging from 1.9 % to 5.3 %). The selected configuration is highlighted in bold.

Table 2: Results of MS-ELS obtained with different parameter configurations

|    | $N_t$      | $NI_t$     | $N_e$     | $N_c$     | $\pi_{min}$ | $\pi_{max}$ | $\rho_{max}$ | $k_{max}$ | Gap        | cpu          |
|----|------------|------------|-----------|-----------|-------------|-------------|--------------|-----------|------------|--------------|
| 1  | 1000       | 800        | 10        | 15        | 1           | 5           | 3            | 3         | 1.9        | 313.8        |
| 2  | 1000       | 800        | 10        | 15        | 1           | 3           | 3            | 3         | 2.8        | 283.2        |
| 3  | 1000       | 800        | 10        | 10        | 1           | 5           | 3            | 3         | 2.9        | 296.1        |
| 4  | 1000       | 800        | 10        | 10        | 1           | 3           | 3            | 3         | 2.9        | 291.1        |
| 5  | 500        | 500        | 10        | 10        | 1           | 3           | 3            | 5         | 3.0        | 381.4        |
| 6  | 1000       | 500        | 10        | 15        | 1           | 3           | 3            | 3         | 3.0        | 227.3        |
| 7  | 1000       | 500        | 10        | 10        | 1           | 5           | 3            | 3         | 3.0        | 236.2        |
| 8  | 1000       | 500        | 10        | 15        | 1           | 5           | 3            | 3         | 3.0        | 260.3        |
| 9  | <b>800</b> | <b>500</b> | <b>10</b> | <b>15</b> | <b>1</b>    | <b>5</b>    | <b>3</b>     | <b>3</b>  | <b>3.0</b> | <b>215.2</b> |
| 10 | 800        | 500        | 10        | 15        | 1           | 3           | 3            | 3         | 3.0        | 225.8        |
| 11 | 800        | 500        | 10        | 10        | 1           | 5           | 3            | 3         | 3.1        | 218.2        |
| 12 | 1000       | 500        | 10        | 10        | 1           | 3           | 3            | 3         | 3.1        | 257.2        |
| 13 | 500        | 200        | 10        | 10        | 1           | 3           | 3            | 5         | 3.3        | 252.4        |
| 14 | 1000       | 300        | 10        | 10        | 1           | 5           | 3            | 3         | 3.3        | 160.5        |
| 15 | 1000       | 300        | 10        | 15        | 1           | 5           | 3            | 3         | 3.3        | 162.2        |
| 16 | 500        | 500        | 10        | 10        | 1           | 5           | 3            | 3         | 3.3        | 158.9        |
| 17 | 500        | 500        | 10        | 15        | 1           | 5           | 3            | 3         | 3.3        | 152.2        |
| 18 | 800        | 500        | 10        | 10        | 1           | 3           | 3            | 3         | 3.3        | 227.0        |
| 19 | 1000       | 300        | 10        | 15        | 1           | 3           | 3            | 3         | 3.3        | 150.3        |
| 20 | 500        | 500        | 10        | 10        | 1           | 3           | 3            | 4         | 3.3        | 252.9        |
| 21 | 500        | 500        | 10        | 15        | 1           | 3           | 3            | 3         | 3.4        | 145          |
| 22 | 500        | 200        | 10        | 10        | 1           | 3           | 3            | 4         | 3.6        | 157.6        |
| 23 | 500        | 500        | 10        | 10        | 1           | 3           | 3            | 2         | 3.7        | 80.5         |
| 24 | 500        | 200        | 10        | 15        | 1           | 3           | 3            | 3         | 3.7        | 101.5        |
| 25 | 800        | 200        | 10        | 15        | 1           | 3           | 3            | 3         | 3.7        | 107.2        |
| 26 | 800        | 200        | 10        | 15        | 1           | 5           | 3            | 3         | 3.7        | 106.6        |
| 27 | 500        | 200        | 10        | 15        | 1           | 5           | 3            | 3         | 3.7        | 102.3        |
| 28 | 200        | 200        | 10        | 10        | 1           | 3           | 3            | 5         | 3.8        | 153.8        |
| 29 | 800        | 200        | 10        | 10        | 1           | 5           | 3            | 3         | 3.8        | 110.3        |
| 30 | 500        | 200        | 10        | 10        | 1           | 5           | 3            | 3         | 3.8        | 106.9        |
| 31 | 1000       | 300        | 10        | 10        | 1           | 3           | 3            | 3         | 3.9        | 161.0        |
| 32 | 500        | 500        | 10        | 10        | 1           | 3           | 3            | 3         | 3.9        | 146.1        |
| 33 | 200        | 200        | 10        | 10        | 1           | 3           | 3            | 4         | 4.1        | 101.6        |
| 34 | 200        | 200        | 10        | 15        | 1           | 5           | 3            | 3         | 4.3        | 59.3         |
| 35 | 200        | 200        | 10        | 15        | 1           | 3           | 3            | 3         | 4.3        | 56.9         |
| 36 | 500        | 200        | 10        | 10        | 1           | 3           | 3            | 3         | 4.3        | 95.2         |
| 37 | 800        | 200        | 10        | 10        | 1           | 3           | 3            | 3         | 4.3        | 104.1        |
| 38 | 500        | 200        | 10        | 10        | 1           | 3           | 3            | 2         | 4.4        | 51.0         |
| 39 | 200        | 200        | 10        | 10        | 1           | 5           | 3            | 3         | 4.6        | 61.9         |
| 40 | 200        | 200        | 10        | 10        | 1           | 3           | 3            | 3         | 4.9        | 59.8         |
| 41 | 200        | 200        | 10        | 10        | 1           | 3           | 3            | 2         | 5.3        | 33.1         |

The selection of parameter values was done with respect to the expected impact on algorithm's performance. Most attention was therefore given to parameters  $N_t$  and  $NI_t$ . Contrarily,  $N_e$ ,  $\pi_{min}$  and  $\rho_{max}$  were always set to the same values.  $N_e$  was fixed to 10 and the reason was that it's impact is strongly correlated with  $N_t$  and  $N_c$ . Hence varying these values provided similar behavior as varying the values of  $N_e$ . Regarding  $\pi_{min}$ , the performance did not change or it was even worse for slightly larger values than 1. The randomization parameter  $\rho_{max}$  was set to 3, since larger values led to initial solutions of poor quality.  $N_c$  alternated between 10 and 15. Lower values than 10 led to worse results while values significantly larger (about 20 or more) did not bring any positive impact. Moreover, it would reduce the number of restarts since the portion of total iterations run on the same starting solution would increase. The variability of tested values of  $\pi_{max}$  and  $k_{max}$  was also quite moderate for similar reasons. The parameter  $\pi_{max}$  was tested only with values 3 and 5 and  $k_{max}$  was tested with values 3, 4, and 5. Larger values of  $\pi_{max}$  led to equal or worse solutions while larger values of  $k_{max}$  increased the computing time significantly.

### 4.3. Computational experiments

Preliminary computational experiments of the stochastic version of the proposed MS-ELS were conducted on instances vrp1 – vrp14 of *Set 2* and *Set 3*. As in Mendoza et al. (2010), the results were obtained with a single of the MS-ELS and the stochastic simulation involved 50 000 scenarios. The algorithm was coded in C++ and compiled with Embarcadero RAD Studio 2010. The tools provided by the Boost library were used for statistical calculations and simulations.

The results obtained for *Set 2* are summarized in Table 3. The table compares the performance of the MS-ELS to the best known solutions including the estimated cost of recourse over 50 000 replications reported in Mendoza et al. (2010). Gap is the percentage difference between the solution obtained with MS-ELS and the BKS. Computing times in the last column of the table are in seconds. The solutions found by our MS-ELS are generally close to the BKSs. The gap ranges from 0.2 % to 3.5 %. Mendoza et al. (2010) reported the results for two versions of the memetic algorithm. The basic version much slower but it calculated most of the best known solutions. A faster version of their memetic algorithm represents uses a spare capacity strategy, which relies on the computation of the deterministic problem and leaving some spare capacity in each vehicle compartment. This modification of the algorithm does not require the costly computations of the recourse probability and hence it is far more effective compared to the stochastic version. In Table 3, our results are compared to the stochastic version of the memetic algorithm.



Table 3: Results of the MS-ELS on the adapted Set 2 instances

| <i>Instance</i> | <i>n</i> | <i>BKS</i> | <i>MS-ELS</i> | <i>Gap</i> | <i>Time</i> |
|-----------------|----------|------------|---------------|------------|-------------|
| vrpnc1          | 50       | 568.7      | 569.9         | 0.2        | 123.3       |
| vrpnc2          | 75       | 976.3      | 978.3         | 0.2        | 198.4       |
| vrpnc3          | 100      | 891.8      | 894.6         | 0.3        | 645.3       |
| vrpnc4          | 150      | 1187.3     | 1198.2        | 0.9        | 1230.5      |
| vrpnc5          | 199      | 1497.3     | 1509.5        | 0.8        | 2056.7      |
| vrpnc6          | 50       | 577.6      | 586.0         | 1.5        | 98.5        |
| vrpnc7          | 75       | 1000.9     | 1020.4        | 1.9        | 103.6       |
| vrpnc8          | 100      | 922.1      | 931.3         | 1.0        | 289.5       |
| vrpnc9          | 150      | 1255.8     | 1273.2        | 1.4        | 432.6       |
| vrpnc10         | 199      | 1572.5     | 1587.3        | 0.9        | 1027.1      |
| vrpnc11         | 120      | 1259.1     | 1302.6        | 3.5        | 1356.4      |
| vrpnc12         | 100      | 999.1      | 1011.4        | 1.2        | 823.6       |
| vrpnc13         | 120      | 1558.8     | 1564.2        | 0.3        | 567.2       |
| vrpnc14         | 100      | 1027.4     | 1036.8        | 0.9        | 431.8       |

The second set of computational experiments was conducted on *Set 3*. The instances in this set involve the service time which is spent by the vehicle in the customer site. Recall that the service is a function of the unloaded quantity. The results are reported in Table 4. BKS is the best solution value obtained during five executions of the algorithm. The column MS-ELS reports the average value over the five runs. The gap ranges from 0.8 % to 4.7 %, which can be considered as an acceptable variance in results obtained in different runs of the algorithm.

Table 4: Results of the MS-ELS on the adapted Set 3 instances

| <i>Instance</i> | <i>n</i> | <i>BKS</i> | <i>MS-ELS</i> | <i>Gap</i> | <i>Time</i> |
|-----------------|----------|------------|---------------|------------|-------------|
| vrpnc1          | 50       | 577.1      | 605.5         | 0.8        | 142.9       |
| vrpnc2          | 75       | 1066.7     | 1075.8        | 3.9        | 231.0       |
| vrpnc3          | 100      | 932.2      | 937.0         | 4.1        | 684.6       |
| vrpnc4          | 150      | 1189.5     | 1241.8        | 3.2        | 729.4       |
| vrpnc5          | 199      | 1642.7     | 1648.9        | 2.1        | 1988.5      |
| vrpnc6          | 50       | 628.1      | 630.7         | 2.9        | 102.2       |
| vrpnc7          | 75       | 1037.7     | 1051.1        | 3.1        | 61.4        |
| vrpnc8          | 100      | 1000.2     | 1015.1        | 4.7        | 401.3       |
| vrpnc9          | 150      | 1355.3     | 1417.0        | 3.6        | 537.1       |
| vrpnc10         | 199      | 1599.0     | 1673.1        | 0.9        | 1020.4      |
| vrpnc11         | 120      | 1357.4     | 1367.6        | 3.4        | 495.8       |
| vrpnc12         | 100      | 1053.0     | 1097.6        | 1.8        | 787.2       |
| vrpnc13         | 120      | 1620.5     | 1680.8        | 1.2        | 463.4       |
| vrpnc14         | 100      | 1079.9     | 1091.4        | 0.9        | 394.8       |

The experiments on *Set 3* are rather preliminary. Detailed analysis of the impact of a stochastic service time will be subject of further research. The time limit  $T_{max}$  in *Set 3* was defined such that the optimal solutions do not differ too much from *Set 2*. The solution values obtained on *Set 3* are slightly above the solutions of *Set 2*.

## CONCLUSIONS

This paper presents a MS-ELS algorithm for the MCVRP with stochastic demands. The solution approach is based on a two-stage stochastic programming formulation proposed in Mendoza et al. (2010). The proposed metaheuristic calculates the estimate of solution total cost when the demand for products is a random variable. The algorithm was tested on benchmarks from the literature. Preliminary results indicate that the MS-ELS is a competitive algorithm compared notably to the memetic algorithm proposed by Mendoza et al. (2010). An extension of the problem formulation presented in the paper involves variable service time depending on the unloaded quantity. This property influences the feasibility of the vehicle route due to the limit of time imposed. The next logical step in the problem model will be the inclusion of customer time windows into the model. The variable service time will then influence the feasibility of arrivals to customer sites.

## ACKNOWLEDGMENTS

The research was fully supported by the Czech Science Foundation, project reg. number P402/12/P635. This financial support is greatly acknowledged.

## REFERENCES

- Avella P., Boccia M., and Sforza A., 2004. Solving a fuel delivery problem by heuristic and exact approaches. *European Journal of Operational Research*, 152 (1), 170 – 179.
- Brown G. G. and Graves G. W., 1981. Real-time dispatch of petroleum tank trucks. *Management Science*, 27 (1), 19 – 32.
- Brown G. G., Ellis C. J., Graves G. W., and Ronen D. 1987. Real-time, wide area dispatch of mobil tank trucks. *Interfaces*, 17 (1), 107 – 120.
- Chajakis E. D. and Guignard M., 2003. Scheduling deliveries in vehicles with multiple compartments. *Journal of Global Optimization*, 26, 43 – 78.
- Dantzig G. B. and Ramser J. H., 1959. The truck dispatching problem. *Management Science*, 6 (1), 80–91.
- El Fallahi A., Prins C., and Wolfler Calvo R., 2008 . A memetic algorithm and a tabu search for the multi-compartment vehicle routing problem. *Computers & Operations Research*, 35 (5), 1725 – 1741.
- Golden B. L., Raghavan S. , and Wasil E. A., 2008. editors. *The Vehicle Routing Problem: Latest Advances and New Challenges*. Springer.
- Lourenço H., Martin O., and Stützle T., 2003. Iterated local search. In Frederick S. Hillier, Fred Glover, and Gary Kochenberger, editors, *Handbook of Metaheuristics*, volume 57 of *International Series in Operations Research & Management Science*, pages 320–353. Springer New York.

Mendoza J. E., Castanier B., Guéret C., Medaglia A. L., and Velasco N., 2010. A memetic algorithm for the multi-compartment vehicle routing problem with stochastic demands. *Computers & Operations Research*, 37 (11), 1886 – 1898.

Merz P. and Wolf S., 2007. Evolutionary local search for the super-peer selection problem and the p-hub median problem. In T. Bartz-Beielstein et al., editors, *Lecture notes in computer science*, volume 4771, pages 1 – 15. Springer Berlin/Heidelberg.

Muyldermans L. and Pang G., 2010. On the benefits of co-collection: Experiments with a multi-compartment vehicle routing algorithm. *European Journal of Operational Research*, 206 (1), 93 – 103.

Prins C., 2009. A grasp x evolutionary local search hybrid for the vehicle routing problem. In F.B. Pereira and J. Tavares, editors, *Bio-inspired algorithms for the vehicle routing problem*, volume 16, pages 35 – 53. Springer Berlin/Heidelberg.

Secomandi N. and Margot F., 2009. Reoptimization approaches for the vehicle-routing problem with stochastic demands. *Operations Research*, 57 (1), 214–230.

Toth P. and Vigo D., editors, 2002. *The vehicle routing problem*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

#### **AUTHORS BIOGRAPHY**

**Jan Melechovsky** is a researcher at University of Economics, Prague. He got his Ph.D. degree in Technology and Management of Transportation Systems from Czech Technical University in Prague in 2009. His research activities are focused to optimization problems in transportation networks.

# INTEGRATED SIMULATION AND OPTIMIZATION IN HEURISTICLAB

Andreas Beham<sup>(a, b)</sup>, Gabriel Kronberger<sup>(a)</sup>, Johannes Karder<sup>(a)</sup>,  
Michael Kommenda<sup>(a, b)</sup>, Andreas Scheibenpflug<sup>(a)</sup>, Stefan Wagner<sup>(a)</sup>, Michael Affenzeller<sup>(a, b)</sup>

<sup>(a)</sup> Heuristic and Evolutionary Algorithms Laboratory  
School of Informatics, Communication, and Media  
University of Applied Sciences Upper Austria,  
Softwarepark 11, 4232 Hagenberg, Austria

<sup>(b)</sup> Johannes Kepler University Linz  
Institute for Formal Models and Verification  
Altenberger Straße 69, 4040 Linz, Austria

<sup>(a)</sup> [andreas.beham@heuristiclab.com](mailto:andreas.beham@heuristiclab.com), [gabriel.kronberger@heuristiclab.com](mailto:gabriel.kronberger@heuristiclab.com), [johannes.karder@heuristiclab.com](mailto:johannes.karder@heuristiclab.com),  
[michael.kommenda@heuristiclab.com](mailto:michael.kommenda@heuristiclab.com), [andreas.scheibenpflug@heuristiclab.com](mailto:andreas.scheibenpflug@heuristiclab.com),  
[stefan.wagner@heuristiclab.com](mailto:stefan.wagner@heuristiclab.com), [michael.affenzeller@heuristiclab.com](mailto:michael.affenzeller@heuristiclab.com)

## ABSTRACT

Process simulation has many applications that are closely related to optimization. Finding optimal steering parameters for the simulated processes is an activity in which the simulation model is often used as an evaluation function to an optimization procedure. Combining optimization and simulation has been achieved in the past already, however optimization procedures implemented in simulation software are often only black box solvers that are difficult to change, extend or parameterize. Optimization software frameworks on the other hand host a range of suitable algorithms, but often lack the ability to describe and run simulation models. Exchange protocols have been proposed in the past, however the interchange has still proven to be complex and work on simplification is ongoing. In this work, we want to pursue a different approach. We intend to integrate simulation capabilities into an optimization framework and thus want to better support applications for simulation-based optimization. We will describe a suitable generic simulation framework and its integration into HeuristicLab. A case study is presented as a demonstration of its usefulness.

Keywords: simulation, optimization

## 1. INTRODUCTION

Work on simulation-based optimization procedures is ongoing and has profited a lot from the steady rise in available computational power. Simulation models are complex functions that are time consuming to run. Often there needs to be a warm-up phase of the model until the process performance has been stabilized and the indicators can be computed. This further lengthens the timespan that the model has to cover and increases the amount of data that has to be processed. We have addressed the problem of simulation-based optimization

(Affenzeller et al. 2007), its integration into HeuristicLab (Wagner et al. 2014) and further developments to generic and extensible protocols (Beham et al. 2012) with the aim of supporting parallel model evaluation in order to speed up the optimization. In this work the aim is to introduce simulation capabilities into the optimization software environment HeuristicLab. Instead of having to connect the optimization software with the simulation environment we want to allow writing and evaluating models in HeuristicLab more easily, especially those that are tied to optimization problems. This should facilitate the application of simulation-based optimization for smaller projects, academic examples, as well as more complex simulation models. For this purpose we have added a simulation core to the HeuristicLab framework that supports modeling and execution of simple and complex simulation models.

## 2. RELATED AND PREVIOUS WORK

This field has been actively researched for several years and a number of solutions have emerged. There are many commercial simulation frameworks that feature embedded optimizers, e.g. AnyLogic or FlexSim which include OptQuest (Laguna and Marti 2002). There are optimization frameworks that must be connected with simulation frameworks through generic interfaces such as HeuristicLab or GenOpt (Wetter 2001). Finally, software is available which allows performing simulation and optimization. For instance, MATLAB (<http://www.matlab.com>) includes SimuLink and also provides optimization algorithms in the global optimization toolbox. SciLab (<http://www.scilab.org>) is an open source environment for numerical computation that includes the Xcos simulator and features basic optimization algorithms such as genetic algorithms and simulated annealing. ECJ is another Java-based optimization framework that connects well with the

MASON simulation framework, both having the same creators. Finally, a large number of optimization frameworks exist which feature real-valued optimization, multi-objective problems and include classical operations research problems from the domain of production, logistics, and supply chain. However often functionality for integration with simulation environments is missing. A review can be found in (Parejo et al. 2012).

### 3. INTEGRATED SIMULATION FRAMEWORK

One major requirement for choosing a simulation framework to integrate in HeuristicLab is the programming language. The framework has to be implemented in a language that can be used within the .Net framework. Otherwise, it is required to cross the language boundary by e.g. serializing to a common format which is precisely what was already realized in (Beham et al. 2012). These languages should ideally be native languages to the .Net Framework such as C#, F# or VB.Net.

From the many existing simulation frameworks that seem suited for integration we chose to go with SimPy (Matloff 2008), more specifically the current version SimPy 3, available at <http://pypi.python.org/pypi/simpy/>. In evaluating the framework we found that the models expressed with SimPy can be short, clear, and self-descriptive. Additionally, researchers have since created more specific layers to the rather slim core, e.g. ManPy “Manufacturing in Python” for manufacturing simulation (Dagkakis et al. 2013) which is developed within the EU FP7 funded project DREAM (<http://dream-simulation.eu/>). A model’s processes can be described in SimPy by writing a program, however the model can be expressed without boilerplate code and without using a complex object-oriented design. As the name suggests the framework is written in Python and can be integrated in HeuristicLab via IronPython (<http://ironpython.net/>) as described in (Beham et al. 2014). However, we noted that some models are rather slow to execute using IronPython and so we created a C# port of SimPy which we termed Sim# and which is available on GitHub (<https://github.com/abeham/SimSharp>). An implementation of an m/m/1 queuing model in Sim# is given in Listing 1.

Processes in this framework are defined as .Net iterators, similarly to SimPy where they are described as Python generators. The process runs until a “yield return” statement is reached where it awaits the yielded event. After the event has been completed the process wakes up and continues to run until the next “yield return” statement or terminates. Termination of processes is also an event so that processes can wait on other processes. In the described case in Listing 1 the process is not yielded, thus process *MM1Q* does not wait on its termination. The environment maintains the current time as well as the event queue in which the events are inserted.

The Sim# framework makes it rather easy to describe simple models with very few lines of code - in SimPy it would be even slightly less code as it does not

require variable declaration, types or visibility declarations. Having to deal with statically typed languages creates some overhead, however as Sim# models are compiled they are faster to execute, something which is relevant for simulation-based optimization.

```
static Environment env;
static Resource server;
static TimeSpan arrival_time = ...
static TimeSpan proc_time = ...
static TimeSpan simulation_time = ...

static IEnumerable<Event> MM1Q() {
 while (true) {
 yield return env.TimeoutExponential(arrival_time);
 env.Process(Item());
 }
}

static IEnumerable<Event> Item() {
 using (var s = server.Request()) {
 yield return s;
 yield return env.TimeoutExponential(proc_time);
 }
}

public static void Main(string[] args) {
 env = new Environment(randomSeed: 42);
 server = new Resource(env, capacity: 1);
 env.Process(MM1Q());
 env.Run(simulation_time);
}
```

Listing 1: Queuing model implemented with Sim#.

### 4. SIMULATION OPTIMIZATION PROBLEM

Writing models in code is a very natural way for programmers and software developers. It is also natural for researchers in the field of mathematical optimization to express fitness functions and constraints as code, such as e.g. a linear program expressed in CPLEX. In a similar sense the integration of such models in HeuristicLab and the task of optimizing their parameters should also remain simple and powerful.

The main use of Sim# simulation models will be in place of an evaluation function. Instead of evaluating a mathematical expression for a given solution, the simulation model should be initialized and run. The performance criteria that result from the model can then be used to form a fitness value which should be either maximized or minimized by the algorithm.

To provide such an integration we have extended HeuristicLab with the introduction of a programmable optimization problem. This problem allows the user to specify the *configuration* in terms of binary, integer, real, and permutation parameters that have to be optimized. There can be an arbitrary number of parameters, each of them having a name that needs to be unique for this configuration. Integer and real parameters are defined in a certain interval, and for integer parameters a step size can be configured for coarse discretization. Permutation parameters define a length for each permutation and a type on how it may be interpreted (absolute, relative directed, or relative undirected). In a second method the

user can then specify the evaluation function which receives a *ParameterVector* that acts as a container for this configuration and that allows to access the parameters' values. The evaluation function also receives a random number generator which can be used for example for stochastic simulation models.

The configuration specifies the solution space, while the parameter vector denotes a concrete solution that the algorithm has identified and should be evaluated. An implementation of a function minimization problem in form of a programmable problem is given in Listing 2. In the evaluation function, any kind of C# code can be executed, including the initialization and execution of simulation models which will be the main interest of this paper.

```
public class Fmin : ISingleObjectiveProblemDefinition {
 public bool IsMaximizationProblem {
 get { return false; }
 }
 public Configuration GetConfiguration() {
 return new Configuration()
 .AddReal("r", length: 5, min: -10.0, max: 10.0);
 }
 public double Evaluate(IRandom random,
 ParameterVector vector) {
 var quality = 0.0;
 quality = vector.Real("r").Select(x => x * x).Sum();
 return quality;
 }
}
```

Listing 2: Implementation of a simple function minimization problem in form of a programmable problem. The parameter configuration and the evaluation function need to be defined.

In addition to the objective function and the encoding, a number of manipulation operations must also be defined that perturb the solution. Different metaheuristics such as genetic algorithms require certain perturbation operators, e.g. crossover and mutation. These operators are automatically configured with default values and can be changed by the user.

This configuration is not trivial however and demands a closer look: The problem class reacts on each change to the script in that it tries to compile the script, instantiate the class and retrieve the configuration from the instance. For each of the parameters defined in this configuration, the problem will then create a solution creator and lookup the relevant operators. It will configure the solution creators in *HeuristicLab* to create the variable with the specified name and all the operators to modify that variable. There are two distinct cases to mention:

1. A single parameter type is chosen, e.g. only a vector of real values.
2. Multiple parameter types are combined, e.g. a vector of real values and a binary vector.

This distinction is important as in the first case the problem may be solved by algorithms which can handle the specific encoding. For a real vector, e.g. the CMAES implementation in *HeuristicLab* could be applied. In the

second case the algorithms would need to support a combined encoding which is realized currently only for variants of genetic algorithms.

## 5. CASE STUDY: SUPPLY CHAIN

Supply chain management is concerned with the flow of goods. Along the supply chain excessive demands and backlog should be avoided, inventory costs and service times should remain low. The famous beer game (Hammond 1994) allows students to explore the effects of demand propagation in a supply chain that consists of 4 echelons and fixed customer demand at the end of the chain. The demand is usually constant for a few periods, but raises after a while. Typically the game is played by students on a table where they should experience the bullwhip effect (Lee et al. 1997) which means that the orders along the chain increasingly start to fluctuate. The factory has to cope with order sizes that bear no relation to the customer demand. This in turn creates a very costly supply chain as inventory levels and backlogs become very high. In this case study, the computer will control the orders using the well-known (S, s) order policy (Scarf 1993) and the parameters will be adjusted by the optimization algorithm. A bad setting of the parameters could result in high costs.

The supply chain consists of 4 echelons which are the factory, the regional distributor, the wholesaler, and the retailer. In each stage the orders will be placed to the next stage. The retailer orders from the wholesaler which in turn orders from the distributor which again orders from the factory. Customers arrive at the retailer and attempt to buy the goods or place an order in case the goods are not available. An order will have to be served as one delivery, order splitting and partial delivery is not considered.

The (S, s) order policy controls the minimum stock level at which an order will be placed (s), as well as the maximum amount of stock that should be kept (S). The difference between S and s basically determines the order size, although already placed orders and the backlog is taken into account. At any given time, if  $Stock + ExpectedDeliveries - BacklogOrders$  becomes smaller or equal to s a demand arises which will be placed to the next stage in the supply chain.

The model, as implemented in *Sim#* is given in Listing 4. Again, an advantage of *Sim#* and also *SimPy* is the textual representation of models which makes it easy to include them together with the publications, e.g. in an appendix if the model is larger.

```
class SupplyChain : SimSharp.Environment {
 public Echelon Factory;
 public Echelon Distributor;
 public Echelon Wholesaler;
 public Echelon Retailer;
 private bool mainSeason;

 public SupplyChain(DateTime start, IntegerVector sS)
 : base(start) {
 Factory = new Echelon(this, 15, null,
 sS[3], sS[3] + sS[7]);
 }
}
```

```

Distributor = new Echelon(this, 15, Factory,
 sS[2], sS[2] + sS[6]);
Wholesaler = new Echelon(this, 15, Distributor,
 sS[1], sS[1] + sS[5]);
Retailer = new Echelon(this, 15, Wholesaler,
 sS[0], sS[0] + sS[4]);

Process(Season());
Process(Simulation());
}

IEnumerable<Event> Simulation() {
 for (int i = 0; i < 50; i++) {
 Process(Round());
 yield return TimeoutD(1);
 }
}

IEnumerable<Event> Round() {
 var demand = mainSeason ? 9 : 5;
 var beer = Process(Retailer.Deliver(demand));
 Retailer.CalculateCosts();
 Wholesaler.CalculateCosts();
 Distributor.CalculateCosts();
 Factory.CalculateCosts();
 Process(Retailer.Order());
 Process(Wholesaler.Order());
 Process(Distributor.Order());
 Process(Factory.Order());

 yield return beer;
}

IEnumerable<Event> Season() {
 yield return TimeoutD(5);
 mainSeason = true;
}
}

class Echelon {
 SupplyChain env;
 Echelon supplier;
 double s, S;
 public Container Stock;
 public double BacklogOrders, BacklogDeliveries;
 public double BacklogCosts, InventoryCosts;

 public Echelon(SupplyChain env, double initialStock,
 Echelon supplier, double s, double S) {
 this.env = env;
 this.Stock = new Container(env, initial: initialStock);
 this.supplier = supplier;
 this.s = s;
 this.S = S;
 }

 public IEnumerable<Event> Order() {
 var newStock = Stock.Level + BacklogDeliveries -
 BacklogOrders;
 if (newStock <= s) {
 var d = S - newStock;
 if (d > 0) {
 if (supplier != null) {
 var shipment = env.Process(supplier.Deliver(d));
 BacklogDeliveries += d;
 yield return shipment;
 var delivery = (ContainerGet)shipment.Value;
 Stock.Put(delivery.Amount);
 } else { // factory
 BacklogDeliveries += d;
 yield return env.TimeoutD(2); // produce
 }
 }
 } else {
 Stock.Put(d);
 }
 BacklogDeliveries -= d;
 }

 public IEnumerable<Event> Deliver(double amount) {
 BacklogOrders += amount;
 var delivery = Stock.Get(amount);
 yield return delivery;
 BacklogOrders -= amount;
 yield return env.TimeoutD(2); // transport
 env.ActiveProcess.Succeed(delivery);
 }

 public void CalculateCosts() {
 BacklogCosts += BacklogOrders;
 InventoryCosts += 0.5 * Stock.Level;
 }
}

```

Listing 3: The model of the beer game supply chain implemented in Sim#. This is the full model description, the  $s$  and  $S$  parameters are given in an IntegerVector which is a HeuristicLab datastructure.

## 6. OPTIMIZATION

The solution to this parameterization problem can be described in terms of 8 integer parameters. We chose the first 4 items to represent  $s$  while the last 4 items denote  $S - s$ . This is chosen in order to avoid a constrained optimization problem, because obviously it must hold that  $s \leq S$ . The bounds have been chosen as  $[2;100]$  for the first 4 items and  $[0;100]$  for the last 4 items. As a step size we chose to go with 2 so that only even values would be evaluated. We have a set up a genetic algorithm to optimize these parameters using a population size of 100. We then proceeded to vary several of the other parameters of the genetic algorithm in order to identify a good combination. These parameters and their variations are shown in Table 1. Because the parameters are integers and some of these operators are known from the real-valued domain, the results are rounded to the nearest feasible integer (also taking step size into consideration).

Table 1: Parameter variation for the genetic algorithm.

| Parameter     | Variation                                   |
|---------------|---------------------------------------------|
| Elites        | 0, 1                                        |
| Crossover     | Discrete, BLX-0.75-0.25, Heuristic, 1-point |
| Mutator       | Uniform, Normal ( $\sigma=2$ )              |
| Selector      | Roulette, 2-Tournament                      |
| Mutation rate | 0.05, 0.15                                  |

The total number of variations was 64 and each possible configuration was evaluated 10 times. The experiment was run on an Intel Core i7 with 4 physical (8 virtual) cores each running at 2.6 Ghz. Using parallel evaluation of the simulation models an average run of the GA took about 17 seconds to evaluate in which around 100.000 simulation runs have been performed. Executing



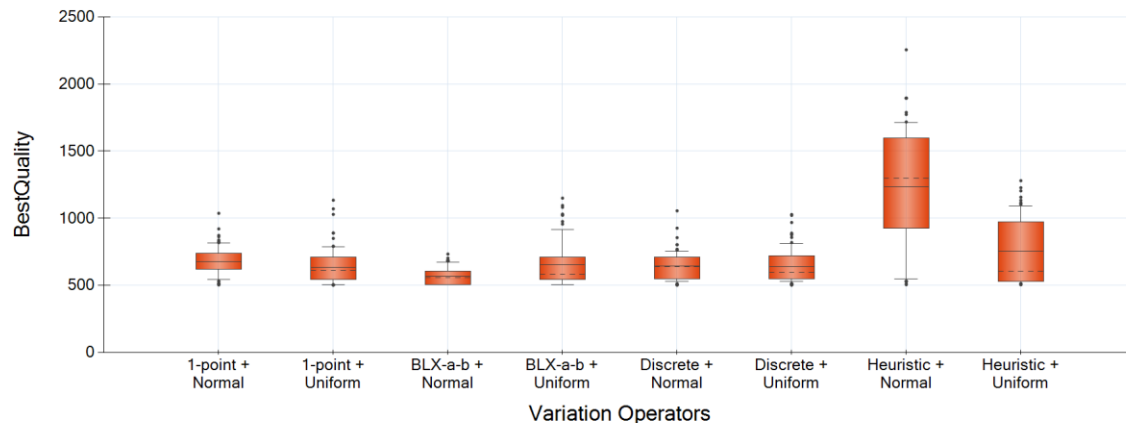


Figure 1: Boxplot of different combinations of crossover and mutation operators for optimizing the order policy parameters.

the algorithm sequentially took about 45 seconds. Parallelizing the evaluation thus presented a nice speed up which HeuristicLab offers out of the box.

We did find a strong dependence on the performance with respect to the crossover and mutator operator that is used. Over the different settings for Elites, Selector, and mutation rates the combination of BLX-a-b and Normal mutation proved most favorably. Figure 1 shows results from this experiment. Each combination of crossover and mutator is given on the x-axis. Because normality cannot be assumed in the present data we used non-parametric statistical significance tests to further evaluate these results. The Kruskal-Wallis analysis of variance is a non-parametric variant of ANOVA which can be used if there is a significant difference among all tested groups. For our experiment we obtained a p-value of 3.01E-30, which suggests a significant difference in the groups. A two tailed Mann-Whitney-U test on the best and second best combination results in a p-value of 0.0004 which also indicates a statistically significant difference.

## 7. CONCLUSIONS

We have shown a new and different approach of mating simulation and optimization in a common framework. The introduced optimization problem allows to define a more complex encoding. We have shown a simple, but powerful and fast simulation framework that can be used to describe models which are run inside an evaluation function.

## ACKNOWLEDGMENTS

The work described in this paper was done within the K-Project Heuristic Optimization in Production and Logistics (HOPL) sponsored by the Austrian Research Promotion Agency (FFG).

## REFERENCES

Affenzeller, M., Kronberger, G., Winkler, S., Ionescu, M., Wagner, S., 2007. Heuristic Optimization Methods for the Tuning of Input Parameters of Simulation Models. In *Proceedings of I3M 2007*, DIPTM University of Genova, 278-283.

- 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. In: *Companion Publication of the 2012 Genetic and Evolutionary Computation Conference, GECCO'12 Companion*, July, Philadelphia, PA, USA, 125-132.
- Beham, A., Karder J., Kronberger, G., Wagner, S., Kommenda, M., Scheibenpflug, S., 2014. Scripting and Framework Integration in Heuristic Optimization Environments. In: *Companion Publication of the 2014 Genetic and Evolutionary Computation Conference, GECCO'14 Companion*, July, Vancouver, CA.
- Dagkakis, G., Heavey, C., Robin, S., Perrin, J., 2013. ManPy: An Open-Source Layer of DES Manufacturing Objects Implemented in SimPy. In: *Proceedings of the 2013 8th EUROSIM Congress on Modelling and Simulation*, September, Cardiff, Wales, UK, 357-363.
- Hammond, J.H., 1994. "Beer Game, The: Board Version." Harvard Business School Background Note 694-104, June 1994 (revised October 1999).
- Laguna, M., Marti, R., 2002. The OptQuest callable library. In *Optimization software class libraries* 193-218. Springer US.
- Lee, H.L., Padmanabhan, V., Whang, S., 1997. The Bullwhip Effect in Supply Chains. *Sloan Management Review* 38 (3), 93-102.
- 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> [accessed 9 July 2014]
- Parejo J. A., Ruiz-Cortés A., Lozano, S., Fernandez, P., 2012. Metaheuristic optimization frameworks: a

survey and benchmarking. *Soft Computing*, 16(3), 527-561.

Scarf, H., 1993. The Optimality of (S, s) Policies in the Dynamic Inventory Problem. *Optimal pricing, inflation, and the cost of price adjustment*, 49-56.

Wagner, S., Kronberger, G., Beham, A., Kommenda, M., Scheibenpflug, A., Pitzer, E., Vonolfen, S., Kofler, M., Winkler, S., Dorfer, V., Affenzeller, M., 2014. Architecture and Design of the HeuristicLab Optimization Environment. In *Advanced Methods and Applications in Computational Intelligence, Topics in Intelligent Engineering and Informatics Series*, 197-261. Springer

Wetter, M., 2001. GenOpt - A generic optimization program. In *Proceedings of the 7th IBPSA Conference*, 601-608, August, Rio de Janeiro, Brazil.

#### **AUTHORS BIOGRAPHY**

**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.

**GABRIEL KRONBERGER** received his PhD in engineering sciences in 2010 from JKU Linz, Austria, and is a professor at the campus Hagenberg. His research interests include genetic programming, machine learning, and data mining. He is a member of the HeuristicLab architects team.

**JOHANNES KARDER** is pursuing his master study of software engineering at the University of Applied Sciences Upper Austria and is a research associate at the Research Center Hagenberg. He is a member of the HeuristicLab development team.

**MICHAEL KOMMENDA** received his Master in bioinformatics in 2007 from the University of Applied Sciences Upper Austria, and is a research associate at the Research Center Hagenberg. His research interests include genetic programming and data mining. He is a member of the HeuristicLab architects team.

**ANDREAS SCHEIBENPFLUG** received his Master in software engineering in 2011 from the University of Applied Sciences Upper Austria, and is a research associate at the Research Center Hagenberg. His research interests include parallel and distributed computing. He is a member of the HeuristicLab architects team.

**STEFAN WAGNER** received his PhD in technical sciences in 2009 from the Johannes Kepler University Linz, Austria. He is a professor at the University of Applied Sciences Upper Austria, 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.

# PRODUCTION LINE MODELING AND BALANCING: COMPARISON OF EXISTING TECHNIQUES AND PROPOSAL OF A NEW METHODOLOGY

Alessandro Silvestri <sup>(a)</sup>, Cristina Cerbaso <sup>(b)</sup>, Gianpaolo Di Bona <sup>(c)</sup>, Antonio Forcina <sup>(d)</sup>, Vincenzo Duraccio <sup>(e)</sup>

(a) (b) (c) (d) Department of Civil and Mechanical Engineering  
University of Cassino and Southern Lazio  
(e) Niccolò Cusano University  
00166 Roma (RM) – Italy  
tel. 0039-776-2994330 - Fax 0039-776-310812

<sup>(a)</sup>[silvestr@unicas.it](mailto:silvestr@unicas.it), <sup>(b)</sup>[c.cerbaso@unicas.it](mailto:c.cerbaso@unicas.it), <sup>(c)</sup>[dibona@unicas.it](mailto:dibona@unicas.it), <sup>(d)</sup>[a.forcina@unicas.it](mailto:a.forcina@unicas.it),  
<sup>(e)</sup>[vincenzo.duraccio@unicusano.it](mailto:vincenzo.duraccio@unicusano.it)

## ABSTRACT

In industrial plants, the heads of production often allocate tasks to operators without following a real methodology, but basing on experience; therefore, there are many problems when a quickly changes of capacity or constraints occur.

In the literature, there are not techniques able to provide unique and effective solutions. The starting point of this study is to explore the main methods of production line balancing, understanding their advantages and disadvantages in assigning tasks to workers with the same amount of time. Well-balanced lines allow to improve productivity and to make production lines enough flexible to absorb external and internal irregularities. The result is a better utilization of workstations, machines and people.

The main aim of the present work is to propose a new technique of balancing, in particular for assembly lines, simple, general and effective.

Keywords: operator, task, saturation, algorithm

## 1. INTRODUCTION

In a flow shop production, it is important to ensure a balanced allocation of the operations among workstations: no overloaded workstations (bottlenecks) rather than shortly saturated. The balancing activity is essential to guarantee the performance of the whole line (in particular, the productivity), whose speed is due to the slowest station, at least in the case of “paced” lines (production rhythm imposed), fixed or continuous (respectively we have the product replaced or in transit in a station for the execution of all tasks assigned to each workstation, during the given time, called station or cycle time).

The criteria for the assignment of tasks to each workstation, are mainly based on efficiency and productivity objectives, fragmentation of operations and specialization of tasks. Therefore, the characteristics of machineries and equipment and the employees required skills are highly specific, in order to realize a short number of actions.

In many case the process modeling through simulation techniques can be realized in order to achieve the production line optimization and improvement (Falcone *et al.*, 2005; Falcone *et al.*, 2010; Falcone *et al.*, 2013).

## 2. WORKLOADS BALANCING ACTIVITY

The balancing activity of a production line means, in general, to allocate the operations among the stations in such a way that the cycle time for each stage (sum of the times of the operations allocated to the station) is as uniform as possible (Aase *et al.*, 2003). In practice, however, the perfect balancing is almost never reached due to constraints such as:

- cycle time;
- space;
- allocation of operations;
- qualification of operators;
- incompatibility between operations;
- management of material flows;
- precedence relationships;
- use of the operators.

The workloads balancing is a very important issue, as it influences the productivity of a line. It is necessary both in the case of designing a new line and in the case of improving the performance of an existing one.

The first step consists in identifying the so-called "bottlenecks", defined as the phases of the process where there is the lowest production capacity (De Carlo *et al.*, 2013; Duraccio *et al.*, 2006). They are due to the accumulation of material to be processed in a single location, which causes a slowdown of the production time of the whole line, and, consequently, a production capacity reduction. The reasons could be the slowness of operators and machines, excessive paths length, non-optimal position of equipment and material, that cause a loss of time due to activities of walking and picking (Falcone *et al.*, 2011; Iannone *et al.*, 2007).

### 3. BALANCING TECHNIQUES

Although many studies have been carried out over the years to try to get solutions to the problem, simple, effective and unambiguous, there is not yet an analytical method able to consider all the different aspects of the problem.

Those difficulties are mainly due to the different structural configurations of the modern industrial plants, which doesn't permit to study a generic approach and solution. Nowadays, also the wide range of realized products contributes to complicate the balancing activity.

Therefore, a large number of models and solutions allowing to solve specific problems of balancing, or, however, a little number of problems, according to their particular application, have been proposed.

However, if there are not suitable models to be applied, customized models are developed, as needed, by the managers who are experts in the production line.

Below the most common methods existing in the literature will be analyzed and a new technique will be proposed, in order to maximize the saturation of the operators and the utilization of the workstations. The new methodology integrates and improves existing methodologies, allowing to obtain a unique configuration of an assembly line (Askin and Zhou, 1997).

#### 3.1. Salveson's Criterion

Sometimes, there are simple problems of balancing production lines. In particular, it happens when there are not typical basic constraints described above, commonly present in practice, especially precedence constraints between operations.

In this specific case, the optimum solution is achieved simply by minimizing the number of stations of the line, choosing the best combinations of operations able to reduce the operator/machine cycle time and the number of utilized stations.

In this case, *Salveson's criterion* is helpful. The iterated procedure allows finding the smallest number of workstations that meets the production requirements.

#### 3.2. Kottas - Lau Method

The *Kottas - Lau* algorithm is a very valid criterion for the industrial reality. This methodology allows to obtain a good balancing degree of the line and seeks to optimize the total cost of production. The algorithm considers the operations execution times as stochastic variables and describes them through a normal distribution with mean value and standard deviation assigned.

The categories of considered costs are:

- labor costs: decrease increasing the amount of work assigned to each operator, thanks to the reduction of the number of workers needed to complete the processing cycle;
- costs of non-completion: if many tasks are assigned to the operator, the likelihood he does

not complete his work and, consequently, the related costs increase.

The correct balancing of a production line permits to reach the best compromise between the two types of costs listed above, such as to minimize the total cost of production (Figure 1). The trends of costs are shown in Figure 2.

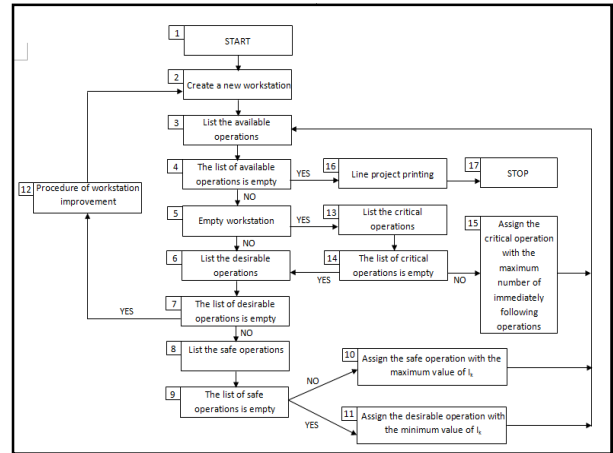


Figure 1: Block diagram of *Kottas - Lau* algorithm -  $I_k$  is the completion out of line total cost of all the operations which, due to precedence constraints, can not be performed if the k-th operation is not completed.

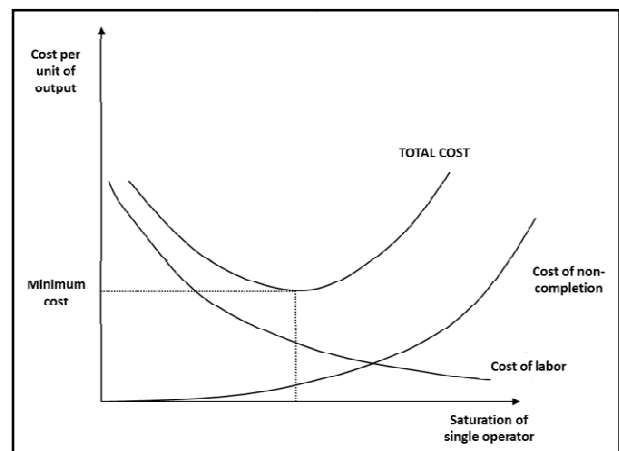


Figure 2: Costs Trends

#### 3.3. LCR and KWM

The acronyms LCR and KWM, or *Largest Candidate Rule* and *Kilbridge & Wester Method*, identify two methods often used in industrial practice to balance production lines, solving relatively simple problems thanks to their simplicity of application. In particular, they use a single parameter, the operations execution time, and the diagram of precedence to assign tasks to stations. Even if the balancing activity is based on only one key factor, the precedence constraints between operations and the maximum cycle time of the station have always to be taken into account.

### 3.3.1. Largest Candidate Rule (LCR)

In the *LCR* method, the iterative process used to search for the balancing solution is only based on the task time, that is the execution time of the considered operations. The procedure is described below and is particularly quick and easy to be applied:

1. build a table containing the tasks, their execution time and their immediate predecessors;
2. order the operations according to the execution time, placing those with longer duration at the top;
3. assign tasks in the obtained order and in accordance with the priorities and the maximum cycle time the station supports;
4. repeat the previous step until the available operations are finished.

### 3.3.2. Kilbridge&Wester Method (KWM)

Heuristic methods can be used to solve the balancing problem (Scholl *et al.*,2006). The *KWM* method is a heuristic algorithm that selects the operations to be assigned to stations, simply according to their position in the diagram of priorities. This method solves a problem of the *LCR* procedure, where the operations at the end of the diagram of priorities may be the first to be assigned just because the value of their execution time is greater.

In the *KWM* method, the operations are positioned in the columns, according to the diagram shown in Figure 3, and, subsequently, are organized and ordered in a table (Table 1) according to their column belonging.

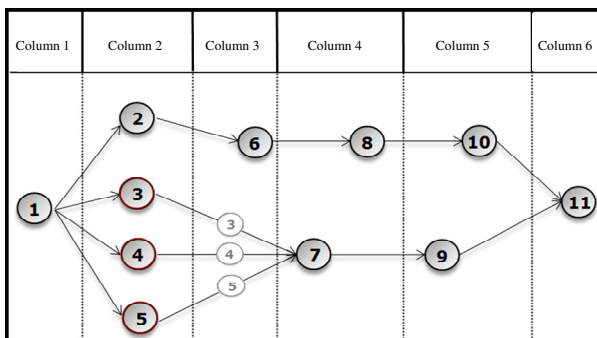


Figure 3: Diagram of priorities

Table 1: KWM operations table

| Operation | Column  | Task Time (min) | Total Task Time of Column (min) |
|-----------|---------|-----------------|---------------------------------|
| 1         | I       | 6               | 6                               |
| 2         | II      | 2               |                                 |
| 3         | II, III | 4               |                                 |
| 4         | II, III | 9               |                                 |
| 5         | II, III | 2               | 17                              |
| 6         | III     | 2               | 2                               |
| 7         | IV      | 3               |                                 |
| 8         | IV      | 6               | 9                               |
| 9         | V       | 5               |                                 |
| 10        | V       | 5               | 10                              |
| 11        | VI      | 3               | 3                               |

If an operation can be included in more than one column, all the columns where the operation appears, must be specified in order to show the portability of that specific operation. Once the table has been completed, the iterative procedure for the tasks assignment is performed, starting with those belonging to the first column until it reaches the maximum cycle time.

If the operations are not assigned in a column due to exceeding the maximum allowable cycle time, it will need to open a new station. Therefore, it cannot be assigned operations belonging to a next column if there is still a previous one containing unassigned operations.

### 3.4. Ranked Positional Weight Method (RPWM)

The *Ranked Positional Weight Method* is a heuristic technique for lines balancing, introduced and developed by Helgeson and Birnie.

As the name implies, this method is based on the so-called "positional weights", consisting of coefficients assigned to all the operations to enable their assignment to the line stations.

The positional weight of an operation is the sum of its task time and all other operations directly following in the diagram of priorities. For example, referring to Figure 4 which shows a simple graph of priorities, the positional weight of operation 2 is given by the sum of its execution time and those of operations 4 and 5 that follow number 2.

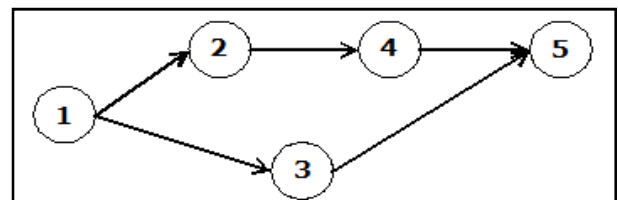


Figure 4: Example of diagram of priorities

Subsequently, the operations are included in a list and ordered according to the decreasing positional weight.

The method assigns to open stations at first the activities with a greater positional weight, always

respecting the maximum cycle time that station can support. If it exceeds that time cycle, a new station will be opened.

According to the heuristic *RPWM*, the greater is the number of operations available for the assignment to the various stations, the greater will be the probability that at least one of them is suitable for a particular workstation. Following this logic, it's possible to exploit the most of the time cycle of each workstation, finding appropriate operations in terms of execution time. This implies a smaller number of stations to be opened and a minimization of associated downtime.

### 3.5. Heuristics COMSOAL

The *Computer Method of Sequencing Operations for Assembly Lines*, alias *COMSOAL*, is another heuristic method able to find possible solutions, but not necessarily optimal, also to balance lines with a great number of operations. The *COMSOAL* method uses a search strategy of assignable operations which allows to obtain a large number of admissible solutions. Randomly and not according to a predetermined criterion it selects and assigns the available tasks. This approach is useful when the line optimization requires solutions which are not necessarily limited only to obtain the minimum number of stations (Becker *et al.*, 2006; Boysen *et al.*, 2008).

### 3.6. FABLE algorithm

The *FABLE* algorithm is structurally similar to *COMSOAL*. Its application involves the construction of a "tree" of possible solutions that also includes the optimal one.

To expedite an acceptable solution, it builds a tree branch at a time. Such construction is called "laser search". In addition, the *FABLE* algorithm uses a technique called "backtracking", usually used to look for solutions in different areas, where it is necessary to comply with specific constraints.

This technique, applied in tree structures such as the *FABLE* algorithm ones, develops the branches of the solutions, keeping track of all the nodes and branches previously visited, so as to be able to go back if the process in the current branch does not lead to valid results.

### 3.7. Elmaghraby criterion

A further criterion, developed by Elmaghraby, is very similar to the *Ranked Positional Weight method*, about logical characteristics and operational.

The heuristic *Elmaghraby* is particularly well suited to solve the balancing problems of simple lines. The development of the algorithm is mainly based on two key points:

- construction of the matrix of priorities;
- assignment of operation position coefficients.

The matrix of precedence "P" consists of the elements  $p(h, k)$ , which, in case of precedence

constraint between the operations "h" and "k", take the value  $p(h, k) = 1$ , whereas in case of absence of sequential constraints, take the value  $p(h, k) = 0$ .

The position coefficient, however, is equivalent to the positional weight of *RPWT* criterion, given by the sum of the execution time of the operation  $i$  ( $i = 1...n$ ) and the time of all following operations bound to it.

### 3.8. Imposed Operators Maximum Degree of Saturation Method

The method called *Imposed Operators Maximum Degree of Saturation Method* considers as binding, the relationships of precedence between the operations and considers also a coefficient, indicating the maximum degree of saturation that a station can support. It defines the *degree of saturation* of the operator (*DS*):

$$DS = \frac{\sum_{i=1}^S T_i}{TC} \quad (1)$$

where:

- $T_i$  is the execution time of the operation "i";
- S is the set of operations assigned to the operator;
- TC is the cycle time.

This methodology is based on the respect for the inequality:

$$DS \leq \alpha \quad (2)$$

and

$$0 < \alpha \leq 1 \quad (3)$$

taking into account the precedence constraints between operations, where  $\alpha$  is the imposed degree of saturation. The logic used in the algorithm is summarized in the block diagram shown in Figure 5.

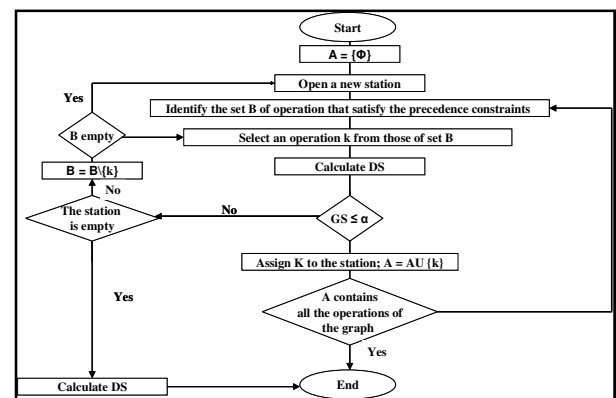


Figure 5: Imposed Operators Maximum Degree of Saturation Method - Block diagram



In the Block diagram, the set A contains the operations already assigned, while the set B contains the operations to be assigned.

#### 4. PROPOSAL OF A NEW METHODOLOGY

The proposed innovative methodology, named *Maximum Degree of Saturation Method (MDS)*, has as its starting point the *Imposed Operators Maximum Degree of Saturation Method* and aims to saturate workstations as much as possible, minimizing a variable that keeps track of the time available to be committed, in accordance with the precedence constraints and other restrictions.

In particular, this methodology could be used for production lines processing a single type or different types of products. An important goal is the reduction of the desaturation percentage and, consequently, the number of stations.

Also in this case, such as for *Imposed Operators Maximum Degree of Saturation Method*, the *Degree of Saturation DS* is defined as:

$$DS = \frac{\sum_{i=1}^S T_i}{TC} \quad (1)$$

*Maximum Degree of Saturation Method (MDS)* finds applications in assembly lines for which the movement of an operator or a robot from a workstation to another is a constraint (Agnesis *et al.*, 2003).

In this case, the maximum available time to complete all transactions within each location corresponds exactly to the “takt time”, dictated by the market demand.

Figure 6 shows the block diagram of the proposed algorithm operation.

The application of the *Maximum Degree of Saturation Method* is quite simple.

First, it starts by opening a new station, being empty the set A of already assigned operations. Then, it identifies the set B of the possible operations sequences, respecting the precedence constraints. Selecting these combinations, it must bear in mind that the available operations are those for which the previous activities have already been assigned to a workstation or have already been taken into account in a possible location.

At this point, it calculates the *Degree of Saturation (DS)* for each identified combination and it compares with the *Imposed Degree of Saturation  $\alpha$* . The choice of the value  $\alpha$  (between 0 and 1) must take into account the constraints and it is also necessary to remember that a high degree of saturation reduces the cost of the line, but increases the cost of non-completion (in this sense it is possible to integrate also the considerations arising from the *Kottas-Lau Method*).

All combinations of operations not satisfying the relation:

$$DS \leq \alpha \quad (2)$$

will be excluded, while the others will be taken into account.

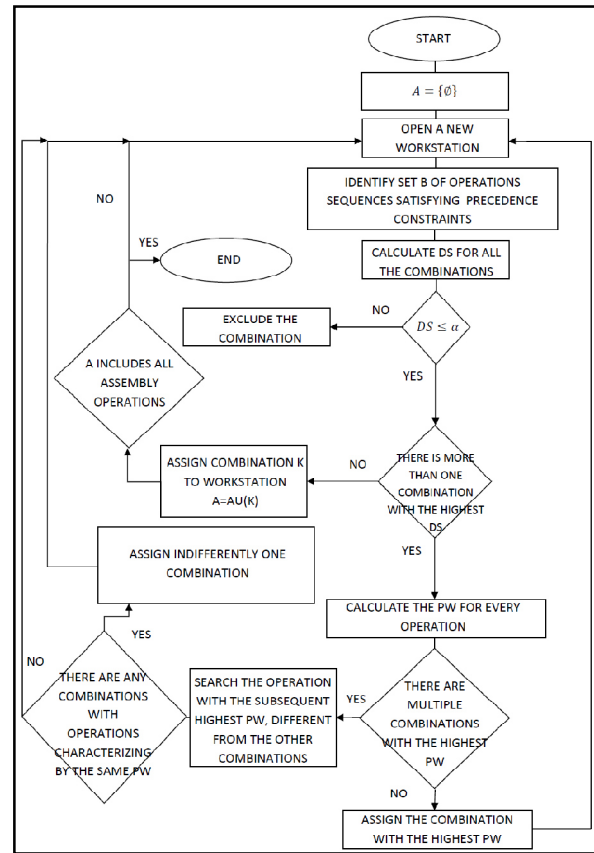


Figure 6: Block diagram of functioning of Maximum Degree of Saturation Method

After selecting combinations of operations satisfying both precedence constraints that the relation, one wonders about the possibility of having two or more sequences of activities with the highest degree of saturation.

If not, it assigns to the current workstation the combination of operations with the highest DS and, later, if the set A still contains operations, it opens a new workstation, otherwise the balancing operation ends.

If so, it calculates the positional weight of each transaction, as the sum of the times of the single operations following the activity, including the operation in question:

$$PW_i = \sum_{i=1}^n TP_i \quad (4)$$

This approach refers to a technique previously presented, namely the *Ranked Positional Weight Method*, in which the operations are assigned to workstations based on the values of the positional weights: the operation to be assigned is the one that

presents a positional weight greater than the others, always respecting the precedence relations.

To give priority to these operations means to reduce the non-completion impact when the production line stops for various reasons.

In particular, stopping the production flow, the operations (in accordance with the precedence diagram and therefore with the technological and assemblability constraints) with a greater execution time could be completed, while the remaining ones, with lower execution times, would be completed out of line at less cost.

Continuing the explanation of the algorithm logic shown in the block diagram, after calculating the positional weight of each operation, it assigns the sequence containing the task with the highest positional weight and then it opens a new workstation.

In case there were multiple combinations of operations with the same saturation degree (the highest), and operations with the same highest weight positional, it gives precedence to the combination containing the operation with the second highest weight positional; whereupon it opens a new station.

If also the latter is equal between two or more stations, it chooses the sequence of operations with the third highest positional weight and so on.

Then it will open a new workstation and then it will continue with the production line balancing. This approach is reasonable, since it prefers to terminate within the line operation in which are related activities engaging more time to be completed, rather than one to which follow operations less onerous in terms of time. This is explained by noting that in the event of a production line shutdown, if it need to finish the job and it has the possibility to complete operations out of the line, surely the activities with a lower duration will have a lower cost than operations with higher running times.

Note that it may happen that not all combinations of operations to be compared for the highest positional weight contain the same number of activities. For example, it might have a combination with a single operation, and another one with two. In this case, both having the same saturation degree and the same highest positional weight, the analyst should look for the sequence with the operation that has the second highest weight, but it is evident that the sequence composed of a single task would be automatically excluded. If a situation like this occurs, it will be taken into account the combination presenting the activity to which is linked the largest number of operations. This reasoning stems from the fact that non-completion of this operation within the line involves higher costs because it should be carried out of the line all the other activities in addition to it and it might need more labor, equipment, resources, etc.

In case the last open station is forcibly shortly saturated, it might delete an already assigned task from a workstation with a very high degree of saturation (in accordance with the precedence constraints) and assign it to the last station. In this way, it would perform a

workloads balancing to avoid that there are stations with a very high degree of saturation, and a station (the last one that has been opened) with a very low degree of saturation. Obviously, a situation like this is inconceivable in real cases, because it would mean, in the case of manual production line, one of the operators with little work, i.e. with a very high percentage of desaturation.

The latter observation is very important because it must always be remembered that balancing a production line means allocate the various operations in such a way as to balance the workload of each operator, minding the goal to pursue.

There will be cases where it will be evident the saturation lack of the last open station. Other times, however, it is uncertain about what to do and the decision will be taken by the person making the line balancing, based on his own experience and on the production cycle knowledge.

Wanting to make a direct comparison between the examined methods, it can refer to Table 2, which summarizes the main features of each technique, including the developed new one.

Table 2 – Comparison between the different balancing techniques

|                                                              | Precedence relations | Cycle time | Production costs | Takt time | Backtracking Task | Same probability of tasks selection | Probabilistic distribution of task time | Generating multiple solutions |
|--------------------------------------------------------------|----------------------|------------|------------------|-----------|-------------------|-------------------------------------|-----------------------------------------|-------------------------------|
| <b>Kottas Lau</b>                                            | √                    | √          | √                | √         | -                 | -                                   | √                                       | -                             |
| <b>K&amp;W</b>                                               | √                    | √          | -                | √         | -                 | -                                   | -                                       | -                             |
| <b>LCR</b>                                                   | √                    | √          | -                | √         | -                 | -                                   | -                                       | -                             |
| <b>RPW M</b>                                                 | √                    | √          | -                | √         | -                 | -                                   | -                                       | -                             |
| <b>Comsoal</b>                                               | √                    | √          | -                | √         | -                 | √                                   | -                                       | √                             |
| <b>Fable</b>                                                 | √                    | √          | -                | √         | √                 | -                                   | -                                       | √                             |
| <b>Salveson</b>                                              | -                    | √          | -                | √         | -                 | -                                   | -                                       | √                             |
| <b>Elmaghraby</b>                                            | √                    | √          | -                | √         | -                 | -                                   | -                                       | -                             |
| <b>Imposed Operators Maximum Degree of Saturation Method</b> | √                    | √          | -                | √         | -                 | √                                   | -                                       | √                             |

|                              |   |   |   |   |   |   |   |   |
|------------------------------|---|---|---|---|---|---|---|---|
| Maximum Degree of Saturation | √ | √ | √ | √ | - | - | - | - |
|------------------------------|---|---|---|---|---|---|---|---|

The *Maximum Degree of Saturation Method* (MDS) provides as output a single solution, which turns out to be the optimal one, as it aims to minimize the coefficient of balancing delay. Furthermore, for a not excessive number of operations to be placed within the workstations, it may be applied manually. That makes this technique quite practical in the business reality.

As shown from Tab. 2, all of the discussed techniques, including the one developed, present as key points the cycle time and the takt time, ie the cycle time available within each station.

This last parameter is very important as it determines the production rate, set depending on the market demand.

In general, it is very difficult, if not impossible, to state with certainty that a method is better than another in searching the optimal solution for the problem of balancing a production line .

Surely the developed technique combines the easiness of application to the effectiveness of the result.

## 5. CONCLUSIONS

The main features of the *Maximum Degree of Saturation Method* are the following:

- it is easy to be applied, since at each step it only requires the verification of compliance with the precedence constraints and relation (2). Also, the evaluation of operations positional weights is very easy, being a simple sum of the activities task time;
- for a limited number of operations, the technique does not require to be implemented using the computer, since it is even easy to be used manually;
- using deterministic times, it does not consider the randomness of the operations execution times. This might seem like a disadvantage, but many business realities are based on unique values of operations task time, deriving from studies carried out by Times and Methods office. These deterministic times then become imposed times for the operator to perform his activities;
- requires a proper choice of the coefficient  $\alpha$  (which not necessarily must be the same for all the workstations);

The *Maximum Degree of Saturation Method* has considerable advantages, as:

- it determines a unique and optimal solution, by selecting in advance the best combination of operations which saturates the available cycle time for each workstation;

- it reduces the number of stations to be opened and the costs associated with this operation, having regard that the number of stations appears in the numerator of the *coefficient of delay balance* formula, defined as:

$$D = \frac{(N_s \times TC_{\max}) - T_k}{(N_s \times TC_{\max})} \times 100 \quad (5)$$

where

- $N_s$  is the number of stations obtained by applying a specific balancing technique;
- $TC_{\max}$  is the maximum cycle time of the line, among all stations;
- $T_k$  is the total task time of the operations;
- it presents a method for choosing the operations to be assigned to workstations based on multiple levels, giving priority to saturation and then to the activity position within the priorities diagram, integrating multiple techniques in a single balancing procedure (*Kottas – Lau; Ranked Positional Weight Method; Imposed Operators Maximum Degree of Saturation Method*).

A disadvantage in the use of this technique arises when the operations to be placed in the various stations are quite numerous. In this case, the searching of the combinations of operations available to assign is very laborious, since the number of sequences achievable grows with the number of activities.

The next step will be the application of the developed method (MDS) to real case studies in order to validated is goodness and generality.

## REFERENCES

- Aase, G.R., Schniederjans, M.J., Olson, J.R., 2003. U-OPT: An analysis of exact U-shaped line balancing procedures. *International Journal of Production Research*.
- Agnetis, A., Alfieri, A., Nicosia, G. 2003 Part batching and scheduling in a flexible cell to minimize setup costs. 2003. *Journal of Scheduling* 6 (1), pp. 87-108.
- Askin, R.G., Zhou, M., 1997. A parallel station heuristic for the mixed-model production line balancing problem. *International Journal of Production Research*.
- Becker, C., Scholl, A., 2006. A survey on problems and methods in generalized assembly line balancing. *European journal of operational research*.
- Boysen, N., Fliedner, M., Scholl, A., 2008. Assembly line balancing: Which model to use when? *International journal of production economics*.

- De Carlo, F., Arleo, M.A., Borgia, O., Tucci, M., 2013. Layout design for a low capacity manufacturing line: A case study. *International Journal of Engineering Business Management*. Volume 5, Issue SPL.ISSUE
- Falcone, D., Duraccio, V., Silvestri, A., Di Bona, G., 2005. Technical and economical analysis of the layout of a palletization plant through simulation techniques. *Summer Computer Simulation Conference 2005, SCSC 2005, Part of the 2005 Summer Simulation Multiconference, SummerSim 2005*, pp. 387-391.
- Duraccio, V., Falcone, D., Silvestri, A., Di Bona, G., 2006. Improvement of performances of an optical system for defectiveness survey in a company of the automotive field. *Summer Computer Simulation Conference 2006, SCSC'06, Summer Simulation Multiconference, SummerSim'06*, pp. 314-318.
- Falcone, D., Silvestri, A., Di Bona, G., Forcina, A., Pacitto, A., 2010. Study and modelling of very flexible lines through simulation. (2010) *Proceedings of the 2010 Spring Simulation Multiconference - Emerging M and S Applications in Industry and Academia Symposium, EAIA*, pp. 11-16.
- Falcone, D., Silvestri, A., Forcina, A., 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 Modelling and Applied Simulation), Rome, 12-14 September 2011*.
- Falcone, D., Silvestri, A., Di Bona, G., Forcina, A., Cerbaso C., Duraccio V., 2013. Technical and economic verification of the convenience in reengineering a production line using simulation techniques. *25th European Modeling and Simulation Symposium, EMSS 2013*, pp. 327-332, September 25-27, Athens, Greece. Code 100159.
- Iannone, R., Miranda, S., Riemma, S., 2007. Supply chain distributed simulation: An efficient architecture for multi-model synchronization. *Simulation Modelling Practice and Theory*, pp. 221-236, Volume 15, Issue 3.
- Scholl, A., Becker, C., 2006. State-of-the-art exact and Heuristic solution procedures for simple assembly line balancing. *European Journal of Operations Research*.

# MAINTENANCE CRITICAL ANALYSIS AND PRIORITY INDEX: A NEW MODEL FOR MAINTENANCE POLICY

Alessandro Silvestri <sup>(a)</sup>, Cristina Cerbaso <sup>(b)</sup>, Domenico Falcone <sup>(c)</sup>, Antonio Forcina <sup>(d)</sup>, Vincenzo Duraccio <sup>(e)</sup>

(a) (b) (c) (d) Department of Civil and Mechanical Engineering  
University of Cassino and Southern Lazio  
(e) Niccolò Cusano University  
00166 Roma (RM) – Italy  
tel. 0039-776-2994330 - Fax 0039-776-310812

<sup>(a)</sup> [silvestr@unicas.it](mailto:silvestr@unicas.it), <sup>(b)</sup> [c.cerbaso@unicas.it](mailto:c.cerbaso@unicas.it), <sup>(c)</sup> [falcone@unicas.it](mailto:falcone@unicas.it), <sup>(d)</sup> [a.forcina@unicas.it](mailto:a.forcina@unicas.it),  
<sup>(e)</sup> [vincenzo.duraccio@unicusano.it](mailto:vincenzo.duraccio@unicusano.it).

## ABSTRACT

Industrial processes are realized according to production levels established in the design phase.

Proper maintenance of machines involved in the process guarantees the compliance of these levels over time, ensuring adequate availability standards of plant and machinery. Conversely, inadequate maintenance can affect a decrease of production levels. In this case, the company may not respect the demands of customers, resulting in economic losses, which may affect its survival.

In this paper, a new procedure is proposed, in order to develop a useful tool to support the design of maintenance.

The proposed procedure is based on the identification of the main factors influencing availability and maintenance of industrial plants, then summarized through appropriate indices.

Indices analysis allows the identification of the guidelines, to optimize the maintenance of the units.

This tool allows an appropriate allocation of budgets to maintenance activities, right for machines or units that are strategic to ensure production.

Keywords: design of maintenance, budgets allocation

## 1. INTRODUCTION

The maintenance plan of a plant is primarily designed to ensure a high level of availability, so as to complete a given production.

The overall availability of an industrial plant depends itself from the wards or by machines that constitute it, and how the machines are linked together. From this point of view, the plant is similar to a complex system, whose target of reliability results from those of its components. In fact, it is possible to establish a plant target value of availability, according to the elements composing it.

In this case the target depends on the quantities set out in the master production schedule and the service

level established in supply contracts (Falcone *et al.*, 2011).

Units or machines have different criticalities depending on many factors that identify their importance in the production process.

The proposed methodology is the result of a careful bibliographical study about the main research areas in the field of maintenance.

The development of a maintenance program must be based on information collected by monitoring the condition of a system or a production process, which can be classified into two categories: direct information, and indirect information.

In direct information, the parameter measures the fault of the process or wear condition (for example, the thickness of brake pads). On the other hand, indirect information give indications on the dynamics of failure, but they are not a direct measure. Example of indirect information can be an analysis of the characteristics of refrigerant oil (Christer and Wang, 1995), (Raheja *et al.*, 2006).

The works of applied research can be divided into three groups as follows.

The first group of research is only focused on the determination of the optimal inspection times or intervals (Chen and Trivedi, 2002), (Wang, 2003), (Falcone *et al.*, 2004), (Kalleh and Van Noortwijk, 2006), (Wang and Jia, 2007).

A common assumption is that the used information are direct; for this reason, the wear condition of the system can be identified by monitoring it and then, after an inspection, it is possible to provide appropriate maintenance action. The aim is to determine the best time of inspection, depending on the optimization of a parameter such as the maximum availability or the minimum cost per unit of time.

The second group of research concerns the dynamic determination of the inspection time and the time of maintenance or replacement (Castanier *et al.*, 2003), (Chen and Trivedi, 2005), (Ghasemi *et al.*, 2007), (Wang *et al.* 2009). Future inspections and the

best maintenance actions are scheduled in this case, according to several performance criteria, such as the "long run system availability" and "long-run expected maintenance cost". For this purpose, information on the wear level of the system result from non-periodic inspections.

The third group is focused on the determination of the optimal level maintenance or replacement (Banjevic *et al.*, 2001), (Chen and Wu, 2007), (Lu *et al.*, 2007). A common assumption is that the information obtained from monitoring the conditions are indirect. The inspections are carried out periodically and aperiodically but with a predetermined schedule. The aim is to determine a level that optimizes a given performance criterion.

Makis and Jardine (1992), propose the definition of the limit level of a system, for a random fault. In their approach, the checks are carried out at fixed interval, determined for each. The equipment is replaced each time it fails. In addition, after each inspection and according to the results of the inspection, if the fault cost reaches or exceeds a predetermined limit, a preventive replacement is provided.

The main techniques in literature are based on the assumption, (implicit but real) you have an unlimited budget and always sufficient to ensure the operations that each methodology identifies.

This hypothesis is often in conflict with the increasing difficulty that companies have to dispose of adequate capital for the ideal maintenance of their plants.

The above considerations suggest the development of a method, in order to quantify units criticalities, identifying the most relevant to ensure the production. The basic hypothesis is that a greater strategic importance of the unit or machine, must signify greater efforts in terms of maintenance activity.

For this purpose we introduce some indices, to take into account factors that can quantify the importance of each unit in a manufacturing process.

## 2. FACTORS DETERMINING THE AVAILABILITY OF MACHINES

For the application of the proposed methodology, factors considered critical to the maintenance of the machines were analyzed.

### 2.1. Level of Use

Not all the units have the same relevance, in a plant. This essentially depends on how one unit or machine is involved in the production process. It is clear that units more "stressed" by the process have a greater influence, showing for this reason, the highest levels of criticality.

In terms of availability, higher performance is required to most involved units. The unavailability of a highly used machine, in fact, risks to compromise the whole process. The level of use depends on:

- working time of machine or unit, compared to running time of the plant;

- flow of material crossing the unit, compared to the overall flow of the plant.

### 2.2. Maintainability

The operational availability of a working system, when uptime and downtime are known, can be precisely calculated as the ratio between the real working time and the theoretical one and depends on the mean values of the distributions of faults and repairs in time. Therefore maintainability greatly influences the availability of the plant (Falcone *et al.*, 2014).

Maintainability characterizes the simplicity of maintenance activities. For the same number of faults and theoretical working time, the greater will be the time required to restore the system in working condition, the lower will be the availability.

These concepts are also analytically reflected, as is well known, in the operational availability  $A(t)$ :

$$A(t) = \frac{MUT}{MUT+MDT} \quad (1)$$

where the Mean Up Time (MUT) and the Mean Down Time (MDT) are average values in a time range, typically one year.

A poor maintainability of the system corresponds to high values of the MDT and consequently a reduction of the system availability, for equal MUT.

The maintenance times (MDT) are therefore an indirect measure of the availability.

Units or machines with a high mean down time, are the critical issues in terms of overall availability of the system, with consequences also in terms of safety (Falcone *et al.*, 2007), (Di Bona *et al.*, 2014).

### 2.3. Working time

Another way to evaluate the importance of the unit in the process, is to estimate how much time it is used for producing a single unit of product, compared to its takt time.

More work times in fact, generally correspond to complex and relevant operations. In case of multi-product companies, the assessment must be repeated for all products which the unit is involved for.

### 2.4. Maintenance cost

In the economic sphere, maintenance is similar to a particular type of investment, depending on production volumes and on the importance of the machine in the production cycle.

The costs related to maintenance activities are numerous and different. The main types of cost are:

- Cost of spare parts and equipment;
- Cost of internal labor;
- Cost of external labor;
- Non-profit due to break of production;
- Costs of image.



Investment in maintenance, which are often very large in modern in company, must be weighed and focus on those departments whose unavailability can compromise the whole production process.

### 2.5. Variety of faults

To make more complex the maintenance operations, influencing the availability of a beam, is not only the quantity (measured in terms of time) of failure events that occurs, but also their variety. Different faults are generally synonym of a high complexity in maintenance operations, as well as a structural complexity of unit or machine.

### 3. STRUCTURE OF INDICES

The influencing factors and critical issues, analyzed so far, have been formalized through appropriate technical and economic indicators, used to measure the criticality of units in terms of maintenance and consequently production.

The mathematical structure of the indices is as follows.

#### 3.1. Flow index

Flow Index ( $FL_i$ ) for the machine  $i$ -th is equal to the ratio between the flow processed by the unit ( $F_i$ ) and the total materials flow of the plant ( $F_{tot}$ ).

$$Fl_i = \frac{F_i}{F_{tot}} \quad (2)$$

This technical indicator is intended as a measure of the importance that the unit or the machine covers in the production process.

It always takes positive values between 0 and 1; in particular take values close to one, in case of indispensable units for production.

The flow index takes into account, even if indirectly, the production process and the functional connections between the machines.

For companies with a single product line layout, the flow rate will be the same for each machine with value always equal to 1 (unless by-pass along the process).

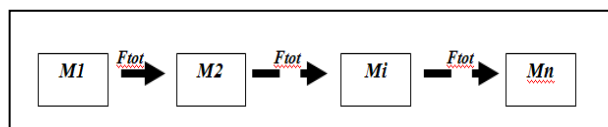


Figure 1: Line layout flow

In case of more complex layout configurations, with more production lines or machines in parallel, the flow index can be calculated as in the example shown in Figure 2.

Being  $F_i$  the flow of materials in the machine  $i$ -th, and assuming that the takt times are equal, is possible to assume values of flow as follows:

- $F_1 = F_7 = F_{tot}$ ;

- $F_2 + F_3 = F_1$
- $F_3 = 2F_2$
- $F_4 = F_5$
- $F_2 = F_6 = F_4 + F_5$

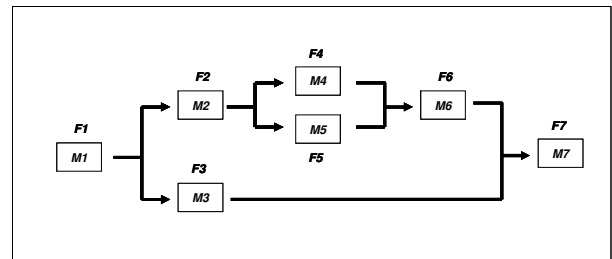


Figure 2: Example of complex layout

It will result for example, the following values of the flow index:

- $Fl_1 = 1$
- $Fl_3 = 2/3$
- $Fl_4 = 1/6$

In calculating the index, it is necessary to measure the flow of materials in a way suited to the specific production process (number of pieces, weight, capacity, MAG).

#### 3.2. Time index

Time Index ( $T_i$ ) for the unit or machine  $i$ -th is the ratio between the working time of the machine for a unit of product and the takt time of the product.

In case of multi-product company, where the same machine is used for the production of different products, the index formula is different.

$P_1, \dots, P_n$  are the products made by the company. If we denote  $TC_1, \dots, TC_n$  as the takt times for each product and  $T_{ij}$  as the time in which the  $i$ -th machine works the product  $j$ -th; the machine time index will be:

$$T_i = \frac{\sum_{j=1}^n T_{ij}}{\sum_{j=1}^n TC_j} \quad (3)$$

Because of its formulation, time index value is between 0 and 1.

#### 3.3. Maintenance Index

Maintenance Index ( $M_i$ ) is given by the ratio between the average number of hours spent in maintenance in the  $i$ -th station ( $TM_i$ ) and the average number of hours of maintenance time, calculated for the unit with the highest maintenance time ( $TM_{max}$ ).

$$M_i = \frac{TM_i}{TM_{max}} \quad (4)$$

The introduced index can take only positive values up to 1, in the case of a unit with the highest maintenance time.

### 3.4. Cost Index

Cost Index ( $C_i$ ) is the ratio between the annual cost of maintenance for the  $i$ -th unit ( $CM_i$ ) and total cost of maintenance of the system ( $CM_{tot}$ ) in a year.

$$C_i = \frac{CM_i}{CM_{tot}} \quad (5)$$

Differently from the previous, it is a purely economic indicator, to estimate the importance of the unit in terms of cost in maintenance.

Also in this case the index takes values greater than 0 and can be 1 if the entire cost of maintenance of the company is used for the analyzed unit. The index only takes into account the costs for industrial maintenance.

### 3.5. Failure Index

Failure Index ( $Fa_i$ ) is the ratio between the failure modes occurring in the  $i$ -th machine ( $Fm_i$ ) and all the different failure modes ( $Fm_{tot}$ ) of the plant (in a year).

$$Fa_i = \frac{Fm_i}{Fm_{tot}} \quad (6)$$

The term "failure mode" refers to the diversity of faults and not to their number. In fact, if the same fault occurs again, in the index it should be counted as one.

## 4. MCA (MAINTENANCE CRITICAL ANALYSIS)

The proposed indices, are the base in a method of analysis, able to constitute a real tool for maintenance management of an industrial plant.

The introduced methodology, called **Maintenance Critical Analysis (MCA)**, aims to identify units or machines that have the most critical maintenance and require major investments. The MCA method can be summarized in some stages.

### 4.1. Plant breakdown

It is necessary preliminary to define what are the units or machines that break down the system to be analyzed. At this stage of breaking down, some basic rules must be observed:

- The number of components has to be such as to adequately describe the specifics of the system, however, does not create problems of data management, due to a too detailed breakdown;
- For chosen parts, the parameters used for the analysis must be uniquely and easily identified.

### 4.2. Data collection

For the selected units, the data needed to calculate the indicators for MCA, should be collected. This operation may require, in some cases, an extended period of time.

### 4.3. Indices calculation

Once the necessary data are been collected, the indices for the analyzed units are calculated.

### 4.4. Maintenance Priority Index (M.P.I.) calculation

The calculated parameters are summarized in a single index, called the Maintenance Priority Index (MPI) to be calculated for each machine.

The mathematical formulation of MPI is as follows:

$$MPI_i = Fl_i \times T_i \times M_i \times Fa_i \times C_i \quad (7)$$

The introduced index can take values between 0 and 1, directly proportional to the criticality related to the unit or the machine.

### 4.5. Analysis of results

By ordering the units according to a decreasing value of MPI, you can get the list of priority maintenance priorities.

According to these priorities, maintenance must be designed and budgets must be allocated.

Units with higher values of MPI are the ones who more resources should be used in terms of maintenance activities. Their unavailability may in fact seriously undermine the success of the production process.

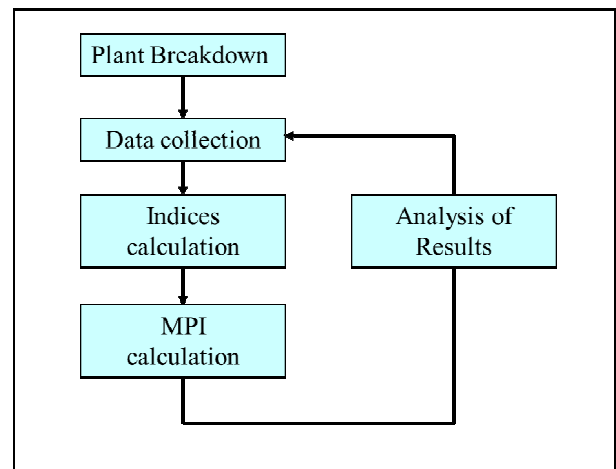


Figure 3: MCA flowchart

### 4.6. Design of the maintenance plan

Once defined which production units are prioritized in terms of maintenance and the overall budget to be allocated, the obtained results can be used to design the maintenance plan.

For each unit, all maintenance operations necessary for its survival must be defined; these operations constitute the minimum level of maintenance. Calculating the corresponding cost for each unit you get the minimum cost of maintenance for the overall plant.

In order to apply the methodology in a profitable way, the budget that the company has set aside for maintenance, has to be greater than the minimum cost previously defined. In this case, by subtracting the

second to the first, it gets the remaining budget that can be allocated, according to the MPI values.

#### 4.7. Results Checking

The nature of the proposed methodology is iterative, since the analysis must be periodically repeated with two different purposes.

The first one is to check the results. The MPI values of those units, which the remaining budget was allocated to, will be reduced in comparison to those previously calculated.

The second reason is to monitor any changes in terms of the importance of the departments, due to possible changes in production.

A further verification of the results, can be done by comparing the break down times of the system before and after the application of the method, expecting a decrease.

### 5. CONCLUSIONS

After explaining the phases of the proposed method, it is appropriate to highlight some important features.

It should first be pointed out that since this is a tool for preliminary analysis, it favors simplicity, both in the mathematical treatment, as in the application. This allows a profitable use, without excessive use of resources.

The generality of the indicators determines the MCA applicability to any type of industrial plant, regardless of its peculiarities.

By repeating the maintenance analysis on selected units after the first application, the method can be also used as comparison and to check the results.

Finally, it is important to note the similarities between MCA and FMECA. The two methods have a very similar logic, although using different indicators.

While FMECA is focused on the analysis of failure modes, MCA takes into account a larger number of factors and is aimed to the analysis of units and machines in general. It should also be highlighted that the introduced tool, can be used as a preliminary stage of more specific analysis on the machines.

Currently the method is tested in two different production systems. The results of the testing phase will be processed in order to identify possible criticalities or improvements.

### REFERENCES

- Banjevic, D., et al., 2001. A control-limit policy and software for condition-based maintenance optimisation. *INFOR*, 39 (1), 32–49.
- Castanier, B., Berenguer, C., and Grall, A., 2003. A sequential condition-based repair/replacement policy with non-periodic inspections for a system subject to continuous wear. *Applied Stochastic Models in Business and Industry*, 19 (4), 327–347.
- Chen, D. and Trivedi, K.S., 2002. Closed-form analytical results for condition-based maintenance. *Reliability Engineering and System Safety*, 76 (1), 43–51.
- Chen, D. and Trivedi, K.S., 2005. Optimization for condition-based maintenance with semi-Markov decision process. *Reliability Engineering and System Safety*, 90 (1), 25–29.
- Chen, A. and Wu, G.S., 2007. Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration. *International Journal of Production Research*, 45 (15), 3351–3379.
- Christer, A.H. and Wang, W., 1995. A simple condition monitoring model for a direct monitoring process. *European Journal of Operational Research*, 82 (2), 258–269.
- Di Bona, G., Duraccio, V., Silvestri, A., Forcina, A., 2014. Validation and application of a safety allocation technique (integrated hazard method) to an aerospace prototype. *Proceedings of the IASTED International Conference on Modelling, Identification and Control*.
- Falcone, D., De Felice, F., Di Bona, G., Silvestri, A., 2004. R.A.M.S. Analysis in a sintering plant by the employment of a new Reliability Allocation Method. *Modelling and Simulation*. Marina del Rey, CA, USA.
- Falcone, D., De Felice, F., Di Bona, G., Duraccio V., Silvestri, A., 2007. Risk assessment in a cogeneration system: Validation of a new safety allocation technique. *Applied Simulation an Modelling*. Palma de Mallorca, Spain.
- Falcone, D., Silvestri, A., Forcina, A., 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 Modelling and Applied Simulation)*. September 12-14, Rome.
- Falcone, D., De Felice, F., Di Bona, G., Duraccio, V., Forcina, A., Silvestri, A., 2014. Validation and application of a reliability allocation technique (advanced integrated factors method) to an industrial system. *Proceedings of the IASTED International Conference on Modelling, Identification and Control*.
- Ghasemi, A., Yacout, S., Ouali, M.S., 2007. Optimal condition based maintenance with imperfect information and the proportional hazards model. *International Journal of Production Research*, 45 (4), 989–1012.

- Kalleh, M.J. and van Noortwijk, J.M., 2006. Optimal periodic inspection of a deterioration process with sequential condition states. *International Journal of Pressure Vessels and Piping*, 83 (4), 249–255.
- Lu, S., Tu, Y.C., Lu, H., 2007. Predictive condition-based maintenance for continuously deteriorating systems. *Quality and Reliability Engineering International*, 23 (1), 71–81.
- Makis, V. and Jardine, A.K.S., 1992. Optimal replacement in the proportional hazards model. *INFOR*, 30 (1), 172–183.
- Raheja, D., et al., 2006. Data fusion/data mining-based architecture for condition-based maintenance. *International Journal of Production Research*, 44 (14), 2869–2887.
- Wang, W., 2003. Modelling condition monitoring Intervals: A hybrid of simulation and analytical approaches. *Journal of the Operational Research Society*, 54 (3), 273–282.
- Wang, L., Chu, J., Mao, W., 2009. A condition-based replacement and spare provisioning policy for deteriorating systems with uncertain deterioration to failure. *European Journal of Operational Research*, 194 (1), 184–205.
- Wang, W. and Jia, X., 2007. An empirical Bayesian based approach to delay time inspection model parameters estimation using both subjective and objective data. *Quality and Reliability Engineering International*, 23 (1), 95–105.

# TOWARDS A METAMODEL FOR AIRPORT MODELING AND SIMULATION

Deniz Çetinkaya <sup>(a)</sup>, İsmet Camcı <sup>(b)</sup>

<sup>(a)</sup> Electrical and Electronics Engineering Department, University of Turkish Aeronautical Association

<sup>(b)</sup> Institute of Science and Technology, University of Turkish Aeronautical Association

<sup>(a)</sup> [dcetinkaya@thk.edu.tr](mailto:dcetinkaya@thk.edu.tr), <sup>(b)</sup> [ismet.camci@stu.thk.edu.tr](mailto:ismet.camci@stu.thk.edu.tr)

## ABSTRACT

Simulation is an effective method to reflect real life situations as well as possible future solutions. Airport modeling and simulation provides great benefits to decision makers to identify potential bottlenecks and operational constraints of an airport via simulating a model of the airport. This work will present a metamodel which will help airport designers to develop airport plans and evaluate alternative options prior to implementation. The metamodel will be used to draw airport design models, and to transform these models into discrete event simulation models automatically. This work will initiate the research on model driven development of airport simulation models.

Keywords: metamodeling, airport modeling, airport simulation, discrete event simulation

## 1. INTRODUCTION

Modeling and simulation is an effective method for analyzing and designing systems and it is of interest to scientists and engineers from all disciplines. Airport modeling and simulation is the activity of developing either the current or the future model of an airport and simulating the model over time in order to identify potential bottlenecks and operational constraints of an airport. Processes in an airport involve many moving objects, and require a good performance that can be measured in several different performance indicators (Verbreack and Valentin 2002). For example, the discrete event simulation of passenger or baggage flows at airport terminals is a very common example in this field (Takakuwa 2003).

Airport modeling and simulation does not only provide benefits to decision makers but also support airport designers, airport engineers, aircraft manufacturers, etc. For example, aircraft manufacturers can utilize simulation to analyze and optimize the relationships between aircrafts and the airport system. Airport design is an important and widely studied research topic. There are various organizations and research groups which are dedicated on this subject. For example, the Airport Design Studio at the University of California, Berkeley provides a diverse group of researchers, students, and faculty with a unique facility dedicated to accommodating research in contemporary

design, planning and management for airport and aviation systems (Airport Design Studio, 2014). A consortium of 8 universities contracted by the Federal Aviation Administration to provide research support for a wide variety of aviation issues is another example. The consortium was awarded the NEXTOR II contract in 2011, and builds upon the success of the original NEXTOR consortium. For example, short courses about airport systems planning and design are conducted for many years.

This work will present a metamodel and a model driven approach which will help airport designers to develop airport models and evaluate alternative options prior to implementation. Analyzing the performance of existing or planned facilities in an airport and assessing the layout of the airport prior to implementation can potentially reduce the implementation costs and increase the chance of successful airport systems.

In the latest global passenger survey conducted by IATA (International Air Transport Association) in June/July 2013, it is reported that %55 of respondents had experienced travel disruptions in the past 12 months (Behan, 2013). Besides, three out of four (%77) experienced severe delays of more than 2 hours. Delays occur with parking, checking in, security screening, passenger identification, limited luggage weight control, safety procedures at boarding, etc. Although passengers acknowledge the need for increased security, delayed boarding, cancelled flights, long waiting times have created an environment of passenger dissatisfaction (Guizzi, Murino and Romano, 2009). Therefore operational efficiency at an airport can have a direct impact on user and customer satisfaction. We believe that modeling and simulation can help to analyze the existing systems to understand the current problems, as well as to evaluate the future airport designs.

The proposed metamodel will be used to draw airport design models, and to transform these models into discrete event simulation models automatically. This work is the initial part of a research project about model driven development of airport simulation models. Following two sections provide more information about airport systems and model driven development respectively. Section 4 presents the metamodel and Section 5 illustrates how the metamodel will be used. Section 6 presents the future work and conclusions.

## 2. AN AIRPORT SYSTEM

An airport system consists of many integrated components to provide facilities and services for air transportation. Each of these components and operations need to be carefully considered during the design of an airport. An airport basically consists of terminals, runways, taxiway, apron, access roads, parking facilities, hangar, fire station and emergency services, etc. For example, Figure 1 shows a simple airport layout with a single terminal. Planning the airport layout, deciding the positions and connections of each component is a very important design problem.

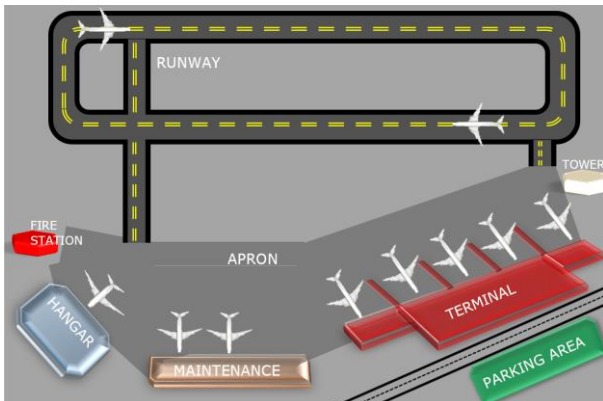


Figure 1: An example airport layout.

For example, the length of runway, the position of parking facilities, the location of the tower, wind analysis for the runway, etc. are very important and the design should conform to the national and international standards such as the airport design guidelines of Federal Aviation Administration of U.S. Department of Transportation (FAA, 2014), or standards by ICAO (International Civil Aviation Organization). Each airport needs to include different means of inland transport (such as road and rail system) and provide various services to passengers (Guizzi, Murino and Romano, 2009). As stated in the previous section, the overall experience of a passenger at an airport can be time consuming. Hence, a systems thinking approach for airport systems can provide an in-depth knowledge of airport structure and operations.

The number of airports in Turkey, as in many countries, has increased in the last decade. While 33 airports were actively used in 2001, there are 52 airports in Turkey as of 2014. Figure 2 shows the increase in number of passengers for Turkish air transportation. The number of passengers increased %233 in 10 years (DHMI, 2014).

Today, due to the increase in aircraft and passenger traffic, improving the efficiency of airport operations and passenger satisfaction has become very important. Therefore, some requirements need to be taken into account when building airports such as reliability, safety, flexibility, comfort and high-quality. Airport modeling is an effective method for airport analysis and design to satisfy these requirements.

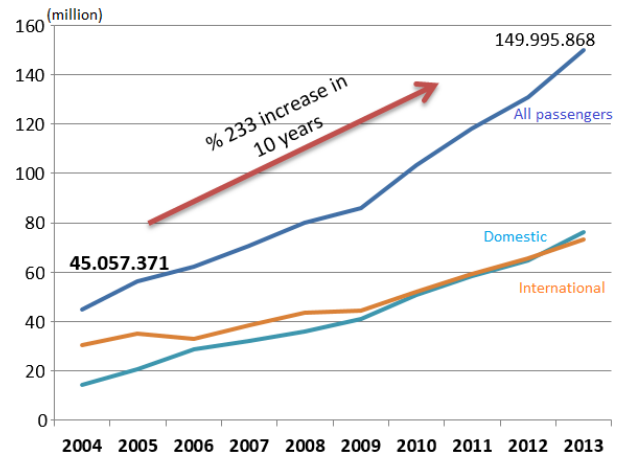


Figure 2: Passenger traffic in Turkish airports.

### 2.1. Passenger Terminals

Due to the fact that passengers spend most of their time before boarding at the terminals, airport terminal modeling has been widely studied in the literature to optimize passenger flows (Takakuwa 2003; Guizzi, Murino and Romano, 2009). An example airport terminal layout is given in Figure 3. A typical terminal includes check-in counters, gates, security check points, shops, restaurants, baggage claim, customs office, etc. Similar to the overall airport design, terminal design is also very important and each of these components and operations need to be carefully planned.

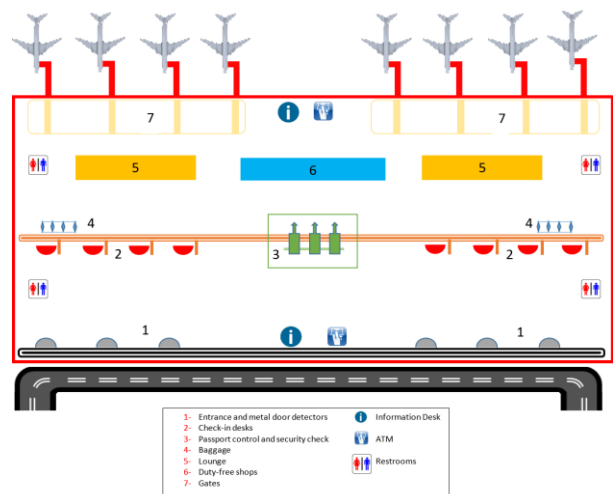


Figure 3: A sample airport terminal layout.

The aviation organizations define the standards and minimum requirements for efficient and safe terminal applications. For example, IATA (International Air Transport Association) defines industry norms for various components and services such as walking distances of passengers, aided or unaided. Similarly, airport passenger terminal planning and design guidebook of the Airport Cooperative Research Program presents the requirements and the guidelines for terminal design (ACRP, 2010). Any airport design and implementation should take care of these kind of regulations.



During the first stage of the project, the focus of the metamodel is currently on airport terminal modeling due to fact that the available knowledge and the simulation models in the literature can provide a sound basis. In the later stages, a comprehensive metamodel for airport modeling will be developed.

## 2.2. Airport Modeling and Simulation

Today, with the emerging computer technologies and simulation research, major airport developments are conducted with the use at some level of airport modeling and simulation and GIS techniques. Especially, to comply with the more stringent security measures at airports airport terminal simulation became essential and critical (Ashford, Mumayiz, and Wright, 2011). Airport simulations generally address and analyze the following problems and features: airport-airspace capacity studies, airport impacts on the environment, airport's functions and services, landside simulation, vehicle and people movements, baggage flows, etc. In the future, the demand for air transportation of passengers is expected to increase. This means that airport design and engineering discipline will need more tools and methods to optimize the processes and functions at the airports.

There are various tools to support airport design and to perform airport simulations such as SIMMOD and TAAM (Total Airspace and Airport Modeller). In order to achieve reasonable outputs various input data is required such as flight schedules, number of passengers, passenger characteristics, security control service rates, and so on.

Recently, more formal approaches in airport modeling and simulation have emerged. For example, Skorupski (2013) presents the concept of airport traffic modelling using coloured, timed, stochastic Petri nets. By the example of an airport with one runway and simultaneous takeoff and landing operations, the applicability of such models in analysis of air traffic processes is shown. Davidrajuh and Lin (2011) presents a similar effort as well. They show how to explore airport traffic capability using a Petri net based model.

In this work, the metamodel for airport modeling can be transformed into formal models such as Petri Nets or DEVS (Discrete Event System Specification). The formal models and formalism based simulation is explained in (Zeigler, Praehofer and Kim, 2000) and (Mittal and Risco-Martin, 2013).

## 3. APPLYING MDD4MS

Model driven development (MDD) approaches have been commonly used in the last decade in the simulation field. Çetinkaya, Verbraeck and Seck (2011a, 2011b) propose a general MDD framework for modeling and simulation (MDD4MS). The MDD4MS framework presents an integrated approach to bridge the gaps between different steps of a simulation study by using metamodeling and model transformations.

Metamodeling, in MDD context, is the process of complete and precise specification of a modeling

language in the form of a metamodel. A model transformation is the process of converting one or more source models into one or more target models according to a set of transformation rules. The MDD4MS framework defines three metamodels for conceptual modeling, simulation model specification, and simulation model implementation. Metamodel based model transformations provide the continuity throughout the M&S life cycle.

Applying MDD4MS in airport terminal modeling and simulation will provide new design mechanisms for airport planning. This study brings Software Engineering, Systems Engineering and Aeronautics disciplines together. The relationships between these three disciplines are shown in Figure 4. In particular, this work combines model driven development and airport modeling and simulation. Although there are some research in MDD literature that uses air transportation or airports as application domains or for illustrative purposes (Oetsch, 2011; Shaw, 2013), to the best of our knowledge, there are no metamodeling and MDD research proposed for airport modeling and simulation.

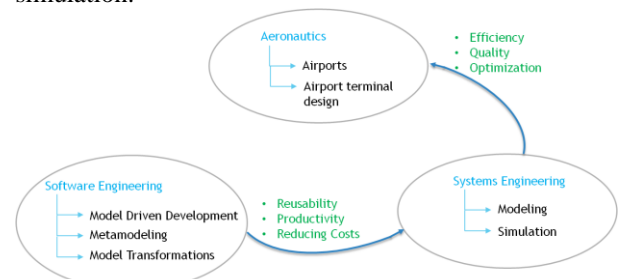


Figure 4: Interdisciplinary research on airport modeling and simulation.

Although simulation modeling is commonly used in this field, in most cases the development of a simulation model for an airport is expensive and takes time. However, with MDD approach the model generation can be automatic and simulation models can be developed more quickly.

In other words, once a conceptual model of the existing or proposed airport is developed, it can be transformed to a platform independent simulation model. The remaining steps can be automatic for a specific simulation platform. In this study, the proposed metamodel for airport conceptual modeling will provide means to determine some parameters for discrete event simulation. In this way, the airport design models can be transformed into simulation models easily.

Figure 5 shows the application of MDD4MS for airport modeling and simulation. There will be two model-to-model transformations and one model-to-code transformation. In case of transforming into a simulation language which is supported by a visual tool, such as the simulation programs like Arena or Simio, the model-to-code transformation is not needed. Our research will focus on transforming models into a simulation language instead of a general programming language.

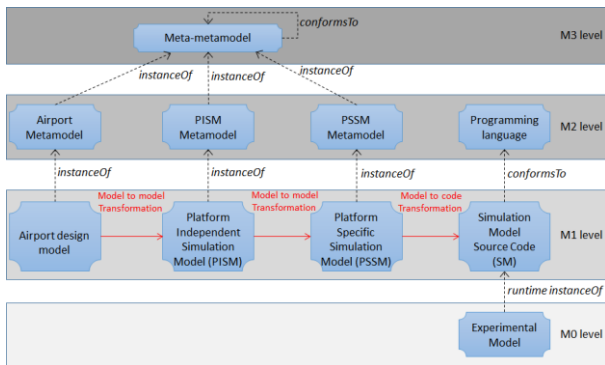


Figure 5: MDD4MS based airport modeling and simulation.

#### 4. METAMODEL FOR AIRPORT MODELING

In this section, we present a metamodel for airport modeling and simulation. The high level view of the metamodel is given in Figure 6. The metamodel is developed in GME (Generic Modeling Environment) version 14.

GME is a free and open source meta programming tool developed by Vanderbilt University. In GME, metamodels are defined as modeling paradigms. Upon loading a modeling paradigm, the MetaGME (the meta-metamodel) interpreter automatically creates an environment for model development in the specified modeling language. GME has a decorator facility for nicer visualization of the models. It is also possible to define some constraints on the metamodel with OCL.

Passengers, airlines and flights are modeled separate from the airport. All other components are modeled as part of an airport. Basically, an airport can have:

- Terminal
- Runway
- Apron
- Hangar
- ARFF (Airport rescue and firefighting station)
- Parking area
- Taxiway

As well as, a terminal can have:

- Entry point (metal door detectors)
- Passport control
- Security check
- Check-in
- Customs office
- Information desk
- Luggage conveyor
- Lift
- Other airport services
  - Car rental company
  - Change office
  - Shop
  - Restaurant
  - ATM
  - Ticket office
  - Restrooms

The metamodel includes two aspects, one for structural representation and one for activity modeling. Both diagrams form an airport design model. As an example, the diagram in Figure 7 shows passenger flow in an international departure terminal. Once an airport design model is ready, it can be transformed to a simulation model.

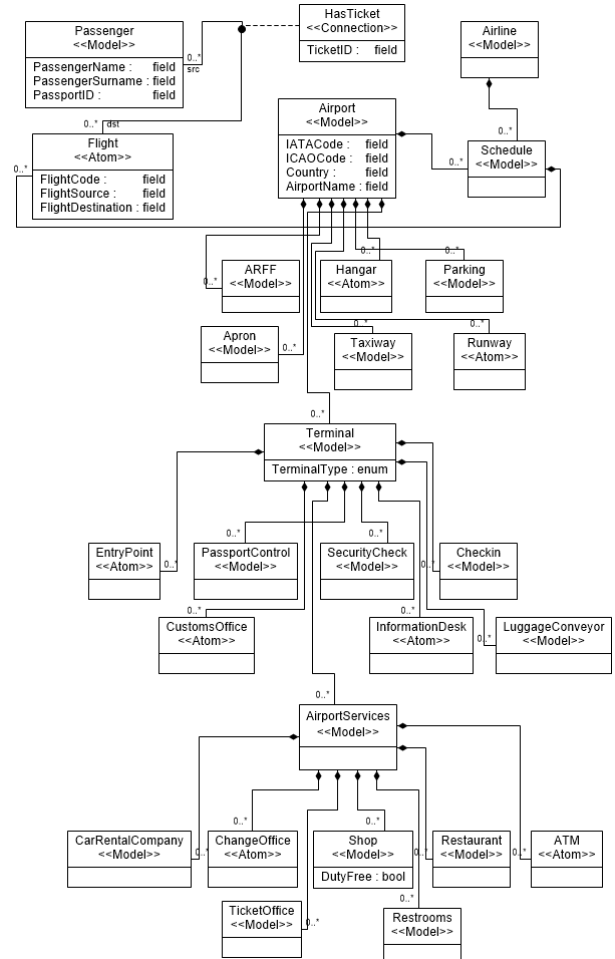


Figure 6: High level view of the airport metamodel.

In order to verify the feasibility and the correctness of the metamodel, the official airport planning and design documents of the Directorate General of Civil Aviation in Turkey will be used (SHGM, 2014). These documents provide the requirements and the constraints for airports, similar to Airport Development Reference Manual by IATA. As a practical case, the metamodel is aimed to be used to model and analyze the activities in one of the Turkish airports. The planned case study will help to evaluate and validate the metamodel.

#### 5. FUTURE WORK

The project has started with terminal modeling and it will continue with the other components and functions of an airport. Our research is going on for developing the metamodel and generating the modeling editor. Model transformation rules will be written next. Currently, the simulation platform is chosen as one of the commercial discrete event simulation tools.

However, in the later stages of the project, the airport design models will be transformed into a more common formalism such as DEVS to provide platform independence.

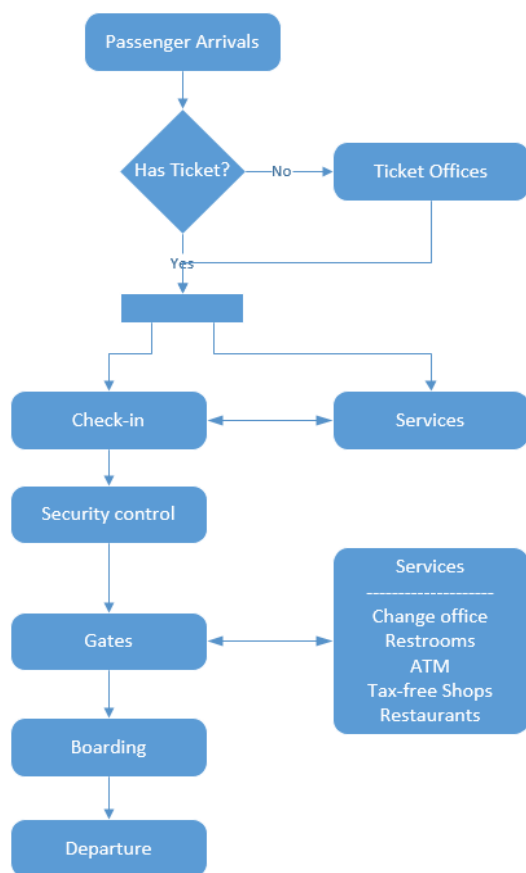


Figure 7: Passenger flow in an international departure terminal.

## REFERENCES

ACRP (The Airport Cooperative Research Program), Transportation Research Board of the National Academies, 2010, *Airport Passenger Terminal Planning and Design*, vol. I and II.

Airport Design Studio, University of California Berkeley, Available from: <http://airportdesign.berkeley.edu/> [accessed 14 May 2014]

Ashford, N. J., Mumayiz, S. and Wright, P. H., 2011, *Airport Modeling and Simulation*, in *Airport Engineering: Planning, Design, and Development of 21st Century Airports*, Fourth Edition, John Wiley & Sons, Inc.

Behan, P., 2013, *Improving the passenger experience*, IATA.

Cetinkaya, D., Verbraeck, A., 2011a. *Metamodeling and Model Transformations in Modeling and Simulation*. *Proceedings of the Winter Simulation Conference*.

Cetinkaya, D., Verbraeck, A. and Seck, M.D., 2011b, *MDD4MS: A Model Driven Development Framework for Modeling and Simulation*,

*Proceedings of the 2011 Summer Computer Simulation Conference (SCSC 2011)*.

Davidrajuh, R., Lin, B., 2011, Exploring airport traffic capability using Petri net based model, *Expert Systems with Applications*, volume 38(9), pp.10923–10931.

DHMI, Devlet Hava Meydanları İşletmesi Genel Müdürlüğü, 2014, *Uçak, yolcu ve yük istatistikleri*. Available from: <http://www.dhmi.gov.tr/istatistik.aspx> [accessed 15 July 2014].

FAA (Federal Aviation Administration), U.S. Department of Transportation, 2014, *Airport Design*. Available from: <http://www.faa.gov/> [accessed 14 May 2014]

Guizzi, G., Murino, T. and Romano, E., 2009, A discrete event simulation to model passenger flow in the airport terminal, *Proceedings of the 11th WSEAS International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering*, pp. 427-434.

Mittal, S. and Risco-Martin, J.L., 2013, *Netcentric System of Systems Engineering with DEVS Unified Process*, CRC Press, Taylor & Francis Group.

Oetsch, J., Pührer, J., Seidl, M., Tompits, H., and Zwickl, P., 2011, VIDEAS: A Development Tool for Answer-Set Programs Based on Model-Driven Engineering Technology, *Proceedings of the 11th International Conference on Logic Programming and Nonmonotonic Reasoning*, Lecture Notes in Computer Science, vol. 6645, pp. 382-387.

Risco-Martin, J.L., de la Cruz, J.M., Mittal, S. and Zeigler, B.P., 2009, eUDEVS: Executable UML with DEVS Theory of Modeling and Simulation, *Simulation*, 85, pp. 750-777.

Shaw, G. D., 2013, *Reliable model-driven engineering using IEC 61499*, PhD thesis, The University of Auckland.

SHGM, Sivil Havacılık Genel Müdürlüğü (Directorate General of Civil Aviation in Turkey), Available from: <http://web.shgm.gov.tr/> [accessed 15 July 2014]

Skorupski, J., 2013, *Airport Traffic Simulation Using Petri Nets*, in *Activities of Transport Telematics, Communications in Computer and Information Science*, vol. 395, pp 468-475.

Takakuwa, S. and Oyama, T., 2003, Modeling people flow: simulation analysis of international-departure passenger flows in an airport terminal, *Proceedings of the 2003 Winter Simulation Conference*, pp. 1627-1634.

Verbraeck, A. and Valentin, E., 2002, Transportation applications of simulation: simulation building blocks for airport terminal modeling, *Proceedings of the 2003 Winter Simulation Conference*, pp. 1199-1206.

Zeigler, B.P., Praehofer, H. and Kim, T.G., 2000, *Theory of modelling and simulation: integrating discrete event and continuous complex dynamic systems*, 2nd edition, Academic Press.

## **AUTHORS BIOGRAPHY**

**Dr. Deniz Çetinkaya** is an Assistant Professor at University of Turkish Aeronautical Association in Ankara, Turkey. She graduated from Department of Computer Engineering at Hacettepe University, Ankara, Turkey with honors in 2002. She received her M.Sc. degree in Computer Engineering from Middle East Technical University, Ankara, Turkey in 2005. She received her Ph.D. degree in Systems Engineering from Delft University of Technology (Technische Universiteit Delft) in the Netherlands in 2013. Her research focuses on model driven development, modeling and simulation, discrete event simulation and conceptual modeling.

**İsmet Camcı** is a M.Sc. candidate in the Electrical and Computer Engineering Department of the Institute of Science and Technology at University of Turkish Aeronautical Association. He graduated from the Department of Computer Systems Education in the Faculty of Technical Education at Gazi University, Ankara, Turkey in 2009. He worked as a teacher in the Department of Computer Technology at Balıkesir Petty Officer Vocational School in 2011 during his military service. Since 2012, he has been working in Ministry of National Education - Fatih Sultan Mehmet Anatolian Vocational High School, Ankara, Turkey.

# SIMULATION BASED ANALYSIS AND DEVELOPMENT OF DECISION SUPPORT SYSTEM FOR VIRTUAL NETWORK BANDWIDTH MANAGEMENT

J. Asmuss<sup>(a)</sup>, G. Lauks<sup>(b)</sup>

Institute of Telecommunication, Faculty of Electronics and Telecommunication,  
Riga Technical University, Azenes str. 12, LV-1048, Riga, Latvia

<sup>(a)</sup> [Julija.Asmuss@rtu.lv](mailto:Julija.Asmuss@rtu.lv), <sup>(b)</sup> [Gunars.Lauks@rtu.lv](mailto:Gunars.Lauks@rtu.lv)

## ABSTRACT

We develop fuzzy logic based methodology of decision making on bandwidth allocation in a substrate network with DaVinci architecture, according to which the physical substrate network is divided into virtual networks. We show that fuzzy game theory can be effectively used as a framework of decision making tools for bandwidth allocation mechanisms. The effectiveness of this methodology is evaluated and improved within simulation experiments realized by using Coloured Petri Nets Tools.

Keywords: bandwidth allocation problem, decision making system, game theory, simulation, CPN Tools

## 1. INTRODUCTION

We consider the problem of resource allocation in a substrate network with DaVinci architecture (He, Zhang-Shen, Li, Lee, Rexford and Chiang 2008), according to which the substrate physical network is divided into multiple virtual networks, and we develop a fuzzy logic based methodology of its analysis. Each virtual network is logically separated and can be modified for a particular class of traffic. All resources that are offered by the substrate network are shared between virtual networks. Finding the suitable bandwidth resource allocation to virtual networks is one of the main problems for virtualization-enabled networking infrastructures (Chowdhury and Boutaba 2010; Haider, Potter and Nakao 2009; Zhang, Wang and Gao 2009; Zhou, Li, Sun, Jin, Su and Zeng 2010).

We deal with the problem, which can be considered as a maximization problem for the aggregate utility of all virtual networks. Accordingly to DaVinci principles individual virtual networks do not have the knowledge of other virtual network's conditions and cannot cooperate with each other. We suppose that at a smaller period of time each virtual network acts individually to maximize its own performance. The main question here is whether optimization of virtual networks combined with the bandwidth share adaptation scheme performed by the substrate network, which does not know the performance objectives of virtual networks, actually maximizes the aggregate utility.

Under such assumptions the resource allocation issue can be modelled as a non-cooperative game. In the context of game theory it can be modelled how virtual

networks interact with each other in an effort to achieve their own goals. In the resource allocation game each virtual network is self-interested and is striving to maximize his utility function, where the utility function represents the virtual network's performance and controls the outcomes of the game. Many researchers have used game theory to model different decision processes in networks (see, e.g., Cao, Shen, Milito and Wirth 2002; Zhou, Li, Sun, Jin, Su and Zeng 2010; Han, Niyato, Saad, Basar and Hjørungnes 2012). However, classical mechanisms usually are not capable of making decisions in uncertain conditions, which are prevailing in the modern networks. In this paper for the first time we combine the concepts of game theory and fuzzy logic theory to involve the uncertainties of information in the model of network bandwidth allocation. Networks having a centralized authority that controls the entire network will give way to self-optimizing networks. We suppose that the substrate network does not know the utility functions of all virtual networks. The aim of this research is to develop an efficient decision support scheme simulated in the form of a non-cooperative game with imprecise information, using fuzzy logic concepts that can assist network management in making an effective decision on bandwidth allocation.

In this paper we develop a fuzzy approach for network bandwidth management presented at the last EUSFLAT conference (Asmuss and Lauks 2013). This methodology was suggested for two virtual networks and was based on fuzzification and defuzzification principles and the expert knowledge database of fuzzy rules. The simulation experiments were conducted for two traffic classes – delay sensitive and throughput sensitive traffic. The main aim of the present paper is to develop this approach for the case of multiple traffic classes. We show the way it will be done by extending the decision making system and simulation scheme with the game theory based module that fairly and efficiently share resources among multiple traffic classes. We apply fuzzy game theory based approach by considering fuzzy players, which use fuzzy rules to make strategic decisions.

## 2. BANDWIDTH ALLOCATION PROBLEM

The problem of bandwidth allocation in a substrate network is a maximization problem for the aggregate

objective of multiple virtual networks with different requirements. According to the DaVinci approach (He, Zhang-Shen, Li, Lee, Rexford and Chiang 2008) each traffic class is carried on its own virtual network with customized traffic management protocols. The substrate network by assigning resources to each virtual link gives each virtual network the illusion that it runs on a dedicated physical infrastructure.

Let the topology of a substrate network be given by a graph  $G_s = \{V_s, E_s\}$  with a set  $V_s$  of nodes (or vertices) and a set  $E_s$  of links (or edges). We suppose that links  $l \in E_s$  are with finite capacities  $C_l$ . Correspondingly to  $G_s$  we consider DaVinci model with  $N$  virtual networks, indexed by  $k$ , where  $k = 1, 2, \dots, N$ . Let the key notations be the following:

- $\mathbf{y}^{(k)}$  – bandwidth of virtual network  $k$ ,  $k = 1, 2, \dots, N$ ;
- $\mathbf{z}^{(k)}$  – path rates for virtual network  $k$ ,  $k = 1, 2, \dots, N$ ;
- $\lambda^{(k)}$  – satisfaction level degree of virtual network  $k$ ,  $k = 1, 2, \dots, N$ ;
- $O^{(k)}$  – performance objective for virtual network  $k$ ,  $k = 1, 2, \dots, N$ .

In the case when different types of traffic coexist over the same network substrate, each virtual network could control only a subset of resources at each node and link. This means that at a smaller timescale, each virtual network runs according to a distributed protocol that maximizes its own performance objective independently. Due to the bandwidth limitation, it is important, but impossible for virtual networks to communicate and cooperate with each other, so as to optimally utilize the resources. Under such conditions in a dynamically changing virtual network environment a fundamental issue of resource allocation is the design of dynamically adaptive bandwidth allocation mechanisms.

Bandwidth values  $\mathbf{y}^{(k)} = (y_l^{(k)})_{l \in E_s}$  are assigned by the substrate network, taking into account such local information as current satisfaction indicators and performance objectives. The substrate network periodically reassigns bandwidth shares for each substrate link  $l \in E_s$  between its virtual links. Thus, values  $\mathbf{y}^{(k)}$  and  $O^{(k)}$  are periodically updated. Usually it is supposed that the objective function  $O^{(k)}$  depends on both virtual link rates  $\mathbf{z}^{(k)}$  and virtual link capacity  $\mathbf{y}^{(k)}$ .

We associate objective or utility function with each virtual network, and we refer to a resource allocation scheme as being socially optimal if it maximizes the sum of utilities of all users in the network. Each virtual network tries to maximize its individual interest, while the system strives to increase its efficiency, i.e. the overall system performance. The goal of the substrate network is to optimize the aggregate utility of all virtual

networks (see, e.g., Lin and Shroff 2006; He, Zhang-Shen, Li, Lee, Rexford and Chiang 2008). So, we formulate the optimization problem for the aggregate utility:

$$\begin{aligned}
 & \text{maximize} && \sum_{k=1}^N O^{(k)}(\mathbf{z}^{(k)}, \mathbf{y}^{(k)}) \\
 & \text{subject to} && \mathbf{H}^{(k)} \mathbf{z}^{(k)} \leq \mathbf{y}^{(k)}, \quad k = 1, 2, \dots, N, \\
 & && \sum_{k=1}^N \mathbf{y}^{(k)} \leq \mathbf{C}, \\
 & && g^{(k)}(\mathbf{z}^{(k)}) \leq 0, \quad k = 1, 2, \dots, N, \\
 & && \mathbf{z}^{(k)} \geq \mathbf{0}, \quad k = 1, 2, \dots, N, \\
 & \text{variables} && \mathbf{z}^{(k)}, \mathbf{y}^{(k)}, \quad k = 1, 2, \dots, N.
 \end{aligned}$$

The maximization is subject to the capacity constraint and possibly other constraints described in terms of  $g^{(k)}(\mathbf{z}^{(k)})$ . The capacity constraint requires the link load  $\mathbf{H}^{(k)} \mathbf{z}^{(k)}$  to be no more than the allocated bandwidth. To compute the link load we use routing indexes

$$H_{lj}^{(k)i} = \begin{cases} 1, & \text{if path } j \text{ of source } i \text{ in virtual} \\ & \text{network } k \text{ uses link } l, \\ 0, & \text{otherwise,} \end{cases}$$

and path rates  $z_j^{(k)i}$  that determine for source  $i$  the amount of traffic directed over path  $j$  of virtual network  $k$ .

An optimization scheme follows directly from DaVinci principles. First, the substrate network determines how satisfied each virtual network is with its allocated bandwidth. Satisfaction level degree  $\lambda_l^{(k)}$  is an indicator that a virtual network  $k$  may want more resources on link  $l$ . Next, the substrate network determines how much bandwidth virtual network  $k$  should have on link  $l$ : the substrate network increases or decreases value  $y_l^{(k)}$  in dependence on the satisfaction level  $\lambda_l^{(k)}$ .

Provided that each virtual network is acting independently, the question is whether individual virtual networks without communication and cooperation with each other, together with the bandwidth share adaptation performed by the substrate network, which does not know the performance objectives of virtual networks, actually maximize the overall performance objective.

### 3. DECISION MAKING METHODOLOGY

Our work focuses on decision making on the substrate network level on bandwidth shares for link  $l \in E_s$ . We denote by  $t_j$ , where  $j = 1, 2, \dots$ , moments of system



adaptation, i.e. moments of decision on bandwidth shares  $y_l^{(k)}(t_j)$  for  $k=1,2,\dots,N$ . Such decision is based on the results of monitoring the performance of the system during interval  $[t_{j-1}, t_j]$  and values  $y_l^{(k)}(t_{j-1})$ ,  $k=1,2,\dots,N$ . We suppose that values  $y_l^{(k)}(t_0)$  for  $k=1,2,\dots,N$  are given and consider the behaviour of the system for  $t \geq t_0$ .

We consider the substrate network shared by many users, where the goal is to share the network resources in an optimal manner. Since the substrate network does not know the utility functions of all users, the network is not able to evaluate the optimal resource allocation. Networks having a centralized authority that controls the entire network will give way to self-optimizing networks.

Game theory is a tool used for modelling and analyzing conflict and cooperation between decision makers called players. Such a situation occurs when multiple decision makers with different objectives act in a system or share resources. Game theory provides a natural framework for the modelling of the bandwidth allocation mechanism in this paper. In the context of this work, we use non-cooperative games due to the competitive nature of the players (virtual networks or users in our case). Two or more virtual networks compete for a limited bandwidth of links of the substrate network.

Every virtual network attempts to maximize its utility in the game. A significant aspect of the game theory is that each virtual network's utility is based on the decision of every other virtual network and hence, each virtual network is able to optimize its utility only with respect to every other virtual network's decision. In our model we consider the non-cooperative game of  $N$  players (virtual networks) with stages  $S_l(t_j)$ ,  $j=1,2,\dots$ , accordingly to decision moments  $t_j$ ,  $j=1,2,\dots$ . A strategy of each virtual network is a decision rule that specifies: does virtual network  $k$  want more resources or not. The result of bandwidth reallocation (i.e. bandwidth shares  $y_l^{(k)}(t_j)$  for  $k=1,2,\dots,N$ ) depends on the strategies of all virtual networks.

A crucial aspect of the specification of a game involves the information that players have when they choose strategies. The simplest games (from the perspective of logical structure) are those in which agents have perfect information, meaning that at every point where each agent's strategy tells him to perform an action, he knows everything that has happened in the game up to that point. In contrast, we consider a game in which virtual networks do not know everything what has happened in the game up to that point when they take actions (such games are classified as games with imperfect information). Given that each virtual network

is acting independently. This means that at moment  $t_j$  player  $k$  has information on behavior during interval  $[t_{j-1}, t_j]$  only for virtual network  $k$ . Games with imperfect information can be modelled by using fuzzy logic based systems. We apply fuzzy game theory based approach by considering fuzzy players, which use fuzzy rules to make strategic decisions. Such generalization of business games was introduced and effectively applied in business management (Oderanti and De Wilde 2010; Oderanti and De Wilde 2011; Oderanti and De Wilde 2012).

#### 4. GAME DECISION SUPPORT SYSTEM

We suppose that each player (virtual link) uses fuzzy rules to make strategic decisions. The procedures necessary for designing the proposed decision support system (Ross 1995; Piegat 2001) are listed below.

1. List fuzzy variables (factors) that will be considered in taking the decision.
2. Determine the strategies of virtual networks (players).
3. Develop membership functions for all variables.
4. Formulate decision rules for the rule base.
5. Describe the fuzzy inference technique.
6. Play the game.

##### 4.1. Input fuzzy variables

The first step of fuzzy logic based decision making system design is to define fuzzy variables. We use one output variable  $F_l^{(k)}(t_j)$  and three input variables for each link  $l$  and each virtual network  $k$ :

$U_l^{(k)}(t_j)$  – the average utilization of the virtual link;

$L_l^{(k)}(t_j)$  – the average length of queue;

$D_l^{(k)}(t_j)$  – the average delay of packets.

Input variables are used to describe the system state and the starting point of adaptation decision. These variables are evaluated for the period  $[t_{j-1}, t_j]$  and have linguistic values according to their membership functions. We use the denotations  $LU_l^{(k)}(t_j)$ ,  $LL_l^{(k)}(t_j)$  and  $LD_l^{(k)}(t_j)$  for the linguistic values of  $U_l^{(k)}(t_j)$ ,  $L_l^{(k)}(t_j)$  and  $D_l^{(k)}(t_j)$  correspondingly.

##### 4.2. Strategy variable

We introduce one fuzzy variable for each virtual network  $k$  to denote an output variable in the strategy decision process with respect to virtual link  $l$  at moment  $t_j$ :  $F_l^{(k)}(t_j)$ .

The strategy of virtual network  $k$  at moment  $t_j$  depends on a defuzzified output value of the fuzzy

inference module. We use the denotation  $LF_l^{(k)}(t_j)$  for the linguistic value of this variable.

### 4.3. Membership functions

For all input variables we consider three linguistic values „Low”, „Medium” and „High” (we use the notation L, M and H as indexes). But for the strategy variable linguistic values are: „More” (meaning that the corresponding virtual network wants more bandwidth resources), „Constant” and „Less” (meaning that the corresponding virtual network wants less bandwidth resources).

The membership function for each linguistic value is given by a triangular or trapezoidal fuzzy number.

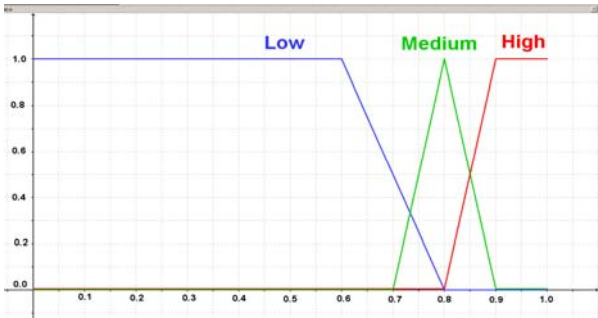


Figure 1: Membership Functions of the Linguistic Values for the Utilization Variable

For example, Fig.1 shows the graphs of the membership functions of linguistic values  $U_L$ ,  $U_M$ ,  $U_H$  for the utilization variables. These membership functions are considered to be independent on  $l$ ,  $k$  and  $j$ . Obviously, such assumption is not true for the membership functions of other linguistic values: the membership functions for delay and queue variables depend on traffic class.

### 4.4. Decision rules

The knowledge base for our model is made as rule database, which is based only on the “expert knowledge”, where the experts are the authors of the paper, and the rules are assumed as logical assumptions, e.g.:

if  $LU_l^{(k)}(t_j) = \text{“High”}$  and  $LL_l^{(k)}(t_j) = \text{“Low”}$  and  $LD_l^{(k)}(t_j) = \text{“Low”}$ , then  $LF_l^{(k)}(t_j) = \text{“Constant”}$ ;

if  $LU_l^{(k)}(t_j) = \text{“High”}$  and  $LL_l^{(k)}(t_j) = \text{“Low”}$  and  $LD_l^{(k)}(t_j) = \text{“Medium”}$ , then  $LF_l^{(k)}(t_j) = \text{“More”}$ ;

if  $LU_l^{(k)}(t_j) = \text{“Low”}$  and  $LL_l^{(k)}(t_j) = \text{“Low”}$  and  $LD_l^{(k)}(t_j) = \text{“Low”}$ , then  $LF_l^{(k)}(t_j) = \text{“Less”}$ ;

if  $LU_l^{(k)}(t_j) = \text{“Medium”}$  and  $LL_l^{(k)}(t_j) = \text{“Low”}$  and  $LD_l^{(k)}(t_j) = \text{“Medium”}$ , then  $LF_l^{(k)}(t_j) = \text{“More”}$  ...

The rule database depends on the type of traffic and it can be freely modified as well as the membership function definitions. The assumed fuzzy rules and the membership functions of input and output values were used for the proposed approach performance evaluation by simulation process and the impact of their modification is considered as the field for the future research. Our future work will mainly focus on modify our approach by applying training (learning) algorithms to the decision process.

### 4.5. Fuzzy inference technique

We list the following parameters of decision making, which are used in our model in order to achieve the advanced goals (see, e.g., Ross 1995; Piegat 2001):

- the number of linguistic values and the membership functions of linguistic values for each input and output variable;
- the base of if-then rules;
- the method of defuzzification of output parameters;
- the type of decision making system.

In our model we apply the fuzzy inference technique, introduced by Mamdani (Mamdani 1974; Mamdani and Assilian 1975; also see Ross 1995; Piegat 2001). The COG (centre of gravity) defuzzification method was used in this decision making process.

### 4.6. Dynamic game

Our game is a repeated game. It is a special case of dynamic games, when players interact by playing a similar stage of the game numerous times. The next section is devoted to the description of stages of the game.

## 5. STAGES OF THE GAME

The game stage  $S_l(t_j)$  for link  $l \in E_s$  is represented as vector

$$S_l(t_j) = (y_l^{(k)}(t_j), P_l^{(k)}(t_j) \mid k = 1, 2, \dots, N),$$

where  $y_l^{(k)}(t_j)$  represents the bandwidth value for virtual network  $k$ , obtained at moment  $t_j$ , and  $P_l^{(k)}(t_j)$  gives the game payoff for virtual network  $k$ , which is evaluated for interval  $[t_{j-1}, t_j]$  accordingly to virtual network’s strategies at the stage  $S_l(t_{j-1})$ ,  $k = 1, 2, \dots, N$ ,  $j = 1, 2, \dots$ .

Initial stage of the game is

$$S_l(t_0) = (y_l^{(k)}(t_0), P_l^{(k)}(t_0) \mid k = 1, 2, \dots, N),$$

where

$$P_l^{(k)}(t_0) = 0, \quad k = 1, 2, \dots, N,$$

but values  $y_l^{(k)}(t_0)$  are given. We can take

$$y_l^{(k)}(t_0) = C_l / N, \quad k = 1, 2, \dots, N.$$

We suppose that in the starting stage all virtual networks choose the strategy „Constant” (of course, it is possible to take into account the choice of virtual network’s own strategies in the initial stage, also). This means that during the first time interval  $[t_0, t_1]$  the distribution of the link bandwidth will be uniform.

At moment  $t_1$  payoffs  $P_l^{(k)}(t_1)$ ,  $k = 1, 2, \dots, N$ , are evaluated and all virtual links choose the strategies for time interval  $[t_1, t_2]$ , taking into account the behavior of the system during the first time interval  $[t_0, t_1]$ . It is done accordingly to the general game algorithm, which is described below for  $t_j$ ,  $j = 1, 2, \dots$ . In order to achieve the efficiency of the substrate network, our primary concern is to design the payoff functions, which will motive individual virtual networks to adopt a social behaviour and to improve the system performance by sharing the resources. In this context we apply a pricing technique (see, e.g., Zhou, Li, Sun, Jin, Su and Zeng 2010).

The payoff value  $P_l^{(k)}(t_j)$  is obtained by three parts:

$$P_l^{(k)}(t_j) = U_l^{(k)}(t_j) - A_l^{(k)}(t_j) - B_l^{(k)}(t_j),$$

where  $U_l^{(k)}(t_j)$  represents the utility value of virtual network  $k$ ,  $A_l^{(k)}(t_j)$  is the pricing part and  $B_l^{(k)}(t_j)$  gives the congestion value for virtual network  $k$  for interval  $[t_{j-1}, t_j]$  and link  $l$ . The payoff function has the physical meaning of utility value minus costs.

We denote by  $p_l$  the price of the bandwidth on the physical link  $l$  for one time unit. Virtual network  $k$  should pay

$$A_l^{(k)}(t_j) = p_l y_l^{(k)}(t_{j-1})(t_j - t_{j-1})$$

as the total price for the assigned bandwidth on link  $l$  during time interval  $[t_{j-1}, t_j]$ . The value  $B_l^{(k)}(t_j)$  for virtual network  $k$  is the measure of the congestion cost according to the assigned bandwidth and actual link rate on the base of congestion price  $\beta_l$  of link  $l$ . If the total link rate of all virtual networks is less than the capacity of the physical link, no congestion cost will be charged. In our model we assume that the real rate is not the reason that causes the congestion, and we simplify the formula by taking  $\beta_l = 0$ .

According to our approach all virtual networks use fuzzy rules to make strategic decisions. As it was described in the previous section each virtual network is considering as a fuzzy player, which apply the fuzzy inference technique to make strategic decisions. The suggested fuzzy solution requires one output variable for each virtual network  $k$ , which determines a fuzzy logic based decision on its strategy at moment  $t_j$ :  $F_l^{(k)}(t_j)$ . A defuzzified output value belongs to the interval  $[0, 1]$ . The value 0.5 corresponds to the strategy „Constant”, the minimal value corresponds to the strategy „Less” and the maximal value corresponds to the strategy „More”, accordingly.

Bandwidth shares  $y_l^{(k)}(t_j)$ ,  $k = 1, 2, \dots, N$ , for all virtual networks at the game stage  $S_l(t_j)$  are evaluated proportionally to the requested values  $b_l^{(k)}(t_j)$ ,  $k = 1, 2, \dots, N$ , which are calculated by the formula:

$$b_l^{(k)}(t_j) = y_l^{(k)}(t_{j-1}) (1 + 0.2(2F_l^{(k)}(t_j) - 1)).$$

This formula realizes our policy: the requested value should not differ from the previous one more than 20 percent. It is clear that this policy can be modified by modifying the formula.

Taking into account all requested values, new bandwidth shares  $y_l^{(k)}(t_j)$ ,  $k = 1, 2, \dots, N$ , now are calculated by the following way:

$$y_l^{(k)}(t_j) = b_l^{(k)}(t_j) \text{ if } \sum_{k=1}^N b_l^{(k)}(t_j) \leq C_l;$$

$$y_l^{(k)}(t_j) = C_l b_l^{(k)}(t_j) : \sum_{i=1}^N b_l^{(i)}(t_j) \text{ if } \sum_{k=1}^N b_l^{(k)}(t_j) > C_l.$$

Now all components of the stage  $S_l(t_j)$  are described.

The assigned values  $y_l^{(k)}(t_j)$ ,  $k = 1, 2, \dots, N$ , will be applied for the next time interval  $[t_j, t_{j+1}]$ . At moment  $t_{j+1}$  we will start the next stage  $S_l(t_{j+1})$  of the game.

## 6. SIMULATIONS

For a practical realization of the proposed approach we extend the design and simulation scheme of dynamically adaptive bandwidth allocation method, which was proposed for two traffic classes (Asmuss, Zagorskis and Lauks 2012; Asmuss and Lauks 2013). The simulation experiments in our previous papers were conducted for two traffic classes - delay sensitive and throughput sensitive traffic. Now we describe the extension done by including the fuzzy game theory based module that fairly and efficiently share resources among multiple traffic classes.

Our simulation scheme is based on Coloured Petri Nets (CPN) Tools (Jensen 1992-1997; Ratzert, Wells, Larsen, Laursen, Qvortrup, Stissing, Westergaard,

Christensen and Jensen 2003; Jensen, Kristensen and Wells 2007; Jensen and Kristensen 2009; Gehlo and Nigro 2010). CPN Tools is a discrete event modelling computer tool for CPN models supporting interactive and automatic simulations, state spaces and performance analysis and combining Coloured Petri Nets and the functional programming language CPN ML which is based on Standard ML. Colours can be effectively used for modelling virtual networks accordingly to the DaVinci architecture. CPN model of a substrate network describes the states of each virtual network of the system and the events (transitions) that can cause the system to change state. By making simulations of the network CPN model with CPN Tools it is possible to investigate different scenarios and explore the behaviours of the system, to use simulation-based performance analysis for decision making and adaptation processes.

To simplify the issue at the present stage of our research we simulate the adaptive bandwidth allocation on the link level. Traffic FIFO queues and virtual links are separated due to colours. The transmission time depends on the size of a packet and the bandwidth of the corresponding virtual link and is calculated for each packet transmission. Flows of packets for all virtual links are parallel controlled and analyzed. For more details we refer to our previous papers (Asmuss, Zagorskis and Lauks 2012; Asmuss and Lauks 2013).

The extended simulation scheme involves bandwidth shares adaptation module, which involves the proposed above fuzzy game theory based decision making submodule. The bandwidth allocation adaptation module is specially designed to update the virtual link resources allocation for dynamically changing traffics. Special observation functions and data collection monitors are included for monitoring the performance of the system. A monitoring mechanism is used not only to control, but also to modify simulations of the net. The decision making system is based on data collection monitors that allow calculating the system performance measures such as the delay in each queue, the length of each queue, the utilization of each link.

We experiment with two and three traffic classes. Simulation results allow us to improve adaptation mechanism by modification of fuzzy rules and membership functions. Bandwidth of substrate links in our experiment is determined as 100 Mbps and 150 Mbps, correspondingly. Packets are generated with exponentially distributed arrival time. In our experiments the initial resource allocation is uniform, i.e. we set the capacity of each virtual link as 50 Mbps. By changing traffic parameters we observe the resource allocation adaptation process (see Table 1 with simulation results for three traffic classes).

Our simulation results demonstrate that the adaptive bandwidth allocation mechanism can efficiently react to traffic changes for  $k=2$  when  $j=7$  and for  $k=1$  when  $j=12$ . The simulations show that the proposed adaptation technique has good adaptation results: after three iterations we obtain

bandwidth shares corresponding to virtual network's traffic.

Table 1: Bandwidth Allocation Results for Three Traffic Classes

|                  |  | $b_l^{(k)}(t_j)$ |    |    | $y_l^{(k)}(t_j)$ |    |    | Utility |
|------------------|--|------------------|----|----|------------------|----|----|---------|
| $j \backslash k$ |  | 1                | 2  | 3  | 1                | 2  | 3  |         |
| 0                |  |                  |    |    | 50               | 50 | 50 | 0.48    |
| 1                |  | 60               | 40 | 46 | 60               | 40 | 46 | 0.74    |
| 2                |  | 66               | 33 | 48 | 66               | 33 | 48 | 0.78    |
| 3                |  | 74               | 39 | 52 | 67               | 36 | 47 | 0.92    |
| 4                |  | 72               | 36 | 51 | 68               | 34 | 48 | 0.90    |
| 5                |  | 66               | 41 | 52 | 62               | 39 | 49 | 0.93    |
| 6                |  | 64               | 38 | 48 | 67               | 40 | 51 | 0.91    |
| 7                |  | 65               | 48 | 51 | 59               | 44 | 47 | 0.84    |
| 8                |  | 68               | 54 | 53 | 58               | 46 | 46 | 0.88    |
| 9                |  | 70               | 52 | 52 | 60               | 45 | 45 | 0.91    |
| 10               |  | 68               | 54 | 51 | 59               | 47 | 44 | 0.90    |
| 11               |  | 68               | 53 | 52 | 59               | 46 | 46 | 0.88    |
| 12               |  | 55               | 52 | 53 | 52               | 49 | 49 | 0.73    |
| 13               |  | 48               | 53 | 52 | 47               | 52 | 51 | 0.82    |
| 14               |  | 48               | 52 | 53 | 47               | 51 | 52 | 0.90    |
| 15               |  | 47               | 53 | 53 | 46               | 52 | 52 | 0.92    |

Of course, these results have been obtained through simulations using very strict assumptions and settings rather than real life scenarios. But they can be considered as an illustration of the proposed approach.

## 7. CONCLUSION

We are applying the concepts of fuzzy logic and game theory to model the decision process on bandwidth allocation. We are suggesting a model and a simulation scheme that can serve as a good tool for network bandwidth management and future investigation. Our future work will mainly focus on improving and modification of our approach by applying learning mechanisms. What we are trying to show is that the uncertainty in network bandwidth management can be suitably modelled or represented using fuzzy logic and game theory concepts.

**Acknowledgments.** This work is supported by the project "Applications of mathematical structures based on fuzzy logic principles in the development of telecommunication network design and resource control technologies" (No.2013/0024/1DP/ 1.1.1.2.0/13/APIA/ VIAA/045) of the European Social Fund.

## REFERENCES

- Asmuss, J., Zagorskis, V. and Lauks, G., 2012. Simulation of dynamically adaptive bandwidth allocation protocols using Coloured Petri Nets. *Proceedings of the 24th European Modeling and Simulation Symposium EMSS*, 408–413, September 2012, Vienna, Austria.

- Asmuss, J. and Lauks, G., 2013. A fuzzy approach for network bandwidth management. *Proceedings of the 8th Conference of the European Society for Fuzzy Logic and Technology EUSFLAT, Advances in Intelligent Systems Research*, 32: 722–727, September 2013, Milan, Italy.
- Cao, X., Shen, H.-X., Milioto, R. and Wirth, P., 2002. Internet pricing with a game theoretical approach: concepts and examples. *IEEE/ACM Transactions on Networking*, 10(2): 208–216.
- Chowdhury, M. and Boutaba, R., 2010. A survey of network virtualization. *Computer Networks*, 54(5): 862–876.
- Gehlo, V. and Nigro, C., 2010. An introduction to system modeling and simulation with coloured Petri nets. *Proceedings of Winter Simulation Conference*, 104–118, December 2010, Baltimore, USA.
- Haider, A., Potter, R. and Nakao, A., 2009. Challenges in resource allocation in network virtualization. *The 20th ITC Specialist Seminar on Network Virtualization – Concept and Performance Aspects*, May 2009, Hoi An, Vietnam.
- Han, Z., Niyato, D., Saad, W., Basar, T. and Hjørungnes, A., 2012. *Game theory in wireless and communication networks: theory, models, and applications*. Cambridge University Press.
- He, J., Zhang-Shen, R., Li, Y., Lee, C.-Y., Rexford, J. and Chiang, M., 2008. DaVinci: dynamically adaptive virtual networks for a customized Internet. *Proceedings of the International Conference on Emerging Networking Experiments and Technologies ACM CoNEXT*, December 2008, Madrid, Spain.
- Jensen, K., 1992–1997. *Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use*. Volumes 1–3. Springer Verlag.
- Jensen, K. and Kristensen, L., 2009. *Coloured Petri nets. Modelling and validation of concurrent systems*. Springer-Verlag.
- Jensen, K., Kristensen, L. and Wells, L., 2007. Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems. *International Journal on Software Tools for Technology Transfer*, 9: 213–254.
- Lin, X. and Shroff, N., 2006. Utility maximization for communication networks with multi-path routing. *IEEE Transactions on Automatic Control*, 51, 766–781.
- Mamdani, E.H., 1974. Applications of fuzzy algorithms for control of simple dynamic plant. *Proceedings of the IEEE*, 121(12): 1585–1588.
- Mamdani, E.H. and Assilian, S., 1975. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7(1): 1–13.
- Oderanti, F. O. and De Wilde, P., 2010. Dynamics of business games with management of fuzzy rules for decision making. *International Journal of Production Economics*, 128(1): 96–109.
- Oderanti, F. O. and De Wilde, P., 2011. Automatic fuzzy decision making system with learning for competing and connected businesses. *Expert Systems with Applications*, 38: 14574–14584.
- Oderanti, F. O. and De Wilde, P., 2012. Application of strategic fuzzy games to wage increase negotiation and decision problems. *Expert Systems with Applications*, 39(2): 11103–11114.
- Piegat, A., 2001. *Fuzzy modeling and control*. Physica-Verlag, Heidelberg.
- Ratzer, A.V., Wells, L., Larsen, H.M., Laursen, M., Qvortrup, J.F., Stissing, M.S., Westergaard, M., Christensen, S. and Jensen, K., 2003. CPN Tools for editing, simulating and analysing coloured Petri net. *Lecture Notes in Computer Science*, 2679: 450–462.
- Ross, T.E., 1995. *Fuzzy logic with engineering applications*. McGraw-Hill, New York.
- Zhang, Y., Wang, C. and Gao, Y., 2009. A virtualized network architecture for improving QoS. *International Journal on Information Engineering and Electronic Business*, 1: 17–24.
- Zhou, Y., Li, Y., Sun, G., Jin, D., Su, L. and Zeng, L., 2010. Game theory based bandwidth allocation scheme for network virtualization. *Proceedings of the Global Communications Conference IEEE GlobeCom*, December 2010, Miami, USA.

# USING TABLETS IN DISTRIBUTED SIMULATION

Josef Brožek<sup>(a)</sup>, Martin Jakeš<sup>(b)</sup>, Lumír Gago<sup>(c)</sup>

<sup>(a)</sup>Faculty of Electrical Engineering and Informatics, University of Pardubice, Czech republic

<sup>(b)</sup>Faculty of Electrical Engineering and Informatics, University of Pardubice, Czech republic

<sup>(c)</sup>Faculty of Electrical Engineering and Informatics, University of Pardubice, Czech republic

<sup>(a)</sup>[mail@jobro.cz](mailto:mail@jobro.cz), <sup>(b)</sup>[jakesmar@gmail.com](mailto:jakesmar@gmail.com), <sup>(c)</sup>[lumir.gago@student.upce.cz](mailto:lumir.gago@student.upce.cz)

## ABSTRACT

The current trend in the sale of computer equipment is a dramatic growth in sales of tablets and smart phones and a gradual decline in the sale of personal computers (PC). Since this market change marks a change in the users' focus, it is desirable to reflect this change in the preferences and advantageously use it for simulation applications.

This article seeks in particular to:

- explain the advantages and disadvantages of the change in the user's preferences and its implications for the development of computer simulation,
- familiarize the reader with the basic advantages and disadvantages of tablets and their comparison with PCs,
- familiarize the user with the classes of tasks, for which it is possible to advantageously use tablets,
- familiarize the reader with the basic methodologies, architectures and techniques of simulation and distributed simulation for tablets,
- give the reader a concrete solution that utilizes the above-mentioned technologies and advantages.

Keywords: Simulation, tablet, smartphone, distributed simulation, heterogenic simulation, HLA.

## 1. MOTIVATION

### 1.1. Introduction to the Issue

When creating simulation systems of some classes, we need to take into account user preferences - this is especially true for simulation applications of the simulator (trainer) type. For this type of applications it is assumed that the user efficiently handles the control device which serves as input-output simulation device.

By generalizing this principle, we achieve a thesis which says that for an simulation control/operation to be effective, the user environment must be user friendly as much as possible. At present, however, in the field of user-friendliness, methods of software engineering and software design are ceasing to play the first fiddle, and

so does the focus on the hardware device that is to serve for operation.

The current development in the information technology market, especially the one focused on the end user, shows that there is a dramatic increase in the sales of tablets at the expense of standard computers and laptops. At the same time, it should be noted that most modern mobile phones sold works on the same principle as tablets. Most of the information in the following text that implicitly assume that a tablet is used can be generalized and applied also to smart mobile phones (for example: prototype testing and verification of most of the principles mentioned in the text were, in addition to the tablets, tested at the same time on the mobile phone Samsung Galaxy S4.).

In general, however, we can draw a conclusion that the end users are increasingly becoming experienced users of devices controlled by the touchscreen. To illustrate the issue, Figure 1 shows the development of the market in each period.

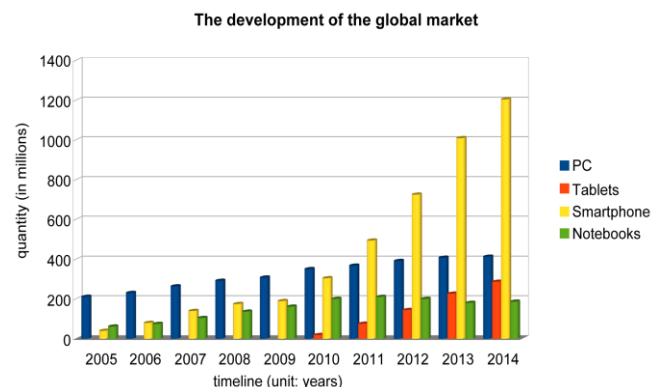


Figure 1: Development of the computer market

### 1.2. Comparison of Tablet PC Platform with IBM PC

If we want to seriously consider using tablets as primary runtime components for distributed simulation, we need to realize the fundamental differences and define the problems that may arise. For more information you can look to Schön (2013), Kovač (2012), Ku (2012), Mocný (2009) or The simulation interoperability standards organization (2001).



### 1.2.1. Performance Comparison

The first characteristic that we naturally focus to is the performance aspect of individual devices. It is certainly a legitimate requirement, but it is important to realize that only for a certain class of tasks. Tablet, PC or IBM PC is not a device that we choose for simulations for their brute force effectiveness, but for the user friendliness. Therefore, this article takes into account that the class of simulation tasks requiring enormous compute performance will run directly on a cluster, or a cloud.

A bit more specifically to the performance, performance of a standard PC processor is about 80 GFLOPS<sup>1</sup>, in contrast to tablets, where the standard performance is about 80 MFLOPS. At first glance, the absolute value says that the performance of the tablet should be approximately 1000 times lower, compared to PC. At this point, it should be stressed that both architectures use completely different processors. A standard IBM PC uses the x86 architecture processors that may appear to be considerably ineffective (information on the very principle of CISC, internal RISC structure and the resulting information can be obtained from [1]). On the contrary, the ARM processors (for more information see [2]) used in Tablet PCs have better transfer of compute performance to instructions applicable in programmes.

The possibility of an illustrative comparison is presented to reader in the following figure.

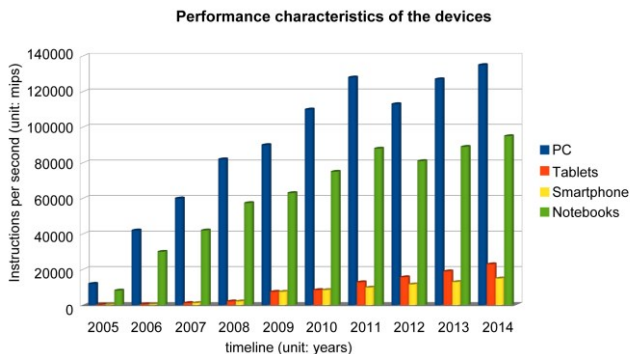


Figure 2: Various theoretical performances

The comparison clearly shows that tablets are not appropriate means for the class of tasks requiring brute force (on the other hand, below, the reader will be acquainted with the fact that even this problem is solvable with relatively easy approaches.)

### 1.2.2. Comparison with Regard to Peripherals

Although this directly begs comparison of the physical peripherals and input-output devices, motivation for comparing peripherals is a bit different. Therefore, let's

<sup>1</sup> Flops is a measure of computer performance, indicating a count of basic floating-point operations per second.

move from the intuitive comparison of user input-output devices to the network communication technologies.

A standard PC network connection is carried out through a cable line, which has a (relatively) high level of reliability. In contrast, a standard tablet will be most often connected during the calculation through a wireless connection, which has a high quality only under ideal conditions - otherwise, there is an increase in latency, or loss of parts of the data. Therefore, the creation of models for tablets is more challenging (on the other hand, it is about as challenging as creating models that are synchronized in, for example, the internet, via web services, etc.).

Wireless communication (either Wi-Fi or Bluetooth, at worst, GSM) has its own merits, which is the portability of the device.

### 1.2.3. Physical Comparison

Since the comparison is very intuitive, just a few words about a strength that we can advantageously use.

Easy portability of tablets allows using them outdoor in support of ad-hoc decision-making (which is the topic of the practical application presented in this article). The risk is the limited battery operating time, but this problem is gradually eliminated. In addition, for certain classes of tasks do not require a long-lasting battery life.

The physical portability implies another major advantage and that is localization - because if we have an active Wi-Fi or GSM receiver, using certain principles (detailed in [3]) we are able to perform accurate localization of the device (this is possible even with GPS, but not in enclosed spaces).

A result of the portability and localization is, for example, that if we create online simulation of factory operation, the tablet will automatically visualize only the part of the plant near to our current location, i.e. the one where a malfunction is being dealt with, for example. Likewise, these principles can be used to collect data from relevant sensors (as the tablet can also serve as a central point for the collection of values collected with the Bluetooth technology and external sensors that support this technology).

### 1.3. Practical Programmer's Perspective

This chapter is rather intended for software engineers, it may not be interesting for other professionals engaged in the simulation. The chapter can be skipped without losing the context of the entire thesis of the article.

For the purpose of implementation, we need to take into account several important criteria, such as the demands of development (particularly in terms of the programming language support), the existence of libraries, support, and size of the community engaged in the development.

In order to discuss these issues, it is important to realize that there are (essentially) two main representatives of tablets, namely Tablet PC (with the most common operating system being Android,

followed by Windows) and the Apple platform – i.e. iPad and iPhone (running on iOS).

### 1.3.1. Development for Android

First, let us focus on (at least in Europe) the most widespread variant, devices that run on Android. The programming language that is used for programming is, de facto, the programming language Java; we just need to use different libraries. The development itself is relatively trivial. Part of external libraries, written in the Java programming language and designed for standard computers, are also applicable in the solutions for Android.

The community of general programmers is relatively broad and focuses mainly on the graphical environment issue, which cannot be considered a problem as most architectural problems is the same as in the standard JAVA.

However, a major disadvantage and a great risk in creating applications for Android is the issue of operating system versions. It is because it happens very often that the things that worked with ten different versions in a standard and predictable way will fail to run with an eleventh version. Therefore, with every new update of the operating system we should re-verify the functionality of the solution. The actual verification is performed though verifying the functionality of each module. It does not happen that the application would not work and return invalid results. The need for testing a new version is still a major inconvenience.

### 1.3.2. Development for iOS

Even for the actual development for iOS it is necessary to have the relevant physical device - emulation is very difficult, practically impossible.

Another major disadvantage is the programming language. It is an Apple company's own language, which is a certain mutation and an extension of the programming language C. For programmers accustomed to standard programming languages (Java, C#, C++), the actual syntax is relatively counterintuitive and we need some time to get accustomed to the language (as well as to the paradigms that the programming follows in that language).

The community is relatively large, but the collective problem solving takes place only to a limited extent. This is probably due to the fact that currently there is a transition to another programming language, so the community is slightly fragmented.

Another major problem is the non-existence of libraries that could be correctly linked to the program that will run on iOS devices, so it would be necessary to create the runtime environment for the simulation virtually from scratch.

### 1.3.3. Development for Windows

A brief mention should also be given to the development for tablets running on the Microsoft Windows operating system. The development is performed in a standard way that developers are used to,

so it will not be discussed in the article, for more information you can go to [4].

A moderate risk of this solution is the fact that it is not that easy to disseminate applications or to network owing to the intricate settings and security policies in this operating system. Thus, even though the programming is simpler, we can often encounter various types of runtime problems caused by the settings.

### 1.3.4. Recommendation

The results of our work show that it is preferable for the simulation to use devices that run on Android OS or Windows. With the experience, we cannot recommend creating simulators for iOS.

## 2. WAYS TO USE TABLETS

The actual tablet can be used in simulations in several different ways, some of which are very intuitive, while some on the contrary deserve more attention (for example, due to their considerable development potential).

### 2.1. Used to Run the Monolithic Simulation

Currently, this method of use is purely a marginal issue. The performance is still insufficient and, even with calculations running on graphics accelerators, tablets experience memory issues. Therefore, the simulation thus triggered would be relatively undemanding and it would be rather a trivial demonstrator without much practical use.

On the other hand, it can be expected that with the increasing computer performance of tablets and modifications in the architecture of processors (number of kernels, higher parallelisms on graphics accelerators), even this approach will be broadly applied in the future.

### 2.2. Distributed Simulations solely on Tablets

Distributed simulations solely on tablets can be run in two different modes, namely:

#### 2.2.1. Based on the Sharing of Performance

If we have a certain amount of tablet users who have access to the network and enabled a specific application, it is possible to run a computationally oriented simulation with a part of distributed computing running in the background of each tablet (even more logical processes in one tablet).

This solution is especially useful for calculations with a high degree of parallelism and it is an alternative to the solution in the cloud. The advantage, however, is advanced debugging options of the application.

#### 2.2.2. Standard Distributed Simulation

In the case of this mode, there is just one logical process running on each tablet, in the foreground of the simulation. Most commonly the method is used in an application of the simulator/trainer type. Each tablet thus serves not only as a computational node for the simulator, but also as an input-output device, which can

be used to parameterize the simulation calculation directly during its run.

### **2.3. Combined Operation of Distributed Simulations**

It is the combined (according to some sources, hybrid) methods of the operation of distributed simulation that currently have the greatest potential for practical use. Specific usable solutions are listed below.

What all solutions have in common is that part of the distributed computation is performed on standard computers, servers, or even in the cloud (it is with the cloud solution that it achieves the best results). The combination of the computing performance of more robust computers with the advantages of tablets (portability, intuitiveness) is also used in the demonstrator presented in this paper in Chapter 3.

Only four basic methodologies are listed, which can be combined.

#### **2.3.1. Used as a Visualizer**

Simple use the tablet as a visualizer is beneficial for those classes of problems where it is desirable to animate the simulation calculation. This has practical applications rather for presentations of the simulation process to a customer and for validation of processes, where a simulation specialist consults realism of the established processes with an expert from the modelled field.

#### **2.4. Used as a Driver**

If we want to use the tablet as a distributed simulation driver, we just need to create such a simulator that will be able to parameterize the conditions prior to the start of the simulation, and then also the conditions for termination of the simulation run.

The solution is useful in situations where it is necessary to parameterize the simulation according to the information that can be obtained only in the field, i.e. beyond the standard stationary device running the simulation.

The ideal is to use this principle in decision support systems, where we can relatively easily parameterize the simulator, without being limited by our physical location.

#### **2.4.1. Used as an Output Device**

The second possibility is to use a tablet as an output device. Since it is not necessary to perform stable synchronization as in the case of online animation (2.3.1), this solution is relatively trivial. Like the previous approach, this solution is advantageous for decision support systems. Users are often not interested in the simulation, but only in the result.

#### **2.4.2. Used as an Input Simulator**

The most demanding method of using the tablet occurs when there is a request to implement the DIS methods (Distributed Interactive Simulation, IEEE1278, more in [8]).

It is especially challenging to secure such transmission problems in the network in order to make sure that all interactive interventions have been performed correctly and at the right time. Synchronization methodology is equally challenging. It should be noted, however, that despite some complications in the development we can achieve excellent results due to the high user-friendliness of such simulators.

### **2.5. Tablet Used as a Preprocessor for Data Processing**

Being a portable computer that has a number of wireless connections (Bluetooth and Wi-Fi), the tablet can be advantageously used as a preprocessor for data processing. It is thus possible to obtain relatively cheaply a complex measuring station. If we have a number of sensors or limit sensors that support the Bluetooth technology, it is possible to retrieve data from these sensors through the tablet and process it into such a form that is required for distributed simulation. The actual distributed simulation is then connected to this node.

## **3. DEMONSTRATION SOLUTIONS**

### **3.1. Description of the Problem**

The chemical metal coating plant is characterized by several facts. The whole plant is automated (human labour is only at the beginning of the process to place the products to be coated on hooks and then during the storage process). The high degree of automation allows the factory to process a large number of products. The factory itself contains several lines for different types of chemical metal coating (bluing, electrolytic and galvanic lines), while the actual chemical bath are subject to change. It is also true that one chemical bath can be used for a number of metal coating lines (i.e. if the chemical bath has a size of 20 x 4 x 2 meters, it can be used for 4 metal coating lines and each line can process different types of products with the same surface finish).

The problem of the plant, however, is that due to the chemicals and enormous currents (thousands A at 3-5 V) there are frequent malfunctions on the lines. The chance that a fault occurs in one line, in any shift is up to 7 percent (up to 14 percent with lines fully loaded). If the factory has 40 lines, it is a relatively high number.

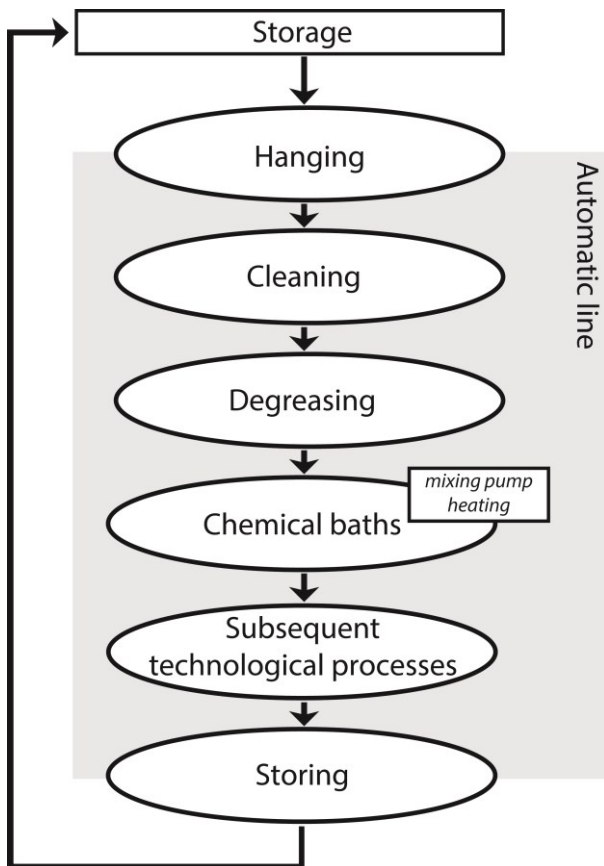


Figure 1: Chemical Plant Line Diagram

### 3.2. Motivation for Solving Simulations

If there is a malfunction, is it better to stop the operation and remove the error, is it advisable to have the production completed, or is it ideal to switch the production to another line? And what happens with the products? Will there not be unreasonable queues? Will there not be a plethora of rejects? Alternatively, will the products degrade that have gone through the preparatory process and are still waiting for a galvanic bath?

For such complex problems encountered analytical methods to their limits, and better way is to use simulation, as is noted by Manling (2009).

At present, all the problems that do not require acute repair are postponed until the production is completed and only then they are dealt with. However, the problem is that some of these errors reduce the efficiency of the process (type restrictions of the functionality of some pumps for mixing the baths) when it is necessary to slow the production by about 20 percent and there is an increased risk of poorly metallised parts.

By applying online simulation (i.e. simulation that uses actual operating data - statuses of the input queues, warehouse and individual lines - but from the start, the simulation runs with a time base much faster than the real time) it is possible to evaluate the best solution method immediately after detecting a problem.

It is because if we know the malfunction, due to the frequent repetition of identical malfunctions we know the time needed for the repair. Based on the parameters

of the line (temperature, type, degree of metal coating, input queue, availability of other lines), the simulation can then be used to create an ideal scenario for the repair (so far, service engineers have usually not been able to evaluate all parameters only by expert judgment, as there are many variable parameters).

The simulation then can recommend redirecting the production to another line for the time need for repair (this solution minimizes the loss of unprocessed elements that have already undergone preparatory processes), suspending the line production and repairing (there will be an increase in the line capacity, if there are large input queues), or recommend continuing the production without any repair (with small input queues, or expected change in the production).

By including the simulation model to support decision making in the work of service engineers, we can increase efficiency of the total production under standard conditions by up to 2-5 percent (in excess of orders and with busy input warehouses and all lines, it is possible to increase efficiency and production throughput 7-10 percent).

### 3.3. Physical Simulation

The actual simulation is distributed and usually uses 4 types of nodes.

- The computing node is located on a powerful computer that is designed just for this single activity,
- The node for parameterizing, defining scenarios, launching the simulation and displaying the results is running on the tablet so that the operator can use any device in any production hall, without having to go back to the service centre once a malfunction has been detected
- The node for data collection is connected to the main automation computer and receives online data from the production,
- The node for obtaining historical data is placed in the company's data warehouse and is used to collect historical data on production processes, repairs, etc.

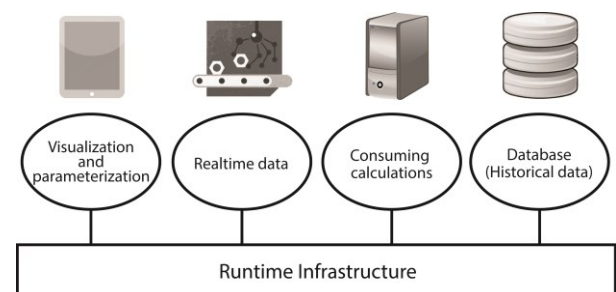


Figure 2: Architecture of Simulation Model

Individual nodes can run in the simulation in a larger number (e.g. two independent nodes for

parameterization running in tablets are a natural part of the solution).

Anyway, the demonstration example is a great example of how a distributed simulation using tablets can effectively support decision making and thus improve the efficiency of industrial production.

### 3.4. Software Solution

Since it is a distributed solution that is to run on heterogeneous hardware platforms and it was assumed that it would be necessary to program various parts of the system in various programming languages (e.g.: the tablet requires the JAVA language, the data acquisition system in the automation computer requires the C programming language coding), the use of HLA seemed to be ideal as it has its internal processes adapted to all these contingencies.

For more information on HLA, its principles and potential usage, we recommend to consult Fujimoto (2000), Robelo, Sala-Diakanda and col. (2013), Kuhl, Dahmann, Weatherly (2000) or IEEE1516:2010 standard (2010).

The software architecture thus uses the central computer not just to run the simulation calculations, but also to run RTI.

It uses synchronization provided directly by HLA - specifically its conservative variant.

## 4. POTENTIAL OF FUTURE DEVELOPMENT

Currently, the prototype works only for a limited class of malfunctions – the article seeks to present an idea that the systems can deal with using tablets.

It is desirable to extend the system itself with the capability of automatic detection and conjecture of malfunctions. Ideally also with neural networks to deliver the learning ability so that the system is better able to work with various malfunctions, their diagnostics and solutions and able to create scenarios for malfunctions of a new type.

In general it may be said that further development is rather a low-level issue, and it does not change anything significant on the very ideas and principles of using combined simulations and tablets for decision support.

## REFERENCES

- Fujimoto, Richard M., c2000, *Parallel and distributed simulation systems*. New York; John Wiley & Sons. ISBN 04-711-8383-0.
- Kovač, Pavel, 2012. Reflection: *Where are we going tablets?*. Available from: <http://www.svethardware.cz/uvaha-kam-smeruji-tablety/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/thinkpad-x230t-review-benchmark,3229-2.html> accessed 15 July 2014].
- Kuhl, Frederick, Dahmann, Judith, Weatherly, Richard, *Creating Computer Simulation Systems: An*

*Introduction to the High Level Architecture*, c2000, Upper Saddle River, NJ; Prentice Hall PTR. ISBN 01-302-2511-8.

- 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.
- 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].
- 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-pro-pc-i-iphone-a-android> [accessed 15 July 2014].
- 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.



# SPEEDSIM.NET - AN OPEN .NET BASED SIMULATION SYSTEM

Wiedemann, Thomas <sup>(a)</sup>, Wendt, Karsten <sup>(b)</sup>

<sup>(a)</sup> University of Applied Science Dresden

<sup>(b)</sup> DUALIS GmbH IT Solution

<sup>(a)</sup> [wiedem@informatik.htw-dresden.de](mailto:wiedem@informatik.htw-dresden.de), <sup>(b)</sup> [KWendt@dualis-it.de](mailto:KWendt@dualis-it.de)

## ABSTRACT

As a result of the contradictory characteristics of existing simulation systems, there is a gap between component based simulators and programming language based simulators. This paper will discuss possible approaches and tool candidates for filling this gap.

The presented approach and the developed simulation system SpeedSim.Net are flexible and powerful options for defining future, complex simulation environments. The system can be used in a standalone mode inside a larger software system or together with a predefined GUI-framework like a traditional simulation system. All components are managed by tools of the Visual Studio, like the property editor or tree list dialog boxes, so experienced VS-developers can start modeling very fast.

Interested simulation experts can participate in the future development of the tool.

Keywords: .NET component based standalone

## 1. INTRODUCTION

The main algorithms and mathematical foundations of discrete simulation systems are well defined and efficient (Wiedewitsch and Heusmann 1995) (Wiedemann, T., 2002), (Zeigler, B. P. (1990). Nevertheless, the real application of simulation systems is still difficult (Kuljis and Paul 2000), because there is no optimal system. Especially the modelling comfort and the performance and software flexibility are contradictory, comparing different levels of existing types of discrete simulation tools (see colored triangles in fig. 1).

As a result of the contradictory characteristics of the simulation systems, there is a gap between component based simulators and programming language based simulators. This paper will discuss possible candidates for filling this gap. If such an optimal system could be defined, it should combine all the good characteristics of the different system classes (see right site of fig. 1).

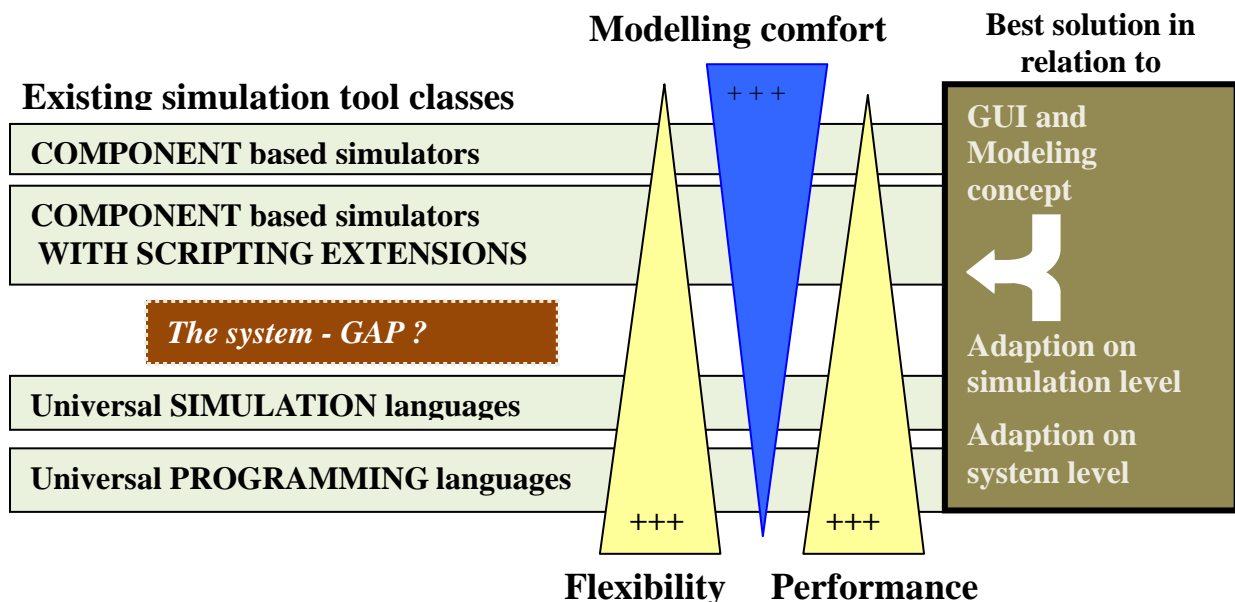


Figure 1: The actual system classes in discrete event simulation and their main characteristics



## 2. DEFINITION OF MAIN SYSTEM REQUIREMENTS

Based on the discussed characteristics of existing discrete simulation systems, we define the following requirements for an optimal discrete simulation system:

1. Use a standard programming language and its integrated developing environment (IDE) for modeling and simulation control.
2. Provide a comfortable and affordable modeling environment similar to modern component based simulation systems.
3. Use the existing standard programming language(s) for scripting at the model level.
4. Provide high simulation performance for complex simulation scenarios and long optimization cycles.
5. Apply sophisticated software technologies for defining simulation components, basic simulation functions like random number generators and statistical modeling and analysis tools (autofit tools, empirical function definitions).
6. Use databases for storing all models and model input and output data.
7. Provide state-of-the-art interfaces for data import and export and for system integration and communication.

This “wish list” will never be complete and should add new software technologies like web-based standards such as HTML5 for user interfaces or web services for system integration and data exchange.

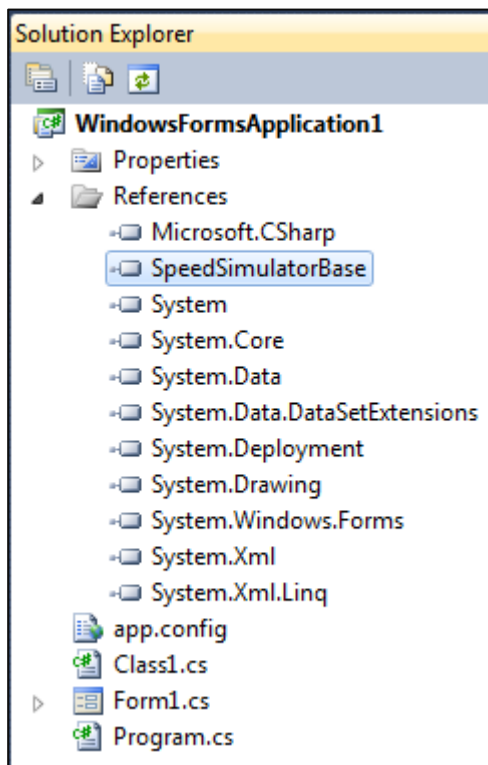


Figure 2 The software libraries

## 3. DISCUSSION OF IMPLEMENTATION OPTIONS

### 3.1. Selection of the IDE

Comparing all existing integrated developing environments, there are two main options: the Visual Studio by Microsoft (VS2013) and the Eclipse IDE from the Eclipse Foundation (Eclipse Foundation 2013). Both options have their advantages and disadvantages:

- Microsoft’s Visual Studio is a very complex environment based on the .NET-architecture and supports all relevant IT-technologies in the desktop and web area. The main problem is the focus on Microsoft operating systems, which limits the application area. But if the customer or end-user uses MS-OS, it is a very good solution.
- The Eclipse IDE is based on Java and runs on any operating system, especially on Unix- and Linux-based systems. If a customer runs such OS, Eclipse is the only solution.

Both environments do not generate native code for the processor but some Intermediate code (MSIL / Java Byte Code). In general, this is a disadvantage under performance considerations, but the actual state of the art of code generators and the optimized code interpreters at the execution level allow execution rates at the level of native code. In some special cases the execution speed of the MSIL code was higher than the speed of native code. In conclusion, the intermediate code is no longer seen as a performance issue, but as a good option for migrating the compiled simulation model to different platforms.

The question of the overall development speed with the IDE’s was tested with two groups of students, working with Visual Studio and the Eclipse IDE on a standard Database oriented web-application. It was impressive that the Visual Studio group could finish the task in only 25% of the time and with a better end user quality than the Eclipse group. The main reason for this huge difference was the high quality of supporting solution assistants and a seamless integration in the Microsoft webserver environments. The Eclipse group finished the main Java programming in nearly the same time, but then lost a lot of time with configuration issues and run-time errors in the web server environment.

As a result of this development test and already existing .NET code at the partner software company DUALIS, the Visual Studio option was selected. In general, the main approach, which will be shown on the following pages, could also be migrated to the Eclipse IDE. In Eclipse, there would be the option for using Java for the GUI, but C++ should be used in the simulation core to get a good performance and to reuse the code from the Visual Studio version (Skeet 2013).

### 3.2. Selection of programming and language

Taking the selection of Visual studio (VS) into account and comparing all possible .NET programming languages for their usability, flexibility and user

acceptance, C# was the logical solution. The old C / C++ were used in an older version of the system, but they are no longer supported by the managed code system of Microsoft and there are increasing limitations and problems (Skeet 2013)..

### 3.3. Standalone user interface

Using the VS IDE as a basic modeling tool, a first simple .NET-like modeling is possible. All the simulation knowledge and functionality is provided by one DLL with a number of predefined simulation classes and additional supporting libraries (see fig. 2). The simulation model is created as a standard .NET-Windows form. The GUI-buttons call the simulation control methods and simulation results are displayed in the log field (see fig. 3).

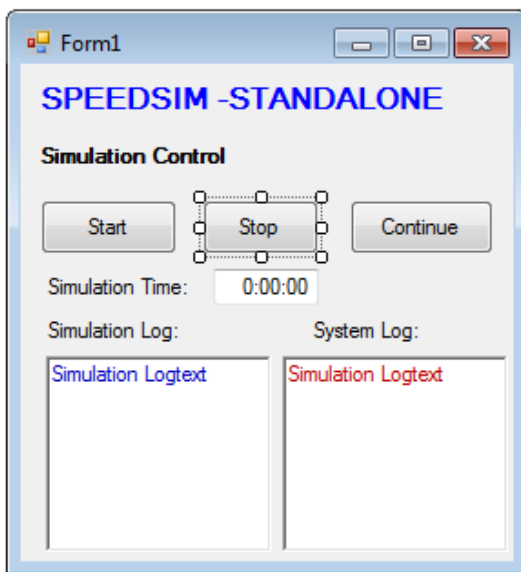


Figure 3 A simple Stand-alone simulation model in Visual Studio

The source code behind the form window of a very simple simulation model is shown in fig. 5. After defining the model, the simulation is started by calling the `_Model.ResumeSimulation();` method with the start button from the user interface.

The standalone modeling option is a good solution for small and quick simulation experiments. It takes about 15 minutes for experienced VS-developers to build a running .NET simulation model.

Further types of applications are highly integrated simulation models. The code of the model from fig. 3 could be migrated to any .NET application, like ERP-backend interfaces or business intelligence tools. By using web based technologies like web services, the input data for the simulation run and the simulation results can be distributed over the web.

### 3.4. A traditional desktop user interface

Simulation users without advanced programming knowledge will encounter problems with the programming orientated approach. For this group of

users the system offers a second method of designing and manipulating models.

To meet specific simulation needs, an experienced VS-programmer will prepare a set of simulation components. The resulting component library is compiled as a DLL file and then a precompiled complex user interface – the SpeedSim.Net framework - is started. The simulation user will load the component library into the SpeedSim.Net framework. Afterwards, he can design and configure complex simulation models like in traditional simulation systems (see fig. 4)

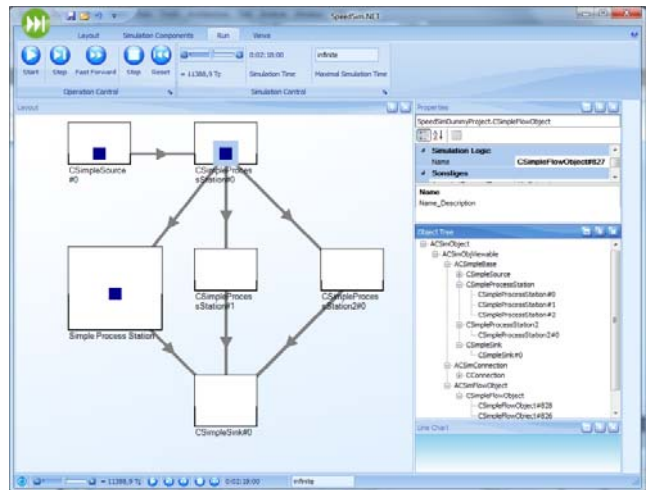


Fig. 4: The SpeedSim.Net user interface framework



Fig. 5: The GUI-view of a loaded component library from the user component DLL

```
public Form1()
{
 InitializeComponent(); // Init Form
 // Define Model for Simulation
 model1=new SpeedSimulatorBase.Simulation.CSimModel();
 model1.Initialize();
 // Register SimObjects in Model
 sim1object1 = new mySimObject();
 sim1object1.Initialise(model1,true,false);
 sim1object2 = new mySimObject2();
 sim1object2.Initialise(model1,true,false);
 // Define Simulation
 simulation1 = new CSimulation();
 simulation1.Initialize(model1);#
 // Define Simulation Duration
 simulation1.MaxSimTime = 100;
 simulation1.SecondsPerTick = 0.01;
}
```

Figure 6: The source of a stand alone model

All model components can be managed by the well-known Visual Studio property editor (see upper dialog box of fig. 7). Newly designed methods of the simulation components are automatically part of the property list (see properties “MaxContent” in fig. 7).

During the simulation all dynamic objects like products or other moving objects are shown in the dynamic object tree list (see lower dialog box of fig. 7). The values of any dynamic object can be displayed and also changed during runtime.

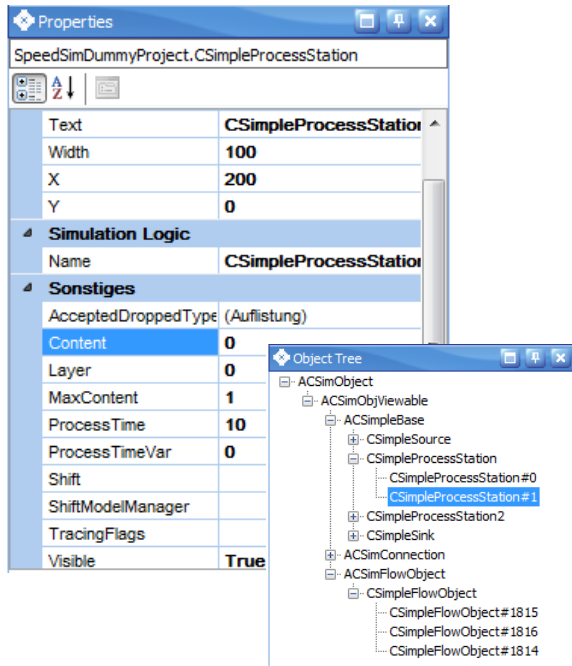


Fig. 7: The VS-based property editor and the list of dynamic object during simulation runtime

Results are written in external files and can be displayed also during run time in different chart windows, which are interactive (see fig. 8)

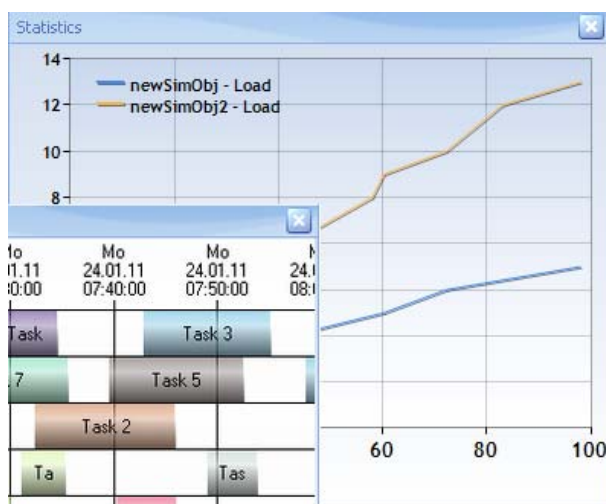


Fig. 8: Visualization of run-time statistics and results

## 4. THE DEVELOPERS VIEW

### 4.1 The SpeedSim.Net OBJECT model

The full object model of the developed simulator is shown in fig. 8. Beside the core modules for building the simulation model, there are a lot of other modules for simulation control and interfacing (see fig. 9).

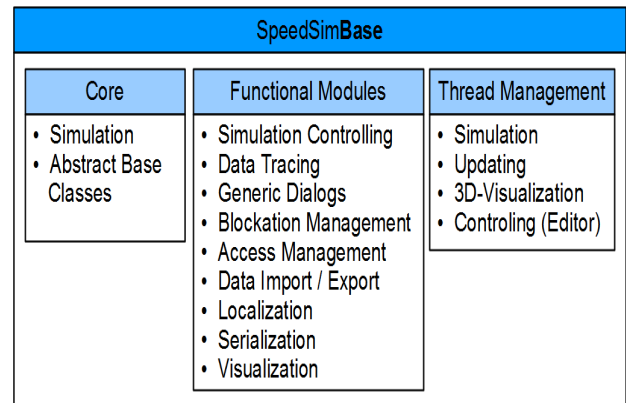


Fig. 9: The S SpeedSim.Net Object Model and basic system architecture (from (Wendt 2013))

This basic system architecture is concentrated in a single DLL. By uploading this DLL to a Visual Studio environment (VS), the VS becomes a simulation environment.

Customer specific changes to the basic algorithm could be added in two ways:

- Single changes of parameter values can be done with the SpeedSim.Net editor according to fig. 7.
- Much more complex and long-term modifications can be concentrated in customer specific libraries (see Fig. 10). It will be possible to build up a system of application specific libraries e.g. for logistics, IT-network or traffic models.

It should be mentioned again, that all changes of model or simulation algorithm will be done with one of the .NET-programming languages and the resulting project will be still a standard .NET-project! This allows the integration of all existing and future software technologies from Microsoft, like Web Services, SOAP--and database interfaces to Data Mining and Business Intelligence tools.

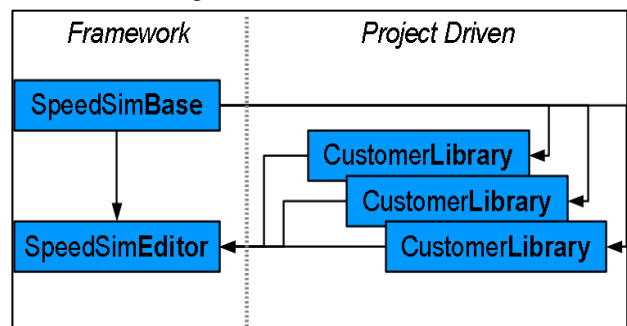


Fig. 10: Framework and customer driven project linnaries (from Wendt, 2013)

The simulation object life cycle is shown in fig. 11. After the setup phase the object state is scheduled between the active states, like “Action” or “Event handler”, and the delay state, where the object is waiting in the future event list of the simulation control for the correct simulation time to be activated again. It is very important that all object states and their corresponding event handlers can be changed by the core or model developer. Again, changes of the algorithms could be saved in the model specific project or if useful in general also in the customer specific libraries.

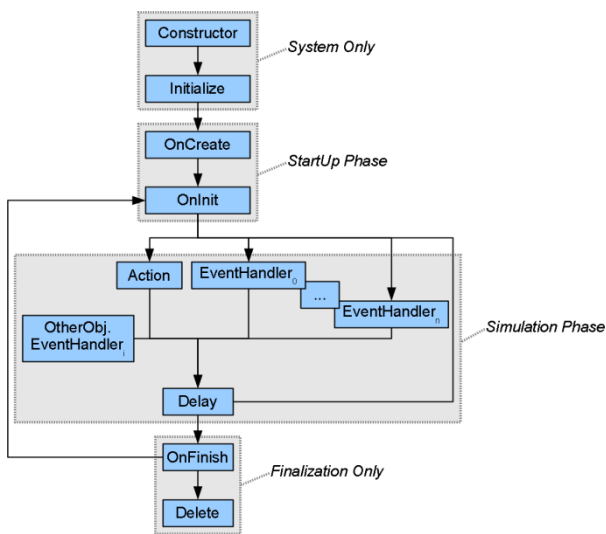


Fig. 11: The simulation object life cycle (Wendt 2013))

#### 4.2 Optimization of visualization and animation

The SpeedSim.Net framework for non-programmers supports a multi-threaded module architecture and will also provide 3D visualization options in the future.

In order to optimize simulation speed, not all simulation events are displayed in the front end view.

Time measurements by the authors show that the time needed for displaying a 5 digit number on the screen is about 1000..2000 longer than the time for any single calculation for getting this number.

In conclusion, there should also be the option to switch off nearly all visualization options in order to achieve maximum speed. The decision of the visualization level is not coded in the simulation core, but is defined on a multi-level control structure (see Fig. 12.) The different levels of this message chain can also be changed for non- desktop purposes.

If the simulation core is running as a standalone program with no visible windows (or only an icon in the system tray), all user input and simulation output could be managed by web based technologies or web services.

Therefore, a remote simulation service is possible, which could be integrated in larger IT-environments like ERP- or business intelligence solutions.

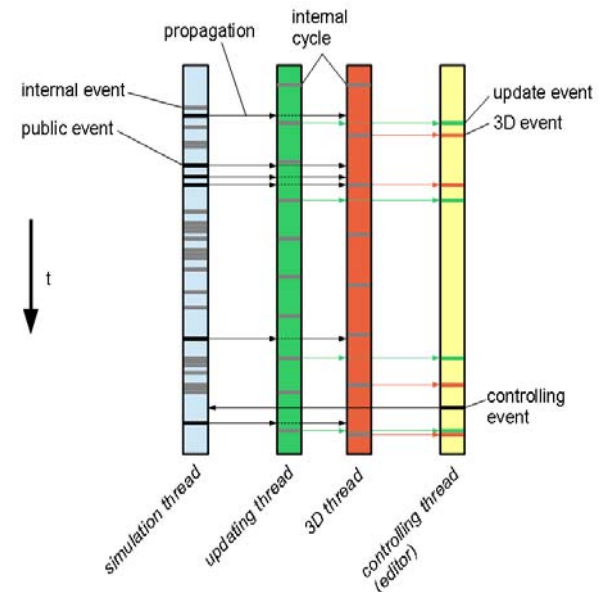


Fig. 12: Thread- and visualization Management

#### 5. ACTUAL STATE AND DEVELOPMENT

The shown modeling options – standalone and framework based, are fully operational and could be used by experienced customers.

The system was already introduced in a master course “Discrete events simulation” at the University of Applied Sciences Dresden. The simulation with SpeedSim.Net was introduced to computer science students with one 90 minutes lecture and one double hour of student exercises. After some small, predefined example projects the students were able to successfully model and simulate a model with 6 model objects.

The actual work focusses on defining complex libraries for customer oriented production models and other typical application areas such as network and transport simulations.

Interested simulation experts can participate in the future development of the tool. Although the development is not yet finished, the shown approach and the existing tools could fill the gap in the actual hierarchy of simulation tools (see fig. 13).



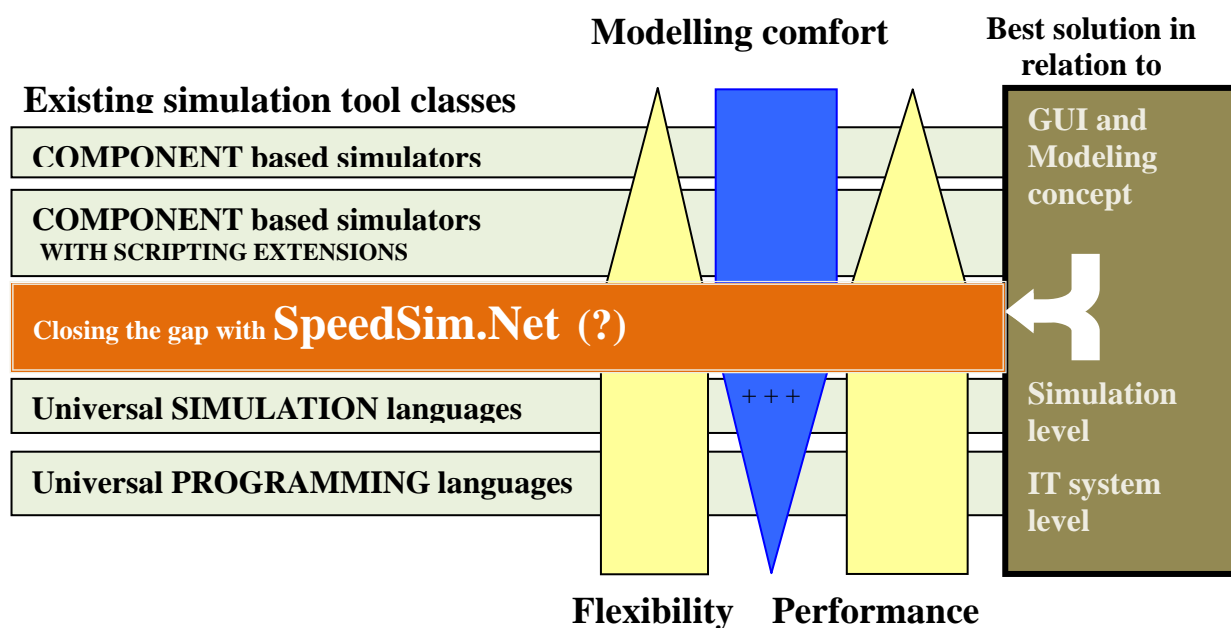


Figure 13: SpeedSim.Net as a candidate for closing the gap in <sup>+++</sup> in discrete event simulation

## REFERENCES

- Kuljis, Jasna and Ray J. Paul, 2000: A Review of web based simulation: whiter we wander?, *Proceedings of the 2000 Winter Simulation Conference*, Orlando Florida, page 1872-1881
- Schriber T., Brunner D., Smith J.: Inside Discrete Event Simulation Software: How It Works and Why It Matters. *Proceedings of the 2013 Winter Simulation Conference*, Pages 424-438
- Skeet, Jon: *C# in Depth*. Manning Publications; 3rd revised edition. 2013
- Wendt, K. (2013) : SPEEDSIM Documentation Version 1.3. Dualis IT Solutions GmbH Dresden, 2013
- Wiedemann, T., 2002. Next generation simulation environments founded on open source software and XML-based standard interfaces, *Proceedings of the 2002 Winter Simulation Conference*
- Wiedewitsch J. and Heusmann J. 1995. "Future Directions of Modeling and Simulation in the Department of Defense", *Proceedings of the SCSC'95*, Ottawa, Ontario, Canada, July 34-26, 1995
- Zeigler, B. P. . Object-oriented simulation with hierarchical, modular models. Academic Press, Boston. 1990

## AUTHOR BIOGRAPHIES

**THOMAS WIEDEMANN** is a professor at the Department of Computer Science at the University of Applied Science Dresden (HTWD). He has finished a study at the Technical University Sofia and a Ph.D. study at the Humboldt-University of Berlin. His research interests include simulation methodology, tools and environments in distributed simulation and manufacturing processes, intranet solutions and database applications.

Email : [wiedem@informatik.htw-dresden.de](mailto:wiedem@informatik.htw-dresden.de)

**KARSTEN WENDT**, born 1980 in Dresden; 2007 diploma in Information System Technology at the TU Dresden; 2007, is -now a scientist and PhD student at the Chair of Highly-Parallel VLSI-Systems and Neuromorphic Circuits, Faculty of Electrical Engineering, TU Dresden, implementing operating systems and multi-criteria cluster algorithms for large-scale neuromorphic hardware; 2008-now senior developer at the Dualis GmbH IT Solution Dresden, responsible for Speedsim.Net

Email : <sup>b)</sup> [KWendt@dualis-it.de](mailto:KWendt@dualis-it.de)

# APPLYING AN ADAPTIVE PETRI NET TO CALCULATE THE MAKESPAN IN THE JOB SHOP SCHEDULING PROBLEM

Joselito Medina-Marin<sup>(a)</sup>, Juan Carlos Seck-Tuoh-Mora<sup>(b)</sup>, Norberto Hernandez-Romero<sup>(c)</sup>,  
Nayeli Jazmin Escamilla-Serna<sup>(d)</sup>

<sup>(a,b,c,d)</sup>Autonomous University of Hidalgo State, Advanced Research Centre in Industrial Engineering,  
Pachuca de Soto, Hidalgo, México

<sup>(a)</sup>[jmedina@uaeh.edu.mx](mailto:jmedina@uaeh.edu.mx), <sup>(b)</sup>[jseck@uaeh.edu.mx](mailto:jseck@uaeh.edu.mx), <sup>(c)</sup>[nhromero@uaeh.edu.mx](mailto:nhromero@uaeh.edu.mx), <sup>(d)</sup>[nj\\_esser@hotmail.com](mailto:nj_esser@hotmail.com)

## ABSTRACT

The Job Shop Scheduling Problem (JSSP) is one of the typical problems that engineers face on designing Flexible Manufacturing Systems (FMS). In this problem, it is important to find the optimal scheduling to perform a set of tasks in the minimum time. Moreover, the JSSP has some restrictions, such as the tasks order and the number of shared resources where the tasks are carried out. To find the optimal tasks sequence it is necessary to obtain the makespan for each sequence. On this way, FMS can be modeled with Petri nets (PNs), which are a powerful tool that have been used to model and analyze discrete event systems. So, the JSSP can be analyzed in a PN representation of the FMS, and the makespan can be calculated by using the PN model. In this work we propose an adaptive PN to obtain the makespan by applying PN analytical tools.

Keywords: job-shop scheduling problem, makespan, Petri nets, state equation.

## 1. INTRODUCTION

A Flexible Manufacturing System (FMS) is a discrete event dynamic system that is composed by jobs and shared resources (Zhou and Venkatesh 1999). The typical problem that engineers faced when they are either designing a Flexible Manufacturing System or planning the master production plan for the FMS, is how they should make the best sequence of jobs in the FMS in order to carry all operations out in the minimum time (Pinedo 2012; Lenstra 1977).

This problem is called the Job Shop Scheduling Problem (JSSP), which is a combinatorial problem classified as NP-Complete (Lenstra et al., 1977). There have been published several research papers about finding the minimum value of makespan in the JSSP (Shuai, and Zhi-Hua 2014; Qing-dao-er-ji, and Wang 2012; Zhao, Zhang, and Bing 2011; Ardakan, Hakimian, and Rezvan 2014). The makespan is the time that all the jobs are processed in the FMS, and it depends on the order that all the tasks are performed.

Several exact methods have been analyzed to find the minimum value for the makespan (Dey, Sarkar and

Basu 2010; Wang, Cai, and Feng 2010; Wang and Zou 2003). These methods such as branch and bound, linear programming and Lagrangian relaxation can find the global minimum value, however for problems with a bigger number of resources and jobs they need a huge amount of computational time to have the final result.

On the other hand, there are also research papers that apply meta-heuristics and/or evolutionary computing to find the minimum makespan time (Qing-dao-er-ji, and Wang 2012; Zhao, Zhang, and Bing 2011). In this case, these proposals can find reasonable results in less time than exact methods. The main drawback of these methods is that the global minimum could not be found, but good approximations are obtained in a short time.

FMSs have been modeled via Petri Nets (PNs) in order to simulate and analyze them. PN theory is adequate to represent in a graphical and mathematical way Discrete Event Systems (DES) such as FMSs, because their dynamic behavior based on event occurrence can be modeled by PN elements (places and transitions) (Murata 1989; Zhou and Venkatesh 1999). Moreover, PN theory offers analytical tools to study the modelled systems, based on the relationship among the FMS resources denoted as PN elements.

One important point in search methods is the calculus of the makespan, taking into account a certain processing order of the tasks. In this paper, we propose the use of an adaptive PN to calculate the makespan by means of the PN state equation.

## 2. JOB SHOP SCHEDULLING PROBLEM

Scheduling tasks in a FMS is a typical combinatorial problem where it is needed to organize the processing of a set of jobs divided in operations, and each operation is carried out in a shared resource (Gonzalez-Hernandez 2011; Quen-dao-er-ji and Wang 2012).

In the JSSP there are  $n$  jobs, and each job consists of  $m$  operations, and each operation is processed in a shared resource or machine during a fixed time. Some restrictions should be considered: operations of the same job have a sequence established previously, a job can visit each machine only once, each machine can



process only one job at any time, and there are not restrictions about the precedence among operations of different jobs.

The aim of JSSP is to find a sequence order for operation processing with the minimum value for the makespan.

For instance, Table 1 shows a FMS taken from (Zhang 2010) with some modifications to respect the JSSP restrictions. This example has three machines, four jobs, and each job has three serial operations.

Table 1: FMS configuration with operation times.

| Items               | Jobs  | Operation Serial Number |                |                |
|---------------------|-------|-------------------------|----------------|----------------|
|                     |       | 1                       | 2              | 3              |
| Machine utilization | $J_1$ | $M_3$                   | $M_1$          | $M_2$          |
|                     | $J_2$ | $M_2$                   | $M_3$          | $M_1$          |
|                     | $J_3$ | $M_3$                   | $M_1$          | $M_2$          |
|                     | $J_4$ | $M_2$                   | $M_1$          | $M_3$          |
| Operation time      | $J_1$ | $O_{1,1,3} 96$          | $O_{1,2,1} 90$ | $O_{1,3,2} 35$ |
|                     | $J_2$ | $O_{2,1,2} 74$          | $O_{2,2,3} 57$ | $O_{2,3,1} 91$ |
|                     | $J_3$ | $O_{3,1,3} 13$          | $O_{3,2,1} 5$  | $O_{3,3,2} 7$  |
|                     | $J_4$ | $O_{4,1,2} 71$          | $O_{4,2,1} 23$ | $O_{4,3,3} 38$ |

Where  $O_{i,j,k}$  denotes the  $j$ -th operation of the  $i$ -th job to be carried out by the  $k$ -th machine.

### 3. PETRI NETS CONCEPTS

A PN is a graphical and mathematical tool that has been used to model concurrent, asynchronous, distributed, parallel, non-deterministic, and/or stochastic systems.

The graph of a PN is directed, with weights in their arcs, and bipartite, whose nodes are of two types: *places* and *transitions*. Graphically, places are depicted as circles and transition as boxes or bars. PN arcs connect places to transitions or transition to places; it is not permissible to connect nodes of the same type. The state of the system is denoted in PN by the use of *tokens*, which are assigned to place nodes.

A formal definition of a PN is presented in table 2 (Murata 1989).

Table 2: Formal definition of a PN

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>A Petri net is a 5-tuple, <math>PN = (P, T, F, W, M_0)</math> where:<br/> <math>P = \{p_1, p_2, \dots, p_m\}</math> is a finite set of places,<br/> <math>T = \{t_1, t_2, \dots, t_n\}</math> is a finite set of transitions,<br/> <math>F \subseteq \{P \times T\} \cup \{T \times P\}</math> is a set of arcs,<br/> <math>W = F \rightarrow \{1, 2, 3, \dots\}</math> is a weight function,<br/> <math>M_0 = P \rightarrow \{0, 1, 2, 3, \dots\}</math> is the initial marking,<br/> <math>P \cap T = \emptyset</math> and <math>P \cup T \neq \emptyset</math>.</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

The token movement through the PN represents the dynamical behaviour of the system. In order to change the token position, the following transition firing rule is used (Murata 1989):

1. A transition  $t \in T$  is enabled if every input place  $p \in P$  of  $t$  has  $w(p,t)$  tokens or more.  $w(p,t)$  is the weight of the arc from  $p$  to  $t$ .

2. An enabled transition  $t$  will fire if the event represented by  $t$  takes place.
3. When an enabled transition  $t$  fires,  $w(p,t)$  tokens are removed from every input place  $p$  of  $t$  and  $w(t,p)$  tokens are added to every output place  $p$  of  $t$ .  $w(t,p)$  is the weight of the arc from  $t$  to  $p$ .

A Timed Place Petri Nets (TPPN) is an extended PN, where a new element is added. It is a six-tuple  $TPPN = (P, T, F, W, M_0, D)$ , where the first fifth elements are similar to PN definition presented above, and  $D = \{d_1, d_2, \dots, d_m\}$  denotes the time-delay for each place  $p_j \in P$  (Zhao, Zhang, and Bing 2011). Output transitions  $t_i$  for each  $p_j$  will be enabled once the time indicated in  $p_j$  is reached.

### 3.1. Analysis methods

In this paper, we are applying the matrix equation approach as the analytical method of PN theory in order to calculate de makespan of the FMS modelled.

#### 3.1.1. Incidence matrix and state equation

A PN with  $n$  transitions and  $m$  places can be expressed mathematically as an  $n \times m$  matrix of integers  $A = [a_{ij}]$ . The values for each element of the matrix are given by:  $a_{ij} = a_{ij}^+ - a_{ij}^-$ , where  $a_{ij}^+$  is the weight of the arc from  $t_i$  to  $p_j$ , and  $a_{ij}^-$  is the weight of the arc from  $p_j$  to  $t_i$ .

The state equation is used to determine the marking of a PN after a transition firing, and it can be written as follows:

$$M_k = M_{k-1} \times A^T U_k, k=1,2,\dots \quad (1)$$

where  $u_k$  is a  $n \times 1$  column vector of  $n - 1$  zeros and one nonzero entries, which represents the transition  $t_j$  that will fire. The nonzero entry is located in the position  $j$  of  $u_k$ .  $A^T$  is the transpose of incidence matrix.  $M_{k-1}$  is the marking before the firing of  $t_j$ . And  $M_k$  is the reached marking after the firing of  $t_j$  denoted in  $u_k$ .

### 4. ADAPTIVE TIMED PLACE PETRI NET

In this paper we propose an adaptive TPPN (ATPPN), which adds some arcs according to tasks sequence of the FMS.

The formal definition of an ATPNN is as follows: An ATPNN is a seven-tuple  $(P, T, F, W, M_0, D, F_d)$ , where the first six elements are similar to TPNN elements, and the last one,  $F_d$ , is the set of dynamic arcs that change depending on the job operations order.  $F_d \subseteq \{P \times T\} \cup \{T \times P\}$ .  $F \cap F_d = \emptyset$ .

#### 4.1. One operation modelling

The ATPNN to model one operation  $O_{ijk}$  of a job  $J_i$  processed by machine  $M_k$  is depicted in figure 1.

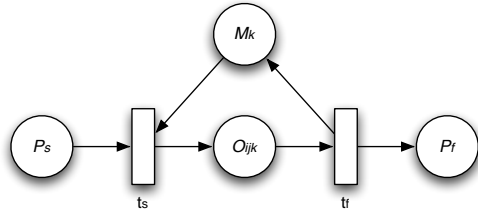


Figure 1: Operation  $O_{ijk}$  processed by machine  $M_k$  denoted as a PN model.

#### 4.2. One job modelling

As we mentioned above, one job  $J_i$  is composed of operations  $O_{ijk}$ , so the PN model for each  $J_i$  is a concatenation of its operations  $O_{ijk}$  (Figure 2).

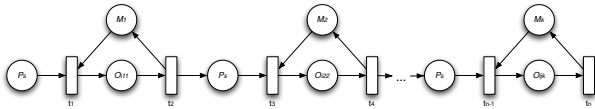


Figure 2: Operations  $O_{ijk}$  of job  $J_i$  processed by machines  $M_k$  denoted as a PN model.

#### 4.3. FMS modelling

In order to model the whole FMS, we add the PN structure for each job  $J_i$  and connect every  $M_k$  place with its corresponding input (output) transition from (to) operation  $O_{ijk}$ . Figure 3 shows the PN model for the FMS described in Table 1.

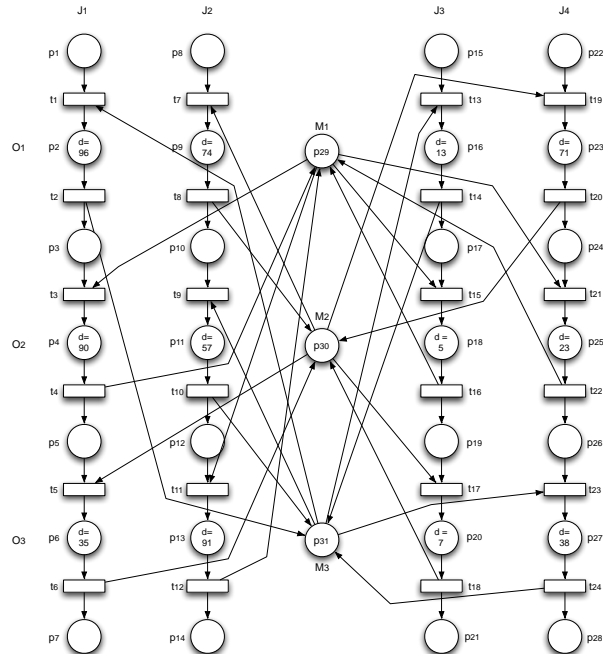


Figure 3: PN model for the FMS described in Table 1.

In Figure 3, each column corresponds to each job  $J_i$ , and some places have a label  $d$ , which denotes the time delay for processing an operation  $O_{ijk}$  in the connected machine  $M_k$ .

##### 4.3.1. Algorithm to convert a TPPN into an ATPPN

At this time, PN model of figure 3 only represents the FMS, but it is also necessary to set the priority in the operations processing by means of arcs connection in

the PN model. So, we need to define the elements of  $F_d$  to denote this priority.

First of all, the operations sequence is defined in a row vector  $OS = [os_1 os_2 \dots os_{ixj}]$ , where each  $OS$  value corresponds to one operation  $O_{ijk}$ . The following algorithm receives as inputs the row vector  $OS$  and the TPPN model, as output of the algorithm we obtain the ATPPN.

##### Algorithm TPPN\_into\_ATPPN

Input: TPPN, OS

Output: ATPPN

```

1. For q=1 to ixj
 $k = \text{machineOf}(OS(q))$
 $\text{add}(M_o(k), OS(q))$
End For
2. For $k = 1$ to NumberOfMachines
 For $i = 1$ to $\text{NumberOfJobs} - 1$
 $p_1 = \text{placeOf}(M_o(k, i))$
 $p_2 = \text{placeOf}(M_o(k, i+1))$
 $t_1 = p_1 \bullet$
 $t_2 = \bullet p_2$
 $p_3 = \bullet t_2$
 $W(t_1, p_3) = 1$
 $W(p_3, t_2) = 2$
 End For
End For

```

In Step 1, a  $k \times i$  matrix called  $M_o$  is created. Operations  $os \in OS$  ( $O_{ijk}$ ) carried out by the same machine  $M_k$  are added in the row  $k$  of  $M_o$ . The sequence order for the same machine is taking into account.

In Step 2, new arcs  $(t, p) \in F_d$  are created, which connect output transitions of places representing operations  $O_{ijk}$  with the input place of next operation  $O_{ijk}$  in the sequence order defined in  $M_o$ . Moreover, a value 2 is assigned to weight  $W(p_3, t_2)$ , to assure the order in the operations processing.

To illustrate the algorithm result, figure 4 shows the ATPPN obtained, based on the operations denoted in figure 3 and following the order:  $OS = [O_1J_2, O_1J_4, O_2J_2, O_1J_3, O_2J_4, O_1J_1, O_3J_2, O_3J_3, O_2J_1, O_3J_1, O_2J_3, O_3J_3]$ .

The ATPPN model presented in figure 4 is used to calculate the makespan for the sequence defined in vector  $OS$ .

## 5. ALGORITHM TO OBTAIN THE MAKESPAN

The proposed algorithm takes into account the mathematical representation of the ATPPN. In particular, the incidence matrix and the state equation are utilised to obtain the time delay for each  $O_{ijk}$ .

As input data the algorithm needs the ATPPN which includes its input arcs matrix  $(a_{ij}^-)$ , the output arcs matrix  $(a_{ij}^+)$ , the time delays column vector  $D$ , and the initial marking  $M_0$ , the total number of jobs ( $n_j$ ), the total number of operations per job ( $no$ ), and the total number of shared machines ( $nm$ ).

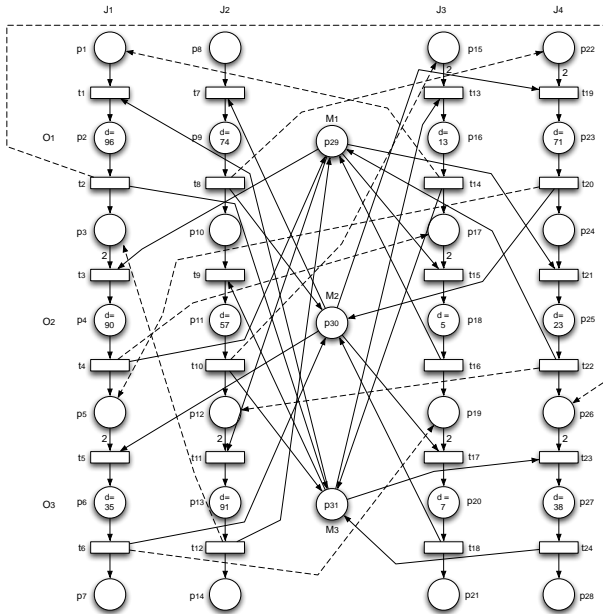


Figure 4: ATPPN model obtained applying the algorithm TTPN\_into\_ATPPN.

#### Algorithm Calculate\_Makespan

Input: ATPPN,  $a_{ij}^-$ ,  $a_{ij}^+$ ,  $D$ ,  $M_0$ ,  $n_j$ ,  $n_o$ ,  $nm$

Output: *makespan*

1. Initialise variables:

$$txj = |T| / n_j$$

$$pxj = (|P| - nm) / n_j$$

$$AT = [0 \ 0 \ \dots \ 0]'_{|P| \times 1}$$

$$TV = [0 \ 0 \ \dots \ 0]_{1 \times (n_j)}$$

2.  $ET = \text{enabledTransitions}(M_0, a_{ij}^-)$

3.  $\forall et \in ET, U_k(et) = 1$

4. While  $|ET| > 0$

For each  $t \in ET$

$$\text{indexT} = \text{ceil}(t/txj)$$

$$Uk_{tmp}^T = [0 \ 0 \ \dots \ 0]'_{|T| \times 1}$$

$$Uk_{tmp}(t) = 1$$

$$\tau = D' \times (a_{ij}^-)' \times Uk_{tmp}$$

$$\tau_{accum} = AT' \times (a_{ij}^-)' \times Uk_{tmp}$$

$$\max\_tau_{accum} = \max(TV(\text{indexT}), \tau_{accum}/2) + \tau$$

For each  $p \in t^*$

$$\text{indexP} = \text{ceil}(p/pxj)$$

If  $\text{indexT} == \text{indexP}$

$$TV(\text{indexP}) = \max\_tau_{accum}$$

Else

$$AT(p) = \max\_tau_{accum}$$

End if

End For

End For

$$M_i = M_{i-1} + (a_{ij}^-)' \times Uk$$

$$ET = \text{enabledTransitions}(M_i, a_{ij}^-)$$

$$Uk^T = [0 \ 0 \ \dots \ 0]'_{|T| \times 1}$$

$\forall et \in ET, U_k(et) = 1$

End While

5. *makespan* =  $\max(TV)$

In step 1, four variables are initialised: the number of transitions per job ( $txj$ ), the number of places per job ( $pxj$ ), a column vector  $AT$  to assign the accumulative time for each place, and a row vector  $TV$  utilised to save the time used for each job.

Step 2 obtains the enabled transitions (ET) for an initial marking  $M_0$ . Step 3 creates the  $Uk$  vector from ET transitions.

Step 4 makes an iterative process while the ATPPN is alive, i.e., while there exists at least one enabled transition in the current marking. So, for every enabled transition  $t$ , we identify the job  $J_i$  where the transition belongs (indexT), initialise a temporal  $Uk$  ( $Uk_{tmp}$ ) to fire transition  $t$ . The time delay  $\tau$  corresponding to current operation  $O_{ijk}$  is calculated multiplying the transpose of the time delay vector  $D'$  by the transpose of the input arcs matrix ( $a_{ij}^-$ ), and the result is multiplied by the firing vector  $Uk_{tmp}$  taking into account only transition  $t$ .

The accumulated time, denoted as  $\tau_{accum}$ , represents the time that the needed machine  $M_k$  has been busy previous to the current operation  $O_{ijk}$ . And it is calculated in a similar way that  $\tau$ , but in this case we use an auxiliary vector  $AT$  where the accumulative time for each place is stored, instead of the time delay vector  $D$ .

Then, we compare both times, the time when the machine  $M_k$  is ready and the time when the operation  $O_{ijk}$  is also ready to be processed. The maximum time plus the time delay for operation  $O_{ijk}$  is assigned to  $\max\_tau_{accum}$ . For every  $p \in t^*$ , if  $p$  and  $t$  are in the same job line then  $\max\_tau_{accum}$  is assigned to the time vector variable  $TV$ . On the other hand, if  $p$  and  $t$  belong to different job lines,  $\max\_tau_{accum}$  is assigned to the vector  $AT$ .

Finally, the ATPPN marking  $M_i$  changes according to the result of the equation state. From this new marking  $M_i$ , the new enabled transitions are assigned to vector  $ET$  and vector  $Uk$  is generated from them.

## 6. APPLICATION EXAMPLE

In order to show the applicability of this approach, the FMS configuration showed in Table 1 is taken. This FMS has four jobs, and each job is divided in three operations performed in three different machines.

Operation times showed in Table 3 come from the transposed matrix corresponding to operation times from Table 1.

Table 3: Processing time for each operation per job.

| Operation serial number | $J_1$ | $J_2$ | $J_3$ | $J_4$ |
|-------------------------|-------|-------|-------|-------|
| 1                       | 96    | 74    | 13    | 71    |
| 2                       | 90    | 57    | 5     | 23    |
| 3                       | 35    | 91    | 7     | 38    |

Furthermore, to identify every operation in the whole FMS, an operation number was assigned to each one. (Table 4).

Table 3: Number assigned to each operation.

| Operation serial number | $J_1$ | $J_2$ | $J_3$ | $J_4$ |
|-------------------------|-------|-------|-------|-------|
| 1                       | 1     | 4     | 7     | 10    |
| 2                       | 2     | 5     | 8     | 11    |
| 3                       | 3     | 6     | 9     | 12    |

Finally, the machine utilization in the FMS is showed in Table 5.

Table 5: Machine utilization.

| Operation serial number | $J_1$ | $J_2$ | $J_3$ | $J_4$ |
|-------------------------|-------|-------|-------|-------|
| 1                       | 3     | 2     | 3     | 2     |
| 2                       | 1     | 3     | 1     | 1     |
| 3                       | 2     | 1     | 2     | 3     |

Based on the operation numbers from Table 4, we take the following operation sequence:  $OS = (4\ 10\ 5\ 7\ 11\ 1\ 6\ 12\ 2\ 3\ 8\ 9)$ , which is a valid sequence.

From this sequence, the algorithm TPPN\_into\_ATPPN is executed which has as input data the operation sequence  $OS$  and the TPPN of figure 3, obtaining as output the ATPPN showed in figure 4. The dashed arcs represent the sequence of the machine operation on the PN model.

Next, the algorithm Calculate\_Makespan needs the following data:

1. ATPPN (figure 4)
2.  $a_{ij}^+$  (figure 4)
3.  $a_{ij}^-$  (figure 4)
4.  $D_{31 \times 1}$ , is a column vector with 31 elements, denoting the processing time in the corresponding place (figure 4). For instance, place  $p_2$  is holding a processing time of 96 time units, so  $D(2) = 96$ , which represents the operation time of machine 3 for the operation number 1 (Table 4), and so on.
5.  $M_0$  is the initial marking, where  $M_0(i) = 1$ , for  $i = 1, 8, 15, 22, 29, 30, 31$ . And  $M_0(j) = 0$  for  $j \neq i, 1 \leq j \leq 31$ . This marking denotes the starting of FMS operations ( $p_1, p_4, p_{15}, p_{22}$ ) and the availability of the machines  $M_1, M_2$  and  $M_3$  ( $p_{29}, p_{30}, p_{31}$ ).
6.  $n_j = 4$  (number of jobs)
7.  $no = 3$  (number of operations per job)
8.  $nm = 3$  (number of machines)

Under these conditions, the algorithm Calculate\_Makespan starts the transition firing and the token game animation. In order to obtain the processing time for each transition that is fired, the Eq. (2) is computed.

$$\tau = D' \times a_{ij}^{-c} \times u_k \quad (2)$$

The accumulated time for each job line is saved in vector  $TV_{1 \times 4}$ . Table 6 shows the TV values for each transition firing. There are some cases where two transitions were fired simultaneously.

Table 6: Time vector state after transition firings.

| Transitions fired | Time Vector (TV) |       |            |       |
|-------------------|------------------|-------|------------|-------|
|                   | $J_1$            | $J_2$ | $J_3$      | $J_4$ |
| $t_7$             | 0                | 0     | 0          | 0     |
| $t_8$             | 0                | 74    | 0          | 0     |
| $t_9\ t_{19}$     | 0                | 74    | 0          | 74    |
| $t_{10}\ t_{20}$  | 0                | 131   | 0          | 145   |
| $t_{13}\ t_{21}$  | 0                | 131   | 131        | 145   |
| $t_{14}\ t_{22}$  | 0                | 131   | 144        | 168   |
| $t_1\ t_{11}$     | 144              | 168   | 144        | 168   |
| $t_2\ t_{12}$     | 240              | 259   | 144        | 168   |
| $t_3\ t_{23}$     | 259              | 259   | 144        | 240   |
| $t_4\ t_{24}$     | 349              | 259   | 144        | 278   |
| $t_5\ t_{15}$     | 349              | 259   | 349        | 278   |
| $t_6\ t_{16}$     | 384              | 259   | 354        | 278   |
| $t_{17}$          | 384              | 259   | 384        | 278   |
| $t_{18}$          | 384              | 259   | <b>391</b> | 278   |

At the end, the maximum value of TV (391 time units) represents the makespan for this operation sequence.

The Gantt chart obtained, is showed in figure 5, where the maximum time is 391 time units. The value #n over each bar denotes the operation number assigned in Table 4.

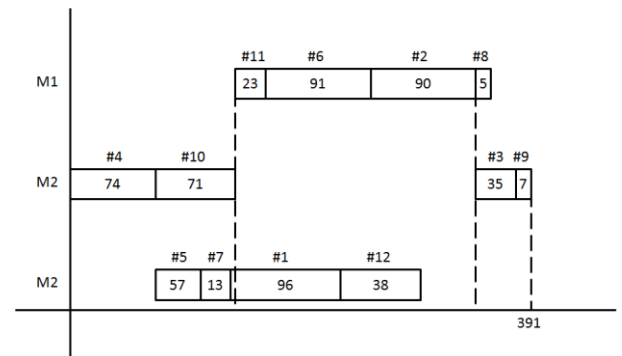


Figure 5: Scheduling chart Gantt for the operation sequence (4 10 5 7 11 1 6 12 2 3 8 9).

The makespan obtained from the scheduling Gantt chart is equal to the makespan obtained from the proposed approach.

## 7. CONCLUSIONS AND FUTURE WORK

JSSP is a NP-hard problem that has been analysed applying different kinds of techniques, such as exact models and heuristics strategies. One important calculus in the JSSP is the makespan value, which depends on the sequence of operations for each job and the order of machine utilisation.

In this paper, we propose a different way to calculate the makespan by means of mathematical tools

of Petri nets, such as the equation state and the incidence matrix. Firstly, we describe an algorithm to create an ATPPN from a FMS description. The ATPPN arcs indicate the order in which operations  $O_{ijk}$  must be done in each job. Moreover, arcs are connected adequately to set the operations order for each machine. And secondly, the marking evolution by using the state equation is taken into advantage to calculate the makespan. We added a time delay vector in order to consider the processing time for every operation involved in the FMS, and it is included in the matrix operations to be able to obtain the makespan for each job.

As future work, we are including these algorithms as part of a study based on evolutionary computing. Moreover, we are interested in analyse the feasibility of PN tools as part of an heuristic to obtain de minimum makespan in the JSSP.

## REFERENCES

- Ardakan, M. A., Hakimian, A., Rezvan, M. T., 2014. A branch-and-bound algorithm for minimising the number of tardy jobs in a two-machine flow-shop problem with release dates. *International Journal of Computer Integrated Manufacturing*, 27(6): 519-528
- Dey, S., Sarkar, D., Basu, A., 2010. A tag machine based performance evaluation method for job-shop schedules. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 29(7):1028-41.
- Gonzalez-Hernandez, M.A., 2011. *Metaheuristics solutions for "Job-Shop Scheduling Problem with sequence-dependent setup times"*. PhD Thesis. Universtiy of Oviedo.
- Lenstra, J.K., Kan, A.H.G., Brucker, P., 1977. Complexity of machine scheduling problem. *Annals of Discrete Mathematics*, 1: 343 – 362.
- Murata, T., 1989. Petri Nets: Properties, Analysis and Applications. *Proceedings of the IEEE*, 77(4), 541 – 580.
- Pinedo, M.L., 2012. *Scheduling: Theory, Algorithms, and Systems, Fourth Edition*, New York:Springer.
- Qing-dao-er-ji, R., Wang, Y., 2012. A new hybrid genetic algorithm for job shop scheduling problem. *Computers and Operations Research*, 39:2291-2299.
- Shuai, J., Zhi-Hua, H., 2014. Path-relinking Tabu search for the multi-objective flexible job shop scheduling problema. *Computers and Operations Research*, 47: 11-26.
- Wang, L., Cai, N., Feng, H.Y., 2010. An adaptive setup planning approach for dynamic machine assignments. *IEEE Transactions on Automation Science and Engineering*, 7(1): 2-14.
- Wang, S.F., Zou, Y.R., 2003. Techniques for the job shop scheduling problem: a survey. *Systems Engineering – Theory & Practice*, 23: 49-55.
- Zhao, Z., Zhang, G., Bing, Z., 2011. Scheduling Optimization for FMS Based on Petri Net Modeling and GA. *Proceedings of the IEEE International Conference on Automation and Logistics*, pp. 422-427. August 2011, Chongqing, China.
- Zhou, M.C., and Venkatesh, K., 1999. *Modeling, Simulation, and Control of Flexible Manufacturing Systems*. New York: World Scientific.

## AUTHORS BIOGRAPHY

**Joselito Medina-Marin.** He received the M.S. and Ph.D. degrees in electrical engineering from the Research and Advanced Studies Centre of the National Polytechnic Institute at Mexico, in 2002 and 2005, respectively. Currently, he is a Professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State at Pachuca, Hidalgo, México. His current research interests include Petri net theory and its applications, active databases, simulation, and programming languages.

**Juan Carlos Seck-Tuoh-Mora.** He received the M.S. and Ph.D. degrees in electrical engineering (option: Computing) from the Research and Advanced Studies Centre of the National Polytechnic Institute at Mexico, in 1999 and 2002, respectively. Currently, he is a Professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State at Pachuca, Hidalgo, México. His current research interests include cellular automata theory and its applications, evolutionary computing and simulation.

**Norberto Hernandez-Romero.** He received the M.S. degree from Department of Electrical Engineering, Laguna Technological Institute at México, in 2001 and the Ph. D. from Autonomous University of Hidalgo State at México in 2009. Currently, he is a professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State at Pachuca, Hidalgo, México. His current research interests include system identification, feedback control design, genetic algorithms, fuzzy logic, neural network and its applications

**Nayeli Jazmin Escamilla-Serna.** She is a postgraduate student in Industrial Engineering in the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State at Pachuca, Hidalgo, México.

# MODELING AND OPTIMIZATION OF THE EXTRACTION OF LIGNOSULFONATE FROM BARLEY STRAW BY USING ARTIFICIAL NEURAL NETWORKS

Maria Guadalupe Serna-Diaz<sup>(a)</sup>, Joselito Medina-Marin<sup>(b)</sup>, Ainhoa Arana-Cuena<sup>(c)</sup>,  
Juan Carlos Seck-Tuoh-Mora<sup>(d)</sup>, Alejandro Tellez-Jurado<sup>(e)</sup>, Yuridia Mercado-Flores<sup>(f)</sup>,  
Angelica Jimenez-Gonzalez<sup>(g)</sup>, Norberto Hernández-Romero<sup>(h)</sup>

<sup>(a)(c) (e) (f) (g)</sup> Polytechnic University of Pachuca, Pachuca de Soto, Hidalgo, México

<sup>(b) (d) (h)</sup> Autonomous University of Hidalgo State, Advanced Research Centre in Industrial Engineering,  
Pachuca de Soto, Hidalgo, México

<sup>(a)</sup>[sedg0210@yahoo.com.mx](mailto:sedg0210@yahoo.com.mx), <sup>(b)</sup>[jmedina@uaeh.edu.mx](mailto:jmedina@uaeh.edu.mx), <sup>(c)</sup>[ainhoa@upp.edu.mx](mailto:ainhoa@upp.edu.mx),  
<sup>(d)</sup>[jseck@uaeh.edu.mx](mailto:jseck@uaeh.edu.mx), <sup>(e)</sup>[alito@upp.edu.mx](mailto:alito@upp.edu.mx), <sup>(f)</sup>[yuridiamercado@upp.edu.mx](mailto:yuridiamercado@upp.edu.mx),  
<sup>(g)</sup>[ajimenez@upp.edu.mx](mailto:ajimenez@upp.edu.mx), <sup>(h)</sup>[nhromero@uaeh.edu.mx](mailto:nhromero@uaeh.edu.mx)

## ABSTRACT

Artificial Neural Networks (ANNs) have been applied to model dynamic systems from a wide range of areas. In particular, we are interested in the application of ANN in fitting data sets to model the extraction process of lignosulfonates from barley straw. In order to know real data about the percentage of delignification taking into account three variables from the experimental process: cooking time of the barley straw, percentage of sulfite concentration applied to the barley straw, and barley size. With the data gathered from 75 experiments performed three times each one, we created an ANN optimal with a minimum mse, and we calculate an optimal value from this ANN model.

Keywords: Artificial neural network, Optimization, Lignosulfonate extraction, Data fitting.

## 1. INTRODUCTION

Lignocellulosic biomass is considered as a key element to avoid the fossil fuel dependence and encourage the use of raw material friendlier with the environment (Zhu, Sun, Su, Zhao, Ma, Zhu, Shi, and Gu, 2013).

The lignocellulosic biomass decomposition is based on three main compounds: cellulose, hemicellulose and lignin, which are joined in microfibrillar grouped in a chain of crystalline cellulose, covered with structures of hemicellulose and holocellulose protected by lignin (Abud, Costa, Wanderley, Souza, and Sant'Anna 2013). Nevertheless, these biopolymers are better utilized individually (Kahar, Taku, and Tanaka 2013).

Lignocellulosic material has a broad range of applications, but the use of each compound depends on the treatments utilized to extract them and how easy and cheap is the separation of the three compounds. These treatments can be physical (Subhedar and Gogate 2014;

Iskalieva, Yimmou, Gogate, Horvath, Horvath, and Csoka 2012), chemical (Yang, Kuittinen, Zhang, Keinänen, and Pappinen, 2013; Duque, Manzanares, Ballesteros, Negro, Oliva, Saez, Ballesteros, 2013) or biological (El-Zawawy, Ibrahim, Abdel-Fattah, Solimanb, Mahmoud, 2011; Wang, Manley, and Feldmant, 2011).

Nowadays, the utilization of biomass extracted from agricultural residuals is often applied to obtain polymers. In the case of Barley (*Hordeum vulgare*), one kilogram of barley corresponds to 0.750 of straw (Carrera and Mateo 2005). For this reason, barley straw is a good option for extracting Lignocellulosic material (Singh, Shukla, Tiwari, and Srivastava, 2014). Moreover, lignin has many applications in industry, such as dispersant agent for mixture of cement and gypsum (Stráněl and Sebök, 1997), as emulsifier (Weis and Bird, 2001), in polymers production (Effendi, Gerhauser, and Bridgwater, 2008), and a new trend is the use of lignin in agricultural industry as soil conditioner (Deng et al., 2011) and vegetal growth promoter (Ertani et al., 2011).

The applications mentioned above utilize the lignin in the form of lignosulfonate (LS), which is the sulfonate lignin obtained from the separation with sulfite (Chakrabarty, Saha, and Ghoshal 2009). This process requires high temperature and pressure in order to depressurize the fibers of cellulosic material and reach its disaggregation. Furthermore, it is important to consider the size of the biomass (pieces of barley straw) and the concentration of the sulfite to find the maximum performance of delignification.

In this research work we took into account three independent variables: barley straw size, sulfite concentration, and cooking time. And as dependent variable we considered the performance percentage of delignification.



Experimental results were fitted through an Artificial Neural Network model in order to find an optimal combination of values that produce the maximum delignification performance.

## 2. MATERIALS AND METHODS

The lignocellulosic material of barley straw utilized in this work was taken from crops located in Zempoala, Hidalgo, Mexico. The separation was carried out in a semi-industrial autoclave AV-3580 Prendo®, which was modified to be able to reach 3 atm of pressure.

### 2.1. Variables

The variables for this experiment were: size of straw with a sieve number 8 (2.0 mm), 10 (1.68 mm) and 20 (0.84 mm); cooking time was 15, 30, 45, 60, and 90 minutes; finally, sodium sulfite concentrations with values of 1%, 3%, 5%, 10%, and 15%. The temperature was kept at 137 °C, the pressure at 3 atm, and the relation between straw and sulfite was 1 part of straw for 25 of liquor. Then, a total of seventy five lab experiments were done ( $3 \times 5 \times 5 = 75$ ). Table 1 shows the combination values for each experiment.

Table 1: Combination of values for each experiment

| Experiment number | Sieve number | Cooking time | Sulfite concentration |
|-------------------|--------------|--------------|-----------------------|
| 1                 | 8            | 15           | 1                     |
| 2                 | 8            | 15           | 3                     |
| ...               | ...          | ...          | ...                   |
| 6                 | 8            | 30           | 1                     |
| 7                 | 8            | 30           | 3                     |
| ...               | ...          | ...          | ...                   |
| 26                | 12           | 15           | 1                     |
| 27                | 12           | 15           | 3                     |
| ...               | ...          | ...          | ...                   |
| 31                | 12           | 30           | 1                     |
| 32                | 12           | 30           | 3                     |
| ...               | ...          | ...          | ...                   |
| 51                | 20           | 15           | 1                     |
| 52                | 20           | 15           | 3                     |
| ...               | ...          | ...          | ...                   |
| 56                | 20           | 30           | 1                     |
| 57                | 20           | 30           | 3                     |
| ...               | ...          | ...          | ...                   |
| 74                | 20           | 90           | 10                    |
| 75                | 20           | 90           | 15                    |

### 2.2. Lab experimental method

The straw was ground to separate the fibers. It was winnowed with chain mails of 8 (2.0 mm), 10 (1.68 mm) and 20 (0.84 mm). Then, 2g of straw for each size were mixed with 100 mL of sodium sulfite, in different concentrations (1%, 3%, 5%, 10%, and 15%). Every experiment was done in triplicate.

The digestive process was carried out in a closed system inside glass bottles and the cooking process was done in an autoclave with temperature of 137 °C and pressure of 3 atm, during 15, 30, 45, 60, and 90 minutes. From this process, the cellulose pulp and the “liquor” were produced. Lignosulfonates contained in the liquor were separated by means of vacuum filtration, in filter paper with pore of 1.6 μm.

To quantify the solubilized material, the material retained by the filter paper was burned at 550 °C during 6 hours. Then, the counting was done by weight difference, with a previous correction of the humidity and solubilized material in a sample without sulfite only with water (Northey, 1992; Ekeberg, Gretland, Gustafsson, Braten, and Fredheim, 2006).

The performance of delignification material is calculated with Equation 1.

$$PSM = \frac{DW_s - (DW_f - A - W)}{DW_s} \cdot 100 \quad (1)$$

Where:

*PSM*: is the percentage of solubilized material.

*DW<sub>s</sub>*: is the initial dry weight.

*DW<sub>f</sub>*: is the final dry weight.

*A*: are the residual ashes.

*W*: is the sample without sulfite (only water).

### 2.3. Data fitting method

The data fitting was done through an Artificial Neural Networks (ANNs), which was trained, validated and tested with the 75 experimental results from the combination of 3 sizes of barley straw, 5 cooking times, and 5 sulfite concentrations.

#### 2.3.1. Artificial Neural Network

An ANN is computational technique that performs a multifactorial analysis. It is inspired in biological neural networks, and is composed of nodes distributed in layers, and interconnected through lines with assigned weights.

Nodes represent artificial neurons, which are processing units performing a no-linear sum (Dayhoff, and Deleo, 2001). Connections weights are adjusted according to a training process of the ANN. An adequate training produces an ANN that helps in values forecasting, objects classification, function approximation, pattern recognition with multifactorial data, and modeling and optimization of data fit for different phenomena (Sinhaa, Chowdhurya, Sahaa, and Dab, 2013; Aghav, Kumar, and Mukherjee, 2011; Duque, Manzanares, Ballesteros, Negro, Oliva, Saez, and Ballesteros, 2013; Maache-Rezzoug, Pierre, Nouviaire, Maugard, and Rezzoug, 2011), among other applications.

In this work, we applied the Matlab® Neural Network Toolbox to perform the data fitting. The ANN model was obtained through an iterative process. In it we were changing the number of neurons, and the ANN model was trained ten times, saving the ANN with the less Mean Square Error (mse). The method used for

training was the Levenberg-Marquardt method, and the data division for training, validating, and testing was done randomly.

### 3. RESULTS

#### 3.1. Lab experimental results

Results obtained from straw delignification with different values for sulfite concentration, cooking time, and size of barley straw are showed in figure 1.

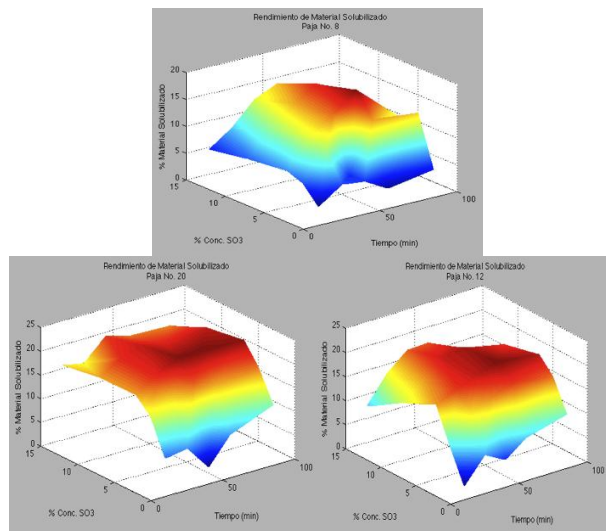


Figure 1: Percentage of solubilized material from straw barley of size 8(a), 12(b) and 20(c).

The lower percentages of delignification are represented by dark blue color and the higher percentages are represented by dark red color.

Figure 1 shows that the maximum value of percentage of delignification is located in the same value for cooking time (60 minutes) and sulfite concentration (5%) for each value of barley straw size. For the barley straw size 8, the maximum value is  $18.7 \pm 0.53$ , for size 12 the maximum value is  $20.4 \pm 0.47$ , and for size 20 the maximum value is  $22.5 \pm 1.3$ .

#### 3.2. ANN model results

The data fitting was performed through an ANN model, which was created with the Neural Network Toolbox included in Matlab®.

Furthermore, a Matlab® script was implemented to find an ANN model with a minimum mse. This script creates ANN models starting from a model with 3 layers; the first one considers the three input variables (barley straw size, cooking time, and sulfite concentration); the second one is the hidden layer where the connections weights are set from the training process; and the last layer contains only one neuron with the value of the percentage of solubilized material.

70% of the experimental data (56 samples) were utilized in the training process, and they were chosen randomly. 15% of the data (11 samples) were utilized to validate the model, and the other 15% (11 samples) of data were utilized to test the ANN model.

The ANNs created in the script were trained with the Levenberg-Marquardt backpropagation algorithm, and we iterate from an ANN with 3 neurons until an ANN with 50 neurons. Each one was trained ten times in order to find a model with the minimum mse.

Taking into account this criterion, we found an ANN model with 20 neurons and the less mse of 1.2214 (Figure 2).

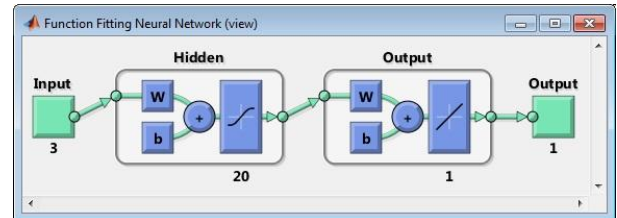


Figure 2: ANN model with minimum mse.

Figure 3 contains the graphic for lab experimental data and the data obtained from the ANN model. The seventy five values obtained from lab experiments are denoted as small red circles, and the dashed line denotes the data forecasted from the ANN model, for the same seventy five combinations. The axis of abscissas represents the experiment number, where each experiment was performed in a combination of values for sieve number, cooking time, and sulfite concentration. The axis of ordinates indicates the percentage of solubilized material obtained from each combination.

The model showed in figure 2 was utilized to estimate the parameters for the optimal value of percentage of delignification. The input value for barley straw size was iterated from 8 to 20, the input value for cooking time was iterated from 15 to 90 minutes, and the input value for sulfite concentration was iterated from 1 to 15. Each parameter was incremented in steps of one, and we obtained a total of 14 820 combinations in order to evaluate the ANN model for each one.

The highest value of solubilized material calculated with this ANN model was 24.6432%, with barley straw size of 16, a cooking time of 42 minutes, and a sulfite concentration of 15%.

Figure 4 shows the graphic generated with all the data forecasted with the ANN model. As in figure 3, the small red circles denotes the seventy five lab experimental data, and the green graphic denotes the 14 820 forecasted data from the ANN model. Similar to figure 3, the axis of abscissas represents the experiment number, and the axis of ordinates indicates the percentage of solubilized material.

The circle filled in red indicates the optimal value of solubilized material.

### 4. CONCLUSIONS

Lignocellulosic material has been applied in different areas, and it has been gotten from different sources. Currently, agricultural residuals are used to extract lignin with good results.

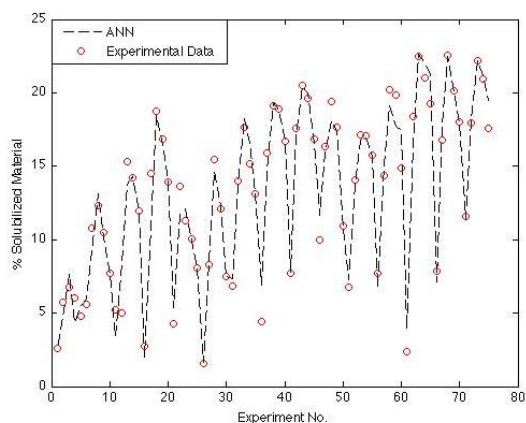


Figure 3: Comparative graphic between experimental data and data predicted with the ANN model.

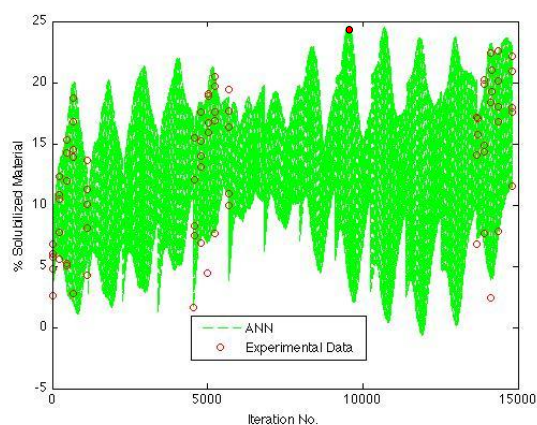


Figure 4: optimal value found in all the data forecasted with the ANN model.

In this work, we used barley straw to get lignin from biomass, and to analyze the effect of factors as barley straw size, cooking time, and sulfite concentration, a total of 75 experiments were carried out. The three barley straw sizes considered were 8, 12, and 20; the five cooking times utilized were 15, 30, 45, 60, and 90 minutes; and the five sulfite concentrations were 1%, 3%, 5%, 10%, and 15%.

Experimental data were used to create ANNs with different number of neurons and trained ten times. The ANN with the lower mse (1.2214) was taken to search the optimal value for percentage of delignification. The maximum value found for the solubilized material with the ANN model was 24.6432%, for a barley straw size of 16, cooking time required of 42 minutes, and sulfite concentration of 15%.

As future work, we are going to take into account some economic considerations, in order to estimate the cost for producing liginosulfonates from barley straw, and to determine if the optimal value found by the ANN model can be obtained with a low cost.

## REFERENCES

Abud, Y., Costa, L.T., Wanderley., Souza, W. and Sant'Anna, C. (2013). Revealing the microfibrillar

arrangement of the cell wall surface and the macromolecular effects of thermochemical pretreatment insugarcane by atomic force microscopy. *Industrial Crops and Products*, 51, 62– 69.

- Aghav, R.M., Kumar, S., Mukherjee, S.N., (2011). Artificial neural network modeling in competitive adsorption of phenol and resorcinol from water environment using some carbonaceous adsorbents. *Journal of Hazardous Materials*, 188, 67-77.
- Carrera-Morales, M., and Mateo-Box, J.M. (2005). *Prontuario de agricultura: cultivos agrícolas. España. Mundi-prensa libros, s.a,pp 73.*
- Chakrabarty, K., Saha, P., and Ghoshal, A.K. (2009). Separation of liginosulfonate from its aqueous solution using supported liquid membrane. *Journal of Membrane Science*, 340, 84-91.
- Dayhoff, J., and Deleo, J.M., (2001). Artificial Neural Networks Opening the Black Box. *CANCER Supplement*, 91(8), 1615-1635.
- Duque, A., Manzanares, P., Ballesteros, I., Negro, M.J., Oliva, J.M., Saez, F., Ballesteros, M., (2013). Optimization of integrated alkaline-extrusion pretreatment of barley straw for sugar production by enzymatic hydrolysis. *Process Biochemistry*, 48, 775-781.
- Effendi, A., Gerhauser, H., Bridgwater, A.V. (2008). Production of renewable phenolic resins by thermochemical conversion of biomass: A review. *Renewable and Sustainable Energy Reviews*. 12, 2092-2116.
- Ekeberg, D., Gretland, K.S., Gustafsson, J., Braten, S.M., Fredheim, G.E., (2006). Characterisation of liginosulfonates and kraft lignin by hydrophobic interaction chromatography. *Analytica Chimica Acta*. 565, 121-128.
- El-Zawawy, W.K., Ibrahim, M.M., Abdel-Fattah, Y.R., Solimanb, N.A., Mahmoud, M.M., (2011). Acid and enzyme hydrolysis to convert pretreated liginocellulosic materials into glucose for ethanol production. *Carbohydrate Polymers*, 84, 865-871.
- Iskalieva, A., Yimmou, B.M., Gogate, P.R., Horvath, M., Horvath, P.G. and Csoka, L. (2012). Cavitation assisted delignification of wheat straw: A review. *Ultrasonics Sonochemistry*, 19, 984–993.
- Kahar, P., Taku, K. and Tanaka., S. (2013). Multiple effects of swelling by sodium bicarbonate after delignification on enzymatic saccharification of rice straw. *Journal of Bioscience and Bioengineering*, 116 (6), 725-733.
- Maache-Rezzoug, Z., Pierre, G., Nouviaire, A., Maugard, T., Rezzoug, S.A., (2011). Optimizing thermomechanical pretreatment conditions to enhance enzymatic hydrolysis of wheat straw by response surface methodology. *Biomass and Bioenergy*, 35, 3129-3138.
- Northey, R.A., (1992). Low-cost uses of lignin, emerging technology of materials and chemicals

from biomass. *ACS Symposium Series 476*. Washington, DC.

- Singh, R., Shukla, A., Tiwari, S., Srivastava, M., (2014). A review on delignification of lignocellulosic biomass for enhancement of ethanol production potential. *Renewable and Sustainable Energy Reviews*, 32, 713-728.
- Sinhaa, K., Chowdhurya, S., Sahaa, P.D., Dab, S. (2013). Modeling of microwave-assisted extraction of natural dye from seeds of *Bixa orellana* (Annatto) using response surface methodology (RSM) and artificial neural network (ANN). *Industrial Crops and Products*, 41, 165-171.
- Stráněl, O., and Sebök, T. (1997). Relationships between the properties of ligninsulphonates and parameters of modified samples with cement binders. *Cement and Concrete Research*. 2 (27), 185-191.
- Subhedar, P.B. and Gogate, P.R. (2014). Alkaline and ultrasound assisted alkaline pretreatment for intensification of delignification process from sustainable raw-material. *Ultrasonics Sonochemistry*. 21, 216–225.
- Wang, J., Manley, J. & Feldmant, D. (1992). Synthetic polymer-lignin copolymers and blends. *Progress in Polymer Science*. 17, 611-646.
- Weis, A. & Bird, M. R. (2001). The influence of multiple fouling and cleaning cycles upon the membrane processing of lignosulphonates. *Chemical Engineering*. 79, 184-187.
- Yang M., Kuittinen, S., Zhang, J., Keinänen, M. & Pappinen, A., (2013). Effect of dilute acid pretreatment on the conversion of barley straw with grains to fermentable sugars. *Bioresource Technology*. 146, 444–450.
- Zhu, Z., Sun, M.M., Su, C., Zhao, H., Ma, X., Zhu, Z., Shi, X., Gu, K., (2013). One-pot quantitative hydrolysis of lignocelluloses mediated by black liquor. *Bioresource Technology*, 128, 229-234.

#### **AUTHORS BIOGRAPHY**

**Maria Guadalupe Serna-Díaz.** She received the M.S. degree in Biotechnology from the Polytechnic University of Pachuca, in Pachuca de Soto, Hidalgo, Mexico, in 2011. Currently, she is a PhD student of the Polytechnic University of Pachuca, Pachuca de Soto, Hidalgo, Mexico.

**Joselito Medina-Marin.** He received the M.S. and Ph.D. degrees in electrical engineering from the Research and Advanced Studies Centre of the National Polytechnic Institute in Mexico, in 2002 and 2005, respectively. Currently, he is a Professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State in Pachuca, Hidalgo, México. His current research interests include Petri net theory and its applications, active databases, simulation, and programming languages.

**Ainhoa Arana-Cuena.** She received the Ph.D. degree in Molecular Biology from the Autonomous University of Madrid in Spain, in 2002. She is working as a full time professor in the Polytechnic University of Pachuca, in Hidalgo State, México. Her research interest is the applications of agricultural residuals.

**Juan Carlos Seck-Tuoh-Mora.** He received the M.S. and Ph.D. degrees in electrical engineering (option: Computing) from the Research and Advanced Studies Centre of the National Polytechnic Institute in Mexico, in 1999 and 2002, respectively. Currently, he is a Professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State in Pachuca, Hidalgo, México. His current research interests include cellular automata theory and its applications, evolutionary computing and simulation.

**Alejandro Jurado-Tellez.** He received the M.S. degree in Biotechnology from the Autonomous Metropolitan University, in México City, México in 1997, and the Ph.D. degree in Molecular Microbiology from the University of Alcalá in Spain, in 2002. He is working as a full time professor in the Polytechnic University of Pachuca, in Hidalgo State, México. His research interest is the applications of agricultural residuals.

**Yuridia Mercado-Flores.** She received the M.S. and Ph.D. degrees in Biotechnology from the National Polytechnic Institute in México, City, México, in 2002 and 2005, respectively. She is working as a full time professor in the Polytechnic University of Pachuca, in Hidalgo State, México. Her research interest is the applications of agricultural residuals.

**Angélica Jiménez-González.** She received the M.S. and Ph.D. degrees in Biotechnology from the Autonomous Metropolitan University, in México City, México, in 2003 and 2008, respectively. She is working as a full time professor in the Polytechnic University of Pachuca, in Hidalgo State, México. Her research interest is the applications of agricultural residuals.

**Norberto Hernandez-Romero.** He received the M.S. degree from Department of Electrical Engineering, Laguna Technological Institute at México, in 2001 and the Ph. D. from Autonomous University of Hidalgo State at México in 2009. Currently, he is a professor of the Advanced Research in Industrial Engineering Centre at the Autonomous University of Hidalgo State at Pachuca, Hidalgo, México. His current research interests include system identification, feedback control design, genetic algorithms, fuzzy logic, neural network and its applications

# A FRIDAY 13<sup>TH</sup> RISK MODEL FOR FAILURE IN CROSS – FLOW MEMBRANE FILTRATION OF PASSION FRUIT JUICE

Wei Zou<sup>(a)</sup>, Kenneth Davey<sup>(b)</sup>

School of Chemical Engineering  
The University of Adelaide, 5005, Australia

<sup>(a)</sup> [wei.zou@adelaide.edu.au](mailto:wei.zou@adelaide.edu.au), <sup>(b)</sup> [kenneth.davey@adelaide.edu.au](mailto:kenneth.davey@adelaide.edu.au)

## ABSTRACT

Membrane filtration for clarification is a widespread unit-operation in the beverage industries. Unexpected membrane fouling (failure) can significantly reduce clarification efficiency. A *Friday 13<sup>th</sup>* (*Fr 13*) risk model for membrane failure is developed and illustrated with independent data for yellow passion fruit juice (*Passiflora edulis* var. *flavicarpa*) and a comparison made with a traditional evaluation. The aim was to gain insight into the risk of stochastic impact on failure in a well-operated cross-flow membrane plant used in juice clarification. A risk factor ( $p$ ) is defined in terms of the design permeate ( $J$ ) and actual critical ( $J_{critical}$ ), flux. A refined Monte Carlo sampling of transmembrane pressure ( $\Delta P$ ) and filtration time ( $t$ ) is used to simulate membrane behavior. Results reveal that some 7 % of all filtrations will fail to achieve  $J_{critical}$  with a practical tolerance and typical commercial conditions over an extended period. This insight is not available from traditional evaluation.

Keywords: membrane fouling, juice clarification, unexpected membrane failure, *Friday 13<sup>th</sup>* risk modelling

## 1. INTRODUCTION

Yellow passion fruit (*Passiflora edulis* var. *flavicarpa*) is a tropical fruit which is widely used as an ingredient in formulated beverages and fruit products. However, traditional filtration methods such as distillation, adsorption and pasteurization, lead to a loss of nutritional and aromatic compounds (Domingues, Ramos, Cardoso and Reis 2014). Recently, non-thermal membrane cross-flow processing is becoming a widely substituted separating technology in juice clarification, because it provides high sensorial quality products with less loss, manpower and reduced processing time and operational costs over traditional methods (D' Souza and Mawson 2005).

However, fouling is a key limitation in cross-flow membrane processing. The transient accumulation of rejected materials at the interface of the membrane can result in a reduced permeate flux or partial blockage. In juice clarification, a low permeate flux can substantially

reduce process efficiency and product (juice) safety (Cassano, Marchio and Drioli 2006).

We are interested in the notion that no matter how good the design and operation of plant, there will be an unexpected, occasional and sudden (surprise) failure. Davey and co-workers have titled this practical phenomenon *Friday 13<sup>th</sup>* syndrome (*Fr 13*) to explain the nature of the event (e.g. Davey, Chandrakash and O'Neill 2013; Davey 2010). Their hypothesis is that: *random variation in plant parameters can accumulate in one direction and leverage significant change in plant or product*. These unexpected failures can be mistakenly blamed on 'human error' or 'faulty fittings' (Davey and Cerf 2003; Davey 2010). According to the recent *Blackett Review* (Anon. 2011) 'high-impact', low-probability' failures are an emerging and practical challenge for various operations of a range of scale (Anon. 2011).

Although several mathematical models for juice filtration have been reported (Nandi, Das and Uppaluri 2012; Razi, Aroujalian and Fathizadeh 2012; Cassano, Marchio and Drioli 2006; Cui, Jiang and Field 2010) there is however no model to meaningfully assess risk of fouling (failure). Further, these do not involve the impact of a stochastic affects.

### 1.1. This study

Against this background we undertook research into synthesis and simulation of a novel *Fr 13* model for vulnerability to stochastic failure of widely used cross-flow membrane filtration plant. The model is illustrated using independent data of Domingues, Ramos, Cardoso and Reis (2014) for yellow passion fruit (*Passiflora edulis* var. *flavicarpa*) juice. A comparison is made with a traditional evaluation.

A *Fr 13* risk model is based on the underlying unit-operations model together with a clear definition of failure and refined Monte Carlo (r-MC) (*Latin Hypercube*) sampling (Zou and Davey 2014; Davey, Chandrakash and O'Neill 2013; Davey 2011).

The aim was to gain new insight that can be used to improve design of operations with reduced vulnerability to failure, together with improved process efficiency (and safety).



## 2. MATERIALS AND METHODS

### 2.1. Synthesis of a Unit-operations Model for Cross-flow Membrane Filtration

An acceptable unit-operations model for a well-operated cross-flow membrane process necessitates integration and synthesis of equations for transmembrane pressure ( $\Delta P$ ), filtration time ( $t$ ), membrane surface area ( $A$ ), kinetic model constant for a particular juice ( $K_c$ ) and an adjust parameter ( $n$ ) according to one of four fouling modes (Domingues, Ramos, Cardoso and Reis 2014; Field, Wu, Howell and Gupta 1995).

Based on the model recently synthesized by Zou and Davey (2014) and, Bowen, Calvo and Hernandez (1995) and Field, Wu, Howell and Gupta (1995), we can write (all symbol used are carefully defined in the Nomenclature) for the design permeate flux ( $J'$ ):

$$J' = J_0 \cdot \left[ 1 + K_c \cdot J_0^{(2-n)} \cdot t \right]^{1/(n-2)} \quad (1)$$

$J_0$  is obtained from initial membrane performance testing with deionized (clean) water (Domingues, Ramos, Cardoso and Reis 2014).

The parameter  $n$  is fixed based on the principal fouling modes of the cross-flow membrane process:  $n = 2$  for complete pore blocking;  $n = 1.5$  for internal pore blocking;  $n = 1$  for partial pore blocking; and,  $n = 0$  for cake filtration layer. Many researchers (e.g. Domingues, Ramos, Cardoso and Reis 2014; Nandi, Das and Uppaluri 2012; Razi, Aroujalian and Fathizadeh 2012; Cassano, Marchio and Drioli 2006; Cui, Jiang and Field 2010) have reported cake filtration with  $n = 0$  is the major mode to describe fouling mechanism with juice processing, including passion fruit, apple, pineapple, mosambi and orange juices.

The permeate flux of the deionized water ( $J_0$ ) is important to obtain the membrane capacity and is defined as (Schafer, Andritsos, Karabelas, Hoek, Schneider and Nystrom 2005):

$$J_0 = \frac{\mu_T}{\mu_{T_{20^\circ\text{C}}}} \cdot \frac{Q}{A \cdot \Delta P} \quad (2)$$

where  $Q$  = clean water flow rate at the treating temperature, and  $\mu_T$  and  $\mu_{20^\circ\text{C}}$  are, respectively, the viscosity of clean water at operation, and a reference  $20^\circ\text{C}$  (Roorda and Graaf 2001). Substitution for  $J_0$  from Eq. (2) and  $n = 0$  into Eq. (1) and simplifying gives:

$$J' = \frac{\mu_T}{\mu_{T_{20^\circ\text{C}}}} \cdot \frac{Q}{A \cdot \Delta P} \cdot \left[ 1 + K_c \cdot \left( \frac{\mu_T}{\mu_{T_{20^\circ\text{C}}}} \cdot \frac{Q}{A \cdot \Delta P} \right)^2 \cdot t \right]^{-0.5} \quad (3)$$

### 2.2. Critical Permeate Flux for Passion Fruit Juice

During cross-flow membrane filtration, there exists a flux below which a decay of flux with time does not occur; above this flux fouling is observed (Field, Wu, Howell and Gupta 1995). This flux is called the critical

flux ( $J_{critical}$ ) which is the design, or operation, level for successful membrane filtration without fouling.

Practical experimental data for cross-flow membrane filtration of passion fruit (*Passiflora edulis var. flavicarpa*) juice have been independently reported by Domingues, Ramos, Cardoso and Reis (2014) at a processing temperature of  $25^\circ\text{C}$ . Hollow fibre membranes with an average pore diameter of  $0.40 \mu\text{m}$  and surface area of  $0.056 \text{ m}^2$  together with  $\Delta P = 100 \text{ kPa}$  and  $5.4 \times 10^{-6} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1} < J < 2.78 \times 10^{-6} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$  were used. These data are conveniently re-plotted as Figure 1.

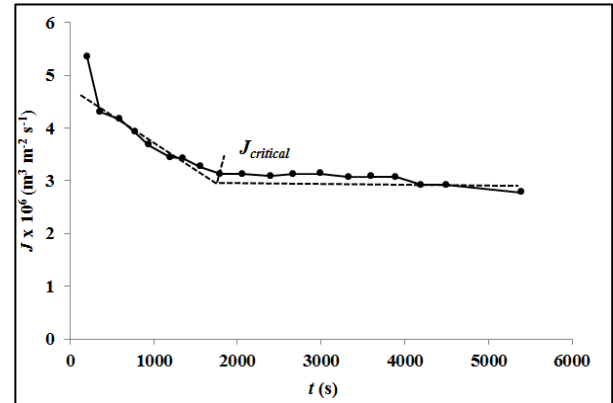


Figure 1: Experimental data for permeate flux ( $n = 19$ ) in cross-flow membrane filtration of passion fruit juice (*Passiflora edulis var. flavicarpa*) at  $\Delta P = 100 \text{ kPa}$  (adapted from Domingues, Ramos, Cardoso and Reis 2014)

From Figure 1, the critical flux,  $J_{critical} = 3.125 \times 10^{-6} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$ . This is determined by the intersection of lines tangent to the initial (0-1600, s) and terminal (3300-5400, s) parts of the curve, as shown on the figure. It is seen in the figure that  $J_{critical}$  for the passion fruit juice will occur at  $t = 1800 \text{ s}$ .

All  $J' < J_{critical}$  indicates membrane fouling (failure). Cleaning or replacement of the membrane is then needed.

The unit-operations model for cross-flow membrane filtration of passion fruit juice is defined by Eq. (1) through (3), together with the value of the critical permeate flux ( $J_{critical}$ ).

### 2.3. Traditional Single Value Assessment Solution

The traditional method for solution of a unit-operations model is called the 'estimated value' or 'single value assessment' (SVA) (Sinnott 2005; Davey 2011).

For typical commercial filtration of passion fruit juice, initial operating conditions include  $\Delta P = 100 \text{ kPa}$ ;  $\mu_T = 0.812 \times 10^6 \text{ pa s}$ ;  $\mu_{20^\circ\text{C}} = 1.002 \times 10^6 \text{ pa s}$ ;  $Q = 1.8 \times 10^{-3} \text{ m}^3 \text{ s}^{-1}$ ;  $A = 0.056 \text{ m}^2$ ;  $K_c = 5.6 \times 10^7 \text{ s m}^{-2}$  and  $t = 1800 \text{ s}$  (Domingues, Ramos, Cardoso and Reis 2014).

Substitution into Eq. (3), gives the SVA value for the permeate flux,  $J' = 3.149 \times 10^{-6} \text{ m}^3 \text{ m}^{-2} \text{ s}^{-1}$ .



## 2.4. Defining Failure Risk

A risk factor ( $p$ ) for vulnerability to failure of the cross-flow membrane can be conveniently defined together with a practical tolerance (Zou and Davey 2014; Davey, Chandrakash and O'Neill 2013) such that:

$$p = -\%tolerance + 100 \left( \frac{J'}{J_{critical}} - 1 \right) \quad (4)$$

where  $J'$  is the instantaneous value of the permeate flux (or more strictly one membrane operation scenario). For an assumed practical tolerance of 2 % (Davey, Chandrakash and O'Neill 2013), Eq. (4) gives:

$$p = -2 + 100 \left( \frac{J'}{J_{critical}} - 1 \right) \quad (5)$$

That is, if the value of permeate flux is less than 1.02 times the critical flux, the filtration has failed.

The risk factor equations are computationally convenient because at all  $J' < J_{critical}$ ,  $p > 0$ .

The novel *Fr 13* model for the cross-flow filtration of passion fruit juice is defined by Eq. (1) through (5), together with the experimentally determined value of  $J_{critical}$ .

## 2.5. Fr 13 Simulations of Cross-flow Membrane Filtration of Passion Fruit Juice

In *Fr 13* the key plant parameters are defined not by a single (best) estimate as with the traditional SVA but by a probability distribution of values, the mean of which generally agrees with the SVA (e. g. Davey, Chandrakash and O'Neill 2013; Zou and Davey 2014; Patil, Davey and Daughtry 2005).

A r-MC sampling, *Latin Hypercube*, is used to ensure sampling covers the entire range of the input distributions (see Vose 2008, Davey, Chandrakash and O'Neill 2013 and Davey 2011 for a detailed explanation). To guarantee the output distribution is *Normal*, a minimum number of samples is needed, this is usually 1,000 to 50,000 (Davey K R – unpublished data). (It is generally a simple matter to visually establish this however by inspection of the output distribution).

In the absence of specific data, it is assumed that during cross-flow filtration the practical variation about the operational mean value of both transmembrane pressure ( $\Delta P$ ) and filtration time ( $t$ ) will be a standard deviation (stdev) = 2 %. An appropriate probability distribution for practical values of  $\Delta P$  and  $t$  are therefore: **RiskNormal** (mean, stdev, **RiskTruncate** (minimum = mean - 2 x stdev), (maximum = mean + 2 x stdev)) (Davey, Chandrakash and O'Neill 2013; Davey 2011). A desired consequence of the 2 x stdev about the mean to obtain the minimum and maximum is that nearly all values (95.45 %) fall in this interval (Sullivan 2004).

For this study, the transmembrane pressure becomes:  $\Delta P = \mathbf{RiskNormal}$  (100, 2, **RiskTruncate** (96, 104)), and; filtration time:  $t = \mathbf{RiskNormal}$  (1800, 36, **RiskTruncate** (1728, 1872)). It is seen that each

distribution has been truncated to limit the range to values that can practically occur (Domingues, Ramos, Cardoso and Reis 2014; Davey, Chandrakash and O'Neill 2013).

The *Fr 13* simulations for the cross-flow membrane were performed in Microsoft Excel<sup>TM</sup> together with a commercial add-on *@Risk* (pronounced *at risk*) version 5.5 (Palisade Corporation). Because spreadsheet tools are used widely this means results can be communicated readily to a range of users of varying sophistication. All values  $p > 0$  indicate a failure of passion fruit cross-flow filtration.

## 3. RESULTS

Table 1 presents a summary comparison of results from the traditional SVA and new *Fr 13* model for cross-flow membrane filtration of passion fruit juice. The membrane parameters are given in column 1 of the table. The traditional SVA is shown in column 2. Computations can be readily read down this column.

For the *Fr 13* model 50,000 r-MC samples of  $\Delta P$  and  $t$  were simulated. This provided a sufficiently *Normal* output distribution (Figure 2).

Column 3 of Table 1 shows one only of the 50,000 filtration scenarios. It can be observed that at  $\Delta P = 98.77$  kPa and corresponding  $t = 1748.78$  s, the value of risk factor,  $p = 0.0851$  ( $> 0$ ), indicating a membrane failure.

The 50,000 process scenarios are conveniently summarized as Figure 2. It can be seen in the right-hand side of the figure that a total of 7.0 % of all filtrations of passion fruit juice over a prolonged period of time would result in  $p > 0$  i.e.  $J' < J_{critical}$ , a membrane failure.

Table 2 presents 10 selected failures for analysis. Row 6 of Table 2 (**bold-text**) is the filtration scenario given in column 3 of Table 1. An advantage of this presentation is that the individual combination of each of the values of the parameters that gave rise to filtration failure can be identified. It is seen in the table for example, Row 11,  $\Delta P = 100.698$  kPa together with  $t = 1728.9$  s will result in failure of cross-flow membrane filtration of passion fruit juice.

## 4. DISCUSSION

If each simulation scenario is considered as a (batch-continuous) processing day, a *Fr 13* failure would occur  $(3,513/(50,000 \text{ days}) \times 365.25 \text{ days / year}) = 26$  times per year, on average i.e. approximately two per month despite good operations and maintenance. These predicted failures would not of course be expected to be equally spaced in time.

It is clear from Figure 2 and Table 2 that apparent steady batch-continuous cross-flow membrane filtration of passion fruit juice should be more correctly regarded as a mix of successful and unsuccessful (failed) operations. Significantly, this information is not available from the traditional SVA.

Table 1: Summary comparison of SVA and *Fr 13* for cross-flow membrane filtration of passion fruit juice

| Parameter                                                        | SVA          | <i>Fr 13</i> model                                              |
|------------------------------------------------------------------|--------------|-----------------------------------------------------------------|
| $\Delta P$ (kPa)                                                 | <b>100</b>   | <b>98.77</b> RiskNormal (100, 2.0, RiskTruncate (96, 104))      |
| $t$ (s)                                                          | <b>1800</b>  | <b>1748.78</b> RiskNormal (1800, 36, RiskTruncate (1728, 1872)) |
| $\mu_{T20} \text{ } ^\circ\text{C}$ (Pa s)                       | 1002000      | 1002000 constant                                                |
| $A$ (m <sup>2</sup> )                                            | 0.056        | 0.056 constant                                                  |
| $K_c$ (s m <sup>-2</sup> )                                       | 56000000     | 56000000 constant                                               |
| $Q$ (m <sup>3</sup> s <sup>-1</sup> )                            | 0.0018       | 0.0018 constant                                                 |
| $J_{critical}$ (m <sup>3</sup> m <sup>-2</sup> s <sup>-1</sup> ) | 0.00000313   | 0.00000313 constant                                             |
| $T$ (°C)                                                         | 25           | 25 constant                                                     |
| $\mu_T$ (Pa s)                                                   | 812000       | 812000 constant                                                 |
| $J'$ (m <sup>3</sup> m <sup>-2</sup> s <sup>-1</sup> )           | 0.0000031494 | 0.000003129 Eq. (3)                                             |
| $p$ (dimensionless)                                              | -            | <b>0.0851</b> Eq. (5)                                           |

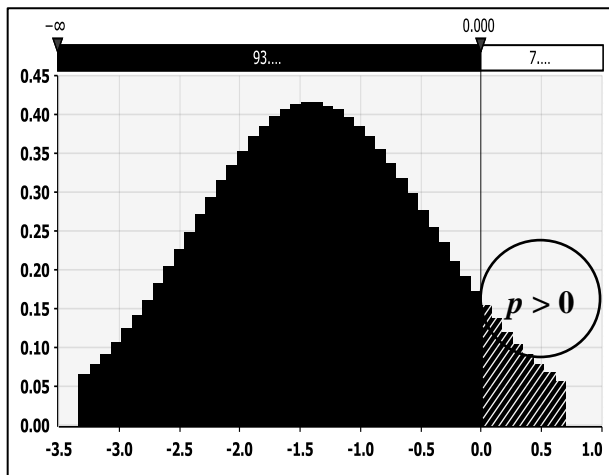


Figure 2: Simulation of the *Fr 13* risk factor ( $p$ ) for cross-flow membrane filtration of passion fruit juice with 50,000 iterations

Table 2: 10 selected *Fr 13* failures from 3,513 in 50,000 scenarios of cross-flow filtration of passion fruit juice

| $\Delta P$ (kPa) | $t$ (s)          | $J' \times 10^6$ (m <sup>3</sup> m <sup>-2</sup> s <sup>-1</sup> ) | $p$           |
|------------------|------------------|--------------------------------------------------------------------|---------------|
| 101.2051         | 1750.8902        | 3.1933                                                             | 0.0232        |
| 100.5250         | 1750.0103        | 3.1941                                                             | 0.0489        |
| 100.3328         | 1749.3012        | 3.1948                                                             | 0.0696        |
| 100.7358         | 1749.1922        | 3.1949                                                             | 0.0727        |
| <b>98.7717</b>   | <b>1748.7761</b> | <b>3.1953</b>                                                      | <b>0.0852</b> |
| 100.7191         | 1745.5379        | 3.1982                                                             | 0.1802        |
| 98.8322          | 1743.4228        | 3.2002                                                             | 0.2423        |
| 99.2111          | 1734.2506        | 3.2086                                                             | 0.5121        |
| 99.4285          | 1729.2371        | 3.2133                                                             | 0.6602        |
| 100.6980         | 1728.9496        | 3.2135                                                             | 0.6681        |

Results from further investigation with the probability distributions used for  $\Delta P$  and  $t$ , together with changes in the stdev used from 2 to 5, % about the means, did not meaningfully affect the number of predicted *Fr 13* failures in the cross-flow membrane filtration of passion fruit juice. The *Fr 13* model is therefore not sensitive to the variance about the mean value of the two key parameters.

Because the value of stdev used in the probability distributions is actually a quantitative measure of the quality, cost and accuracy of relevant filtration control (Davey 2011; Davey, Chandrakash and O'Neill 2013) this result underscores that increased control of  $\Delta P$  and  $t$  does not significantly affect the number of stochastic (random) *Fr 13* fouling failures in cross-flow filtration of passion fruit juice.

The number of *Fr 13* fouling failures would however be expected to be sensitive to the %tolerance used in defining the risk factor,  $p$  i.e. the acceptable process risk.

Repeat simulations were therefore carried out for a range of values  $2 \leq \%tolerance \leq 10$  on the risk factor, Eq. (4), and the results as number of filtration failures in 50,000 scenarios presented as Figure 3.

It is seen in the figure that the number of predicted failures decreases with increasing %tolerance until at a value 8 % there would be no failures in filtration of the passion fruit juice. This result is interpreted as the more stringent the tolerance on practical filtration the more failures in fouling.

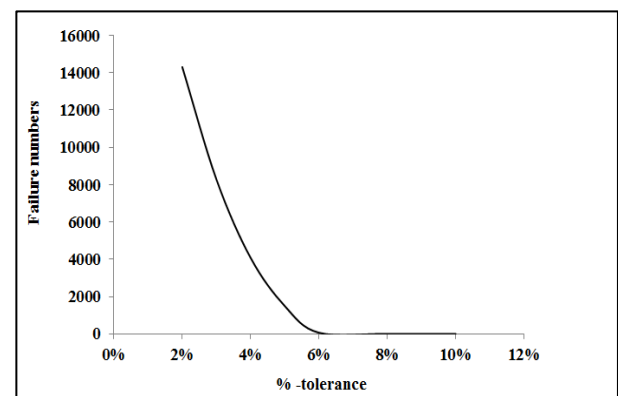


Figure 3: Effect of %tolerance on the number of filtration failures per 50,000 scenarios

A major advantage of the *Fr 13* model over the traditional SVA (with or without sensitivity analyses) (Sinnott 2005) is that all practically realizable process outcomes are actually computed. This results in significantly improved operational insight into

vulnerability to failure of processes. The use of *Fr 13* simulations in second-tier computations to reduce vulnerability to failure due to stochastic effects by making changes to design and operation has been discussed elsewhere (Davey 2011; Davey, Chandrakash and O'Neill 2013).

At present, developments are continuing in *Fr 13* methodology to analyses of two or more inter-connected unit-operations to simulate progressively more complex plant and process interactions (Davey, Chandrakash and O'Neill 2014). A process with two or more inter-connected unit-operations is defined by Davey and co-workers as a global model.

## 5. CONCLUSIONS

A new *Fr 13* quantitative risk assessment has revealed that unexpected failure will occur in 7.0 % of all batch-continuous cross-flow membrane operations in otherwise well-operated plant used in clarification of passion fruit juice.

Cross-flow filtration is revealed to actually be a combination of successful and failed operations. This insight cannot be obtained by traditional methods.

The *Fr 13* risk model provides an advance over traditional methods in that all practical scenarios that can occur are computed.

## NOMENCLATURE

The equation number given after description refers to that in which the symbol is first used or defined.

|                    |                                                                      |
|--------------------|----------------------------------------------------------------------|
| $A$                | Surface area, $m^2$ , Eq. (2)                                        |
| $J'$               | Permeate flux, $m^3 m^{-2} s^{-1}$ , Eq. (1)                         |
| $J_{critical}$     | Permeate flux filtration success, $m^3 m^{-2} s^{-1}$ , Eq. (4)      |
| $J_0$              | Permeate flux clean water, $m^3 m^{-2} s^{-1}$ , Eq. (1)             |
| $K_c$              | Parameter (passion juice = $5.6 \times 10^7$ ), $s m^{-2}$ , Eq. (1) |
| $n$                | Fouling parameter ( $n = 2$ for cross-flow), Eq. (1)                 |
| $n$                | Data (independent), Figure 1                                         |
| $p$                | Risk factor, (dimensionless), Eq. (4)                                |
| $Q$                | Water flow operational temperature, $m^3 s^{-1}$ , Eq. (2)           |
| $t$                | Filtration time, $s$ , Eq. (1)                                       |
| <b>Greek</b>       |                                                                      |
| $\Delta P$         | Transmembrane pressure, kPa, Eq. (2)                                 |
| $\mu_T$            | Viscosity clean-water at temperature, Pa s, Eq. (2)                  |
| $\mu_{20}^\circ C$ | Viscosity clean-water at $20^\circ C$ , Pa s, Eq. (2)                |

## REFERENCES

- Anon., 2011. *Blackett review of high impact low probability risks*. Government Office for Science, UK. <http://www.bis.gov.uk>
- Bowen, W.R., Calvo, J.I., Hernandez, A., 1995. Steps of membrane blocking in flux decline during protein microfiltration. *Journal of Membrane Science*, 101 (5), 153-165.
- Cassano, A., Marchio, M., Drioli, E., 2006. Clarification of blood orange juice by ultrafiltration: analyses of operating parameters, membrane fouling and juice quality. *Desalination*, 212, 15-27.
- Cui, Z.F., Jiang, Y., Field, R.W., 2010. Fundamentals of pressure-driven membrane separation processes. In: Z. F. Cui, H. S. Muralidhara, ed. *Membrane Technology: A Practical Guide to Membrane Technology and Applications in Food and Bioprocessing*. Butterworth-Heinemann, UK, 1-18.
- Davey, K.R., 2010. A novel proposal to advance the discipline and to quantitatively safeguard important hygienic bio-processing. In: *Proc. 40<sup>th</sup> Australasian Chemical Engineering Conference-CHEMECA*. Sept. 28-Oct. 1, Adelaide (South Australia, AUS).
- Davey, K.R., 2011. Introduction of fundamental and benefits of *Friday 13<sup>th</sup>* risk modelling technology for food manufacturers. *Food Australia*, 63 (5), 192-197.
- Davey, K.R., Cerf, O., 2003. Risk modelling - An explanation of *Friday 13<sup>th</sup> syndrome* (failure) in well-operated continuous sterilization plant. In: *Proc. 31<sup>st</sup> Australasian Chemical Engineering Conference-CHEMECA*. Sept. 28-Oct 1, Adelaide (South Australia, AUS).
- Davey, K.R., Chandrakash, S., O'Neill, B.K., 2013. A new risk analysis of Clean-In-Place milk processing. *Food Control*, 29 (1), 248-254.
- Davey K.R., Chandrakash, S., O'Neill, B.K., 2014. A novel *Friday 13<sup>th</sup>* risk analysis of a global food process - Application to pasteurization of milk containing *Mycobacterium avium paratuberculosis*. In: *Proc. 26<sup>th</sup> European Modeling and Simulation Symposium-EMSS 2014*. Sept. 10-12, Bordeaux (FRANCE).
- D'souza, N., Mawson, A.J., 2005. Membrane cleaning in the dairy industry: A review. *Critical Reviews in Food Science and Nutrition*, 45 (2), 125-134.
- Domingues, R.C.C., Ramos, A.A., Cardoso, V.L., Reis, M.H.M., 2014. Microfiltration of passion fruit juice using hollow fibre membranes and evaluation of fouling mechanisms. *Journal of Food Engineering*, 121, 73-79.
- Field, R.W., Wu, D., Howell, J.A., Gupta, B.B., 1995. Critical flux concept for microfiltration fouling. *Journal of Membrane Science*, 100, 259-272.
- Nandi, B.K., Das, B., Uppaluri, R., 2012. Clarification of orange juice using ceramic membrane and evaluation of fouling mechanism. *Journal of Food Processing Engineering*, 35, 403-423.
- Patil, R.A., Davey, K.R., Daughtry, B.J., 2005. A new quantitative risk assessment of a fermenter for *Friday 13<sup>th</sup>* syndrome. In: *Proc. 32<sup>nd</sup> Australasian Chemical Engineering Conference-CHEMECA*. Sept. 25-29, Brisbane (Queensland, AUS).
- Roorda, J.H., Van der Graaf, J.M., 2001. New parameter for monitoring fouling during ultrafiltration of WWTP effluent. *Water Science and Technology*, 43 (10), 241-248.
- Razi, B., Aroujalian, A., Fathizadeh, M., 2012. Modeling of fouling layer deposition in cross-flow microfiltration during tomato juice clarification. *Food and Bioprocesses Processing*, 90, 841-848.
- Snedecor, G.W., Cochran, W.G., 1989. *Statistical Methods*. 8th ed. USA: Iowa State University Press.

- Sinnott, R.K., 2005. *Chemical Engineering Design*. 2nd ed. UK: Butterworth-Heinemann.
- Sullivan, M., 2004. *Statistics-informed decision making using data*. USA: Pearson Education.
- Schafer, A.I., Andritsos, N., Karabelas, A.J., Hoek, E.M., Schneider, R., Nystrom, M., 2005. Fouling in nanofiltration. In: A. I. Schafer, T. D. Waite, A. G. Fane, ed. *Nanofiltration-principles and Applications*. UK: Elsevier, 169-239.
- Vose, D., 2008. *Risk Analysis-A Quantitative Guide*. 3rd ed. UK: John Wiley and Sons.
- Zou, W., Davey, K.R., 2014. A novel *Friday 13<sup>th</sup>* stochastic assessment of failure of membrane filtration in juice clarification. In: *Proc. 44<sup>st</sup> Australasian Chemical Engineering Conference-CHEMECA*. Sept. 28-Oct. 1, Perth (Western Australia, AUS).

### **BIOGRAPHY**

Mr Wei Zou ME(Chem) is a postgraduate research student researching with academic Kenneth Davey PhD FIChemE FIEAust CPEng CEng.

# A NOVEL FRIDAY 13<sup>TH</sup> RISK ANALYSIS OF A GLOBAL FOOD PROCESS - APPLICATION TO PASTEURIZATION OF RAW MILK CONTAINING *MYCOBACTERIUM AVIUM PARATUBERCULOSIS*

Davey K R<sup>(a)</sup>, Chandrakash S<sup>(b)</sup>, O'Neill B K<sup>(c)</sup>

School of Chemical Engineering  
The University of Adelaide, SA 5005, Australia

<sup>(a)</sup>[kenneth.davey@adelaide.edu.au](mailto:kenneth.davey@adelaide.edu.au) <sup>(b)</sup>[saravanan.chandrakash@adelaide.edu.au](mailto:saravanan.chandrakash@adelaide.edu.au)  
<sup>(c)</sup>[brian.oneill@adelaide.edu.au](mailto:brian.oneill@adelaide.edu.au)

## ABSTRACT

A novel *Fr 13* global model for pasteurization of raw milk containing *Mycobacterium avium* subsp. *paratuberculosis* that consists of three unit-operations, heating, holding and cooling, is presented for the first time and compared with traditional methods. A global model is defined by us as two or more inter-connected unit-operations. The aim was to gain quantitative insight into stochastic effects that can lead to surprise failure of an otherwise well-operated pasteurization plant. Failure is defined in terms of criteria for safe operation and a risk factor is developed for each unit-operation in terms of actual and design performance, together with a practical tolerance. Simulations are based on a refined Monte Carlo sampling of key parameters. Results reveal that, with a tolerance overall of 3 %, pasteurization is vulnerable to failure in 8.4 % of all (batch) continuous operations over a prolonged time. This insight cannot be obtained from traditional methods.

Keywords: pasteurization failure, global milk model, process failure, *Fr 13<sup>th</sup>* risk modelling

## 1. INTRODUCTION

The food industries are very important to Australia as a major exporter. In particular, Australia is a major exporter of milk and milk products. Raw milk is an excellent substrate for pathogen growth (Madigan and Martinko 2006). Unexpected (surprise) failure of pasteurization can therefore have a serious impact on consumer health (with or without fatalities). According to the recent *Blackett Review* (Anon 2011) low probability-high impact failures are a growing theoretical and practical challenge for processors in a wide range of industries. In milk pasteurization, unexpected failures are acknowledged and real events.

In recent years Davey and co-workers have illustrated a novel risk assessment titled *Friday 13<sup>th</sup> syndrome* (*Fr 13*) to address acknowledged, unexpected (surprise) failure in otherwise well-operated plant (Davey and Cerf 2003; Davey, Chandrakash and O'Neill 2013; Abdul Halim and Davey 2014; Chandrakash, Davey and O'Neill 2014). Unit-operation

case studies include surprise fermenter washout (Patil, Davey and Daughtry 2005), an unexpected and sudden shift from potable to non-potable water (Davey, Abdul Halim and Lewis 2012) and sudden shift from safe to unsafe Clean-In-Place processing in milk processing (Davey, Chandrakash and O'Neill 2013). Their underlying research hypothesis is that: *random changes in process parameters can accumulate in one direction in amounts sufficient to leverage significant change and thereby make processes vulnerable to sudden and unexpected (surprise) failure.*

A major advantage of a *Fr 13* study is that it provides quantitative insight into plant behaviour that can be used to devise process intervention and re-design strategies to reduce risk of unexpected failure; it can be applied at both the analysis and synthesis stages (Davey 2011). It is timely that this new methodology be applied to several inter-connected unit-operations that define a process. A global model was defined by Davey and co-workers as two or more inter-connected unit-operations.

Chandrakash, Davey and O'Neill (2014) recently presented a simplified global *Fr 13* analysis for the first time to heating, holding and cooling of (batch) continuous pasteurization of milk. The justification was that raw milk is processed universally (and substantial commercial data are available for model validation). Importantly, the multiplicity of manufacturing steps and linkages provided a stringent test for the research hypothesis. Results from this study revealed that pasteurization is actually a continuous mix of successful and failed operations, with about 11.5 % likely to fail over the long term, despite good design and maintenance. However, because their definition for failure was based on plant physical parameters they argued this result could be misleading. It became apparent that failure must take account of microbiological kinetics of any contaminant pathogen.

Here the work of Chandrakash, Davey and O'Neill (2014) is developed further for real-time processing. Failure in the holding unit-operation is defined as a reduction in viable pathogen numbers and not residence time of the milk in the holding tube. It is hoped that insight gained from this study can be applied to improve

pasteurization design, and; findings be more widely generalized for foods processing.

The *Fr 13* global model for raw milk pasteurization is predicated on inter-connection of the underlying unit-operations, heating, holding and cooling, together with an unambiguous definition of failure and a refined Monte Carlo (r-MC) (*Latin Hypercube*) sampling of parameters. A comparison is made with traditional solution methods.

## 2. MATERIALS AND METHODS

### 2.1 Raw milk pasteurization as inter-connected unit-operations of Heating (PHE-1), Holding (H-2) and Cooling (PHE-3)

Figure 1 is a schematic of raw milk pasteurization with heating (PHE-1), holding (H-2) and cooling (PHE-3) unit-operations. Each unit-operation is analyzed separately and then overall. All symbols used are carefully defined in the Nomenclature.

#### 2.1.1. Heating (PHE-1)

An adequate unit-operations model for raw milk heating can be based on a Plate Heat Exchanger (PHE) (Chandrakash, Davey and O'Neill 2014). A PHE consists of a stack of closely-spaced (thin) plates clamped in a frame (Sinnott 2005). Advantages over shell-and-tube exchangers include cost, maintenance and flexibility (Sinnott 2005).

Based on commercial practice  $A_{PHE-1} = 33 \text{ m}^2$  with heat transfer co-efficient  $U_{PHE-1} = 2.5 \text{ kW m}^{-2} \text{ K}^{-1}$  and temperature correction factor  $F_{t-1} = 0.92$  (Kothandaraman and Subramanyan 2007; Sinnott 2005) to give  $\Delta T_{LMTD-1} = 20 \text{ }^\circ\text{C}$ . The individual plates are length  $L_{p,1} = 1.5 \text{ m}$  and thickness  $w_{p,1} = 0.15 \text{ m}$ . The gap between plates is  $b_{p,1} = 0.05 \text{ m}$  (Anon 2014).

Milk (typically) enters PHE-1 at  $T_{i,m-1} = 4 \text{ }^\circ\text{C}$  at a flow  $m_{m-1} = 5.56 \text{ kg s}^{-1}$  and a specific heat  $C_{p,m-1} = 3.99 \text{ kJ kg}^{-1} \text{ K}^{-1}$  (Kessler 2002). On the water-side  $t_{i,w-1} = 90 \text{ }^\circ\text{C}$  and  $m_{w-1} = 6.2 \text{ kg s}^{-1}$  (Kessler 2002). Because the physical properties of water do not vary significantly with temperature over a range of  $5 \text{ }^\circ\text{C} - 120 \text{ }^\circ\text{C}$ , it is assumed that  $C_{p,w-1} = 4.2 \text{ kJ kg}^{-1} \text{ K}^{-1}$  (Perry and Green 1997).

The generalized heat transfer equations for PHE-1 are (Sinnott 2005):

$$\dot{Q} = m_{m-1} C_{p,m-1} (T_{o,m-1} - T_{i,m-1}) \quad (1-1)$$

$$\dot{Q} = m_{w-1} C_{p,w-1} (t_{i,w-1} - t_{o,w-1}) \quad (1-2)$$

$$\dot{Q} = U_{PHE-1} A_{PHE-1} \Delta T_{LMTD-1} F_{t-1} \quad (1-3)$$

The required equivalent plate diameter ( $D_{e,PHE-1}$ ) is given by (Kessler 2002):

$$D_{e,PHE-1} = \frac{2w_{p,1}b_{p,1}}{(w_{p,1}+b_{p,1})} \quad (1-4)$$

The required number of plates ( $n_{PHE-1}$ ) for this PHE-1 can be determined using (Sinnott 2005):

$$n_{PHE-1} = \frac{A_{PHE-1}}{\pi D_{e,PHE-1} L_{p,1}} \quad (1-5)$$

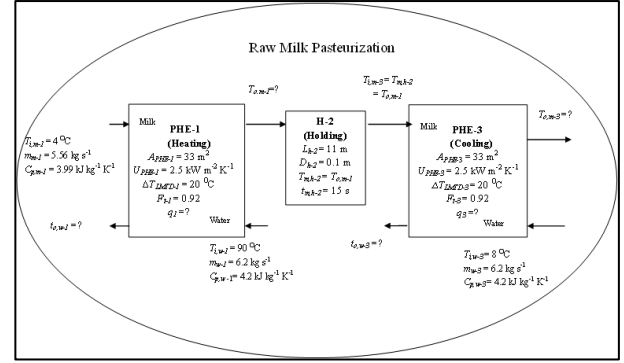


Figure 1: A Global Model for Pasteurization of Raw Milk Containing *Mycobacterium avium paratuberculosis*

The physical properties of milk can be determined as a function of temperature and are obtained using (Al-Hilphy and Ali 2013):

$$T_{m,avg-1} = \frac{T_{i,m-1} + T_{o,m-1}}{2} \quad (1-6)$$

$$\mu_{m,1} = \left( (-0.00445 T_{m,avg-1}) + 0.947 \right) * 10^{-3} \quad (1-6 a)$$

$$\rho_{m,1} = 1033.7 - 0.2308 T_{m,avg-1} - 0.00246 T_{m,avg-1}^2 \quad (1-6 b)$$

Milk velocity ( $v_{m,1}$ ) is estimated from the continuity equation of Kessler (2002):

$$v_{m,1} = \frac{4m_{m-1}}{\rho_{m,1} \pi D_{e,PHE-1}^2} \quad (1-7)$$

The recommended flow rate for raw milk in a PHE is  $0.5 < v_{m,1} < 1.5 \text{ m s}^{-1}$  (Kessler 2002) in turbulent flow ( $Re > 4,000$ ) (Hayhurst 1997; Maroulis and Saravacos 2003). The Reynolds number ( $Re_1$ ) of the milk is given by (Sinnott 2005):

$$Re_1 = \frac{\rho_{m,1} v_{m,1} D_{e,PHE-1}}{\mu_{m,1}} \quad (1-8)$$

The residence time of the milk ( $t_{r,1}$ ) in PHE-1 is computed from (Kessler 2002):

$$t_{r,1} = \frac{n_{PHE-1} L_{p,1}}{v_{m,1}} \quad (1-9)$$

The pressure drop in PHE-1 can be calculated from (Kumbhare and Dawande 2013; Sinnott 2005):

$$\Delta P_{PHE-1} = 8j_{F,1} \frac{L_{p,1}}{D_{e,PHE-1}} \frac{\rho_{m,1} v_{m,1}^2}{2} \quad (1-10)$$

where the term  $j_{F,1}$  is the friction factor of the liquid flowing through the plate; its value depends on the type of plate used. Generally for turbulent flow ( $Re > 4,000$ ) the term  $j_{F,1}$  can be described by (Sinnott 2005):

$$j_{F,1} = 0.6 Re_1^{-0.30} \quad (1-11)$$

Eqs. (1-1) through Eq. (1-11) define the heating unit-operation PHE-1 of the global model for milk pasteurization.

#### 2.1.2. Holding (H-2)

As is seen from Figure 1, milk leaving PHE-1 is held at a constant temperature for a fixed time. The general



design equation for the holding tube (H-2) (Katoh and Yoshida 2009) is:

$$t_{m,h-2} = \frac{L_{h-2}}{v_{m,h-2}} \quad (2-1)$$

The external holding tube (H-2) generally consists of a tube in a spiral pattern (Sinnott 2005; Smith 2011). The most widely used temperature-time combination is  $T_{m,h-2} = 72$  °C and  $t_{m,h-2} = 15$  s (Alfa Laval 1987; Juffs and Deeth 2007; Katoh and Yoshida 2009). It is assumed that the diameter of the holding tube  $D_{h-2} = 0.1$  m (Berk 2009) with length  $L_{h-2} = 11$  m. In steady-state flow, milk will enter the holding tube with a temperature equivalent to  $T_{o,m-1}$  and a flow of  $m_{m-2} = 5.56$  kg s<sup>-1</sup>.

Because the temperature  $T_{m,h-2}$  is known, milk density in the holding tube can be calculated using the correlation of Al Hilphy and Ali (2013):

$$\rho_{m,2} = 1033.7 - 0.2308 T_{m,h-2} - 0.00246 T_{m,h-2}^2 \quad (2-2)$$

Milk velocity ( $v_{m,h-2}$ ) in the holding tube is obtained from the continuity equation (Kessler 2002):

$$v_{m,h-2} = \frac{4m_{m-2}}{\rho_{m,2}\pi D_{h-2}^2} \quad (2-3)$$

There are several microbial species of concern that have the ability to survive pasteurization: *Mycobacterium avium* subsp. *paratuberculosis*, *Bacillus cereus*, *Brucella melitensis*, *Enterobacter sakazakii*, *Staphylococcus aureus*, *Streptococcus agalactiae*, and *Streptococcus zooepidemicus* (Juffs and Deeth 2007). Reliable data are available in the literature that can be used to model *Mycobacterium avium paratuberculosis*. This micro-organism consumed in small dosages can have lethal effects (Hammer, Kiesner and Walte 2014).

The decimal reduction time for this micro-organism at a reference temperature  $T_{ref,h-2} = 72$  °C is  $D_{t,ref,h-2} = 1.2$  s with a z-value  $z_{h-2} = 7.7$  °C (Rademaker, Vissers and Giffel 2007). The decimal reduction time ( $D_{t,h-2}$ ) at any temperature value  $T_{m,h-2} = T_{o,m-1}$  in H-2 can be obtained from (van Asselt and Zwietering 2005):

$$\log_{10} D_{t,h-2} = \log_{10} D_{t,ref,h-2} - \frac{T_{m,h-2} - T_{ref,h-2}}{z_{h-2}} \quad (2-4)$$

The microbial death rate ( $k_{d,h-2}$ ) for *Mycobacterium avium paratuberculosis* in H-2 can be obtained using (Koutchma, Bail and Ramaswamy 2001; Smith 2011):

$$k_{d,h-2} = \frac{2.303}{D_{t,h-2}} \quad (2-5)$$

The logarithmic reduction in *Mycobacterium avium paratuberculosis* ( $\log_{10} (N/N_0)_{h-2}$ ) in H-2 can be obtained from (Ibarz and Barbosa-Canovas 2003):

$$\log_{10} \left( \frac{N}{N_0} \right)_{h-2} = \frac{k_{d,h-2} t_{m,h-2}}{2.303} \quad (2-6)$$

The steps involved in computation of  $D_{t,h-2}$  are detailed in Appendix C. Typically a 5.5 log<sub>10</sub> reduction is required (Hammer, Kiesner and Walte 2014; McDonald et al. 2005; Rademaker, Vissers and te Giffel 2007).

Eqs. (2-1) through Eq. (2-6) define the holding unit-operation H-2 for raw milk pasteurization.

### 2.1.3. Cooling (PHE-3)

Milk leaving H-2 is cooled in a second plate heat exchanger PHE-3. This is defined by  $A_{PHE-3} = 33$  m<sup>2</sup> with  $U_{PHE-3} = 2.5$  kW m<sup>-2</sup> K<sup>-1</sup> and  $F_{t-3} = 0.92$  (Kothandaraman and Subramanyan 2007; Sinnott 2005) to give  $\Delta T_{LMTD-3} = 20$  °C. The individual plates are the same dimensions as PHE-1 (Anon 2014).

From Figure 1 it can be seen that milk enters PHE-3 at  $T_{i,m-3} = T_{m,h-2}$  with  $m_{m-3} = m_{m-2} = 5.56$  kg s<sup>-1</sup>. On the cooling-side (typically) water is supplied at  $t_{i,w-3} = 8$  °C with a flow  $m_{w-3} = 6.2$  kg s<sup>-1</sup>. The specific heat of water is a constant  $C_{p,w-3} = 4.2$  kJ kg<sup>-1</sup> K<sup>-1</sup> (Perry and Green 1997).

The generalized heat transfer design equations for PHE-3 are (Sinnott 2005):

$$3 = m_{m-3} C_{p,m-3} (T_{i,m-3} - T_{o,m-3}) \quad (3-1)$$

$$3 = m_{w-3} C_{p,w-3} (t_{o,w-3} - t_{i,w-3}) \quad (3-2)$$

$$3 = U_{PHE-3} A_{PHE-3} \Delta T_{LMTD-3} F_{t-3} \quad (3-3)$$

The equivalent plate diameter  $D_{e,PHE-3}$  is given by (Kessler 2002):

$$D_{e,PHE-3} = \frac{2w_{p,3}b_{p,3}}{(w_{p,3}+b_{p,3})} \quad (3-4)$$

The required number of plates ( $n_{PHE-3}$ ) for PHE-3 can be determined by using (Sinnott 2005):

$$n_{PHE-3} = \frac{A_{PHE-3}}{\pi D_{e,PHE-3} L_{p,3}} \quad (3-5)$$

The physical properties of milk can be determined as a function of temperature and can be obtained using (Al-Hilphy and Ali 2013):

$$T_{m,avg-3} = \frac{T_{i,m-3} + T_{o,m-3}}{2} \quad (3-6)$$

$$\mu_{m,3} = ((-0.00445 T_{m,avg-3}) + 0.947) * 10^{-3} \quad (3-6 a)$$

$$\rho_{m,3} = 1033.7 - 0.2308 T_{m,avg-3} - 0.00246 T_{m,avg-3}^2 \quad (3-6 b)$$

Milk velocity  $v_{m,3}$  in PHE-3 is obtained from the continuity equation of Kessler (2002):

$$v_{m,3} = \frac{4m_{m-3}}{\rho_{m,3}\pi D_{e,PHE-3}^2} \quad (3-7)$$

Typical values are  $0.5 < v_{m,3} < 1.5$  m s<sup>-1</sup> (Kessler 2002) in turbulent flow ( $Re > 4,000$ ) (Hayhurst 1997; Kessler 2002; Lewis and Heppell 2000; Maroulis and Saravacos 2003). The Reynolds number ( $Re_3$ ) of the milk is obtained from (Sinnott 2005):

$$Re_3 = \frac{\rho_{m,3} v_{m,3} D_{e,PHE-3}}{\mu_{m,3}} \quad (3-8)$$

The residence time of the milk ( $t_{r,3}$ ) is computed from (Kessler 2002):

$$t_{r,3} = \frac{n_{PHE-3} L_{p,3}}{v_{m,3}} \quad (3-9)$$

The pressure drop in the PHE-3 is given by (Kumbhare and Dawande 2013; Sinnott 2005):

$$\Delta P_{PHE-3} = 8j_{F,3} \frac{L_{p,3}}{D_{e,PHE-3}} \frac{\rho_{m,3} v_{m,3}^2}{2} \quad (3-10)$$

where the term  $j_{F,3}$  is the friction factor of the liquid flowing through the plate and can be estimated using (Sinnott 2005):

$$j_{F,3} = 0.6Re_3^{-0.30} \quad (3-11)$$

Eqs. (3-1) through Eq. (3-11) define the cooling unit-operation PHE-3 for milk pasteurization.

## 2.2. Traditional Solution Method (SVA)

The traditional method for solving a unit-operations model is the traditional single point approach or Single Value Assessment (SVA) method (Sinnott 2005).

For PHE-1: From Eq. (1-3)  $q_1 = 1,518 \text{ kJ s}^{-1}$ , using Eq. (1-1),  $T_{o,m-1} = 72.43 \text{ }^\circ\text{C}$  and from Eq. (1-2)  $t_{o,w-1} = 31.71 \text{ }^\circ\text{C}$ . From Eq. (1-4)  $D_{e,PHE-1} = 0.075 \text{ m}$  and from Eq. (1-5)  $n_{PHE-1} = 93$ . Using Eq. (1-6)  $T_{m,avg-1} = 38.18 \text{ }^\circ\text{C}$ . With  $T_{m,avg-1}$  known using Eq. (1-6 a)  $\mu_{m,1} = 0.00078 \text{ Pa s}$  and from Eq. (1-6 b)  $\rho_{m,1} = 1,021.29 \text{ kg m}^{-3}$ . From Eq. (1-7)  $v_{m,1} = 1.2329 \text{ m s}^{-1}$  and from Eq. (1-8)  $Re_1 = 121,073$ . From Eq. (1-9)  $t_{r,1} = 113.15 \text{ s}$ . Using Eq. (1-10) and (1-11)  $\Delta P_{PHE-1} = 2,226.14 \text{ N m}^{-2}$ .

For H-2: Milk enters the holding tube at  $T_{m,h-2} = 72.43 \text{ }^\circ\text{C}$  (same as  $T_{o,m-1}$ ). Using Eq. (2-2)  $\mu_{m,2} = 1,004.13 \text{ kg m}^{-3}$ . From Eq. (2-3)  $v_{m,h-2} = 0.7054 \text{ m s}^{-1}$ . Since the length of the holding tube ( $L_{h-2}$ ) is known, using Eq. (2-1)  $t_{h-2} = 15.59 \text{ s}$ . Since  $D_{t,ref,h-2} = 1.2 \text{ s}$ ,  $T_{ref,h-2} = 72 \text{ }^\circ\text{C}$  and  $z_{h-2} = 7.7 \text{ }^\circ\text{C}$ , using Eq. (2-4) at  $T_{m,h-2} = 72.43 \text{ }^\circ\text{C}$ ,  $D_{t,h-2} = 1.0236 \text{ s}$ . From Eq. (2-5)  $k_{d,h-2} = 2.2499 \text{ s}^{-1}$ . Using Eq. (2-6)  $\log_{10}(N/N_0)_{h-2} = 14.6483 > 5.5$ . The pasteurized milk is therefore "safe".

For PHE-3: From Eq. (3-3)  $q_3 = 1,518 \text{ kJ s}^{-1}$ , using Eq. (3-1)  $T_{o,m-3} = 3.91 \text{ }^\circ\text{C}$  and from Eq. (3-2)  $t_{o,w-3} = 66.3 \text{ }^\circ\text{C}$ . Using Eq. (3-4)  $D_{e,PHE-3} = 0.075 \text{ m}$  and from Eq. (3-5)  $n_{PHE-3} = 93$ . Using Eq. (3-6)  $T_{m,avg-3} = 38.13 \text{ }^\circ\text{C}$ . With  $T_{m,avg-3}$  known, from Eq. (3-6 a)  $\mu_{m,3} = 0.00078 \text{ Pa s}$  and from Eq. (3-6 b)  $\rho_{m,3} = 1,021.29 \text{ kg m}^{-3}$ . Using Eq. (3-7)  $v_{m,3} = 1.2329 \text{ m s}^{-1}$  and from Eq. (3-8)  $Re_3 = 121,073$ . From Eq. (3-9)  $t_{r,3} = 113.15 \text{ s}$ . Using Eq. (3-10) and (3-11),  $\Delta P_{PHE-3} = 2,226.14 \text{ N m}^{-2}$ .

The SVA solution is summarized as Table A-1, Appendix A. The bold-text values in the table underscore that the output from one unit-operation is the input to the next.

## 3. FR 13 RISK MODEL

### 3.1. Defining failure

For PHE-1 a suitable risk factor ( $p_1$ ) can be defined in terms of the outlet milk temperature ( $T_{o,m-1}$ ) together with an acceptable tolerance (Davey 2011; Chandrakash, Davey and O'Neill 2014; Davey, Chandrakash and O'Neill 2013). such that:

$$p_1 = -\%tolerance_1 + 100\left(1 - \frac{T_{o,m-1}'}{T_{o,m-1}}\right) \quad (1-12)$$

where  $T_{o,m-1}'$  is the actual temperature that is obtained due to the impact of stochastic effects (or more strictly, the particular r-MC scenario value). With a practical  $\%tolerance = 3 \%$  (Brigitte Carpentier, Laboratoire de securite sanitaire de Maisons-Alfort, France, *pers. comm.*), Equation (1-12) becomes:

$$p_1 = -3 + 100\left(1 - \frac{T_{o,m-1}'}{T_{o,m-1}}\right) \quad (1-13)$$

That is, if the design temperature plus 3 % is not reached then the heating unit-operation is said to have failed. Eq. (1-13) is computationally convenient (Davey 2011) because all  $p_1 > 0$  are failures.

Eqs. (1-1) through (1-11), together with Eq. (1-13), define the *Fr 13* model for PHE-1.

Similarly, for H-2,  $p_2$  can be defined such that:

$$p_2 = -\%tolerance_2 + 100\left(1 - \frac{\log_{10}\left(\frac{N}{N_0}\right)_{h-2}'}{\log_{10}\left(\frac{N}{N_0}\right)_{h-2}}\right) \quad (2-7)$$

where  $\log_{10}(N/N_0)_{h-2}'$  is the  $\log_{10}$  reduction due random effects. With  $\%tolerance_2 = 3 \%$  on Eq. (2-7) becomes:

$$p_2 = -3 + 100\left(1 - \frac{\log_{10}\left(\frac{N}{N_0}\right)_{h-2}'}{\log_{10}\left(\frac{N}{N_0}\right)_{h-2}}\right) \quad (2-8)$$

That is, if the required 5.5  $\log_{10}$  reduction plus 3 % is not reached then the holding unit-operation is said to have failed since all  $p_2 > 0$ .

Eqs. (2-1) through (2-6), together with Eq. (2-8), is seen to constitute the *Fr 13* model for H-2.

For PHE-3,  $p_3$  can be defined in terms of outlet temperature of milk from PHE-3 ( $T_{o,m-3}$ ) such that:

$$p_3 = -\%tolerance_3 + 100\left(\frac{T_{o,m-3}'}{T_{o,m-3}} - 1\right) \quad (3-12)$$

With  $\%tolerance_3 = 3 \%$ , Eq. (3-12) becomes:

$$p_3 = -3 + 100\left(\frac{T_{o,m-3}'}{T_{o,m-3}} - 1\right) \quad (3-13)$$

That is, if the design temperature plus 3 % is not reached then the cooling unit-operation is said to have failed revealed by all  $p_3 > 0$ .

Eqs. (3-1) through (3-11), together with Eq. (3-13), are seen to constitute the *Fr 13* model for PHE-3.

### 3.2. Fr 13 simulations

In contrast to the traditional SVA simulation, in *Fr 13* simulation each process parameter is defined as a probability distribution of values, the mean of which will agree with the SVA (Patil, Davey and Daughtry 2005; Patil 2006; Davey 2011; Chandrakash, Davey and O'Neill 2014; Davey, Chandrakash and O'Neill 2013; 2011).

In the absence of specific information, the probability distributions for the input parameters for pasteurization of the milk are defined as: **RiskNormal** (mean, standard deviation, **RiskTruncate** (minimum, maximum)). (Some 40 distribution types are available *see* for example Vose 2008; Davey and Cerf 2003). A standard deviation around the mean was assumed at  $stdev = 5 \%$  and the minimum and maximum practical values defined by  $\pm 3 \times stdev$  on this mean value. An advantage of using  $3 \times stdev$  about the mean to obtain the minimum and maximum values is that nearly all values (99.73 %) the parameter can take will fall in this interval (Vose 2008).

For example for the inlet milk temperature to PHE-1 ( $T_{i,m-1}$ ) the risk function becomes: **RiskNormal** (4, 0.2, **RiskTruncate** (3.4, 4.6)).

The other parameters are similarly defined and are shown in the *Fr 13* global model schematic for

pasteurization of raw milk containing *Mycobacterium avium paratuberculosis* as Figure 2.

A r-MC sampling (*Latin Hypercube*) of the probability distributions is used because ‘pure’ MC sampling can over- and under- sample from parts of the distribution (Vose 2008; Davey 2011). To ensure that the output distribution is sufficiently *Normal*, a minimum number of random samples is needed; this is usually 1,000 to 50,000 samples (K R Davey – unpublished data). This is simple to establish visually.

For each of heating, holding and cooling, simulations were carried in Microsoft Excel™ with a commercially available add-on @-Risk (pronounced at risk) version 5.5 (Palisade Corporation™). An advantage is that because spread-sheeting tools are used universally this makes ready communication of results (Davey 2011).

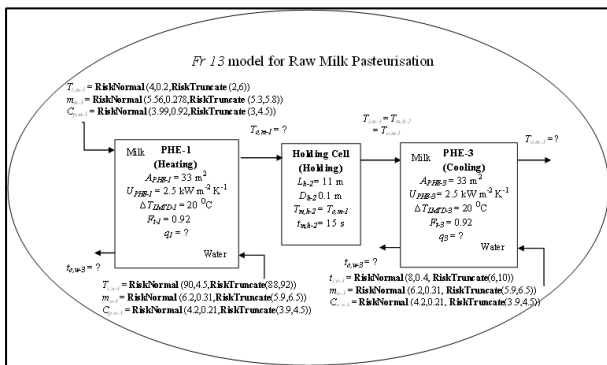


Figure 2: The *Fr 13* Global Model for Pasteurization of Raw Milk Containing *Mycobacterium avium paratuberculosis*

#### 4. RESULTS

Detailed *Fr 13* simulations for each of heating, holding and cooling, and a comparison with traditional SVA, are presented as Tables B-1, B-2 and B-3 in Appendix-B. 1,000 Latin Hypercube samples were needed to ensure that the outputs were sufficiently *Normal* for all three. From Appendix B it can be readily appreciated that one only scenario can be reported using tabular methods for comparison.

However, process scenarios can be conveniently presented as easily digested output distributions (Davey 2011; 2010). For example, Figure 3 presents all 1,000 *Fr 13* output scenarios for heating, PHE-1. The figure is seen to be convenient because all process failures ( $p_1 > 0$ ) can be visually identified. The right-side of the figure shows all 308 failures,  $p_1 > 0$ , identified to meet the required outlet temperature of raw milk  $T_{o,m-1} = 72$  °C due to stochastic effects.

With these 308 temperature values as inputs to the holding tube (H-2) and  $\%tolerance = 3$  % there were 44 failures to meet the required logarithmic reduction of *Mycobacterium avium paratuberculosis*  $\log_{10} (N/N_0)_{h-2} = 5.5$  as was evidenced by all  $p_2 > 0$ . However, with the same 308 output values as inputs to PHE-3, there were 84 failures ( $p_3 > 0$ ) to meet the specified milk outlet

temperature ( $T_{o,m-3} = 4$  °C) due to random effects in PHE-3.

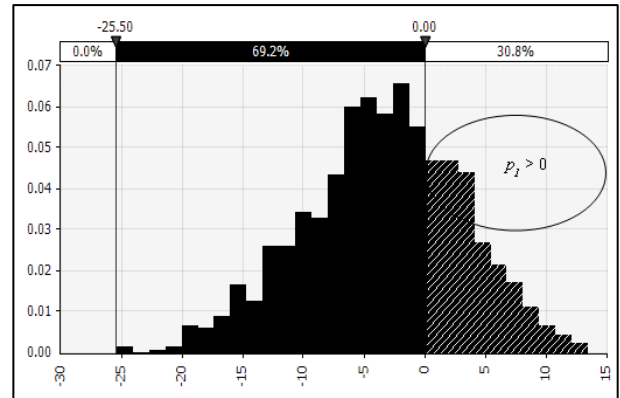


Figure 3: *Fr 13* Simulation of Risk Factor ( $p_1$ ) for Heating Raw Milk Containing *Mycobacterium avium paratuberculosis* with 1,000 Simulated Scenarios

Table 1 represents four (4) out of 308 failures in PHE-1. It can be seen from Table 1 that the outlet temperature of milk ( $T_{o,m-1}$ ) from PHE-1 is  $< 72$  °C and this resulted in  $p_1 > 0$ . A major advantage of this presentation is that the value of each of the contributing parameters to each particular process failure can be identified. For example, failure 2, column 3 (**bold-text**) of the table shows an inlet raw milk temperature of  $T_{i,m-1} = 4.3971$  °C in combination with a milk flow of  $m_{m-1} = 5.8656$  kg s<sup>-1</sup> and  $C_{p,m-1} = 3.96$  kJ kg<sup>-1</sup> K<sup>-1</sup>, together with an inlet heating water temperature of  $t_{i,w-1} = 95.23$  °C in combination with a water flow of  $m_{w-1} = 6.7962$  kg s<sup>-1</sup> and  $C_{p,w-1} = 4.3609$  kJ kg<sup>-1</sup> K<sup>-1</sup>, which gave rise to a heat duty of  $q_1 = 1,518$  kJ s<sup>-1</sup>, a water outlet temperature  $t_{o,w-1} = 44.01$  °C, milk velocity  $v_{m-1} = 1.3001$  m s<sup>-1</sup>,  $Re_1 = 127,398$ , residence time  $t_{r-1} = 107.78$  s, pressure drop  $\Delta P_{PHE-1} = 2,437.70$  N m<sup>-2</sup>, milk outlet temperature  $T_{o,m-1} = 69.75$  °C and  $p_1 = 0.6951$  (failure).

Table 1: Four of 308 Failures ( $p_1 > 0$ ) in PHE-1

| Parameter                                          | <i>Fr 13</i> in PHE-1 (Heating) |                 |          |          |
|----------------------------------------------------|---------------------------------|-----------------|----------|----------|
|                                                    | 1                               | 2               | 3        | 4        |
| $T_{i,m-1}$ (°C)                                   | 4.0303                          | <b>4.3971</b>   | 4.2765   | 3.7058   |
| $m_{m-1}$ (kg s <sup>-1</sup> )                    | 5.6788                          | <b>5.8656</b>   | 5.9014   | 6.2506   |
| $C_{p,m-1}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 4.0422                          | <b>3.96</b>     | 4.2557   | 4.0798   |
| $t_{i,w-1}$ (°C)                                   | 91.93                           | <b>95.23</b>    | 90.09    | 90.66    |
| $m_{w-1}$ (kg s <sup>-1</sup> )                    | 5.6231                          | <b>6.7962</b>   | 5.6392   | 6.8461   |
| $C_{p,w-1}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 3.8557                          | <b>4.3609</b>   | 4.2418   | 4.3988   |
| $q_1$ (kJ s <sup>-1</sup> )                        | 1,518                           | <b>1,518</b>    | 1,518    | 1,518    |
| $t_{o,w-1}$ (°C)                                   | 21.91                           | <b>44.01</b>    | 26.63    | 40.26    |
| $v_{m-1}$ (m s <sup>-1</sup> )                     | 1.2587                          | <b>1.3001</b>   | 1.3067   | 1.3834   |
| $Re_1$                                             | 123,352                         | <b>127,398</b>  | 126,324  | 133,030  |
| $\Delta P_{PHE-1}$ (N m <sup>-2</sup> )            | 2,307.15                        | <b>2,437.70</b> | 2,471.30 | 2,728.60 |
| $t_{r-1}$ (s)                                      | 111.33                          | <b>107.78</b>   | 107.24   | 101.29   |
| $T_{o,m-1}$ (°C)                                   | 70.15                           | <b>69.75</b>    | 64.72    | 63.23    |
| $p_1$                                              | 0.1417                          | <b>0.6951</b>   | 7.6411   | 9.6942   |

Table 2 reports the above same reported four (4) failed scenarios as inputs to H-2 (holding). It can be seen from

Table 2 that not all  $p_2 > 0$ . This underscores that not all scenarios from PHE-1 resulted in failure in H-2 i.e. the viable load ( $\log_{10}(N/N_0)_{h-2}$ ) can be achieved if milk is held for a sufficiently long period in the holding tube. Table 2 presents four (4) of the total 44 failures in H-2.

Table 2: Four Fail Scenarios from PHE-1 Input to H-2

| Parameter                            | Fr 13 in H-2 (Holding) |          |          |          |
|--------------------------------------|------------------------|----------|----------|----------|
|                                      | 1                      | 2        | 3        | 4        |
| $T_{m,h-2}$ ( $^{\circ}\text{C}$ )   | 70.15                  | 69.75    | 64.72    | 63.23    |
| $m_{m-2}$ ( $\text{kg s}^{-1}$ )     | 5.6788                 | 5.8656   | 5.9014   | 6.2506   |
| $\rho_{m-2}$ ( $\text{kg m}^{-3}$ )  | 1,005.40               | 1,005.63 | 1,008.46 | 1,009.27 |
| $v_{m,h-2}$ ( $\text{m s}^{-1}$ )    | 0.7195                 | 0.7430   | 0.7455   | 0.7889   |
| $L_{h-2}$ (m)                        | 11                     | 11       | 11       | 11       |
| $D_{h-2}$ (m)                        | 0.1                    | 0.1      | 0.1      | 0.1      |
| $T_{ref,h-2}$ ( $^{\circ}\text{C}$ ) | 72                     | 72       | 72       | 72       |
| $D_{t,ref,h-2}$ (s)                  | 1.2                    | 1.2      | 1.2      | 1.2      |
| $z_{h-2}$ ( $^{\circ}\text{C}$ )     | 7.7                    | 7.7      | 7.7      | 7.7      |
| $D_{t,h-2}$ (s)                      | 1.3762                 | 1.4497   | 2.7863   | 3.3798   |
| $k_{d,h-2}$ ( $\text{s}^{-1}$ )      | 1.6735                 | 1.5886   | 0.8266   | 0.6814   |
| $\log_{10}(N/N_0)_{h-2}$             | 11.1091                | 10.2120  | 5.2960   | 4.1254   |
| $t_{m,h-2}$ (s)                      | 15.29                  | 14.81    | 14.76    | 13.94    |
| $p_2$                                | -104.98                | -85.67   | 0.7096   | 21.9935  |

Table 3 summarizes the four (4)  $T_{m,h-2}$  outputs from (H-2) as inputs to PHE-3 (cooling). The tabulated data reveal that not all four (4) scenarios will fail to reach the design cooling temperature of  $4^{\circ}\text{C}$  ( $T_{o,m-3}$ ) i.e. not all  $p_3 > 0$ . Table 3 presents four (4) of 84 *Fr 13* failures in PHE-3 i.e. an overall failure of 8.4 % in the global model for pasteurization of raw milk.

Table 3: Four Scenarios from H-2 Input to PHE-3

| Parameter                                         | Fr 13 in PHE-3 (Cooling) |        |        |          |
|---------------------------------------------------|--------------------------|--------|--------|----------|
|                                                   | 1                        | 2      | 3      | 4        |
| $T_{i,m-3}$ ( $^{\circ}\text{C}$ )                | 70.15                    | 69.75  | 64.72  | 63.23    |
| $m_{m-3}$ ( $\text{kg s}^{-1}$ )                  | 5.6788                   | 5.8656 | 5.9014 | 6.2506   |
| $C_{p,m-3}$ ( $\text{kJ kg}^{-1} \text{K}^{-1}$ ) | 4.0428                   | 3.9600 | 4.2557 | 4.0798   |
| $t_{i,w-3}$ ( $^{\circ}\text{C}$ )                | 8.0981                   | 8.1245 | 8.4345 | 7.9953   |
| $m_{w-3}$ ( $\text{kg s}^{-1}$ )                  | 6.48                     | 6.44   | 6.28   | 6.12     |
| $C_{p,w-3}$ ( $\text{kJ kg}^{-1} \text{K}^{-1}$ ) | 4.08                     | 4.43   | 4.41   | 4.08     |
| $q_3$ ( $\text{kJ s}^{-1}$ )                      | 1,518                    | 1,518  | 1,518  | 1,518    |
| $T_{o,w-3}$ ( $^{\circ}\text{C}$ )                | 65.52                    | 61.33  | 63.25  | 68.79    |
| $T_{o,m-3}$ ( $^{\circ}\text{C}$ )                | 4.0303                   | 4.3971 | 4.2765 | 3.7058   |
| $p_3$                                             | -2.2432                  | 6.9264 | 3.9124 | -10.3554 |

## 5. DISCUSSION

The global model for pasteurization of raw milk containing *Mycobacterium avium paratuberculosis* reveals a failure rate, over the long term, of 8.4 % of all (batch continuous) operations. That is pasteurization is actually a mix of successful operations together with unsuccessful ones.

The distribution of these failures would not be expected to be equally spaced in time. If each batch continuous (daily) process is considered one scenario on average there would be  $(8.4/100 \text{ days} \times 365.25 \text{ days/year}) = 31$  failures each year to meet the  $\log_{10}$  reduction in contaminants that could not be attributed to human error or faulty fittings (Davey and Cerf 2003, Davey, Chandrakash and O'Neill 2013).

This significant new insight is not available from traditional methods (with or without sensitivity

analysis). A significant advantage of *Fr 13* simulation is that all possible scenarios are simulated.

The distributions used in the global model parameters are constrained by practical considerations. However, in the absence of specific data the choice of distribution is at present, to some degree, arbitrary. There are some 40 different probability distributions that can be used (Vose 2008). In particular cases, however these might be defined by detailed expert knowledge or process experience (Davey 2010; 2011; Davey, Chandrakash and O'Neill 2013). However, some experimentation with different forms (e.g. with Beta-subjective *see* Davey and Cerf 2003) in the present model resulted in no meaningful change in failure rate i.e. the failure rate more or less remained the same at  $\sim 8.4\%$ . In part it is thought this result is because data used in the model reflects large-scale commercial pasteurization of raw milk (Yapp and Davey 2009).

We believe the ability demonstrated here to quantitatively study the cumulative effect of stochastic changes in a global model is an exciting new development; in particular because this new ability has the potential to be integrated with existing commercial design software such as Aspen Plus<sup>TM</sup> to produce significantly more powerful design and risk assessment tools (Davey 2010; Davey, Chandrakash and O'Neill 2013; Chandrakash, Davey and O'Neill 2014).

Although the global model for pasteurization involves a reduction in the viable numbers of *Mycobacterium avium paratuberculosis* it is clear the generalized form of the *Fr 13* methodology means it could be used for a range of contaminants and/or changes to process equipment. The methodology is flexible because it is predicated on the underlying principles in unit-operations modelling (Sinnott 2005). The particular micro-organism is of concern because of its association with Crohn's disease in humans; it is often inactivated during pasteurization of raw milk but has the ability to survive, if it survives a small dosage can have lethal effect on consumers (Hammer, Kiesner and Walte 2014).

The *Fr 13* global model established and demonstrated here can now be used in second-tier studies (Davey 2010; 2011) to quantitatively simulate intervention strategies and design changes that might be made to minimize unexpected failure due to stochastic effects and to improve process safety in the pasteurization of raw milk. This research is currently being undertaken.

## 6. CONCLUSIONS

A novel *Fr 13* risk analysis of a global model for pasteurization of raw milk containing *Mycobacterium avium paratuberculosis* has been developed and presented for the first time.

Results reveal that batch continuous pasteurization is a mix of successful operations together with unsuccessful ones, with some 8.4 % of all operations resulting in failure due to cumulative impact of stochastic effects. This insight cannot be obtained using

existing risk methodologies (with or without sensitivity analysis).

A significant advantage of *Fr 13* risk modelling over traditional risk approaches is that all possible scenarios that could exist can be quantitatively simulated.

The global model is generalized in form and can be readily used to simulate a range of contaminants and process equipment in pasteurization of raw milk.

## NOMENCLATURE

|                 |                                                                              |
|-----------------|------------------------------------------------------------------------------|
| $A_{PHE-1}$     | Heat transfer area PHE-1 ( $m^2$ ), Eq. (1-3)                                |
| $A_{PHE-3}$     | Heat transfer area PHE-3 ( $m^2$ ), Eq. (3-3)                                |
| $C_{p,m-1}$     | Specific heat milk PHE-1 ( $kJ\ kg^{-1}\ K^{-1}$ ), Eq. (1-1)                |
| $C_{p,m-3}$     | Specific heat of milk PHE-3 ( $kJ\ kg^{-1}\ K^{-1}$ ), Eq. (3-1)             |
| $C_{p,w-1}$     | Specific heat water PHE-1 ( $kJ\ kg^{-1}\ K^{-1}$ ), Eq. (1-2)               |
| $C_{p,w-3}$     | Specific heat of water PHE-3 ( $kJ\ kg^{-1}\ K^{-1}$ ), Eq. (3-2)            |
| $D_{e,PHE-1}$   | Equivalent diameter plate PHE-1 (m), Eq. (1-4)                               |
| $D_{e,PHE-3}$   | Equivalent diameter plate PHE-3 (m), Eq. (3-4)                               |
| $D_{h-2}$       | Diameter holding tube H-2 (m), Eq. (2-3)                                     |
| $D_{t,ref,h-2}$ | Decimal reduction time at $T_{ref,H-2}$ (s), Eq. (2-4)                       |
| $D_{t,h-2}$     | Decimal reduction time in H-2 (s), Eq. (2-4)                                 |
| $F_{r-1}$       | Temperature correction factor PHE-1 (dimensionless), Eq. (1-3)               |
| $F_{r-3}$       | Temperature correction factor PHE-3 (dimensionless), Eq. (3-3)               |
| $L_{h-2}$       | Length holding tube H-2 (m), Eq. (2-1)                                       |
| $L_{p,1}$       | Length plate PHE-1 (m), Eq. (1-5)                                            |
| $L_{p,3}$       | Length plate PHE-3 (m), Eq. (3-5)                                            |
| $Re_1$          | Reynolds number PHE-1 (dimensionless), Eq. (1-8)                             |
| $Re_3$          | Reynolds number PHE-3 (dimensionless), Eq. (3-8)                             |
| $T_{m,avg-1}$   | Average (bulk) milk temperature PHE-1 ( $^{\circ}C$ ), Eq. (1-6)             |
| $T_{m,avg-3}$   | Average (bulk) milk temperature PHE-3 ( $^{\circ}C$ ), Eq. (3-6)             |
| $T_{m,h-2}$     | Holding temperature milk H-2 ( $^{\circ}C$ ), Eq. (2-2)                      |
| $T_{i,m-1}$     | Inlet temperature raw milk PHE-1 ( $^{\circ}C$ ), Eq. (1-1)                  |
| $T_{i,m-3}$     | Inlet temperature milk PHE-3 ( $^{\circ}C$ ), Eq. (3-1)                      |
| $T_{o,m-1}$     | Outlet temperature milk PHE-1 ( $^{\circ}C$ ), Eq. (1-1)                     |
| $T_{o,m-1}'$    | <i>Fr 13</i> scenario value $T_{o,m-1}$ ( $^{\circ}C$ ), Eq. (1-12)          |
| $T_{o,m-3}$     | Outlet temperature pasteurized milk PHE-3 ( $^{\circ}C$ ), Eq. (3-1)         |
| $T_{o,m-3}'$    | <i>Fr 13</i> scenario value $T_{o,m-3}$ ( $^{\circ}C$ ), Eq. (3-12)          |
| $T_{ref,h-2}$   | Reference temperature for microbial growth in H-2 ( $^{\circ}C$ ), Eq. (2-4) |
| $U_{PHE-1}$     | Overall heat transfer coefficient PHE-1 ( $kW\ m^{-2}\ K^{-1}$ ), Eq. (1-3)  |
| $U_{PHE-3}$     | Overall heat transfer coefficient PHE-3 ( $kW\ m^{-2}\ K^{-1}$ ), Eq. (3-3)  |
| $b_{p,1}$       | Gap between plates PHE-1 (m), Eq. (1-4)                                      |
| $b_{p,3}$       | Gap between plates PHE-3 (m), Eq. (3-4)                                      |
| $J_{F,1}$       | Friction factor liquid through plate PHE-1 (dimensionless), Eq. (1-11)       |

|                           |                                                                                                                                     |
|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| $J_{F,3}$                 | Friction factor liquid through plate PHE-3 (dimensionless), Eq. (3-11)                                                              |
| $k_{d,h-2}$               | Microbial death rate H-2 ( $s^{-1}$ ), Eq. (2-5)                                                                                    |
| $\log_{10}(N/N_0)_{h-2}$  | $\log_{10}$ reduction of <i>Mycobacterium avium paratuberculosis</i> in H-2 (dimensionless), Eq. (2-6)                              |
| $\log_{10}(N/N_0)'_{h-2}$ | <i>Fr 13</i> scenario value $\log_{10}(N/N_0)_{h-2}$ , Eq. (2-7)                                                                    |
| $m_{m-1}$                 | Mass flow rate raw milk PHE-1 ( $kg\ s^{-1}$ ), Eq. (1-1)                                                                           |
| $m_{m-2}$                 | Mass flow rate milk H-2 ( $kg\ s^{-1}$ ), Eq. (2-3)                                                                                 |
| $m_{m-3}$                 | Mass flow rate milk PHE-3 ( $kg\ s^{-1}$ ), Eq. (3-1)                                                                               |
| $m_{w-1}$                 | Mass flow rate water PHE-1 ( $kg\ s^{-1}$ ), Eq. (1-2)                                                                              |
| $m_{w-3}$                 | Mass flow rate water PHE-3 ( $kg\ s^{-1}$ ), Eq. (3-2)                                                                              |
| $n_{PHE-1}$               | Number plates required PHE-1 (dimensionless), Eq. (1-5)                                                                             |
| $n_{PHE-3}$               | Number plates required PHE-3 (dimensionless), Eq. (3-5)                                                                             |
| $p_1$                     | Risk factor heating PHE-1 (dimensionless), Eq. (1-12)                                                                               |
| $p_2$                     | Risk factor holding H-2 (dimensionless), Eq. (2-7)                                                                                  |
| $p_3$                     | Risk factor cooling PHE-3 (dimensionless), Eq. (3-12)                                                                               |
| $q_1$                     | Heat duty PHE-1 ( $kJ\ s^{-1}$ ), Eq. (1-1)                                                                                         |
| $q_3$                     | Heat duty PHE-3 ( $kJ\ s^{-1}$ ), Eq. (3-1)                                                                                         |
| $t_{m,h-2}$               | Holding time H-2 (s), Eq. (2-1)                                                                                                     |
| $t_{r,1}$                 | Residence time milk PHE-1 (s), Eq. (1-9)                                                                                            |
| $t_{r,3}$                 | Residence time milk PHE-3 (s), Eq. (3-9)                                                                                            |
| $t_{i,w-1}$               | Inlet temperature water PHE-1 ( $^{\circ}C$ ), Eq. (1-2)                                                                            |
| $t_{i,w-3}$               | Inlet temperature water PHE-3 ( $^{\circ}C$ ), Eq. (3-2)                                                                            |
| $t_{o,w-1}$               | Outlet temperature water PHE-1 ( $^{\circ}C$ ), Eq. (1-2)                                                                           |
| $t_{o,w-3}$               | Outlet temperature water PHE-3 ( $^{\circ}C$ ), Eq. (3-2)                                                                           |
| $v_{m,1}$                 | Milk velocity PHE-1 ( $m\ s^{-1}$ ), Eq. (1-7)                                                                                      |
| $v_{m,3}$                 | Milk velocity PHE-3 ( $m\ s^{-1}$ ), Eq. (3-7)                                                                                      |
| $v_{m,h-2}$               | Milk velocity holding tube H-2 ( $m\ s^{-1}$ ), Eq. (2-1)                                                                           |
| $w_{p,1}$                 | Thickness plate PHE-1 (m), Eq. (1-4)                                                                                                |
| $w_{p,3}$                 | Thickness plate PHE-3 (m), Eq. (3-4)                                                                                                |
| $z_{h-2}$                 | Temperature change ten-fold reduction in $D_{t,h-2}$ for <i>Mycobacterium avium paratuberculosis</i> H-2 ( $^{\circ}C$ ), Eq. (2-4) |
| $\Delta P_{PHE-1}$        | Pressure drop PHE-1 ( $N\ m^{-2}$ ), Eq. (1-10)                                                                                     |
| $\Delta P_{PHE-3}$        | Pressure drop PHE-3 ( $N\ m^{-2}$ ), Eq. (3-10)                                                                                     |
| $\Delta T_{LMTD-1}$       | Log-mean temperature difference PHE-1, Eq. (1-3)                                                                                    |
| $\Delta T_{LMTD-3}$       | Log-mean temperature difference PHE-3, Eq. (3-3)                                                                                    |

## Greek symbols

|              |                                                  |
|--------------|--------------------------------------------------|
| $\rho_{m,1}$ | Density milk PHE-1 ( $kg\ m^{-3}$ ), Eq. (1-6 b) |
| $\rho_{m,2}$ | Density milk H-2 ( $kg\ m^{-3}$ ), Eq. (2-2)     |
| $\rho_{m,3}$ | Density milk PHE-3 ( $kg\ m^{-3}$ ), Eq. (3-6 b) |
| $\mu_{m,1}$  | Viscosity milk PHE-1 (Pa s), Eq. (1-6 a)         |
| $\mu_{m,3}$  | Viscosity milk PHE-3 (Pa s), Eq. (3-6 a)         |

## Other

|                 |                                   |
|-----------------|-----------------------------------|
| $\%tolerance_1$ | Tolerance $p_1$ PHE-1, Eq. (1-12) |
| $\%tolerance_2$ | Tolerance $p_2$ H-2, Eq. (2-7)    |
| $\%tolerance_3$ | Tolerance $p_3$ PHE-3, Eq. (3-12) |

## APPENDIX A – TRADITIONAL SVA COMPUTATIONS FOR RAW MILK PASTEURIZATION

Table A-1 presents a summary of the traditional Single Value Assessment (SVA) computations for heating (PHE-1), holding (H-2) and cooling (PHE-3) unit-operations for raw milk pasteurization. The **bold-text** values in Table A-1 underscores that the output from one unit-operation is the input to the next interconnected unit-operation.

Table A-1: Summary of Traditional Single Value Assessment (SVA) Computations for Raw Milk Pasteurization with Heating (PHE-1), Holding (H-2) and Cooling (PHE-3) Unit-operations

| PHE-1 (Heating)                                    |              |                 | H-2 (Holding)                      |                |                 | PHE-3 (Cooling)                                    |              |                 |
|----------------------------------------------------|--------------|-----------------|------------------------------------|----------------|-----------------|----------------------------------------------------|--------------|-----------------|
| Process Parameter                                  | SVA          | Equation        | Process Parameter                  | SVA            | Equation        | Process Parameter                                  | SVA          | Equation        |
| <b>Constants</b>                                   |              |                 |                                    |                |                 |                                                    |              |                 |
| $U_{PHE-1}$ (kW m <sup>-2</sup> K <sup>-1</sup> )  | 2.5          | constant        | $L_{h-2}$ (m)                      | 11             | constant        | $U_{PHE-3}$ (kW m <sup>-2</sup> K <sup>-1</sup> )  | 2.5          | constant        |
| $A_{PHE-1}$ (m <sup>2</sup> )                      | 33           | constant        | $D_{h-2}$ (m)                      | 0.1            | constant        | $A_{PHE-3}$ (m <sup>2</sup> )                      | 33           | constant        |
| $\Delta T_{LMTD-1}$ (°C)                           | 20           | constant        |                                    |                |                 | $\Delta T_{LMTD-3}$ (°C)                           | 20           | constant        |
| $F_{t-1}$ (dimensionless)                          | 0.92         | constant        |                                    |                |                 | $F_{t-3}$ (dimensionless)                          | 0.92         | constant        |
| <b>Plate Properties</b>                            |              |                 |                                    |                |                 |                                                    |              |                 |
| $w_{p-1}$ (m)                                      | 0.15         | constant        |                                    |                |                 | $w_{p-3}$ (m)                                      | 0.15         | constant        |
| $b_{p-1}$ (m)                                      | 0.05         | constant        |                                    |                |                 | $b_{p-3}$ (m)                                      | 0.05         | constant        |
| $L_{p-1}$ (m)                                      | 1.50         | constant        |                                    |                |                 | $L_{p-3}$ (m)                                      | 1.50         | constant        |
| <b>Inputs</b>                                      |              |                 |                                    |                |                 |                                                    |              |                 |
| $T_{i,m-1}$ (°C)                                   | 4            | input           |                                    |                |                 | $T_{i,m-3}$ (°C)                                   | <b>72.43</b> | <b>input</b>    |
| $m_{m-1}$ (kg s <sup>-1</sup> )                    | 5.56         | input           | $m_{m-2}$ (kg s <sup>-1</sup> )    | 5.56           | input           | $m_{m-3}$ (kg s <sup>-1</sup> )                    | 5.56         | input           |
| $C_{p,m-1}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 3.99         | input           | $D_{t,ref,h-2}$ (s)                | 1.2            | input           | $C_{p,m-3}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 3.99         | input           |
| $t_{i,w-1}$ (°C)                                   | 90           | input           | $T_{ref,h-2}$ (°C)                 | 72             | input           | $t_{i,w-3}$ (°C)                                   | 8            | input           |
| $m_{w-1}$ (kg s <sup>-1</sup> )                    | 6.2          | input           | $z_{h-2}$ (°C)                     | 7.7            | input           | $m_{w-3}$ (kg s <sup>-1</sup> )                    | 6.2          | input           |
| $C_{p,w-1}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 4.2          | input           |                                    |                |                 | $C_{p,w-3}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 4.2          | input           |
| <b>Calculations</b>                                |              |                 |                                    |                |                 |                                                    |              |                 |
| $q_1$ (kJ s <sup>-1</sup> )                        | 1,518        | Eq.(1-3)        |                                    |                |                 | $q_3$ (kJ s <sup>-1</sup> )                        | 1,518        | Eq.(3-3)        |
| $t_{o,w-1}$ (°C)                                   | 31.71        | Eq.(1-2)        |                                    |                |                 | $t_{o,w-3}$ (°C)                                   | 66.29        | Eq.(3-2)        |
| $D_{e,PHE-1}$ (m)                                  | 0.0750       | Eq.(1-4)        |                                    |                |                 | $D_{e,PHE-3}$ (m)                                  | 0.0750       | Eq.(3-4)        |
| $n_{p,PHE-1}$ (dimensionless)                      | 93           | Eq.(1-5)        |                                    |                |                 | $n_{p,PHE-3}$ (dimensionless)                      | 93           | Eq.(3-5)        |
| $T_{m,avg-1}$ (°C)                                 | 38.21        | Eq.(1-6)        |                                    |                |                 | $T_{m,avg-3}$ (°C)                                 | 38.22        | Eq.(3-6)        |
| $\mu_{m-1}$ (Pa s)                                 | 0.00078      | Eq.(1-6 a)      |                                    |                |                 | $\mu_{m-3}$ (Pa s)                                 | 0.00078      | Eq.(3-6 a)      |
| $\rho_{m-1}$ (kg m <sup>-3</sup> )                 | 1,021.29     | Eq.(1-6 b)      | $\rho_{m-2}$ (kg m <sup>-3</sup> ) | 1,004.13       | Eq.(2-2)        | $\rho_{m-3}$ (kg m <sup>-3</sup> )                 | 1,021.29     | Eq.(3-6 b)      |
| $v_{m-1}$ (m s <sup>-1</sup> )                     | 1.2329       | Eq.(1-7)        | $v_{m,h-2}$ (m s <sup>-1</sup> )   | 0.7054         | Eq.(2-3)        | $v_{m-3}$ (m s <sup>-1</sup> )                     | 1.2329       | Eq.(3-7)        |
| $Re_1$ (dimensionless)                             | 121,549      | Eq.(1-8)        | $D_{th-2}$ (s)                     | 1.0236         | Eq.(2-4)        | $Re_3$ (dimensionless)                             | 121,551      | Eq.(3-8)        |
| $t_{r-1}$ (s)                                      | 113.66       | Eq.(1-9)        | $k_{d,h-2}$ (s <sup>-1</sup> )     | 2.2499         | Eq.(2-5)        | $t_{r-3}$ (s)                                      | 113.65       | Eq.(3-9)        |
| $j_{F-1}$ (dimensionless)                          | 0.0179       | Eq.(1-11)       |                                    |                |                 | $j_{F-3}$ (dimensionless)                          | 0.0179       | Eq.(3-11)       |
| $\Delta P_{PHE-1}$ (N m <sup>-2</sup> )            | 2,222.46     | Eq.(1-10)       | $T_{mh-2}$ (°C)                    | <b>72.43</b>   | <b>input</b>    | $\Delta P_{PHE-3}$ (N m <sup>-2</sup> )            | 2,222.45     | Eq.(3-10)       |
| $T_{o,m-1}$ (°C)                                   | <b>72.43</b> | <b>Eq.(1-1)</b> | $t_{h-2}$ (s)                      | <b>15.59</b>   | <b>Eq.(2-1)</b> | $T_{o,m-3}$ (°C)                                   | <b>4.00</b>  | <b>Eq.(3-1)</b> |
|                                                    |              |                 | $\log_{10}(N/N_0)_{h-2}$           | <b>14.6483</b> | <b>Eq.(2-6)</b> |                                                    |              |                 |



## APPENDIX B – COMPARISON SVA WITH *FR 13*

Tables B-1, B-2 and B-3 present a summary comparison of detailed *Fr 13* results with those of traditional Single Value Assessment (SVA) for heating (PHE-1), holding (H-2) and cooling (PHE-3) unit-operations for pasteurization of raw milk containing *Mycobacterium avium paratuberculosis*. The table layout follows the presentation methodology of Davey and co-workers (Chandrakash 2012; Chandrakash, Davey and O'Neill 2014; Davey, Chandrakash and O'Neill 2013; Davey 2011).

Table B-1: Comparative Summary of SVA and *Fr 13* Results for Heating (PHE-1) with %tolerance = 3 %

| Process Parameter                                  | SVA*     | <i>Fr 13 model**</i> |                                                          |
|----------------------------------------------------|----------|----------------------|----------------------------------------------------------|
| <b>Inputs</b>                                      |          |                      |                                                          |
| $T_{i,m-1}$ (°C)                                   | 4        | <b>4.2765</b>        | <b>RiskNormal(4,0.2, RiskTruncate(3.4,4.6))</b>          |
| $m_{m-1}$ (kg s <sup>-1</sup> )                    | 5.56     | <b>5.9014</b>        | <b>RiskNormal(5.56,0.278, RiskTruncate(4.726,6.394))</b> |
| $C_{p,m-1}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 3.99     | <b>4.2557</b>        | <b>RiskNormal(3.99,0.2, RiskTruncate(3.39,4.59))</b>     |
| $t_{i,w-1}$ (°C)                                   | 90       | <b>90.0902</b>       | <b>RiskNormal(90,4.5, RiskTruncate(76.5,103.5))</b>      |
| $m_{w-1}$ (kg s <sup>-1</sup> )                    | 6.2      | <b>5.6392</b>        | <b>RiskNormal(6.2,0.31, RiskTruncate(5.27,7.13))</b>     |
| $C_{p,w-1}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 4.2      | <b>4.2418</b>        | <b>RiskNormal(4.2,0.21, RiskTruncate(3.57,4.83))</b>     |
| <b>Constants</b>                                   |          |                      |                                                          |
| $A_{PHE-1}$ (m <sup>2</sup> )                      | 33       | 33                   | constant                                                 |
| $U_{PHE-1}$ (kW m <sup>-2</sup> K <sup>-1</sup> )  | 2.5      | 2.5                  | constant                                                 |
| $\Delta T_{LMTD-1}$ (°C)                           | 20       | 20                   | constant                                                 |
| $F_{r-1}$ (dimensionless)                          | 0.92     | 0.92                 | constant                                                 |
| <b>Calculations</b>                                |          |                      |                                                          |
| $t_{o,w-1}$ (°C)                                   | 31.71    | 26.63                | Eq. (1-2)                                                |
| $q_1$ (kJ s <sup>-1</sup> )                        | 1,518    | 1,518                | Eq. (1-3)                                                |
| $v_{m-1}$ (m s <sup>-1</sup> )                     | 1.2329   | 1.3067               | Eq. (1-7)                                                |
| $Re_1$ (dimensionless)                             | 121,549  | 126,324              | Eq. (1-8)                                                |
| $t_{r,1}$ (s)                                      | 113.66   | 107.24               | Eq. (1-9)                                                |
| $\Delta P_{PHE-1}$ (N m <sup>-2</sup> )            | 2,222.46 | 2,471.3              | Eq. (1-10), (1-11)                                       |
| $T_{o,m-1}$ (°C)                                   | 72.43    | <b>64.72</b>         | Eq. (1-1)                                                |
| $p_1$ (dimensionless)                              |          | <b>7.6411</b>        | Eq. (1-13)                                               |

Table B-2: Comparative Summary of SVA and *Fr 13* Results for Holding (H-2) with %tolerance = 3 %

| Process Parameter                  | SVA*     | <i>Fr 13 model**</i> |               |
|------------------------------------|----------|----------------------|---------------|
| <b>Inputs</b>                      |          |                      |               |
| $T_{m,h-2}$ (°C)                   | 72.43    | <b>64.72</b>         | $= T_{o,m-1}$ |
| $m_{m-2}$ (kg s <sup>-1</sup> )    | 5.56     | <b>5.9014</b>        | $= m_{m-1}$   |
| <b>Constants</b>                   |          |                      |               |
| $L_{h-2}$ (m)                      | 11       | 11                   | constant      |
| $D_{h-2}$ (m)                      | 0.1      | 0.1                  | constant      |
| $T_{ref,h-2}$ (°C)                 | 72       | 72                   | constant      |
| $D_{t,ref,h-2}$ (s)                | 1.2      | 1.2                  | constant      |
| $z_{h-2}$ (°C)                     | 7.7      | 7.7                  | constant      |
| <b>Calculations</b>                |          |                      |               |
| $\rho_{m,2}$ (kg m <sup>-3</sup> ) | 1,004.08 | 1,008.46             | Eq. (2-2)     |
| $v_{m,h-2}$ (m s <sup>-1</sup> )   | 0.7054   | 0.7455               | Eq. (2-3)     |
| $t_{m,h-2}$ (s)                    | 15       | 14.7559              | Eq. (2-1)     |
| $D_{t,h-2}$ (s)                    | 1.0236   | 2.7863               | Eq. (2-4)     |
| $k_{d,h-2}$ (s <sup>-1</sup> )     | 2.2499   | 0.8266               | Eq. (2-5)     |
| $\log_{10}(N/N_0)_{h-2}$           | 14.6483  | 5.2960               | Eq.(2-6)      |
| $p_2$ (dimensionless)              |          | <b>0.7096</b>        | Eq. (2-8)     |

Table B-3: Comparative Summary of SVA and *Fr 13* Results for Cooling (PHE-3) with %tolerance = 3 %

| Process Parameter                                  | SVA*  | <i>Fr 13 model**</i> |                                                      |
|----------------------------------------------------|-------|----------------------|------------------------------------------------------|
| <b>Inputs</b>                                      |       |                      |                                                      |
| $T_{i,m-3}$ (°C)                                   | 72.43 | <b>64.72</b>         | $= T_{m,h-2} (= T_{o,m-1})$                          |
| $m_{m-3}$ (kg s <sup>-1</sup> )                    | 5.56  | <b>5.9014</b>        | $= m_{m-2}$                                          |
| $C_{p,m-3}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 3.99  | 4.2557               | $= C_{p,m-1}$                                        |
| $t_{i,w-3}$ (°C)                                   | 8     | <b>7.5845</b>        | <b>RiskNormal(8,0.4, RiskTruncate(6.8,9.2))</b>      |
| $m_{w-3}$ (kg s <sup>-1</sup> )                    | 6.2   | <b>6.6789</b>        | <b>RiskNormal(6.2,0.31, RiskTruncate(5.27,7.13))</b> |
| $C_{p,w-3}$ (kJ kg <sup>-1</sup> K <sup>-1</sup> ) | 4.2   | <b>4.3987</b>        | <b>RiskNormal(4.2,0.21, RiskTruncate(3.57,4.83))</b> |
| <b>Constants</b>                                   |       |                      |                                                      |
| $A_{PHE-3}$ (m <sup>2</sup> )                      | 33    | 33                   | constant                                             |
| $U_{PHE-3}$ (kW m <sup>-2</sup> K <sup>-1</sup> )  | 2.5   | 2.5                  | constant                                             |
| $\Delta T_{LMTD-3}$ (°C)                           | 20    | 20                   | constant                                             |
| $F_{r-3}$ (dimensionless)                          | 0.92  | 0.92                 | constant                                             |
| <b>Calculations</b>                                |       |                      |                                                      |
| $t_{o,w-3}$ (°C)                                   | 66.29 | 63.25                | Eq. (3-2)                                            |
| $q_3$ (kJ s <sup>-1</sup> )                        | 1,518 | 1,518                | Eq. (3-3)                                            |
| $T_{o,m-3}$ (°C)                                   | 4     | <b>4.2765</b>        | Eq. (3-1)                                            |
| $p_3$ (dimensionless)                              |       | <b>3.9124</b>        | Eq. (3-13)                                           |

## APPENDIX C – SAMPLE CALCULATION $D_{t,h-2}$

Steps for estimating Decimal Reduction time ( $D_t$ ) in H-2 at  $T_{m,h-2} = T_{o,m-1} = 72.43$  °C for *Mycobacterium avium paratuberculosis*

From literature

$T_{ref} = 72$  °C;  $D_{t,ref,h-2} = 1.2$  s;  $z_{h-2} = 7.7$  °C;  $T_{m,h-2} = T_{o,m-1} = 72.43$  °C;  $t_{m,h-2} = 15$  s

Using Eq. (2-4),

$$\log_{10} D_{t,h-2} = \log_{10} D_{t,ref,h-2} - \frac{T_{m,h-2} - T_{ref}}{z} \quad (2-4)$$

$$\log_{10} D_{t,h-2} = \log_{10} 1.2 - \frac{72.43 - 72}{7.7}$$

Solving we get,

$$\log_{10} D_{t,h-2} = 0.02333$$

$$D_{t,h-2} = \mathbf{1.0236 \text{ s}}$$

Using Eq. (2-5),

$$k_{d,h-2} = \frac{2.303}{D_{t,h-2}} = \frac{2.303}{1.0236} \quad (2-5)$$

Solving we get,

$$k_{d,h-2} = \mathbf{2.2499 \text{ s}^{-1}}$$

Using Eq. (2-6),

$$\log_{10} \left( \frac{N}{N_0} \right)_{h-2} = \frac{k_{d,h-2} t_{m,h-2}}{2.303} \quad (2-6)$$

$$\log_{10} \left( \frac{N}{N_0} \right)_{h-2} = \frac{2.2499 * 15}{2.303}$$

Solving we get,

$$\log_{10} (N/N_0)_{h-2} = \mathbf{14.6483}$$

## REFERENCES

- Abdul Halim, N., Davey, K. R., 2014. A new Friday 13<sup>th</sup> risk assessment of UV irradiation for potable water in turbulent flow in a series annular-reactor. *Food Control*. (submitted-May 2014).
- Anon., 2014. Individual plate characteristics in a Plate Heat Exchanger (PHE): [http://www.armstrongpumps.com/Data/pdfbrochures/Link s/02\\_09\\_003\\_113.12L\\_phe\\_brochure.pdf](http://www.armstrongpumps.com/Data/pdfbrochures/Link%2009%20003%20113.12L_phe_brochure.pdf) [accessed 24<sup>th</sup> March 2014]
- Anon., 2011. Blackett Review of High Impact Low Probability Risks, Government Office for Science (UK): <http://www.bis.gov.uk/assets/goscience/docs/b/12-519-blackett-review-high-impact-low-probability-risks.pdf> [accessed 17<sup>th</sup> May 2014].

- Al-Hilphy, A.R.S., Ali, H.I., 2013. Milk flash pasteurization by the microwave and study its (*sic*) chemical, microbiological and thermo physical characteristics. *Food Processing and Technology*, 4 (7), 1–5.
- Alfa-Laval, 1987. *Dairy Handbook-Food Engineering*. Sweden: Alfa Laval Firm.
- Berk, Z., 2009. *Food Process Engineering and Technology*. 1st ed. USA: Academic Press.
- Cerf, O., Davey, K.R., 2001. An explanation of non-sterile (leaky) milk packs in well operated UHT plant. *Transactions of Institution of Chemical Engineers, Food and Bioproducts Processing (Part C)*, 79 (4), 219–222.
- Chandrakash, S., Davey, K.R., O'Neill, B.K., 2014. A novel risk analysis of failure in a global food process – preliminary application to milk processing. *Australasian Chemical Engineering Conference*, Perth (Australia) (submitted-June 2014).
- Chandrakash, S., 2012. *A New Risk Analysis of Clean-In-Place (CIP) Milk Processing*. Thesis (MEngSc). The University of Adelaide.
- Davey, K.R., 2011. Introduction to fundamentals and benefits of Friday 13<sup>th</sup> risk modelling technology for food manufactures. *Food Australia*, 63 (5), 192–197.
- Davey, K.R., 2010. A novel proposal to advance the discipline and to quantitatively safeguard important hygienic bio-processes. *Proceedings of 40<sup>th</sup> Australasian Chemical Engineering Conference (Engineering at The Edge)*, paper 0495. Sept. 26–29, Adelaide (Australia).
- Davey, K.R., Cerf, O., 2003. Risk modelling—An explanation of Friday 13<sup>th</sup> syndrome in well-operated continuous sterilisation plant. *Proceedings of 31st Australasian Chemical Engineering Conference*, paper 61. Sept 28–Oct 1, Adelaide (Australia).
- Davey, K.R., Abdul Halim, N., Lewis, D., 2012. Friday 13<sup>th</sup> failure modelling: A new quantitative risk assessment of UV irradiation of potable water. *Proceedings of 42nd Australasian Chemical Engineering Conference (Quality of Life Through Chemical Engineering)*, Sept. 23–26, Wellington (New Zealand).
- Davey, K.R., Chandrakash, S., O'Neill, B.K., A new risk analysis of Clean-In-Place milk processing. *Food Control*, 29 (1), 248–253.
- Davey, K.R., Chandrakash, S., O'Neill, B.K., 2011. Friday 13<sup>th</sup> failure of Clean-In-Place operations in a milk process plant. *Proceedings of 41st Australasian Chemical Engineering Conference (Engineering a Better World)*, paper 150. Sept. 18–21, Sydney (Australia).
- Hammer, P., Kiesner, C., Walte, H.G.C., 2014. *Short communication: Effect of homogenization on heat inactivation of Mycobacterium avium subspecies paratuberculosis in milk*. *Journal of Dairy Science*, 97 (4), 2045–2048.
- Hayhurst, A.N., 1997. Introduction to heat transfer. In: P. J. Fryer, D. L. Pyle, C. D. Rielly, ed. *Chemical Engineering for the Food Industry*. 1st ed. London: Blackie Academic and Professional. 105–152.
- Ibarz, A., Barbosa-Canovas, G.V., 2003. *Unit Operations in Food Engineering*. USA: CRC Press.
- Juffs, H., Deeth, H., 2007. *Scientific Evaluation of Pasteurisation for Pathogen Reduction in Milk and Milk Products*. Food Standards Australia and New Zealand (FSANZ).
- Katoh, S., Yoshida, F., 2009. *Biochemical Engineering: A Textbook for Engineers, Chemists and Biologists*. Germany: Wiley-VCH Verlag GmbH and Co.
- Kessler, H.G., 2002. *Food and Bio Process Engineering-Dairy Technology*. 5th ed. Germany: Kessler Publishing House.
- Kothandaraman, C.P., Subramanian, S., 2007. *Heat and Mass Transfer Data Book*. 6th ed. India: New Age International.
- Koutchma, T., Le Bail, A., Ramaswamy, H.S., 2001. Comparative experimental evaluation of microbial destruction in continuous-flow microwave and conventional heating systems. *Canadian Biosystems Engineering*, 43, 3.1–3.8.
- Kumbhare, M.B., Dawande, S.D., 2013. Performance evaluation of Plate Heat Exchanger (*sic*) in laminar and turbulent flow conditions. *International Journal of Chemical Sciences and Applications*, 4 (1), 77–83.
- Lewis, M., Heppell, N., 2000. *Continuous Thermal Processing of Foods-Pasteurization and UHT Sterilization*. USA: Aspen Publishers Inc.
- Madigan, M.T., Martinko, J.M., 2006. *Biology of Microorganisms*. USA: Pearson Prentice Hall.
- Maroulis, Z.B., Saravacos, G.D., 2003. *Food Process Design*. USA: Marcel Dekker Inc.
- McDonald, W.L., O'Riley, K.J., Schroen, C.J., Condron, R.J., 2005. Heat inactivation of *Mycobacterium avium* subsp. *paratuberculosis* in milk. *Applied and Environmental Microbiology*, 71 (4), 1785–1789.
- Patil, R.A., 2006. *Novel Application of Quantitative Risk Assessment Modelling to a Continuous Fermenter*. Thesis (MEngSc). The University of Adelaide.
- Patil, R.A., Davey, K.R., Daughtry, B.J., 2005. A new quantitative risk model of a fermenter for Friday 13<sup>th</sup> Syndrome. *Proceedings of 33<sup>rd</sup> Australasian Chemical Engineering Conference (Smart Solutions-Doing More with Less)*, paper 407. Sept. 25–28, Brisbane (Australia).
- Perry, R.H., Green, D.W., 1997. *Chemical Engineers' Handbook*. 7th ed. USA: McGraw Hill.
- Rademaker, J.L.W., Vissers, M.M.M., te Giffel, M.C., 2007. Effective heat inactivation of *Mycobacterium avium* subsp. *paratuberculosis* in raw milk contaminated with naturally infected feces. *Applied and Environmental Microbiology*, 73 (13), 4185–4190.
- Sinnott, R.K., 2005. *Chemical Engineering Design*. 4th ed. USA: Elsevier Butterworth-Heinemann.
- Smith, P.G., 2011. *Introduction to Food Process Engineering*. 2nd Edition. UK: Springer Science.
- Vose, D., 2008. *Risk Analysis - A Quantitative Guide*. 3rd ed. UK: Wiley and Sons.
- van Asselt, E.D., Zwietering, M.H., 2005. A systematic approach to determine global thermal inactivation parameters for various food pathogens. *International Journal of Food Microbiology*, 107 (1), 73–82.
- Yapp, E., Davey, K.R., 2009. *A Detailed Mass and Energy Audit of a Milk Processing Plant*. Thesis (Hons). The University of Adelaide.

## AUTHORS BIOGRAPHY

**Kenneth R Davey** PhD FIChemE CEng FIEAust CPEng is a Senior Lecturer in the School of Chemical Engineering, The University of Adelaide, Australia. **Saravanan Chandrakash** MEngSc (Research) is a research student working with academic Kenneth R Davey and Associate Professor **Brian K O'Neill** PhD FIChemE CEng.

# A NEW TRANSIENT PREDICTIVE MODEL TO QUANTIFY TAINT AS EITHER GEOSMIN (GSM) OR 2-METHYLISOBORNEOL (MIB) IN RAINBOW TROUT (*ONCORHYNCHUS MYKISS*) FARMED IN RECIRCULATING AQUACULTURE SYSTEMS (RAS)

Davey K R <sup>(a)</sup>, Hathurusingha P I <sup>(b)</sup>

School of Chemical Engineering  
The University of Adelaide, 5005, Australia

(a) kenneth.davey@adelaide.edu.au (b) priyantha.hathurusinghaarachchige@adelaide.edu.au

## ABSTRACT

Taint, as off-flavour, accumulation in fish as either geosmin (**GSM**) or 2-methylisoborneol (**MIB**) is a major concern globally in fish farmed in Recirculated Aquaculture Systems (RAS), especially in European farmed rainbow trout (*Oncorhynchus mykiss*). To aid farm management in RAS growth protocols a new transient (unsteady-state) predictive model for taint was developed. The approach is based on conservation of mass and energy and established thermodynamics. The model is the sum of two exponential terms with growth-time that simulate both taint uptake (through gills) and elimination (through gills, growth dilution, and metabolic-transformations). The model was simulated with growth water concentrations of **GSM** and **MIB** for the typical growth period of 9 months and gave good agreement with independent data. It is concluded the model is therefore free of computational and programming errors. It is planned the model be used to develop specific growth protocols to minimize taint development in fish in RAS systems.

Keywords: taint, geosmin (**GSM**), 2-methylisoborneol (**MIB**), Recirculated Aquaculture Systems (RAS)

## 1. INTRODUCTION

Farming of rainbow trout (*Oncorhynchus mykiss*) in Recirculating Aquaculture systems (RAS) is becoming increasingly popular. However, taint as ‘earthy’ or ‘muddy’ off flavors can accumulate in the fish flesh. This is true also of other species in RAS such as barramundi (*Lates calcarifer*) (Percival et al. 2008), arctic charr (*Salvelinus alpinus*) (Houle et al. 2011), largemouth bass (*Micropterus salmoides*), and; white sturgeon (*Acipenser transmontanus*) (Schrader et al. 2005).

Two major taint molecules have been identified; these are Geosmin (**GSM**), and; 2-methylisoborneol (**MIB**). These are produced by benthic and planktonic cyanobacteria, some fungi and actinomycetes species as byproducts (Wood et al. 2001). Taint due to **GSM** and **MIB** has been a common feature in conventional aquaculture systems e.g. ponds, but in RAS, reduced water consumption, higher nutrient(s) load, and higher

stocking densities have acerbated the problem. These two taint compounds impart earthy or muddy taste off flavors and odor in the growth water which is then taken into the fish flesh via gills or skin (Howgate 2004). The result is strong buyer resistance.

Effective control measures to minimize taint are yet to be established. However, leaching of fish in clean water prior to market is widely used (Tucker and van der ploeg 1999). A drawback is that this can lead to a loss of body mass of the fish. Also, the time taken for leaching may be several days depending on the intensity of the taint (Tucker 2000). During RAS growth, copper sulfate is sometimes used as an algacide to control the growth of taint producing bacteria. There are however some concerns over effectiveness of this treatment (Rimando and Schrader 2003).

### 1.1 This research

Against this background, a new transient-state quantitative model was developed to predict taint as either **GSM** or **MIB** in European farmed rainbow trout. Published mechanistic models have been based on steady-state conditions where chemical uptake into and from the fish is assumed to be zero (e.g. Arnot and Gobas 2004; Gobas 1993). However, RAS is dynamic and its growth conditions continuously change with time. Therefore, it is questionable as to whether steady-state condition can be used when developing predictive models for fish grown in RAS.

Here the approach used is to refine the generalized model for taint accumulation of Hathurusingha and Davey (2013; 2014) for both constant and varying concentration of **GSM** and **MIB** in the RAS growth water. This is then used to simulate taint accumulation up to the consumer rejection threshold. Predictions are compared with best available independent data. The model can be conveniently set-up and solved as a Microsoft Excel™ spread sheet. The widespread use of these tools means that communication of results can be streamlined. It is planned the validated model will be used to investigate a number of practical RAS growth protocols to minimize taint. A typical growth-period for rainbow trout in RAS is 270 days (9 months) (Petersen et al. 2011).

## 2. MATERIALS AND METHODS

### 2.1 Model development

The model of Hathurusingha and Davey is a sum of two exponential terms with growth-time: one for taint uptake (through the gills) and one for taint elimination (through gills plus growth dilution and metabolic-transformations). The model can be conveniently summarized as shown schematically in Figure 1. All the symbols used are carefully defined in the Nomenclature.

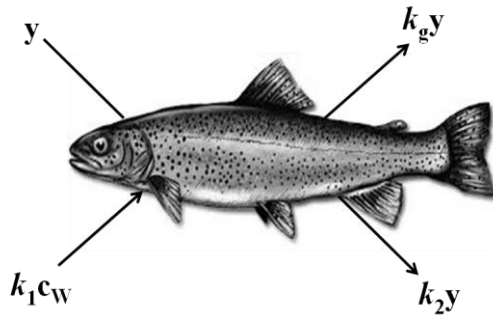


Figure 1: Schematic of rainbow trout fish with uptake ( $k_1 C_w$ ) and elimination ( $k_2 y + k_g y$ ) of taint molecules (as GSM or MIB)

It is assumed that the taint concentration, as either **GSM** or **MIB** in the fish-flesh, is 'y' ( $\mu\text{g kg}^{-1}$ ). The concentration of the taint compound in the RAS growth water at any time ( $t$ ) is  $C_w$  ( $\mu\text{g L}^{-1}$ ). Therefore, the rate of change of taint ( $y$ ) in the fish-flesh with time ( $t$ ) can be expressed by:

$$\frac{dy}{dt} = \text{uptake rate} - \text{elimination rate} \quad (1)$$

The uptake rate of **GSM** or **MIB** will be a function of the rate constant for uptake ( $k_1$ ,  $\text{L kg}^{-1} \text{day}^{-1}$ ), the fish mass ( $m_f$ , kg) and the taint concentration in the RAS water ( $C_w$ ). The elimination rate of **GSM** or **MIB** will be a function of the rate constant for elimination through the gills ( $k_2$ ,  $\text{day}^{-1}$ ), the rate constant for bio-transformation and growth dilution ( $k_g$ ,  $\text{day}^{-1}$ ) and the instantaneous value of  $y$  at time,  $t$ . Substitution of these parameters into Eq. (1) gives:

$$\frac{dy}{dt} = k_1 m_f C_w - (k_2 + k_g) y \quad (2)$$

where  $\left(\frac{dy}{dt}\right)$  is the rate of change of taint in fish-flesh ( $\mu\text{g kg}^{-1} \text{day}^{-1}$ ).

The mass of the fish ( $m_f$ ) is known to be a function of growth-time in the RAS tank. Published growth data as mass of fish (kg) vs growth-time (day) is generally highly exponentially correlated (*see* below Eq. (19) for rainbow trout) and fish mass is therefore given by:

$$m_f = \beta \exp^{\gamma t} \quad (3)$$

Let:

$$b = \beta k_1 C_w \quad (4)$$

and;

$$a = (k_2 + k_g) \quad (5)$$

Substitution for  $b$  and  $a$  into Eq. (2) and rearranging gives:

$$\frac{dy}{dt} + ay = b \exp^{\gamma t} \quad (6)$$

Eq. (6) can be integrated by parts (Evans 2010) to give:

$$y = \frac{b \exp^{\gamma t}}{(a + \gamma)} - \frac{b}{(a + \gamma) \exp^{at}} \quad (7)$$

Eq. (7) can be rearranged to conveniently give:

$$y = \left(\frac{b}{a + \gamma}\right) \left[ \exp^{\gamma t} - \exp^{-at} \right] \quad (8)$$

The model for taint in fish-flesh of Eq. (8) shows that the predicted level is the sum of two exponential terms with time in the RAS growth tanks, namely, uptake and elimination.

There are generally however no published data for  $k_1$  and  $k_2$  or  $k_g$  in the refereed literature (for example for rainbow trout fish) for immediate simulation of taint in the model. For model development these rate constants need to be defined mathematically.

### 2.2 Rate constants $k_b$ , $k_2$ and $k_g$

Arnot and Gobas (2004) found the gill uptake rate constant ( $k_1$ ) is a combination of two processes: gill ventilation, and chemical uptake efficiency across the gills. Accordingly, the chemical uptake rate was expressed as:

$$k_1 = \frac{E_w G_v}{m_f} \quad (9)$$

where  $E_w$  is the gill chemical uptake efficiency (dimensionless fraction) and  $G_v$  is the gill ventilation rate ( $\text{L day}^{-1}$ ). The chemical uptake efficiency was, reasonably, assumed by Arnot and Gobas (2004) to be a function of an octanol-water partition coefficient ( $K_{ow}$ ) of the chemical of interest that can be expressed through the following relationship:

$$E_w = (1.85 + 155 / K_{ow})^{-1} \quad (10)$$

A relationship between gill ventilation rate and oxygen consumption rate of the fish species based on empirical data is given by Aront and Gobas (2004), namely:

$$G_v = \frac{1400 m_f^{0.65}}{C_{ox}} \quad (11)$$

where  $C_{ox}$  is the concentration of dissolved oxygen ( $\text{mg L}^{-1}$ ). This is considered to be a function of temperature and can be computed from the equation given by Neely (1979):

$$C_{ox} = 14.45 - 0.413T + 0.00556T^2 \quad (12)$$

where  $T$  = RAS growth water temperature in degree Celsius.

The chemical elimination rate from the fish gills to the water ( $k_2$ ) is correlated with the chemical transport rate in aqueous and lipid phases of the fish, lipid content of the fish and octanol water partition coefficient of the taint chemical. The relationship can be expressed (Gobas 1993) as:

$$\frac{1}{k_2} = \left(\frac{V_L}{Q_W}\right)K_{OW} + \left(\frac{V_L}{Q_L}\right) \quad (13)$$

where  $Q_W$  is rate of chemical transport in the aqueous phase ( $L \text{ day}^{-1}$ ),  $Q_L$  is the rate of chemical transport in the lipid phase ( $L \text{ day}^{-1}$ ) and  $V_L$  is lipid weight (mass). Gobas and Mackay (1987) derived a relationship between  $Q_W$  and  $m_f$  using experimental data to give:

$$Q_W = 88.3m_f^{0.6(\pm 0.2)} \quad (14)$$

Gobas and Mackay (1987) reported that the chemical transport rate in the aqueous phase is ~100 times higher than in the lipid phase; therefore it can be assumed that:

$$Q_L = 0.01Q_W \quad (15)$$

Lipid mass ( $V_L$ ) of the fish is correlated to the lipid mass ratio ( $e$ ) (dimensionless) and can be conveniently expressed as:

$$V_L = em_f \quad (16)$$

The rate constant for combined metabolic transformation of the taint chemical and growth dilution rate of the taint chemical ( $k_g$ ) is given by:

$$k_g = (k_G + k_m) \quad (17)$$

The rate constant for growth dilution ( $k_G$ ) can be computed using the equation given in Thomann et al. (1992) as:

$$k_G = 0.000502m_f^{-0.2} \quad (18)$$

The rate constant for metabolic transformation of the taint chemical ( $k_m$ ) is available in the refereed literature (Gobas, 1993) to cover a range of fish species of  $k_m = 0.00063 \text{ day}^{-1}$ .

Equations (1) through (18) plus the general value for  $k_m$ , defines the model for taint in RAS farmed fish-flesh.

### 2.3 Simulations for European farmed rainbow trout (*Oncorhynchus mykiss*)

To illustrate the generalized form for rainbow trout (*Oncorhynchus mykiss*), it is necessary to find a suitable growth equation for this species.

A graph of the mass of the fish and the growth-time was obtained (Anon 2010) and correlated (Figure 2) to give:

$$m_f = 0.0346 \exp^{0.0138t} \quad (19)$$

with  $R^2 = 0.94$  (Snedecor and Cochran 1989) indicating a very good fit.

The model simulations can now be carried out using the procedure (Table 1) as follows:

At a mid-range of 150 days RAS growth, from Eq. (21),  $m_f = 0.274 \text{ kg}$ . The average growth water

temperature  $14 \text{ }^\circ\text{C}$  (Petersen et al. 2011) is substituted into Eq. (12) to give  $C_{OX} = 9.75 \text{ mg L}^{-1}$ . Substitution for  $m_f$  into Eq. (11) gives  $G_V = 61.68 \text{ L day}^{-1}$ . Substituting for  $\log_{10}(K_{OW}) = 3.57$  (dimensionless) (Howgate 2004) into Eq. (10) gives  $E_W = 0.528$  (dimensionless).

The uptake rate constant ( $k_1$ ) for **GSM** now can be calculated by substituting the  $E_W$ ,  $G_V$  and  $m_f$  values into Eq. (9) to give  $k_1 = 119.29 \text{ L kg}^{-1} \text{ day}^{-1}$ .

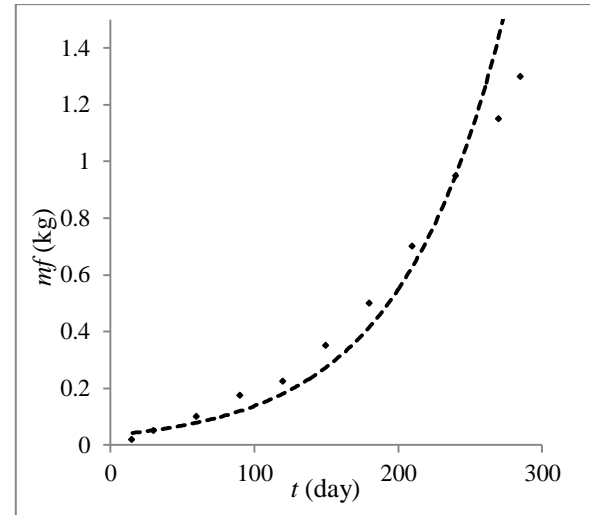


Figure 2: Fit of growth data (◆) for rainbow trout fish farmed in ponds at a water temperature  $15 \text{ }^\circ\text{C}$  (Anon. 2010)  $m_f = 0.0346 \exp^{0.0138t}$ ,  $R^2 = 0.94$

Substitution for  $m_f = 0.274 \text{ kg}$  at a mid-range growth of 150 days into Eq. (14) gives  $Q_W = 40.62 \text{ L day}^{-1}$  which when substituted into Eq. (15) gives  $Q_L = 0.406 \text{ L day}^{-1}$ . A lipid to mass ratio of rainbow trout is assumed to be 0.045 (Robertson et al. 2006) and hence from Eq. (16), lipid weight (mass)  $V_L = 0.0123 \text{ kg}$ . The taint elimination rate  $k_2$  now can be determined by substituting values for  $V_L$ ,  $Q_L$ ,  $Q_W$  and  $\log_{10}(K_{OW})$  into Eq. (13) to give  $k_2 = 0.863 \text{ day}^{-1}$ .

Substitution for  $m_f = 0.274 \text{ kg}$  into Eq. (18) gives  $k_G = 0.00065 \text{ day}^{-1}$ . The published value for  $k_m = 0.00063 \text{ day}^{-1}$  (Gobas 1993). From Eq. (17) the sum of  $k_G$  and  $k_m$  gives  $k_g = 0.00128 \text{ day}^{-1}$ .

A typical  $C_W$  value for **GSM** in growth ponds is reported to be  $0.0066 \text{ } \mu\text{g L}^{-1}$  (Petersen et al. 2011). Substitution for  $C_W = 0.0066 \text{ } \mu\text{g L}^{-1}$  into Eq. (4) gives **GSM** taint uptake in the rainbow trout,  $b = 0.04094 \text{ } \mu\text{g kg}^{-1} \text{ day}^{-1}$  at 150 day in the RAS growth cycle. The level of taint as **GSM** in the rainbow trout fish-flesh at 150 day of growth can now be obtained from Eq. (8) such that  $y = 0.3573 \text{ } \mu\text{g kg}^{-1}$ .

Similar calculations can be readily carried out for **MIB** accumulation in the RAS rainbow trout fish-flesh. A typical value for  $C_W$  for **MIB** is  $0.0032 \text{ } \mu\text{g L}^{-1}$  (Petersen et al. 2011). Substitution for  $\log_{10}(K_{OW}) = 3.31$  (dimensionless) (Howgate 2004) into Eq. (10) gives  $E_W = 0.519$ . Parameters  $k_g$ ,  $C_{OX}$ ,  $G_V$ ,  $Q_W$ ,  $Q_L$  and  $V_L$  are the same as that for **GSM** (as they are dependent on the mass of the fish only). Accordingly, the

calculated values for  $k_1$  and  $k_2$  for **MIB** are, respectively,  $117.17 \text{ L kg}^{-1}\text{day}^{-1}$  and  $1.537 \text{ day}^{-1}$ , and that for taint as **MIB** in the fish-flesh,  $y = 0.099 \mu\text{g kg}^{-1}$ .

### 3. RESULTS

#### 3.1 Predicted GSM levels

The (generic) procedure for model simulation is summarized in Table 1 and is illustrated with **GSM**. The procedure is such that the computations can be read conveniently down column 4 of the table. The value of **GSM** taint at 150 days RAS growth with constant  $C_w = 0.0066 \mu\text{g L}^{-1}$  (Petersen et al. 2011) is shown in the table to be  $y = 0.3573 \mu\text{g kg}^{-1}$ .

Table 2 summarizes the model predictions for taint as **GSM** in RAS rainbow trout fish-flesh for the growth period 30 to 270 days in RAS tanks. The table shows that  $k_1$ ,  $k_2$  and  $k_g$  values decrease with increasing growth-time. The level of predicted **GSM** taint ranges from 0.0508 to  $2.0219 \mu\text{g kg}^{-1}$ , respectively, with growth time, 30 to 270, days. The bold text in Table 2 at 150 days of growth is the detailed simulation illustrated in Table 1. The mass of the fish over the period increases some 34 times from 0.042 to 1.436, kg.

Table 1: Generic simulation procedure showing the model inputs, calculations and outputs. The particular illustration is for **GSM** accumulation in rainbow trout fish-flesh at 150 days of growth in RAS tanks

| Inputs       |          |                                      |                        |
|--------------|----------|--------------------------------------|------------------------|
| line         |          |                                      |                        |
| 1            | $t$      | day                                  | 150                    |
| 2            | $k_{ow}$ | dimensionless                        | 3.57                   |
| 3            | $k_m$    | $\text{day}^{-1}$                    | 0.00063                |
| 4            | $C_w$    | $\mu\text{g L}^{-1}$                 | 0.0066                 |
| 5            | $T$      | $^{\circ}\text{C}$                   | 14                     |
| 6            | $e$      | dimensionless                        | 0.045                  |
| Calculations |          |                                      |                        |
| 7            | $C_{ox}$ | $\text{mg L}^{-1}$                   | 9.75, (Eq.12)          |
| 8            | $G_v$    | $\text{L day}^{-1}$                  | 61.88, (Eq.11)         |
| 9            | $E_w$    | dimensionless                        | 0.528, (Eq.10)         |
| 10           | $k_1$    | $\text{L kg}^{-1}\text{day}^{-1}$    | 119.29, (Eq.9)         |
| 11           | $Q_w$    | $\text{day}^{-1}$                    | 40.62, (Eq.14)         |
| 12           | $Q_L$    | $\text{day}^{-1}$                    | 0.406, (Eq.15)         |
| 13           | $V_L$    | kg                                   | 0.012, (Eq.16)         |
| 14           | $k_2$    | $\text{day}^{-1}$                    | 0.863, (Eq.13)         |
| 15           | $k_G$    | $\text{day}^{-1}$                    | 0.00065, (Eq.18)       |
| 16           | $k_g$    | $\text{day}^{-1}$                    | 0.00128, (Eq.17)       |
| 17           | $m_f$    | kg                                   | 0.274, (Eq.19)         |
| 18           | $a$      | $\text{day}^{-1}$                    | 0.8668, (Eq.5)         |
| 19           | $b$      | $\mu\text{g kg}^{-1}\text{day}^{-1}$ | 0.04094, (Eq.4)        |
| 20           | $y$      | dimensionless                        | 0.0138, (Eq.19)        |
| Output       |          |                                      |                        |
| 21           | $y$      | $\mu\text{g kg}^{-1}$                | <b>0.3573</b> , (Eq.8) |

#### 3.2 Predicted MIB

Table 3 presents a summary of the model predictions for taint as **MIB** in RAS rainbow trout fish-flesh for the growth period 30 to 270 day. The value of taint in the table is seen to range from 0.0140 to  $0.5621 \mu\text{g kg}^{-1}$ .

#### 3.3 Accumulation of GSM and MIB

To visualise clearly the accumulating taint as **GSM** and **MIB** the data of Tables 2 and 3 are presented as Figure 3 together with the consumer rejection threshold

for off-flavour for both taint molecules. From the figure, it can be seen there is a rapid accumulation of **GSM** and **MIB** in the rainbow trout fish-flesh in the period following 100 days of growth. Viewed this way, the data highlight that accumulation of taint as **GSM** and **MIB** in the fish-flesh has an exponential pattern. The consumer threshold for **GSM** is  $0.9 \mu\text{g L}^{-1}$  and for **MIB** is  $0.7 \mu\text{g L}^{-1}$  (Robertson et al. 2005).

Table 2: Predicted taint ( $y$ ) as **GSM** in rainbow trout fish-flesh with growth-time ( $t$ ) in RAS tanks with  $T = 14 \text{ }^{\circ}\text{C}$  and  $C_w = 0.0066 \mu\text{g L}^{-1}$  (Petersen et al. 2011). (The **bold text** at 150 days is the detailed illustrative simulation presented in Table 1)

| $t$<br>(day) | $m_f$<br>(kg) | $k_1$<br>( $\text{L kg}^{-1}\text{day}^{-1}$ ) | $k_2$<br>( $\text{day}^{-1}$ ) | $k_g$<br>( $\text{day}^{-1}$ ) | $y$<br>( $\mu\text{g kg}^{-1}$ ) |
|--------------|---------------|------------------------------------------------|--------------------------------|--------------------------------|----------------------------------|
| 15           | 0.042         | 228.98                                         | 1.818                          | 0.0016                         | 0.0508                           |
| 30           | 0.052         | 212.98                                         | 1.673                          | 0.0015                         | 0.0631                           |
| 60           | 0.079         | 184.25                                         | 1.418                          | 0.0015                         | 0.0973                           |
| 90           | 0.119         | 159.40                                         | 1.201                          | 0.0014                         | 0.1502                           |
| 120          | 0.181         | 137.89                                         | 1.018                          | 0.0013                         | 0.2317                           |
| <b>150</b>   | <b>0.274</b>  | <b>119.29</b>                                  | <b>0.862</b>                   | <b>0.0013</b>                  | <b>0.3573</b>                    |
| 180          | 0.414         | 103.20                                         | 0.731                          | 0.0012                         | 0.5512                           |
| 210          | 0.627         | 89.28                                          | 0.619                          | 0.0012                         | 0.8502                           |
| 240          | 0.949         | 77.24                                          | 0.525                          | 0.0011                         | 1.3112                           |
| 270          | 1.436         | 66.82                                          | 0.444                          | 0.0010                         | 2.0219                           |

Table 3: Predicted taint ( $y$ ) as **MIB** in rainbow trout fish-flesh with growth-time ( $t$ ) in RAS tanks at  $T = 14 \text{ }^{\circ}\text{C}$  and  $C_w = 0.0032 \mu\text{g L}^{-1}$  (Petersen et al. 2011)

| $t$<br>(day) | $m_f$<br>(kg) | $k_1$<br>( $\text{L kg}^{-1}\text{day}^{-1}$ ) | $k_2$<br>( $\text{day}^{-1}$ ) | $k_g$<br>( $\text{day}^{-1}$ ) | $y$<br>( $\mu\text{g kg}^{-1}$ ) |
|--------------|---------------|------------------------------------------------|--------------------------------|--------------------------------|----------------------------------|
| 15           | 0.042         | 224.92                                         | 3.238                          | 0.0016                         | 0.0140                           |
| 30           | 0.052         | 209.20                                         | 2.981                          | 0.0015                         | 0.0174                           |
| 60           | 0.079         | 180.98                                         | 2.526                          | 0.0015                         | 0.0269                           |
| 90           | 0.119         | 156.56                                         | 2.140                          | 0.0014                         | 0.0415                           |
| 120          | 0.181         | 135.44                                         | 1.814                          | 0.0013                         | 0.0641                           |
| <b>150</b>   | <b>0.274</b>  | <b>117.17</b>                                  | <b>1.537</b>                   | <b>0.0013</b>                  | <b>0.099</b>                     |
| 180          | 0.414         | 101.37                                         | 1.302                          | 0.0012                         | 0.1528                           |
| 210          | 0.627         | 87.69                                          | 1.103                          | 0.0012                         | 0.2359                           |
| 240          | 0.949         | 75.86                                          | 0.935                          | 0.0011                         | 0.3641                           |
| 270          | 1.436         | 65.63                                          | 0.792                          | 0.0010                         | 0.5621                           |

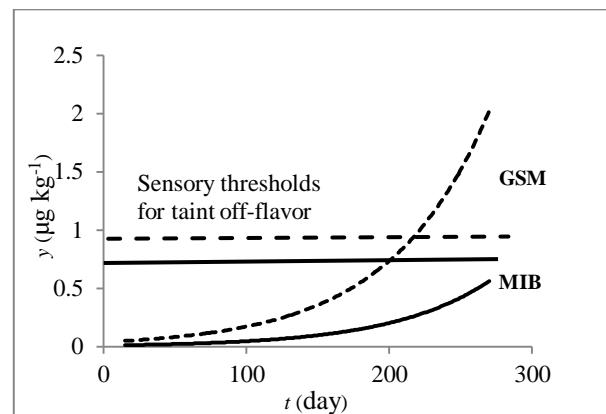


Figure 3: Predicted taint as **GSM** (---) and **MIB** (—) in rainbow trout with RAS growth-time at growth water temperature  $T = 14 \text{ }^{\circ}\text{C}$  with  $C_w$ , respectively, 0.0066 and 0.0032,  $\mu\text{g L}^{-1}$ . The threshold for **GSM** is 0.9, and for **MIB** is  $0.7 \mu\text{g L}^{-1}$  (Robertson et al. 2005)



#### 4. DISCUSSION

It can be seen clearly from the continuous plot of data as Figure 3 that the predicted concentration of **MIB** in rainbow trout fish-flesh is less than that for **GSM** at any growth-time over the period 30 to 270 days. Both curves are however exponential with growth time in the RAS tanks.

The reported sensory threshold for **GSM** in rainbow trout fish-flesh is  $0.9 \mu\text{g kg}^{-1}$  (Robertson et al. 2005). As seen from Figure 3 this threshold is predicted to be reached at about 220 days growth in the RAS tanks.

At harvest (270 days) it is seen from Figure 3 the predicted concentration of **MIB** is  $0.5621 \mu\text{g kg}^{-1}$  (see also Table 2). Given the threshold reported for **MIB** is  $0.7 \mu\text{g kg}^{-1}$  (Robertson et al. 2005) it is concluded that this value will not be reached in a typical RAS growth.

Significantly, **GSM** concentration in the harvested fish is around 3.5 times that predicted for **MIB**. This difference can be attributed to the initial higher  $C_w = 0.0066 \mu\text{g L}^{-1}$  for **GSM** compared with  $C_w = 0.0032 \mu\text{g L}^{-1}$  for **MIB** used in the simulations. Possible reasons for the reported difference in these  $C_w$  values (Petersen et al. 2011) for the two molecules include the diversity of taint causing micro-organisms involved and the different aqueous solubility of the molecules (Jüttner and Watson 2007, Pirbazari et al. 1992). In general, cyanobacteria and actinomycetes are considered to be the major contributors to taint and they are able to produce either **GSM** or **MIB**. Interestingly, certain species are capable of producing both molecules (Jüttner and Watson 2007).

Given the reported consumer sensory threshold for **MIB** in rainbow trout fish-flesh is  $0.70 \mu\text{g kg}^{-1}$  (Robertson et al. 2005) this threshold is not exceeded with typical growth water concentration of  $C_w = 0.0032 \mu\text{g L}^{-1}$ .

The model simulations have been predicted on a constant value of  $C_w$  in the growth water with growth-time ( $t$ ). This may not always be the case and  $C_w$  may vary with growth-time in the tanks due to various factors that include environmental (light, growth water temperature), nutrient(s) load, change(s) of water volume, and the (complex) dynamic nature of the taint causing bacteria. The impact of changing  $C_w$  with time was therefore investigated using simulations of the model.

##### 4.1. Impact of $C_w$ on GSM accumulation in fish-flesh

The impact of varying taint concentration in the growth water ( $C_w$ ) on accumulation in the fish-flesh for both taint molecules can readily be modelled as a function of growth-time,  $t$ , in the tanks.

A suitable form is likely to be increasing  $C_w$  with time in either a linear, or, a moderately-exponential form.

To illustrate this for **GSM**, it is assumed there is a linear dependence with a reasonable estimate for  $30 \leq t \leq 270$  day given by:

$$C_w = 0.0052 + 0.0001t \quad (20)$$

Eq. (20) has been synthesized to predict, respectively, a mean value of  $C_w = 0.0213$ , a minimum, 0.0066 and maximum, 0.0361,  $\mu\text{g L}^{-1}$  at each of 15, 135 and 270 days growth.

Repeat simulations of the model (Table 1) using Eq. (20) to vary  $C_w$  were carried out, and; the predicted results for **GSM** in the flesh of the rainbow trout are presented in Table 4.

Table 4: Predicted taint ( $y$ ) as **GSM** in rainbow trout fish-flesh at varying concentration of  $C_w$  (Eq. (20)) with growth-time ( $t$ ) in RAS tanks

| $t$<br>(day) | $m_f$<br>(kg) | $C_w$<br>( $\mu\text{g L}^{-1}$ ) | $y$<br>( $\mu\text{g kg}^{-1}$ ) |
|--------------|---------------|-----------------------------------|----------------------------------|
| 15           | 0.042         | 0.0067                            | 0.0515                           |
| 30           | 0.052         | 0.0082                            | 0.0776                           |
| 60           | 0.079         | 0.0112                            | 0.1604                           |
| 90           | 0.119         | 0.0142                            | 0.3078                           |
| 120          | 0.181         | 0.0172                            | 0.5638                           |
| 150          | 0.274         | 0.0202                            | 1.0009                           |
| 180          | 0.414         | 0.0232                            | 1.7369                           |
| 210          | 0.627         | 0.0262                            | 2.9627                           |
| 240          | 0.949         | 0.0292                            | 4.9853                           |
| 270          | 1.436         | 0.0322                            | 8.2970                           |

Figure 4 compares and summarizes the impact of assumptions regarding  $C_w$  on the accumulation of **GSM** in the fish-flesh. It is clearly seen in the figure that at harvest taint is 4.1 times if  $C_w$  is assumed to vary with growth-time in the RAS tanks according to Eq. (20). The exact dependence is not known at present. The model simulations have demonstrated a practical utility (and flexibility) in illustrating the impact however.

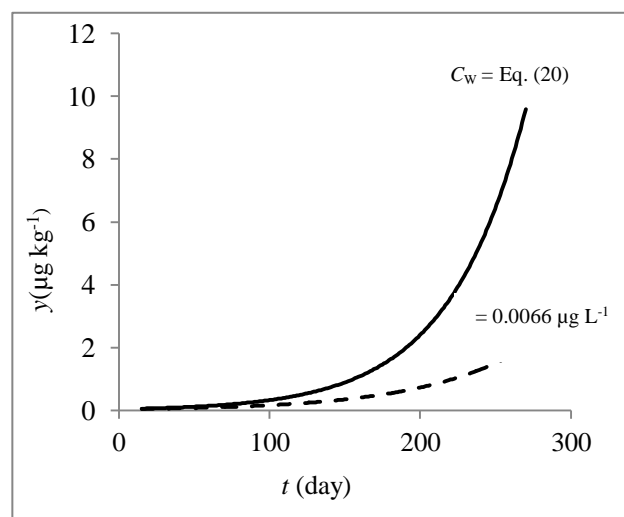


Figure 4: Predicted taint ( $y$ ) as **GSM** in rainbow trout fish-flesh with  $C_w$  constant ( $= 0.0066 \mu\text{g L}^{-1}$ ) and  $C_w$  varying (Eq. (20)), with growth-time ( $t$ )

##### 4.2 Model validation

A model validation was carried out using the independent experimental data of Petersen et al. (2011)

for rainbow trout farmed in open recirculation aquaculture in Denmark.

The mean growth water temperature was 12 °C and the mean mass of the fish at harvest was 0.30 kg. Model inputs and rate constants were computed as described in Section 2.

A comparison of the experimental data ( $n = 14$ ) and model predictions is summarized as Figure 5.

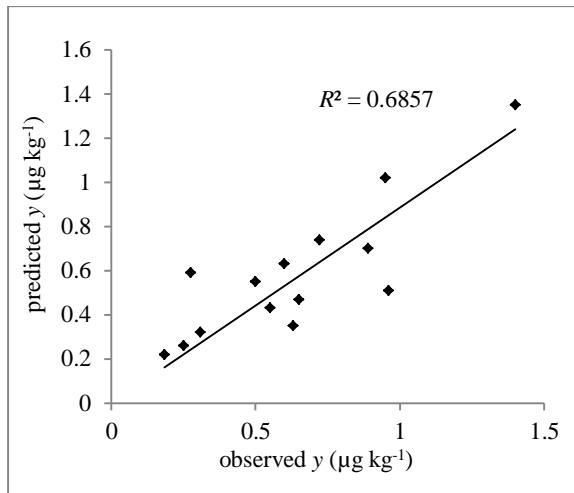


Figure 5: Relationship between observed and predicted **GSM** ( $y$ ) in rainbow trout

The figure shows a good fit to these independent data with  $R^2 = 0.68$  (Snedecor and Cochran 1989). Strictly, the (open) tank system of Petersen et al. (2011) would not be as easily controlled as RAS. The overall good fit suggests the model simulation is free of (gross) computational and programing errors.

#### 4.3 Applying the model to RAS growth protocols to minimise taint taste

A significant advantage of the validated model for taint taste is that it can be used to investigate growth protocols to minimize taint. For example, simulations underscore consumer acceptance of RAS rainbow trout at harvest of 270 day if the RAS water concentration  $C_w$  for **GSM** and **MIB** is maintained at a value, respectively, less than 0.0028, and 0.004  $\mu\text{g L}^{-1}$ .

Clearly, a range of farming practices can be investigated in this way without recourse to further unlimited experimental trials.

#### 4.4 Benefits of the model to the supply chain

Analyses of taint molecules as **GSM** and **MIB** in fish-flesh is difficult, expensive and time consuming and few research institutes have the necessary facilities.

Judicial use of the validated model however can assist famers however to develop growth protocols to result in fish at harvest that have taint at levels lower than the consumer threshold without recourse to extensive analytical analyses. It is clear a key parameter in impacting taint taste is the value of  $C_w$ .

Significantly, the quantitative nature of the model suggests it could be applied through the application of automatic growth controls.

#### 4.5 Experimental studies

Indeed at present, extensive experimental studies are underway by the authors using automatic-controlled dosing of hydrogen peroxide (*ProMinent* Pty Ltd) as an environmental benign treatment to regulate  $C_w$  of the growth-water (*Global Pumps* Pty Ltd) in a commercial farm in Australia (Hathurusingha and Davey 2013).

Although the particular fish is barramundi (*Lates calcarifer*), a premium export protein from Australia, the generalized form of the model will mean findings will be applicable to a range of species.

It is planned the model will be widely used to provide insight into establishing effective growth protocols on resulting taint taste in fish-flesh and result in the elimination of algaecides such as copper sulphate, and periods of purging post-harvest (Tucker 2000). A significant reduction in production costs is expected, together with enhanced consumer acceptance.

### 5. CONCLUSIONS

A generalized, new transient model for predicting taint taste accumulation in flesh of farmed fish has shown a key parameter is the concentration of taint molecule (as either **GSM** or **MIB**) in the growth-water.

For rainbow trout (*Oncorhynchus mykiss*), which is typically harvested at 270 days, the consumer threshold for **GSM** of 0.9  $\mu\text{g kg}^{-1}$  will have been exceeded at 220 days growth with usual commercial practices. For **MIB** the threshold (0.70  $\mu\text{g kg}^{-1}$ ) will not be reached prior to typical harvest.

Model validation against independent (limited  $n = 14$ ) data suggests that accumulation of taint as **GSM** and **MIB** in fish-flesh can be reliably modelled as the sum of an exponential uptake and exponential elimination from the fish-flesh.

Importantly, the generalized form of the model means it can be applied to a range of fish species to investigate a range of growth farming protocols, and; that it could be applied to commercial practice using automatic controls.

The model can be set-up in standard spreadsheeting to facilitate ready communication with a range of users of varying sophistication.

#### NOMENCLATURE

The equation given in parentheses after description refers to that in which the symbol is first used or defined

- $a$  taint elimination coeff.,  $\text{day}^{-1}$  Eq. (5)
- $b$  taint accumulation coeff.,  $\mu\text{g kg}^{-1} \text{day}^{-1}$  Eq. (4)
- $C_{\text{ox}}$  dissolved oxygen,  $\text{mg L}^{-1}$  Eq. (11)
- $C_w$  taint (as **GSM** or **MIB**) in water,  $\mu\text{g L}^{-1}$
- $\frac{dy}{dt}$  rate of change of taint,  $\mu\text{g kg}^{-1} \text{day}^{-1}$  Eq. (1)
- $e$  lipid mass ratio, dimensionless Eq. (16)

$E_w$  chemical uptake efficiency across gill, dimensionless Eq. (9)

**GSM** geosmin

$G_v$  gill ventilation rate, L day<sup>-1</sup> Eq. (9)

$m_f$  mass of the fish, kg Eq. (2)

$k_1$  taint accumulation gills, L kg<sup>-1</sup> day<sup>-1</sup> Eq. (2)

$k_2$  taint elimination gills, day<sup>-1</sup> Eq. (2)

$k_g$  elimination through gills and metabolic transformation, day<sup>-1</sup> Eq. (2)

$k_G$  rate const. growth dilution, day<sup>-1</sup> Eq. (17)

$k_m$  rate const. metabolic transformation, day<sup>-1</sup> Eq. (17)

$K_{OW}$  octanol-water partition coefficient, dimensionless Eq. (10)

**MIB** 2-methylisoborneol

$Q_w$  chemical transport aq. phase, L day<sup>-1</sup> Eq. (13)

$Q_L$  chemical transport lipid phase, L day<sup>-1</sup> Eq. (13)

$R^2$  correlation coefficient

**RAS** Recirculating Aquaculture System

$T$  temperature, °C Eq. (12)

$t$  time, day (Eq. (1))

$V_L$  lipid mass, kg Eq. (13)

$y$  taint (as **GSM** or **MIB**), µg kg<sup>-1</sup> Eq. (1)

#### Greek symbols

$\beta$  pre-exponential constant Eq. (3), Eq. (22)

$\gamma$  exponential constant Eq. (3), Eq. (22)

$e$  lipid mass ratio, dimensionless Eq. (16)

#### ACKNOWLEDGMENT

PIH is grateful to the *CRC Seafood, Australia*, for providing financial assistance and a scholarship to carry out this research.

#### REFERENCES

Anon., 2010. *Feasibility Assessment of Freshwater Arctic Char and Rainbow trout Grow-Out in New Brunswick*.

[http://www.gnb.ca/0027/Aqu/pdfs/NB%20%20FreshwaterTrout%20%20Char%20Study%20-Final%20Report%20\\_Ev\\_-1.pdf](http://www.gnb.ca/0027/Aqu/pdfs/NB%20%20FreshwaterTrout%20%20Char%20Study%20-Final%20Report%20_Ev_-1.pdf) viewed 18 Mar 2014

Arnot, J.A., Gobas, F.A.P.C., 2004. A food web bioaccumulation model for organic chemicals in aquatic ecosystems. *Environmental Toxicology and Chemistry*, 23, 2343-2355.

Evans, L.C., 2010. *Partial Differential Equations (Graduate Studies in Mathematics, Vol. 19)*, 2nd ed. USA: American Mathematical Society.

Gobas, F.A.P.C., 1993. A model for predicting the bio-concentration of hydrophobic organic chemicals in aquatic food-webs: Application to Lake Ontario. *Ecological Modelling*, 69, 1-17.

Gobas, F.A.P.C., Mackay, D., 1987. Dynamic of hydrophobic organic chemical bio-concentration in fish. *Journal of Environmental Toxicological Chemistry*, 6, 495-504.

Hathurusingha, P.I., Davey, K.R., 2013. A new transient-state model for quantitative prediction of taint in farmed barramundi fish (*Lates calcarifer*).

In: *Proc. 34<sup>th</sup> International Conference on Marine Science & Aquaculture-ICMSA 2013*, Amsterdam, The Netherlands, May 15-16, pp. 380-382 (eISSN 2010-3778).

Hathurusingha, P.I., Davey, K.R., 2014. A new quantitative model for predicting taint accumulation as geosmin (GSM) or 2-methylisoborneol (MIB) in farmed-fish in Recirculating Aquaculture Systems (RAS). *Ecological Modelling* (June-in review).

Houle, S., Schrader, K.K., Francois, N.R., Comeau, Y., Kharoune, M., Summerfelt, S.T., Savoie, A., Vandenberg, G.W., 2011. Geosmin causes off-flavour in Arctic charr in recirculating aquaculture systems. *Aquaculture Research*, 42, 360-365. doi: 10.1111/j.1365-2109.2010.02630.x

Howgate, P., 2004. Tainting of farmed fish by geosmin and 2-methylisoborneol: A review of sensory aspects and of uptake/depuration. *Aquaculture*, 234, 155-181. doi:10.1016/j.aquaculture.2003.09.032.

Jüttner, F., Watson, S., 2007. Biochemical and ecological control of geosmin and 2-methylisoborneol in source waters. *Applied and Environmental Microbiology*. 73, 4395-4406. doi: 10.1128/AEM.02250-06

Neely, W.B., 1979. Estimating rate constant for the uptake and clearance of chemicals by fish. *Journal of Environmental Technology*, 13, 1506-1510

Percival, S., Drabsch, P., Glencross, B., 2008. Determining factors affecting muddy flavour taint in farmed barramundi (*Lates calcarifer*). *Aquaculture*, 284, 136-143. doi:10.1016/j.aquaculture.2008.07.056.

Petersen, M.A., Hyldig, G., Strobel, B.W., Henriksen, N.H., Jørgensen, N.O.G., 2011. Chemical and sensory quantification of geosmin and 2-methylisoborneol in rainbow trout (*Oncorhynchus mykiss*) from recirculated aquacultures in relation to concentrations in basin water. *Journal of Agriculture and Food Chemistry*, 59, 12561-12568. doi: 10.1021/jf2033494

Pirbazari, M., Borow, H.S., Craig, S., Ravindran, V., McGuire, M.J., 1992. Physical chemical characterization of five earthy-musty-smelling compounds. *Water Science Technology*. 25, 81-88.

Rimando, A.M., Schrader, K.K., 2003. *Off-flavour in Aquaculture (Vol. 848)*. Washington, DC, USA: American Chemical Society.

Robertson, R.F., Hammond, A., Jauncey, K., Beveridge, M.C.M., Lawton, L.A., 2006. An investigation into the occurrence of geosmin responsible for earthy-musty taints in UK farmed rainbow trout, *Oncorhynchus mykiss*. *Aquaculture*, 259, 153-163. doi:10.1016/j.aquaculture.2004.11.046

Robertson, R.F., Jauncey, K., Beveridge, M.C.M., Lawton, L.A. 2005. Depuration rates and the sensory threshold of geosmin responsible for earthy-musty taint in rainbow trout, *Oncorhynchus*

- mykiss. *Aquaculture*. 245, 89-99.  
doi:10.1016/j.aquaculture.2004.11.045
- Schrader, K.K., Piedrahita, R.H., Rubio., S.A., Rimando, A.G., 2005. Geosmin and 2-methylisoborneol cause off-flavors in cultured Largemouth Bass and White sturgeon reared in recirculating-water systems. *North American Journal of Aquaculture*, 67, 177-180.
- Snedecor, G.W., Cochran, W.G., 1989. *Statistical Methods*, 8th ed. Ames, Iowa, USA: Iowa State University Press.
- Thomann, R.V., Connolly, J.P., Parkerton, T., 1992. Modelling accumulation of organic chemicals in aquatic food-webs. In: Gobas, F.A.P.C., McCorquodale, J.A., *Chemical Dynamics in Aquatic Ecosystems*. Chelsea, USA. : Lewis Publishers, pp. 153-186.
- Tucker, C.S., van der ploeg, M., 1999. Managing off-flavour problem in pond-raised catfish. *Southern Regional Aquaculture Centre (SRAC) Publication*, No 192.
- Wood, S., Williams, S.T., White, W.R., 2001. Microbes as a source of earthy flavours in potable water – A review. *International Biodeterioration and Biodegradation*. 48, 26-40.  
doi.org/10.1016/S0964-8305(01)00064-6

## BIOGRAPHY

Priyantha Hathurusingha is ME(Civil) is a senior research student researching with academic Kenneth R Davey PhD FICHEM CEng FIEAust CPEng in the School of Chemical Engineering, The University of Adelaide.

# A MODIFIED BEER GAME FOR SIMULATION AND OPTIMIZATION TEACHING

Diego Crespo-Pereira<sup>(a)</sup>, Adolfo Lamas-Rodriguez<sup>(b)</sup>, Rosa Rios-Prado<sup>(c)</sup>

<sup>(a, b, c)</sup>University of A Coruña, Spain

<sup>(a)</sup>[diego.crespo@udc.es](mailto:diego.crespo@udc.es), <sup>(b)</sup>[alamas@udc.es](mailto:alamas@udc.es), <sup>(c)</sup>[rrios@udc.es](mailto:rrios@udc.es)

## ABSTRACT

A key aspect that should be considered when planning a simulation course is to develop students' skills for dealing with real cases such as conceptual modelling or validation. However, although desirable, providing real case studies is not always possible. To have a real system available for students to visit and analyze might not be possible due to time or budget constraints. In this circumstances, educational games provide with a suitable means for facing students with realistic case studies. This paper presents a modified version of the classic "Beer Game" developed by the MIT adapted for discrete events simulation (DES) and optimization teaching. The game is played using four computer applications connected through a local area network (LAN). Some aspects of a supply chain that are simplified in the original game are introduced in the game in order to provide with a complex case study in which M&S methodologies application is necessary. Students of a master in industrial engineering participated in this teaching activity achieving in general good results after designing decision making rules by means of simulation.

Keywords: simulation, education, logistics, supply chain, optimization.

## 1. INTRODUCTION AND PREVIOUS WORKS

Modelling and Simulation (M&S) is nowadays a popular tool which finds many applications in various fields, from being a general purpose research methodology (Powers et al., 2012) to a tool for decision making and planning (Cimino et al., 2010). Simulation courses can be found in various educational levels from undergraduate to doctorate studies. It has also been employed as a tool for education, providing a way to design realistic case studies within a controlled environment. One remarkable area of application is operations management since simulation can demonstrate the complexity of a real world scenario (Cleophas, 2012; Costantino et al., 2012).

Although M&S is generally viewed as an accurate and powerful analytic methodology which provides with a deep insight into a system and raises high expectations, misusing it can lead to project failure,

economic losses and a loss of confidence on it as a tool. Thus some authors have analyzed possible causes of failure in simulation practice (McHaney et al., 2002) and provided tips for successful simulation practice (Sturrock, 2011; Clark & Krahl, 2011; Sharda & Bury, 2011; Sadowski, 2007; Schmeiser, 2001). (Robinson, 2002) provides an interesting quality management approach to simulation projects. In this paper, he groups quality features in three categories:

- Content quality, referred to technical aspects of the model and whether it has been properly verified and validated.
- Process quality, referred to how the simulation project is carried out, customer's expectations are fulfilled and there is a fluid communication between the modelers and the customer.
- Results quality, referred to whether simulation results are actually implemented and lead to improvements that can be perceived by the customer.

Although we are not aware of any rigorous study that provides statistics on how often simulation projects fail due to those reasons, (Robinson, 2002) suggests that many projects may fail due to issues from the "process" or "results" category, more than methodological aspects. Many introductory papers (Clark & Krahl, 2011; Shannon, 1998) or simulation textbooks (Banks et al., 2010; Sokolowski & Banks, 2010; Robinson, 2004) remark the importance of project planning, managing the information exchange with process experts, raising reasonable expectations and successful implementation of the proposed measures.

In this regard, simulation teaching must create awareness about possible causes of failure and explain students how to face these problems in a real project. A mistake that must be avoided is to center the course exclusively in coding issues. Coding should ideally represent only a 20% of the simulation time according to the 40-20-40 rule enounced by (Shannon, 1998) that stands for a 40% of effort in data acquisition and conceptual modelling, 20% coding and 40% validation, experimentation and implementation. However, students are often new to simulation software and the

time devoted to model implementation teaching can easily overcome the time allotted for other simulation issues.

The educational game presented in this paper intends to provide with a case study that focuses on the following simulation issues:

- Conceptual modelling.
- Model coding and verification.
- Experimentation and optimization.
- Results implementation.
- Teamwork.

The game rules are described in section 2.

## 2. GAME DESCRIPTION

### 2.1. Rules

The game is based on the classical beer developed by the MIT (Sternan, 1984; Hammond, 1994). This game is intended to demonstrate the bullwhip effect in supply chains and how coordination among the different echelons can help to improve overall profits. The game has become quite popular because it is easy to setup and fast to play (only 45min) but also provides a deep insight in supply chain dynamics. It is so powerful indeed for reproducing the bullwhip effect that it has led to numerous research studies on its causes and its relation to human behavior and cognitive biases (Ancarani et al., 2013; Liu et al., 2009; Bendoly et al., 2006; Bendoly et al., 2009).

The original beer game has been modified and enriched for the purpose of this educational activity. The supply chain in this case is formed by four agents as seen on Figure 1. There is a manufacturer that produces items of a certain product. The manufacturer provides an intermediate warehouse which, in turn, distributes the items between two retailers that receive orders from a set of customers. The flow of orders in the chain is opposite to the flow of products and a rule is imposed so that no item can be sent if there is no supply order issued.

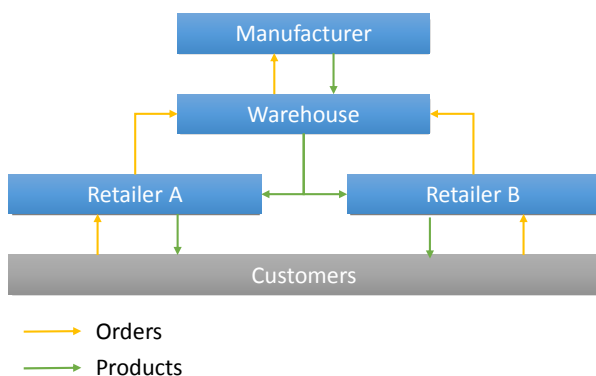


Figure 1. Supply chain diagram for the game.

Each agent has a buffer for storing items with a limited capacity and a list of pending orders received from the next echelon of the supply chain. The manufacturer also has a constrained production capacity

that limits the number of items produced in each turn. Transport capacity among agents is constrained as well but there is no limit imposed on how many orders can be issued.

The decisions to be made by the agents are:

- The manufacturer decides how many items to produce in each term. The pending orders are assumed to be served automatically as stock is available.
- The intermediate warehouse decides how many orders to place to the manufacturer and how many items to serve to retailers A and B.
- The retailers decide upon how many orders to place to the warehouse. Items are automatically served to the customers.

The game is run in 50 turns. A delay of one turn is assumed for transportation and order placement. Thus, for instance, an item served in turn 1 will be received in the next echelon of the supply chain in turn 2. An order issued in turn 1 by a retailer will be received by the warehouse in turn 2. Manufacture is assumed to require a delay of 1 turn with 85% probability and 2 turns with 15% probability. Demand from the final customers is generated randomly according to a certain stochastic process. The mean demand is varied in order to reproduce different possible market trends such as a growing or diminishing demand (Figure 2).

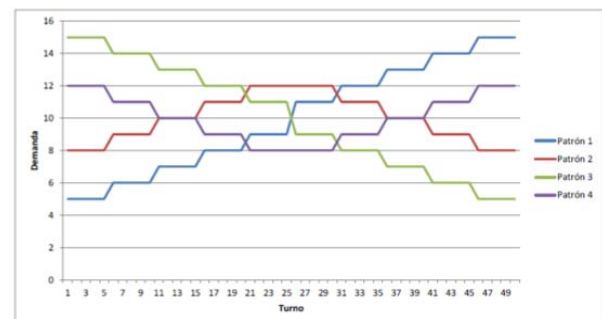


Figure 2. Demand patterns.

### 2.2. Setting

The game has been implemented in four applications developed in Visual Basic, each one for managing each one of the agents. The graphical user interface of each application displays the relevant information for each agent and allows for placing orders or serving customers. The applications are connected via sockets through a local area network using the TCP protocol.



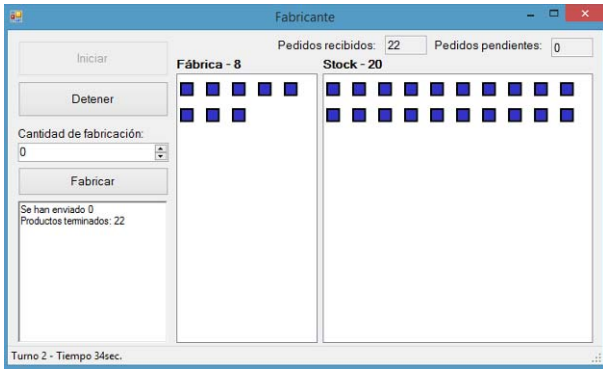


Figure 3. Manufacturer application.

Figure 3 shows the application for the manufacturer. It can decide how many items to produce at each turn. The stock of finished products is displayed as well as the work in progress. Information on the pending orders and incoming orders is provided as well.

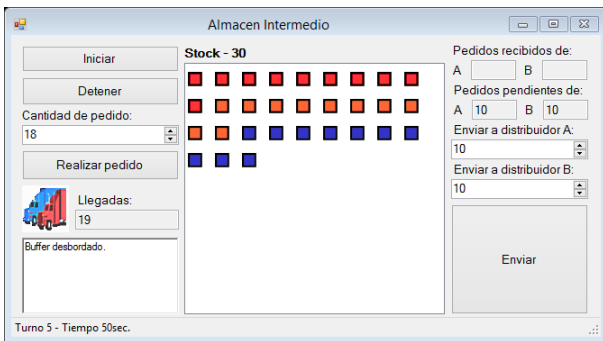


Figure 4. Intermediate warehouse application.

Figure 4 shows the application for the warehouse. It can decide how many items to request to the manufacturer and how many items to serve to each retailer. The stock of products is displayed as well as the number of items received from the manufacturer. Information on the pending orders and incoming orders is provided as well.

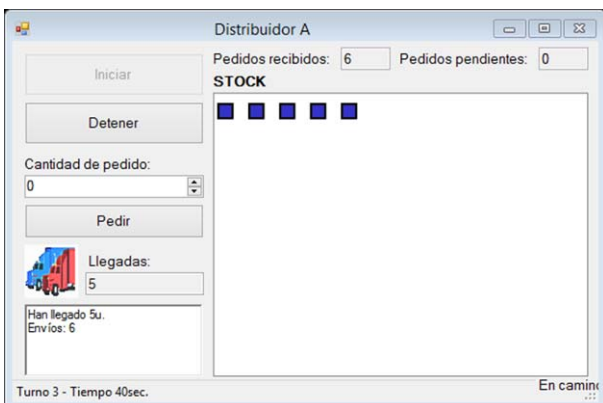


Figure 5. Retailer application.

Figure 5 shows the application for the retailer. It can decide how many items to request to warehouse only. The stock of products is displayed as well as the number of items received from the warehouse.

Information on the pending orders and incoming orders is provided as well.

At each turn, a maximum of 60 seconds is allowed for decision making. Past this time the applications automatically move to the next turn.

### 2.3. Goal

The game goal is to minimize a cost function that depends on the stored items, the back orders and the manufacturer's capacity, transport capacity and buffers capacity. These costs are aggregated by the following equation:

$$\begin{aligned}
 COST = & 23MC + 5 \cdot (B_M + B_W + B_{RA} + B_{RB}) \\
 & + 12(T_{M \rightarrow W} + 2T_{W \rightarrow R}) \\
 & + 50(\overline{S}_M + \overline{S}_W + \overline{S}_{RA} + \overline{S}_{RB}) + 2.4 \\
 & \cdot 50(\overline{BO}_M + \overline{BO}_W + \overline{BO}_{RA} + \overline{BO}_{RB})
 \end{aligned}$$

Where MC stands for the capacity of the manufacturer, B the buffers, T the transporter, S the average stock and BO the average back orders. The subscript M refers to the manufacturer, W to the warehouse and RA and RB to the retailers A and B respectively.

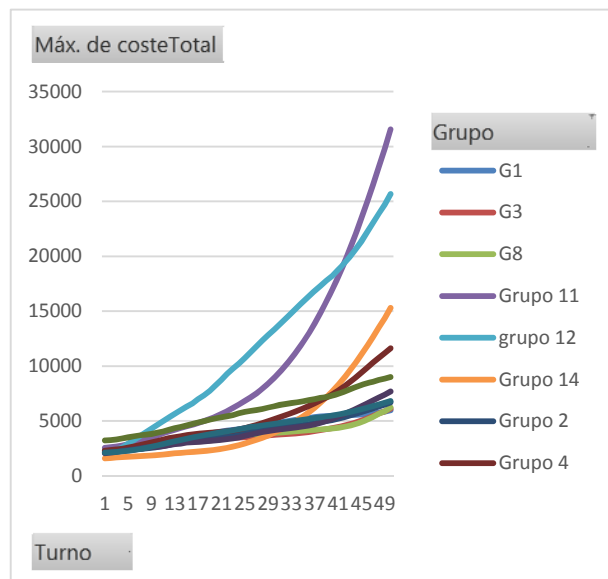
### 3. TEACHING ACTIVITY

The teaching activity consists of developing a simulation model by teams of four and using it to design decision rules and minimize costs competing with other teams. The activity is carried out in the following steps:

1. Students, prior to been given information about the case, participate in a first gameplay in which they make their decisions intuitively. They are explained the basic rules of the game and the established constraints.
2. Project planning. Members of each team must identify the required activities to carry out the work and assign tasks to each member as well as estimate task durations and set up intermediate deadlines which are then supervised by the teachers.
3. The students develop a conceptual model of the game based on their experience and the information provided.
4. Model coding. Extendsim was used as the simulation software in this course, although other software could be employed.
5. The students must plan a set of simulation experiments in order to analyze the sensitivity of the cost function to the different variables that can be modified (such as the capacities) and the parameters used in the ordering policies adopted. These experiments are intended to guide a following optimization step in which they search for a good solution to the problem.



idea that simulation is a useful and practical tool for solving real problems. Students could assess the effect of implementing their simulation results.



**Figure 8. Scores in the final gameplay as a function of turn number.**

## 5. CONCLUSION

The educational game described in this paper constitutes a tool for teaching the practical use of M&S technologies as well as optimization techniques. Students had to develop the conceptual model of a system that they had experienced, code it and design decision rules, conduct experiments to test them and optimize. Their results had to be put into practice in a final gameplay in which they were scored upon their results. Most teams achieved great improvements compared to the initial gameplay, thus successfully applying M&S technology in a realistic environment.

Teams competed intensely for achieving the highest score, being motivated by the score by ranking reward. Students' involvement in the activity was higher than expected and led in general to an enriching experience for both the teachers and the students.

## REFERENCES

- Ancarani, a., Di Mauro, C. & D'Urso, D. (2013) A human experiment on inventory decisions under supply uncertainty. *International Journal of Production Economics*, **142** (1): 61–73. <http://linkinghub.elsevier.com/retrieve/pii/S0925527312003696>, accessed 15 May 2014.
- Banks, J., Carson, J.S., Nelson, B.L. & Nicol, D.M. (2010) *Discrete-Event System Simulation*. Pearson.
- Bendoly, E., Croson, R., Goncalves, P. & Schultz, K. (2009) Bodies of Knowledge for Research in Behavioral Operations. *Production and Operations*

*Management*, **19** (4): 434–452. <http://doi.wiley.com/10.1111/j.1937-5956.2009.01108.x>, accessed 6 May 2014.

Bendoly, E., Donohue, K. & Schultz, K. (2006) Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management*, **24** (6): 737–752. <http://linkinghub.elsevier.com/retrieve/pii/S0272696305001579>, accessed 2 August 2010.

Cimino, A., Longo, F. & Mirabelli, G. (2010) A General Simulation Framework for Supply Chain Modeling: State of the Art and Case Study. *International Journal of Computer Science Issues*, **7** (2): 1–9. <http://arxiv.org/abs/1004.3271>, accessed 16 May 2014.

Clark, R. & Krahl, D. (2011) Roadmap to success: Your first simulation model. In: *Proceedings of the 2011 Winter Simulation Conference (WSC)*. IEEE, pp.1465–1475. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6147865>, accessed 18 October 2012.

Cleophas, C. (2012) Designing serious games for revenue management training and strategy development. In: *Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC)*. IEEE, pp.1–12. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6465154>, accessed 15 May 2014.

Costantino, F., Di Gravio, G., Shaban, A. & Tronci, M. (2012) A simulation based game approach for teaching operations management topics. *Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC)*: 1–12. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6465028>.

Hammond, J. (1994) *The beer game: Board version*. Harvard Business School: Boston.

Liu, H., Howley, E. & Duggan, J. (2009) Optimisation of the Beer Distribution Game with complex customer demand patterns. *2009 IEEE Congress on Evolutionary Computation*: 2638–2645. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4983273>.

McHaney, R., White, D. & Heilman, G.E. (2002) Simulation Project Success and Failure: Survey findings. *Simulation & Gaming*, **33** (1): 49–66. <http://sag.sagepub.com/cgi/doi/10.1177/1046878102033001003>, accessed 9 November 2012.

Powers, M.J., Sanchez, S.M. & Lucas, T.W. (2012) The exponential expansion of simulation in research. In: *Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC)*. IEEE, pp.1–12. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6465125>, accessed 15 May 2014.

Robinson, S. (2002) General concepts of quality for discrete-event simulation. *European Journal of Operational Research*, **138** (1): 103–117. <http://linkinghub.elsevier.com/retrieve/pii/S0377221701001278>.

Robinson, S. (2004) *Simulation□: The Practice of Model*. Wiley and Sons.

Sadowski, D.A. (2007) Tips for successful practice of simulation. In: *2007 Winter Simulation Conference*. IEEE, pp.107–112. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4419593>, accessed 18 October 2012.

Schmeiser, B.W. (2001) Some myths and common errors in simulation experiments. In: *Proceeding of the 2001 Winter Simulation Conference (Cat. No.01CH37304)*. IEEE, pp.39–46. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=977244>, accessed 4 July 2012.

Shannon, R.E. (1998) Introduction to the art and science of simulation. In: *1998 Winter Simulation Conference. Proceedings (Cat. No.98CH36274)*. IEEE, pp.7–14. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=744892>, accessed 8 November 2012.

Sharda, B. & Bury, S.J. (2011) Best practices for effective application of discrete event simulation in the process industries. In: *Proceedings of the 2011 Winter Simulation Conference (WSC)*. IEEE, pp.2315–2324. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6147942>, accessed 18 October 2012.

Sokolowski, J.A. & Banks, C.M. (2010) *Modeling and Simulation Fundamentals*.

Sterman, J. (1984) *Instructions for running the beer distribution game (D-3679)*. Sloan Sch. of Manag, MIT.

Sturrock, D.T. (2011) Tips for successful practice of simulation. In: *Proceedings of the 2011 Winter Simulation Conference (WSC)*. IEEE, pp.1415–1422. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6147861>, accessed 18 October 2012.

## AUTHORS BIOGRAPHY

**Diego Crespo Pereira**, PhD, MSc Industrial Engineer. Diego is Assistant Professor of the Department of Economic Analysis and Company Management of the UDC. He also works in the Integrated Group for Engineering Research (GII) of the UDC as a research engineer and project manager. He is mainly involved in the development and coordination of R&D projects related to industrial and logistic processes optimization.

**Rosa Rios Prado**, MSc Industrial Engineer. Rosa is Adjunct Professor of the Department of Economic Analysis and Company Management of the UDC as well as a research engineer in the GII of the UDC. Rosa is mainly devoted to the development of transportation and logistic models for the assessment of multimodal networks and infrastructures.

**Adolfo Lamas Rodriguez** graduated from the University of Vigo in 1998. He holds an MSc and a PhD in Industrial Engineering. He combines his research activities in the GII and his position as a senior engineer in the Spanish leading shipbuilding company *Navantia*. He is also Associate Professor in the University of A Coruna. His main areas of research are offshore wind and simulation and optimization of manufacturing processes.

# A SIMPLE QUEUE MODEL FOR AN APPOINTMENT SYSTEM AND APPLICATIONS IN A HOSPITAL CT SCAN FACILITY

F. Boenzi

DMMM (Department of Mechanics, Mathematics and Management) – Engineering Faculty at Politecnico di Bari (Italy)

[boenzi@poliba.it](mailto:boenzi@poliba.it)

## ABSTRACT

Key features of modern healthcare services in developed countries, in a general context characterized by strict cost control and aging population, are flexibility in access to care for different types of patient flows (e.g. inpatients, ED patients and outpatients) and capability of adapting to quantitative service demand change, commonly posed by outpatients. An expensive resource, such as a CT scanner or an MRI machine, is often shared by diverse patient flows and this single server has to assure satisfactory performances (in terms of waiting time) to both the random and the planned service demand component, implementing appropriate buffering and priority rules. For outpatients, requesting access to an appointment service in general, waiting time is represented by waiting time for the appointment day and waiting time experienced at the facility. The governing rules determining these performance results can be treated as separate problems and, in the present paper, the first kind of wait, commonly related to the existence of long waiting lists, is considered. A simple general analytical model, adopting statistical considerations, is proposed and, successively, applications in a CT scan facility in a public hospital, supported by simulation, are illustrated.

Keywords: appointment scheduling, queues, health care, simulation

## 1. INTRODUCTION

In many European countries, healthcare services still represent a basic pillar of the welfare system; however, public expenditure for new investment is often unrealizable by the local NHS (National Health Service) Departments/Authorities.

As a consequence, it is not uncommon that, at the examination facility, an expensive resource for digital imaging diagnosis (e.g. a CT scanner, MRI equipment) is unique and shared among diverse patient flows. On the other hand, for patients in critical conditions (e.g. patients from the Emergency Department in a hospital), an immediate response is crucial, but, ideally, it would require dedicated equipment. As a trade-off solution, in a single-server system, appropriate buffering and priority rules have to be implemented in order to assure satisfactory performances (in terms of waiting time) to

both the random and the planned service-demand component.

The last consists of outpatients who are given appointments well in advance with respect to the examination date; for them a booking method, in most cases supported by IT tools, is in place. The coexisting presence of a random component can be considered beneficial in order to attain a “satisfactory” utilization level of expensive resources, but, on the other hand, as well-known, “emergency medical (ambulance) and fire service systems operate at only 10-15% utilization. Short response time is essential for these systems and hence excessive waiting times are simply not allowable. Since variability is unavoidable (people don’t schedule their heart attacks), the only way to keep waiting to a minimum is to have a great deal of excess capacity” (Hopp 2008). Therefore, two opposite needs must be conciliated. On the one hand “hospital managers have every incentive to keep these machines fully utilized” and “this is often done by filling most, if not all, examination slots during the day with outpatient appointments” (Green et al. 2006). On the other hand, this is in contrast with the outlined features of an emergency service, since, in order to keep short waiting time, urgent cases should find, ideally, the server idle upon their arrival (null waiting time) or working at low utilization levels (which means the server is found idle most of the times). Because of the presence of appointed patients and/or inpatients (when the service facility is inside a hospital), priority rules must necessarily be adopted and, inevitably, longer waiting times for outpatients result.

Extensive literature exists, regarding the analysis of system performance in terms of waiting time at the facility for different patient classes and the risk of overtime (prolonged working time beyond the daily shift), in relation to appointment scheduling and other problem parameters. However, the investigation on specific characteristics of healthcare appointment systems in relation to the existence of long waiting lists for outpatients appear to be overlooked.

The present paper is organized as follows: in Section 2, a brief literature overview is given; in Section 3, some general considerations about the behavior of a generic appointment system in steady-state conditions are illustrated. In Sections 4 and 5 a possible analytical



approach, adopting statistical results, for the description of the outpatient buffer in unsteady-state conditions is proposed. In Section 6, basic simulation models, adopting the illustrated concepts for the specific case-study of a CT scan facility, are introduced and results are discussed. Finally, remarks and general conclusions are drawn in Section 7.

## 2. LITERATURE SURVEY

The kind of system illustrated in the introduction can be treated of as a single-server queue model with a non-preemptive priority rule (Adan and Resing 2002).

In fact, even though ED patients are served as soon as possible (and the FCFS, first-come-first-served, rule is adopted for patients of the same class), the process cannot be interrupted once started for any patient, even if he belongs to a lower priority class. Using Kendall's notation, it isn't possible to describe the system in a short comprehensive manner, because of the diverse nature of these patient flows. For the appointed patients, an appropriate representation could be in the form  $D/G/I$  (deterministic arrivals based on the appointment schedule in place and general *pdf* for service times). However, the mathematical model becomes easily untreatable when service time moves away from a Markovian distribution or when it is not assumed to be deterministic, or when phenomena like no-shows or unpunctuality must be taken into account ( $D + noise/M/I$  models, cited, for example, in (Hassin and Mendel 2008). In (Kuiper 2012), the author illustrates the solution method for finding the optimal appointment schedule (which minimizes a cost function of the sum of expected patients' waiting times and server's idle times) by means of a phase-type distribution approximation of service time ( $D/PH/I$  setting), neglecting the mentioned disturbances. (Kaandorop and Koole 2007) find the optimal outpatient appointment schedule, with no-shows and exponential service time, for the minimization of an objective function which includes tardiness (time exceeding the given session period), discretizing the time session and employing a local search algorithm. Exploring over a limited number of time points for appointments, derived dividing the continuous time session into discrete intervals, makes sense for practical applications and doesn't limit the general principles behind the optimal solution, found to be in "dome shapes", as observed in literature. (Hassin and Mendel 2008) use a continuous time analytical model  $D/M/I$ , including no-shows and an objective function to be minimized (linear combination of waiting time cost and server's availability time cost). They obtain "what is believed to be an optimal solution by applying sequential quadratic programming (SQP)".

The possibility of including in a mathematical model additional types of patients aside the appointed ones often implicates the assumption of deterministic service times and rigidly fixed slot durations. For example, (Green et al. 2006) formulate the real-time capacity allocation problem for a medical diagnostic facility (MRI) accessed by outpatients, inpatients and

emergency clients (with associated arrival probabilities and a daily expected total reward function, with overtime implicitly included in the form of penalty costs for unseen patients), as a finite horizon dynamic programming problem. They highlight the relationship between the "strategic" decision level of the system represented by the Appointment Schedule (AS) rule, and the "tactical" decision level regarding the set of dynamic priority rules for admitting patients into service. For the latter, they investigate the properties of the optimal capacity allocation policy, and propose a simpler linear approximation (LA) heuristic too. Results are illustrated for different heuristic appointment policies ("fill all slots", "balanced" and "newsvendor") and problem parameters. The mentioned decision levels could also be framed in the more general context of healthcare planning-and-control as "offline" and "online" operational levels (Hans et al. 2011), in which the first "concerns the in-advance planning of operations" and the second "involves control mechanisms that deal with monitoring the process and reacting to unforeseen or unanticipated events". (Kolisch and Sickinger 2008) examine the same maximization mathematical model, in the form of a Markov Decision Process, for a CT scan facility, serving the above mentioned patient classes. They illustrate the results of the optimal tactical decision policy (found by the "backward induction algorithm") and compare them with simpler decision rules, which can be applied manually (Linear approximation, FCFS and random choice), under three commonly adopted AS schemes. (Sickinger and Kolisch 2009) illustrate the results of the objective function, obtained optimally solving the associated stochastic dynamic decision problem, for a Generalized Bailey-Welch (GBW) schedule, compared to optimal and Neighborhood Search (NS) heuristic, under various combinations of problem parameters. (Patrick 2012) develops a Markov Decision Process model of an outpatient clinic and "demonstrates that a short booking window does significantly better than open access". The model takes into account the no-show rate, but assumes a deterministic service-time.

In order to remove the numerous assumptions that a mathematical model poses, extensive literature, adopting discrete event simulation models, exists and the authors follow an experimental approach. (Klassen and Rohleder 1996) carry out a full factorial ANOVA analysis on a simulation model (with lognormal service time and no-shows), taking as decision variables 1) three possible positions of a couple of open slots to accommodate urgency; 2) a total of 10 known pre-defined AS rules. They illustrate results for server idle time, total client waiting time and a linear combination of the two, for various environmental factors. In (Rohleder and Klassen 2002), attention is shifted to the observation of the system from the wider perspective of the scheduler of appointments, considering a rolling-horizon environment with fluctuating demand loads. The model takes into account the additional client



performance measure of the number of days between call and appointment, and the server oriented measure of the ending time of the session (possible overtime). Different trade-off solutions for the combination of overloading rules and rule delay choices are illustrated, under six examination demand patterns, highlighting that the managerial choice depends on which performance measure is considered to be the most important. In (Hutzschenreuter 2004) a simulation model, which accounts for no-shows and unpunctuality of patients, is utilized to test the performances of a general outpatient clinic (in terms of mean patient waiting time and mean physician's idle time), adopting three well-known AS rules and under different combinations of mean service time and variability. (Zhu et al 2009) find empirically, on a simulation model of a specialist outpatient clinic, the optimal number of scheduled patients per session for three system configurations, considering four performance measures (queue length, patient waiting time, overtime and doctors' utilization) and a previously established AS rule for each configuration. They point out how the result can vary according to the chosen performance measure. (Wijewickrama 2006) examines the simulation model of an outpatient department with different patient flows (appointed and not, with priority rule for the first ones), not including no-shows and unpunctuality. He illustrates the efficient frontier of results in terms of weighted average of patient waiting time and servers' idle time, adopting four basic AS rules and some combinations.

### 3. GENERAL MODEL FRAMEWORK

As initially stated, healthcare services open to different patient flows are usually subject to demand variations. These variations are quite ordinary for the random component in the course of a day (e.g. ED patients) and in simulation models are usually averaged as *HPP*.

In the longer term, also the appointed component of the demand can exhibit variations, whereas the number of appointment slots could not have been adjusted adequately in the course of time. As a result, in case of average demand greater than the number of available appointments in a time period (for example, a week), the backlog represented by the list of waiting outpatients will increase and, correspondingly, mean waiting time for an appointment. Attention can also be focused on the opposite case, when a long waiting list exists and, with the aim of reducing it, agreed decisions between managers and medical staff have to be taken about possible increments of the number of appointments (exceeding the current demand). A question in this case could regard how long time it is necessary to maintain the system's workload greater than normal to attain the objective and to reach a balanced situation.

It's to keep in mind that short waiting time for an appointment may be an external constraint, set at the legislative level for specific patient classes. The last classification is often based on urgency criteria

(outpatient priority classes) and must be adopted by family or specialty doctors, when they prescribe a diagnostic examination, as in the case of CT scans (for example, for patients to be followed-up).

Setting a greater number of appointments, when needed, should be framed in the general context of a demand-offer problem, disjointed from the assignment, which could be addressed on a second step, regarding how to displace additional appointments appropriately (avoiding excessive waiting time at the facility and/or overtime).

Looking at the entire appointment process, at the origin, we can find the service request generation, for example phone-calls at call-centers, which, as a whole, constitute the demand. Service requests are differentiated for each outpatient class and enqueued into the corresponding buffer, usually processed on a FAFS (first-appointment-first-serve) basis. A patient is given an appointment date corresponding to the first available slot (for that priority class/buffer) and slots in the future are filled so on. Therefore, any appointment system, at a macroscopic scale, can be thought of as a "black-box", which processes the random demand with a set "productive capability", equal to the number of appointments given in a time period, according to the current schedule in place.

Typically, in healthcare systems, a repetitive scheduling scheme is specified over a week, with possible variations from day to day, setting time and number of appointments for each class of outpatients, when separated booking agendas are in place (commonly corresponding to the mentioned priority classes). The approach presented herein treats each buffer (making up a particular booking agenda) separately. Therefore, if we consider the time measurement unit equal to the repetitive schedule time horizon (for example, one week) and we renounce the exact quantification of patient waiting time within one time unit (the week during which he will be given the appointment), we can look at each buffer in the system as a *G/D/1* single server queue. Our attention is focused on waiting time for an appointment and on how it is related to the buffer length.

The arrival process of clients (with an associated *pdf* for inter-arrival times in general form) is represented by the service request generation. Service time is not related to real machine times, but it's determined by the set "productive capability", or number of appointments. It can be considered a fictitious deterministic time, under the hypothesis that, during each period (time unit), the stated number of appointments will be processed for sure and no patient will be re-scheduled. Therefore, disregarding the way in which appointments are really allocated in the course of the scheduling period of time length  $T=1$  [time unit], we are just interested in their total number per period, said  $N_p$ . Assuming an equally spaced distribution of the appointments, the fictitious deterministic service time  $t_0$  is simply equal to  $T/N_p$  [time units], with constant service rate  $\mu = N_p/T$ . Of course, the real utilization of

the server isn't related to  $t_0$ , because, for the feasibility of the scheduled appointments, the mean processing time must necessarily be smaller (e.g. mean time for an examination / a visit). Furthermore, the real overall utilization of the service will be the result of sharing the resource among all the different patient flows. Nevertheless, such a system  $G/D/I$  can describe appropriately what happens in the outpatient-buffer (not beyond it) and the above defined service rate is the fundamental parameter to be compared to the average arrival rate of clients during a time period. If we define  $\overline{N_A}$  as the average number of service requests (arrivals) in  $T$  and, correspondingly,  $\lambda$  as the arrival rate  $\overline{N_A}/T$ , it is well-known that the system can remain in steady-state conditions as far as the condition  $\lambda < \mu$  is met, that's to say  $\overline{N_A} < N_p$ . An approximate expression of the expected waiting time in the buffer is given writing Kingman's equation in the following form:

$$\varphi_b \cong \frac{c_a^2 + c_0^2}{2} \frac{U}{1-U} t_0 = \frac{c_a^2}{2} \frac{U}{1-U} \frac{T}{N_p} = \frac{c_a^2 \overline{N_A}}{2} \frac{T}{N_p N_p - \overline{N_A}}. \quad (1)$$

$$U = \lambda/\mu = \overline{N_A}/N_p;$$

$c_a$ : coefficient of variation of the arrival process;  
 $c_0$ : coefficient of variation of the fictitious service process; being the process deterministic = 0;  
 $T = 1$  [time unit].

The  $U$  ratio, although isn't the real utilization of the server, represents the utilization of the capacity at disposal. Under the assumption of a purely random arrival process with average  $\overline{N_A}$ , a Poisson distribution can be adopted, the system can be denoted as  $M/D/I$  and in this case  $c_a=1$ . Then, it's possible to observe that, since  $N_p$  must be an integer number (scheduled number of appointments), adopting the following very simple rule that assures the demand/offer balance of the system (steady-state conditions):

$$N_p = \text{upper int} \{ \overline{N_A} \} + 1. \quad (2)$$

the calculated value of  $\varphi_b$  would always remain lower than  $0.5 T$  [time units], as shown in Fig.1.

Furthermore, from Little's law, the average number of clients in the buffer  $\varphi_b \cdot \lambda$ , which represent the waiting list length, would remain lower than half  $\overline{N_A}$ . Numerically, it is also possible to verify that for  $\overline{N_A} > 7.2$  and for  $\overline{N_A} > 17.1$  the value of  $U$  always results above  $0.8$  and  $0.9$ , respectively, which means that, adopting (2), the system would be over-capacitated in percentages from +25% to +11.1%, respectively, with reference to the demand.

Average waiting time for an appointment within one period  $T$  (for example, one week) implies that for the scheduler it's possible to implement an approach very similar to "open access", since service demand is met quite immediately. Of course, it will always be possible to set appointments far in the future (e.g. for

examinations which require previous particular treatments), but the deriving waiting time becomes a scheduler's choice and is not determined by the governing laws of the system.

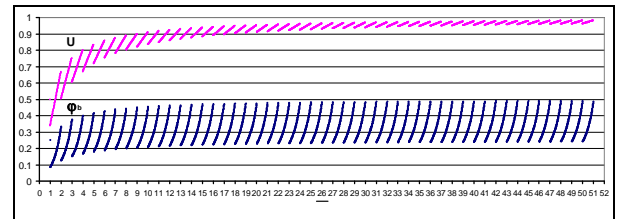


Figure 1: Expected Waiting Time in the buffer  $\varphi_b$  and  $U$  versus  $\overline{N_A}$  for the  $M/D/I$  system adopting (2)

In general, it's possible to affirm that for such appointment systems (in steady-state conditions), which can be summarized by the  $G/D/I$  notation, filtering the demand has the advantage of isolating just one source of variability, determined by the arrival process. Moreover, as observable from Kingman's equation, assumed a Markovian arrival process, the term accounting for stochasticity is  $1/2$  and, differently from what happens for other generic queue systems, average waiting time can't reach elevated values, unless  $\overline{N_A}$  is very close to  $N_p$ . In fact, conversely, from equation (1) and defined  $\Delta$  as the difference ( $N_p - \overline{N_A}$ ), it's possible to derive limiting values of  $\Delta$  for which  $\varphi_b$  is  $k$  times smaller than the period  $T$ . In mathematical form:

$$\varphi_b \leq kT : \Delta \geq \frac{1}{1/N_p + 2k}. \quad (3)$$

For high values of  $N_p$ , neglecting  $1/N_p$ ,  $\Delta \geq 1/(2k)$ .

In Fig. 2, example values for  $\Delta$  and  $U$  are reported for  $k=1,2,3$  (sub-index of labels).

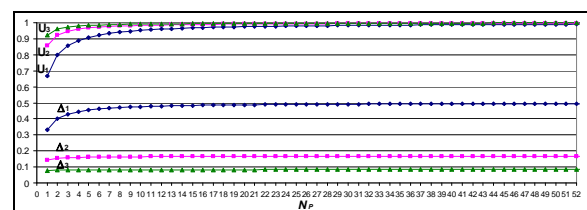


Figure 2:  $\Delta$  and  $U$  derived from (3) versus  $N_p$

As an example of a practical application, set  $N_p=10$ , the average waiting time in the buffer will never exceed one period  $T$  ( $k=1$ ) as far as  $\Delta \geq 0.4762$ , that's to say  $\overline{N_A} \leq 9.5238$ . It's interesting to note that, for  $k=1$ ,  $\Delta$  is always less than  $0.5$  and, practically, becomes independent from  $N_p$ , for increasing values.

The elements outlined so far highlight the importance of monitoring the behavior of the demand of a healthcare service in a systematic manner, in order to take corrective actions and keep backlog under control.

In the following two sections, the behavior of the buffer of a generic appointment system in unsteady-state conditions will be investigated by means of an

analytical approach. Firstly, the case with increasing buffer ( $\overline{N}_A > N_P$ ) and secondly the case with initial backlog and decreasing buffer ( $\overline{N}_A < N_P$ ) will be discussed.

#### 4. CASE WITH INCREASING BUFFER

In the following, the unbalanced situation with  $\lambda > \mu$  is examined, starting from an empty system and considering the instantaneous number of patients/clients in waiting list  $N_b(t)$  and the instantaneous value of  $\varphi_b(t)$ , defined as waiting time a patient exiting at time  $t$  has experienced. Successively, average values are calculated.

##### 4.1. Instantaneous values

Adopting the discretization of time in intervals of length  $T$ , the following notation is introduced:

- $N_{bi}$ : number of waiting patients in the buffer at the end of the  $i$ -th period;
- $N_{Ai}$ : number of requests for appointments arrived at the system and taken into consideration for processing during the  $i$ -th period (indeed collected at call-centers in the course of the previous  $(i-1)$ -th period, for practical reasons of scheduling assignment); neglecting waiting time within one period of time, as a “physiological” wait, it’s then possible to assume the number of requests  $N_{Ai}$  as arrived during the  $i$ -th period and the server always able to process them, within the limit of its productive capability  $N_P$  for each period.

At the end of each period:

$$N_{b1} = \max\{0, N_{A1} - N_P\};$$

$$N_{b2} = \max\{0, N_{b1} + N_{A2} - N_P\};$$

....

$$N_{bi} = \max\{0, N_{b(i-1)} + N_{Ai} - N_P\}; \quad \forall i.$$

Assuming that the buffer does not empty (it always remains a positive integer), also during the initial intervals, from the recursive expression of  $N_{bi}$ , it’s possible to derive:

$$N_{bi} = \sum_{k=1}^i N_{Ak} - iN_P. \quad (4)$$

From (4), assuming that arrivals belong to a Poisson distribution, applying the *Central Limit Theorem* (see Appendix A.1 for details), it’s possible to derive:

$$N_{bi} \cong i(\overline{N}_A - N_P) \pm \sqrt{\overline{N}_A} \cdot t_{\text{value}\{i-1\}} \sqrt{i}; \quad (\text{for } i \geq 2). \quad (5)$$

(of course, because of the statistical uncertainty and approximation,  $N_{bi}$  isn’t anymore an integer number and negative values are not consistent). In Fig. 3, the factor in (5) due to uncertainty and depending solely on  $i$  is reported, together with power interpolation functions (dashed lines) in the form  $a \cdot i^b$ , in the interval

0-500, for the two confidence levels 90% and 95%. In particular, setting lower limits for  $i$  so that the percentage difference between the function and its interpolation is lower than 20% also for initial values of  $i$  (then, it rapidly decreases to around 1%), it’s possible to find ( $a=1.8255$ ;  $b=0.4814$ ) for  $\gamma=90\%$ , with  $i \geq 5$  and ( $a=2.2206$ ;  $b=0.4778$ ) for  $\gamma=95\%$ , with  $i \geq 6$ .

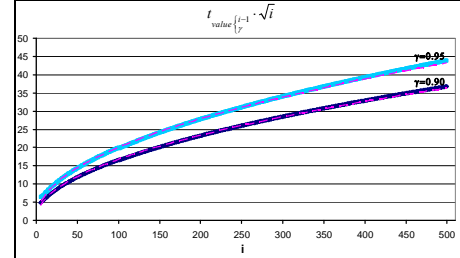


Figure 3: Values of  $t_{\text{value}\{i-1\}} \cdot \sqrt{i}$  versus  $i$  for  $\gamma=90\%$  and  $\gamma=95\%$  and respective interpolating functions (dashed).

Considering now  $N_b$  as a time-continuous function, being  $t=i \cdot T$  and substituting time as variable in the above reported interpolating function, it’s possible to write:

$$N_b(t) \cong (\overline{N}_A - N_P) \frac{t}{T} \pm \sqrt{\overline{N}_A} \cdot a \left(\frac{t}{T}\right)^b. \quad (6)$$

with  $t \geq 5T$  and  $t \geq 6T$  for  $\gamma=90\%$  and  $\gamma=95\%$  respectively. The second quantity to be analyzed is waiting time in the buffer for each patient and the following quantities are considered:

- $\varphi_i$ : total flow-time time ( $t_{out\ i} - t_{in\ i}$ ) for the  $i$ -th patient, assuming that the time instant of the appointment request generation is shifted one period later (requests must be collected and scheduled at least in a previous period and waiting time within one period of time can be considered unavoidable and neglected). In particular, it is assumed  $t_{in\ 1} = 0$ .
- $\varphi_{bi}$ : waiting time in the buffer for the  $i$ -th patient; assuming that the server is always busy (which is the same previous assumption regarding the buffer never empty) and patients are served on a FCFS basis, exit time  $t_{out\ i}$  is deterministic and equal to  $i \cdot t_0 = i \cdot T/N_P$ ; moreover,  $\varphi_{bi} = \varphi_i - t_0$ .

Observing that  $t_{in\ i}$  (time instant of the appointment request) can be viewed as the sum of the preceding  $(i-1)$ , for  $i \geq 2$ , inter-arrival times  $t_{ak}$  between two generic contiguous requests, it’s then possible to write:

$$\varphi_{b1} = \varphi_1 - t_0 = t_{out\ 1} - t_{in\ 1} - t_0 = 0;$$

<sup>1</sup> Indeed, also for the function  $N_b(t)$ , it should be considered  $\max\{0, \text{calculated value}\} \forall t$ , but a resulting empty buffer at any time instant would invalidate the expression because it would be equivalent to shifting time back to zero.

$$\begin{aligned}
\varphi_{b2} &= \varphi_2 - t_0 = 2t_0 - t_{a1} - t_0 = t_0 - t_{a1}; \\
\varphi_{b3} &= \varphi_3 - t_0 = 3t_0 - t_{a1} - t_{a2} - t_0 = 2t_0 - t_{a1} - t_{a2}; \\
&\dots \\
\varphi_{bi} &= (i-1)t_0 - \sum_{k=1}^{i-1} t_{ak}; \quad (\text{for } i \geq 2). \tag{7}
\end{aligned}$$

Adopting analogous considerations illustrated for the buffer and applying the *Central Limit Theorem* (see Appendix A.2 for details), waiting time can be rewritten as:

$$\begin{aligned}
\varphi_{b_i} &\cong (i-1)(t_0 - \bar{t}_a) \pm \bar{t}_a \cdot t_{\text{value}} \left\{ \frac{i-2}{\gamma} \sqrt{i-1} \right. \\
&= (i-1) \frac{T}{N_p} \left( 1 - \frac{N_p}{N_A} \right) \pm \frac{T}{N_A} \cdot t_{\text{value}} \left\{ \frac{i-2}{\gamma} \sqrt{i-1} \right. \tag{8}
\end{aligned}$$

This discrete values function could be approximated by a time-continuous function recalling that exit time for the  $i$ -th patient  $t_{out_i}$  is equal to  $i \cdot T/N_p$ . Confounding waiting time of the  $i$ -th patient with waiting time of the preceding  $(i-1)$ -th patient, then, in (8),  $i$  could be considered in place of  $(i-1)$  and for the term  $t_{\text{value}} \left\{ \frac{i-2}{\gamma} \sqrt{i-1} \right.$ , which is the same reported in Fig. 3, the same interpolation function of  $i$  can be adopted. Finally, the time-continuous function  $\varphi_b(t)$ , representing waiting time for the patient exiting from the system at time  $t = i \cdot T/N_p$ , can be written as:

$$\varphi_b(t) \cong \left( 1 - \frac{N_p}{N_A} \right) t \pm \frac{T}{N_A} \cdot a \left( \frac{t}{T} N_p \right)^b. \tag{9}$$

with the reported couple of coefficients  $(a,b)$  for the considered values of  $\gamma$  (90% and 95%) and lower limits  $t \geq 5 T/N_p$  and  $t \geq 6 T/N_p$  respectively.

Considering now the ratio  $N_b(t)/\varphi_b(t)$  and taking the same sign for the term due to uncertainty for numerator and denominator one at a time, it's possible to write:

$$\frac{N_b(t)}{\varphi_b(t)} \cong \frac{\bar{N}_A}{T} \frac{(\bar{N}_A - N_p) \frac{t}{T} \pm \sqrt{\bar{N}_A} \cdot a \left( \frac{t}{T} \right)^b}{(\bar{N}_A - N_p) \frac{t}{T} \pm (N_p)^b \cdot a \left( \frac{t}{T} \right)^b}. \tag{10}$$

The multiplication factor of the term  $\bar{N}_A/T$ , with the above reported values of the coefficients  $(a,b)$  for  $\gamma=90\%$ , is reported in Fig. 4 as a function of  $t$  (assumed  $T=1$  [time unit]) and for various combinations of  $N_p$  and  $\Delta = \bar{N}_A - N_p$ . It's possible to observe that in the deterministic case (deterministic arrivals)  $N_b(t)/\varphi_b(t) = \bar{N}_A/T = \lambda$ , whereas uncertainty, represented by the reported factor, tends to unity for larger values of  $t$ . It assumes relatively low scattered values around 1, except for initial periods, for increasing values of  $N_p$  and/or  $\Delta$ . From the knowledge of the number of patients in the queue at a generic time, taking into account the previous

assumptions, then it could be possible to esteem grossly current waiting time of an exiting patient as  $N_b(t)/\bar{N}_A \cdot T$ .

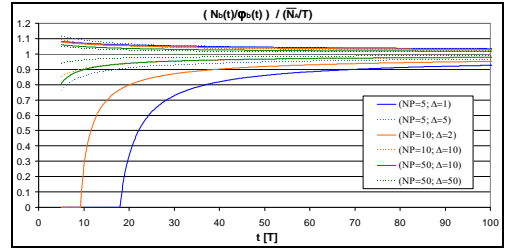


Figure 4: Values of the multiplication factor of  $\bar{N}_A/T$  in (10) as a function of  $t$  for  $\gamma=90\%$ .

#### 4.2. Average values

The time-average  $\bar{N}_{bi}$  of the first  $i$  terms  $N_{bk}$  ( $k=1,2,\dots,i$ ) can be calculated assuming, for the sake of simplicity, that each value  $N_{bk}$  is maintained over the time interval of the  $k$ -th period. From the definition of average (see Appendix A.3 for details), it's possible to write:

$$\bar{N}_{bi} \cong \frac{1}{2} (i+1) (\bar{N}_A - N_p) \pm \sqrt{\bar{N}_A} \cdot \frac{1}{i} \sum_{k=2}^i t_{\text{value}} \left\{ \frac{k-1}{\gamma} \sqrt{k} \right. \tag{11}$$

In Fig. 5, the right hand term in (11) depending solely on  $i$  and  $t$ -value and due to uncertainty is reported, together with power interpolation functions (dashed lines) in the form  $c \cdot i^d$ , in the interval  $0-500$ , for the two confidence levels 90% and 95%. In particular, setting lower limits for  $i$  so that the percentage difference between the function and its interpolation is lower than 15% also for initial values of  $i$  (then, it rapidly decreases to around 2% in the considered interval), it's possible to find  $(c=1.3993; d=0.4585)$  for  $\gamma=90\%$ , with  $i \geq 8$  and  $(c=1.802; d=0.4456)$  for  $\gamma=95\%$ , with  $i \geq 14$ .

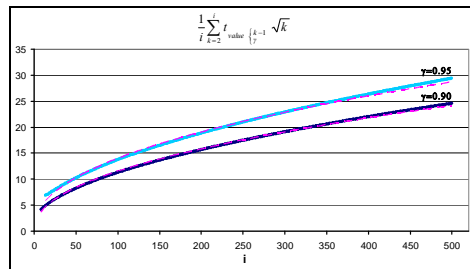


Figure 5: Values of the term depending on  $i$  and  $t$ -value in (11) versus  $i$  (for  $\gamma=90\%$  and  $\gamma=95\%$ ) and respective interpolating functions (dashed).

Finally, considering for  $\bar{N}_b(t)$  a time-continuous function, confounding the term  $(i+1)$  with  $i$ , which is equal to  $t/T$ , and taking into account the interpolation functions, it's possible to write:

$$\bar{N}_b(t) \cong \frac{1}{2} (\bar{N}_A - N_p) \frac{t}{T} \pm \sqrt{\bar{N}_A} \cdot c \left( \frac{t}{T} \right)^d; \tag{12}$$

with the shown values of  $(c, d)$  for  $\gamma=90\%$  and  $\gamma=95\%$  and with  $t \geq 8T$  and  $t \geq 14T$  respectively.

As regards the average value  $\overline{\varphi_{bi}}$  of the first  $i$  terms  $\varphi_{bk}$  ( $k=1, 2, \dots, i$ ), it's possible to write (see Appendix A.4 for details):

$$\overline{\varphi_{bi}} \cong \frac{1}{2}(i-1) \frac{T}{N_p} \left(1 - \frac{N_p}{N_A}\right) \pm \frac{T}{N_A} \cdot \frac{1}{i} \sum_{k=3}^i t_{\text{value}} \left\{ \begin{matrix} k-2 \\ \gamma \end{matrix} \right\} \sqrt{k-1}; \quad (\text{for } i \geq 3).$$

Adopting the same previous approximation consisting of considering  $i$  in place of  $(i-1)$ , for the factor depending on  $t$ -values and  $i$ , the same previously found interpolation function can be used<sup>2</sup>. Moreover, considering  $\overline{\varphi_b}$  as a time-continuous function:

$$\overline{\varphi_b(t)} \cong \frac{1}{2} \left(1 - \frac{N_p}{N_A}\right) \frac{T}{N_p} t \pm \frac{T}{N_A} \cdot c \left(\frac{t}{T} N_p\right)^d; \quad (13)$$

with the lower limits  $t \geq 8 T/N_p$  and  $t \geq 14 T/N_p$  for  $\gamma=90\%$  and  $\gamma=95\%$ , respectively.

Considering the ratio of the average values:

$$\frac{\overline{N_b(t)}}{\overline{\varphi_b(t)}} \cong \frac{\overline{N_A}}{T} \frac{\frac{1}{2}(\overline{N_A} - N_p) \frac{t}{T} \pm \sqrt{\overline{N_A}} \cdot c \left(\frac{t}{T}\right)^d}{\frac{1}{2}(\overline{N_A} - N_p) \frac{t}{T} \pm (N_p)^d \cdot c \left(\frac{t}{T}\right)^d} \quad (14)$$

The multiplication factor of the term  $\overline{N_A}/T$ , with the above reported values of coefficients  $(c, d)$  for  $\gamma=90\%$ , is reported in Fig. 6 for various combinations of  $N_p$  and  $\Delta = \overline{N_A} - N_p$  as a function of  $t$ . As observable, also the ratio of average values tends to  $\overline{N_A}/T = \lambda$ , for increasing  $t$  and the same considerations for the ratio of instantaneous values apply. Of course, this result derives from the linearity with  $t$  of the part of the functions  $N_b(t)$  and  $\varphi_b(t)$  not related to uncertainty, which makes also their average values, apart from the second term, proportional to  $1/2 t$ .

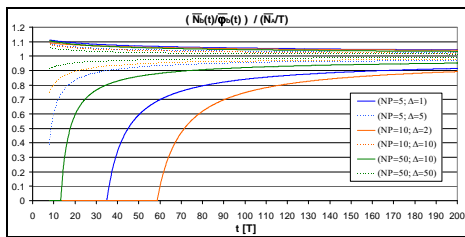


Figure 6: Values of the multiplication factor of  $\overline{N_A}/T$  in (14) as a function of  $t$  for  $\gamma=90\%$ .

## 5. CASE WITH DECREASING BUFFER

In this section, the opposite unsteady-state situation with  $\mu > \lambda$  (that's to say  $N_p > \overline{N_A}$ ) is discussed, starting

<sup>2</sup> In detail,  $\frac{1}{i} \sum_{k=3}^i t_{\text{value}} \left\{ \begin{matrix} k-2 \\ \gamma \end{matrix} \right\} \sqrt{k-1} = \frac{1}{i} \sum_{j=2}^{i-1} t_{\text{value}} \left\{ \begin{matrix} j-1 \\ \gamma \end{matrix} \right\} \sqrt{j}$

and  $(i-1)$ , upper limit of the summation term, is substituted by  $i$ .

with a buffer containing a backlog of  $N_{b0}$  patients. Of course, once the backlog has been cancelled, the system will be in the condition illustrated in Section 3, with fluctuating instantaneous values, but constant limit values (for  $t \rightarrow \infty$ ) for averages. Therefore the following analysis is limited to the period during which the buffer level remains above zero.

### 5.1. Instantaneous values

With the same meaning of the above defined quantities:

$$\begin{aligned} N_{b1} &= \max \{0, N_{b0} - (N_p - N_{A1})\}; \\ N_{b2} &= \max \{0, N_{b1} - (N_p - N_{A2})\}; \\ &\dots \\ N_{bi} &= \max \{0, N_{b(i-1)} - (N_p - N_{Ai})\}; \end{aligned}$$

under the assumption that the server is always busy. Moreover, as far as the second term in the brackets isn't negative, the recursive expression  $N_{bi}$  can be written as:

$$N_{bi} = N_{b0} - iN_p + \sum_{k=1}^i N_{Ak} \quad \left( \text{for } N_{bi} \geq 0 \Rightarrow i \leq \left( N_{b0} + \sum_{k=1}^i N_{Ak} \right) / N_p \right).$$

Adopting the same above considerations:

$$N_{bi} \cong N_{b0} - i \left( N_p - \overline{N_A} \right) \pm \sigma_{POP} \cdot t_{\text{value}} \left\{ \begin{matrix} i-1 \\ \gamma \end{matrix} \right\} \sqrt{i}; \quad \left( \text{for } 2 \leq i \leq \left( N_{b0} + \sum_{k=1}^i N_{Ak} \right) / N_p \right).$$

The function  $t_{\text{value}} \left\{ \begin{matrix} i-1 \\ \gamma \end{matrix} \right\} \sqrt{i}$  could be interpolated by the same power function  $a \cdot i^b$ , already shown, for given values of  $\gamma$ . Finally, stated that  $t = i \cdot T$  and for Poisson arrivals  $\sigma_{POP} = \sqrt{\overline{N_A}}$ , considering  $N_b$  as a time-continuous function:

$$N_b(t) \cong N_{b0} - \left( N_p - \overline{N_A} \right) \frac{t}{T} \pm \sqrt{\overline{N_A}} \cdot a \left( \frac{t}{T} \right)^b. \quad (15)$$

Since it's relevant to find, in an immediate way, values of  $i$  (or of  $t$ ) for which  $N_b=0$  and observing that the coefficient  $b$  is around 0.5, the slightly different interpolation function  $a \cdot i^{0.5}$  can be adopted (with neglectable differences in the same interval)<sup>3</sup>. In particular,  $a'=0.675$  for  $\gamma=50\%$ , with  $i \geq 4$ ,  $a'=1.65$  for  $\gamma=90\%$ , with  $i \geq 8$ , and  $a'=1.97$  for  $\gamma=95\%$ , with  $i \geq 7$  (percentage difference between the function and its interpolation lower than 15%). Therefore, it's possible to find, under the illustrated assumptions, the roots of the equation representing the time interval extremes (for a given confidence level) within which the buffer is expected to become empty:

$$t \cong T \cdot \left[ \frac{N_{b0}}{N_p - \overline{N_A}} + \frac{1}{2} \frac{a'^2 \overline{N_A}}{(N_p - \overline{N_A})^2} \left( 1 \pm \sqrt{1 + \frac{4N_{b0}(N_p - \overline{N_A})}{a'^2 \overline{N_A}}} \right) \right]. \quad (16)$$

In Fig. 7, two simple examples of  $N_b(t)$  are reported, both starting from  $N_{b0}=200$  and with  $\overline{N_A}=38$ , for two

<sup>3</sup> This means assuming a constant value of  $t_{\text{value}} \left\{ \begin{matrix} i-1 \\ \gamma \end{matrix} \right\}$  for a given  $\gamma$ .



possible values of  $N_p$  (40 and 45). These last parameters are realistic values, implemented in the successive simulation model.

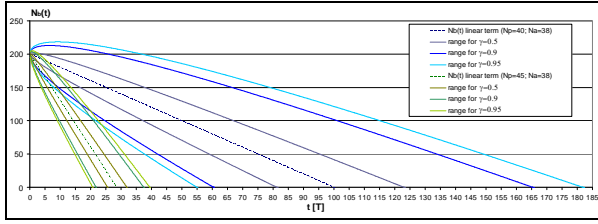


Figure 7:  $N_b(t)$  and Time Limits from (16) for  $N_b(t)=0$  for different  $\gamma$  ( $N_{b0}=200$ ,  $\bar{N}_A=38$ ,  $N_p=40$  and  $N_p=45$ )

It's possible to observe how, for each alternative, the limit values of the interval aren't symmetric around the zero for the linear term of (15) (dashed lines), but are shifted to the right. Moreover, comparing the two alternatives, the range of the uncertainty interval, with the same  $\gamma$ , is lower for greater values of  $N_p$ . Therefore, if the objective of management is the abatement of the waiting list, doing so in a shorter period of time could be preferable also for the reduced uncertainty.

As regards waiting time in the buffer, with the same assumptions of the case with an increasing buffer (the server is always busy and patients are served on a FCFS basis), exit time  $t_{out\ i}$  is deterministic and equal to  $(N_{b0}+i)t_0$ . Also in this case,  $t_{in\ i}$  can be written as the sum of the preceding  $(i-1)$  inter-arrival times  $t_{ak}$ , for  $i \geq 2$ , and therefore:

$$\varphi_{b1} = \varphi_1 - t_0 = t_{out1} - t_{in1} - t_0 = N_{b0}t_0;$$

$$\varphi_{b2} = N_{b0}t_0 + t_0 - t_{a1};$$

....

$$\varphi_{bi} = N_{b0}t_0 + (i-1)t_0 - \sum_{k=1}^{i-1} t_{ak}; \quad (\text{for } i \geq 2).$$

Recalling (20) in Appendix A.2:

$$\varphi_{bi} \cong N_{b0} \frac{T}{N_p} + (i-1) \frac{T}{N_p} \left(1 - \frac{N_p}{N_A}\right) \pm \frac{T}{N_A} \cdot t_{value} \Big|_{\gamma} \sqrt{i-1}; \quad (\text{for } i \geq 3).$$

Confounding waiting time of the  $i$ -th patient with waiting time of the  $(i-1)$ -th patient, then  $i$  could be used in place of  $(i-1)$  and for the term linked to uncertainty the same previous interpolating function can be adopted. Considering  $\varphi_b$  as a time-continuous function and deriving  $i$  as a function of exit time  $t$  as  $(t/t_0 - N_{b0})$ , it's possible to write:

$$\varphi_b(t) \cong N_{b0} \frac{T}{N_A} - \left(\frac{N_p}{N_A} - 1\right) t \pm \frac{T}{N_A} \cdot a \left(\frac{t}{T} N_p - N_{b0}\right)^b.$$

with the lower limits  $t \geq (N_{b0}+5)t_0$  and  $t \geq (N_{b0}+6)t_0$  for  $\gamma=90\%$  and  $\gamma=95\%$  respectively.

Finally, taking into consideration the ratio  $N_b(t)/\varphi_b(t)$ :

$$\frac{N_b(t)}{\varphi_b(t)} \cong \frac{\bar{N}_A}{T} \frac{N_{b0} - (N_p - \bar{N}_A) \frac{t}{T} \pm \sqrt{\bar{N}_A} \cdot a \left(\frac{t}{T}\right)^b}{N_{b0} - (N_p - \bar{N}_A) \frac{t}{T} \pm a \left(\frac{t}{T} N_p - N_{b0}\right)^b} \quad (17)$$

The multiplication factor of the term  $\bar{N}_A/T$ , with the above reported values of the coefficients  $(a,b)$  for  $\gamma=90\%$ , is reported in Fig. 8 as a function of  $t$  ( $T=1$  [time unit]), for various combinations of  $\bar{N}_A$ ,  $\Delta = N_p - \bar{N}_A$  and different values of  $N_{b0}$ , set in such a way that the liner term of (15) is zero at  $t=200$ .

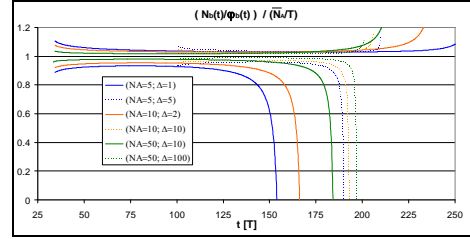


Figure 8: Values of the multiplication factor of  $\bar{N}_A/T$  in (17) as a function of  $t$  for  $\gamma=90\%$ .

It's possible to observe that in the deterministic case (deterministic arrivals)  $N_b(t)/\varphi_b(t) = \bar{N}_A/T = \lambda$ , whereas uncertainty, represented by the reported factor, is around 1 in the previous period approaching the empty buffer condition.

## 5.2. Average values

For the average value of the buffer, it's possible to write:

$$\bar{N}_{bi} \cong N_{b0} - \frac{1}{2}(i+1)(N_p - \bar{N}_A) \pm \sqrt{\bar{N}_A} \cdot \frac{1}{i} \sum_{k=2}^i t_{value} \Big|_{\gamma} \sqrt{k}; \quad (\text{for } i \geq 2).$$

With the same previous assumptions and interpolation function for the term depending on  $t$ -values and  $i$  and considering a time-continuous function:

$$\bar{N}_b(t) \cong N_{b0} - \frac{1}{2}(N_p - \bar{N}_A) \frac{t}{T} \pm \sqrt{\bar{N}_A} \cdot c \left(\frac{t}{T}\right)^d.$$

As regards the average value  $\bar{\varphi}_{bi}$  of the first  $i$  terms  $\varphi_{bk}$  ( $k=1, 2, \dots, i$ ), proceeding analogously as before:

$$\bar{\varphi}_{bi} \cong N_{b0} \frac{T}{N_p} - \frac{1}{2}(i-1) \frac{T}{N_p} \left(\frac{N_p}{N_A} - 1\right) \pm \frac{T}{N_A} \cdot \frac{1}{i} \sum_{k=3}^i t_{value} \Big|_{\gamma} \sqrt{k-1}; \quad (\text{for } i \geq 3).$$

With the same previous assumptions, considering a time-continuous function:

$$\bar{\varphi}_b(t) \cong \frac{1}{2} N_{b0} T \left(\frac{1}{N_p} + \frac{1}{N_A}\right) - \frac{1}{2} \left(\frac{N_p}{N_A} - 1\right) t \pm \frac{T}{N_A} \cdot c \left(\frac{t}{T} N_p - N_{b0}\right)^d.$$

with the lower limits  $t \geq (N_{b0}+8) \cdot T/N_p$  and  $t \geq (N_{b0}+14) \cdot T/N_p$ , for  $\gamma=90\%$  and  $\gamma=95\%$  respectively.



Considering the ratio of the two above quantities, we have:

$$\frac{\overline{N_b(t)}}{\overline{\varphi_b(t)}} = \frac{\overline{N_A}}{T} \frac{N_{b0} - \frac{1}{2}(N_p - \overline{N_A})\frac{t}{T} \pm \sqrt{\overline{N_A}} \cdot c \left(\frac{t}{T}\right)^d}{\frac{1}{2}N_{b0} \left(1 + \frac{\overline{N_A}}{N_p}\right) - \frac{1}{2}(N_p - \overline{N_A})\frac{t}{T} \pm c \left(\frac{t}{T}N_p - N_{b0}\right)} \quad (18)$$

Numerically, it's possible to verify that the multiplication factor of  $\overline{N_A}/T$  is always greater than one (as shown in Fig. 9 for  $\gamma=90\%$ ) and therefore an upper limit for the instantaneous average waiting time in the buffer  $\overline{\varphi_b(t)}$  can be assumed as  $(\overline{N_b(t)}/\overline{N_A}) \cdot T$ .

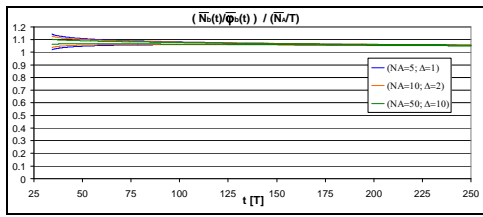


Figure 9: Values of the multiplication factor of  $\overline{N_A}/T$  (18) as a function of  $t$  for  $\gamma=90\%$ .

## 6. SIMULATION MODELS

In this section two simple DES models, coded with the process algebra language  $\chi$  (Hofkamp and Rooda 2007), are presented and applied to real cases of a CT scan facility inside a public hospital.

Outpatient access to the diagnostic service (one multi-slice CT scanner) is managed during two six hour work shifts from 8:00 a.m. to 8:00 p.m., from Monday to Friday. During this weekly period  $T$ , normally 8 appointments each day are scheduled. The first model reproduces exactly an M/D/1 system and a deterministic machine, with a fictitious service time  $t_0$ , is implemented in order to simulate a scheduling process with  $N_p$  served patients for each period  $T$  ( $t_0=T/N_p$ ). Therefore, in such a model, detailed information about the really adopted scheduling plan (i.e. information regarding appointment times) is not needed, with the assumptions already illustrated in Section 3, since a deterministic service time is equivalent to assuming the scheduled patients' access uniformly distributed during the time period. Two cases are examined, respectively with  $N_p=40$  and, adding one further slot per day,  $N_p'=45$  weekly scheduled patients, starting, in both situations, with an assumed initial buffer  $N_{b0}=200$  patients and an hypothesized value of  $\overline{N_A}=38$  weekly average demand.

Attention is focused on the decreasing buffer case because it meets more practical interest than the opposite one. The aim is verifying that the resulting range of time necessary to empty the buffer is within the values derived from (16). Successively, once the backlog is over and the system has reached steady-state conditions (since  $\overline{N_A} < N_p$ ), average waiting time is tracked in order to verify relation (1). It should be remarked that a huge backlog could have been caused,

during preceding unbalanced periods, by greater demand and/or less planned appointments. Moreover, a generic buffer level  $N_{b0}$  implies that the first appointment slot for a patient entering the system at that time will be available in  $N_{b0}/N_p \cdot T$  periods (value of  $\varphi_{b1}$ ). If this level were a constant, with  $N_{b0}=200$ , waiting time would always be 5 (with  $N_p=40$ ) or 4.4 (with  $N_p'=45$ ) weeks; with  $N_{b0}=100$  it would be reduced to 2.5 or 2.2 weeks, which can be considered generally acceptable. Therefore, also a partial drop down of  $N_{b0}$  from 200 to 100 patients could represent a managerial objective (in (16) the difference of the two values should be used in place of  $N_{b0}$  to esteem the needed time). Successively, under the hypothesis  $\overline{N_A}$  is not variable,  $N_p$  should vary in order to keep the buffer level fluctuating around the desired level; this could be advantageous for freeing resource time for other purposes, fixing periods with less appointments than normal (and, correspondingly, buffer increasing).

The second model, reported in Fig. 10, is a simplified adaptation of a more complete and detailed model of a CT scan facility open to diverse patient flows, described in (Boenzi et al. 2013), in which the scheduling process of outpatients is simulated by means of a controlled release mechanism from the buffer.

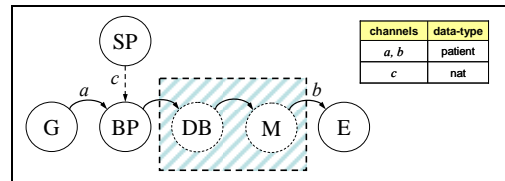


Figure 10: Complete Model Processes, truncated portion (hatch box) and resulting Simplified Model

In the simplified examined case, the illustrated general framework refers to one buffer only (for a specific class of outpatients). Moreover, for the aims of this paper, a truncated model can be considered. Patients are generated by process  $G$  (adopting a  $HPP$ ) and stored in  $BP$ . Their access to the service is regulated by means of a scheduling process  $SP$ , which repeats a daily cycle, according to the particular schedule in place.  $SP$  controls  $BP$  requesting the release of established numbers of patients at particular time instants to a daily working-list buffer  $DB$ . From process  $DB$  (in which, normally, priority rules are adopted to manage diverse patient flows) a patient can access service  $M$ , when it is not busy (representing a common shared resource in the complete model). Finally, patients leave the system in process  $E$ . However, since we are interested, in terms of simulation results, only in waiting time in  $BP$  (time interval from the examination request to the appointment day) and not in waiting time in  $DB$  (at the scan facility), the processes involved beyond  $BP$  are out of the scope of the model. For this reason  $BP$  can be linked directly to the exit process  $E$ , in which data are collected.

### 6.1. Simulation results for the first model

The unity of measurement of time is the week ( $T = 1 \text{ week}$ ) and, in the following figures, results for two groups of 10 simulation runs (with the mentioned parameter values and the two values of  $N_p$ , shown in different colors) are reported. In Fig. 11, the time-history of the instantaneous buffer level is reported, terminated when the buffer becomes empty.

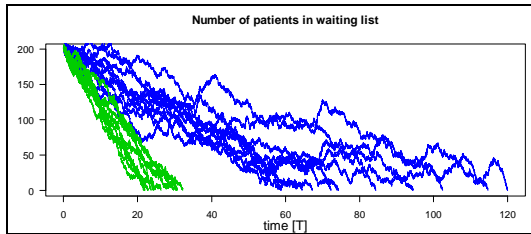


Figure 11: Buffer Level time-history for ten simulation runs for  $N_p=40$  (blue) and  $N_p=45$  (green)

In particular, Fig. 12 shows in detail how the resulting values of time are within the expected confidence interval limits calculated from (16) for the respective case, reported as box edges. Finally, in Fig. 13, the time-history of the average waiting time is reported for 10000 weeks. It can be observed that, after a transition period, the value stabilizes around the limiting values  $\cong 0.25 T$  and  $\cong 0.07 T$  for  $N_p=40$  and  $N_p=45$  respectively, in agreement with (1).

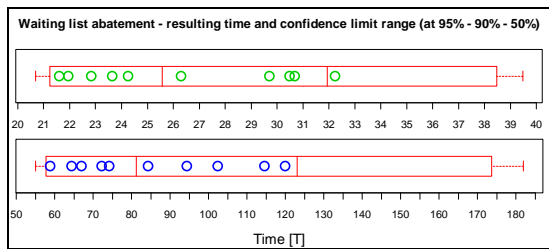


Figure 12: Time for which  $N_b=0$  in ten simulation runs, ( $N_p=40$ : blue,  $N_p=45$ : green) compared to interval limits derived from (16) for  $\gamma=50\%-90\%-95\%$

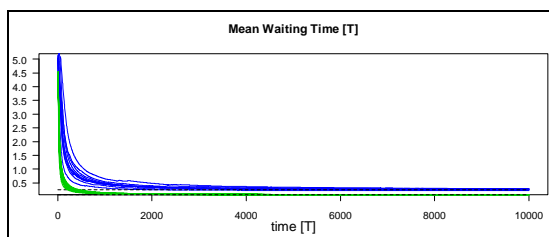


Figure 13: Time-history of the Mean Waiting Time in the Buffer for ten simulation runs for  $N_p=40$  (blue) and  $N_p=45$  (green)

### 6.2. Simulation results for the second model

In the second model, two slightly different appointment schedules have been specified in more detail and implemented. In general, at the facility, outpatients are admitted one at a time (each patient is assigned a specific slot). The time interval between two consecutive appointments is set at one hour in the

morning (because of the presence of a more intense internal service request in the hospital), starting at 8:00, and half a hour in the afternoon, starting at 15:00. Maintaining the same weekly number of appointments  $N_p$  adopted for the first model, the total number of daily patients in the first schedule is equally divided as four patients in the morning shift and four patients in the afternoon shift (last appointment at 16:30). In the second schedule, an additional appointment has been set at 17:00. It's to remark that the feasibility of this new plan, in particular with respect to the risk of over-time, could be assessed by a more detailed simulation model, but is beyond the scope of the present work. The unity of measurement of time is one hour, but results are presented in weeks ( $T = 1 \text{ week} = 60 \text{ hours}$  in the simulation). The following Figures from 14 to 16 are analogous to those referring to the first model and the same considerations could be made. A slight difference between the two models can be found with regard to average waiting time in the buffer. In fact, considering each group of ten simulations, time terminated after 10000 weeks, this value is  $\cong 0.283 T$  when  $N_p=40$  and  $\cong 0.087 T$  when  $N_p=45$ . This is due to the different distribution of patient appointments, which is not perfectly uniform in the course of time in the second model.

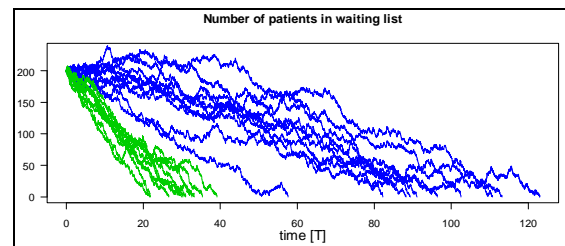


Figure 14: Buffer Level time-history for ten simulation runs for  $N_p=40$  (blue) and  $N_p=45$  (green)

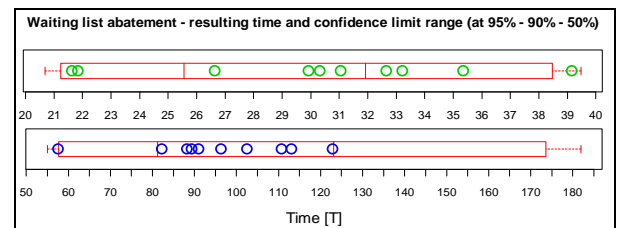


Figure 15: Time for which  $N_b=0$  in ten simulation runs, ( $N_p=40$ : blue,  $N_p=45$ : green) compared to interval limits derived from (16) for  $\gamma=50\%-90\%-95\%$

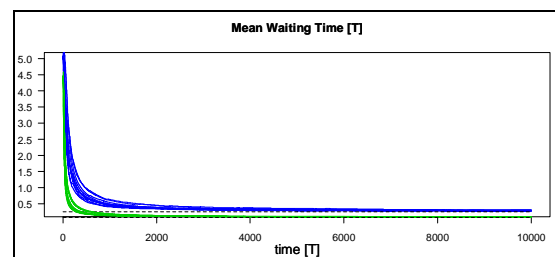


Figure 16: Time-history of the Mean Waiting Time in the Buffer for ten simulation runs for  $N_p=40$  (blue) and  $N_p=45$  (green)

Finally, in the model adopting the first schedule, a buffer control mechanism has been introduced, through the monitoring of the current waiting list length. The aim is maintaining the buffer at a constant level, which can be beneficial both to the facility staff (in order to manage possible cancellations with other appointments) and to the customers, who can plan their examinations well in advance (sometimes this is unavoidable because of additional examinations and/or preparation). In Fig. 17, the time-history (terminated at 210 weeks) of the instantaneous number of patients in the buffer and of current waiting time (for the patient currently exiting from the system) for three simulation runs (in different colors) are reported, starting with  $N_{bo}=200$  patients.

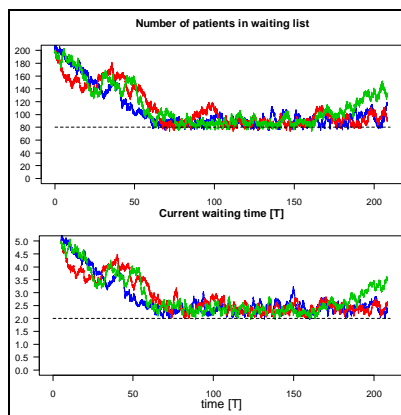


Figure 17: Time-history of Current Buffer Level (above) and Current Waiting Time (below) for three simulation runs for  $N_p=40$  adopting Buffer Control

The lower limit for the buffer has been set at 80 patients because this implies a lower waiting time limit equal to 2 weeks (assuming  $N_p=40$  patients/week), which can be considered a good compromise. In the simulation model, the buffer level control is done at the end of each day and, if the level is below the set limit, no appointments are scheduled for the subsequent day. Of course this is a simplification, being unfeasible in a real appointment system. In the last, a day without appointments could be set within the already booked period. However a problem is that as this time is set far in the future, as the lag-time between the signal and the corrective action increases. Anyway, this day could be better utilized for other patient flows (for example, additional appointments for a different booking agenda could be placed). Solutions different from control reveal to be ineffective. For example, implementing a cyclic alternate schedule, made of 19 days with 8 appointments per day and one day without appointments, would realize, for the assumed value  $\overline{N_A}=38$  patients/week, the exact balance between demand and offer in a four week period. However, since this cycle is completely unrelated to variability of the demand, the system state variables are highly fluctuating. Therefore, some type of control loop in these systems should be enforced. Once adopted, demand variability could be turned into an opportunity, with the appropriate flexibility in the work organization.

## 7. CONCLUSION

In the present work, some common features of an appointment system in general have been highlighted, by means of an analytical approach for non-steady state conditions; successively, applications to a real case of a CT scan facility have been presented. It can be observed that, at a macroscopic level, the main factors characterizing the system are on the one side the appointment demand and, on the other side, the offer of planned appointments.

At this level, details regarding the acceptability of a scheduling plan in terms of overtime and/or excessive waiting time experimented at the facility can be put apart and analyzed successively. The only source of variability is represented by the appointment demand, since setting the number of appointments in a period of time  $T$  means specifying the system “productive capability”. Therefore, it can be treated as a G/D/1 queue. In particular, in the M/D/1 case, from Kingman’s equation it’s possible to observe that waiting time in the buffer can’t exceed one period  $T$ , as far as the difference between the set number of appointments and the expected number of requests in  $T$  remains above  $0.5$ . Therefore, in such systems, the existence of long waiting time for appointments can only be explained by current demand greater than offer (increasing buffer and waiting time) or by former relevant backlog, created in preceding unbalanced periods. An analytical approach has been proposed in order to study the two situations represented by increasing buffer and decreasing buffer with initial backlog. In particular, in the latter case, statistical considerations permit to define confidence intervals for estimation of time necessary to clear the backlog, which could be useful in managerial decisions. It should be noticed that also an intermediate solution, aiming at maintaining a set number of patients in the waiting list, could be an objective. However, it would require a very strict control of the buffer and of the demand. Finally, two simulation models, adopting parameter values of a CT scan facility, have been illustrated. Results obtained from the two models are similar and show, on the one hand, agreement with the predicted statistical intervals; on the other hand, results show that the M/D/1 model can well describe the behavior of the buffer as with a more refined model. Future work will be focused on the detailed investigation, by means of simulation, of the appointment schedule in place at the CT facility inside a public hospital, differentiated for each day of the week, in order to verify, in particular, the impact of different alternatives with increased numbers of appointments on the probability of overtime.

## APPENDIX A

**A.1** The summation term of (4) can be thought of as the sample average  $\overline{N_{A_i}}$  of the first  $i$  arrivals multiplied by  $i$ . Applying the *Central Limit Theorem*, this sample average can be approximated by the population average  $\overline{N_A}$ , with a specified confidence level  $\gamma$ , expressed by the amplitude of a variability range, which is the

product of the critical  $t$ -value of the  $t$ -probability distribution, for  $\gamma$  and  $(i-1)$  degrees of freedom, and  $\sigma_{POP}/\sqrt{i}$ , being  $\sigma_{POP}$  the population standard deviation:

$$\sum_{k=1}^i N_{Ak} = i \overline{N_{Ai}} \cong i \left( \overline{N_A} \pm t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{i}} \right\} \right) = (19)$$

$$= i \overline{N_A} \pm \sigma_{POP} \cdot t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{i}} \right\}; \quad (\text{for } i \geq 2).$$

In particular, when arrivals belong to a Poisson distribution,  $\sigma_{POP} = \sqrt{\overline{N_A}}$ .

**A.2** For the summation term of (7), applying the *Central Limit Theorem*, the sample average of the first  $(i-1)$  terms can be approximated by the population average  $\overline{t_a}$ :

$$\sum_{k=1}^{i-1} t_{ak} \cong (i-1) \left( \overline{t_a} \pm t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{i-1}} \right\} \right) = (20)$$

$$= (i-1) \overline{t_a} \pm \sigma_{POP} \cdot t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{i-1}} \right\}; \quad (\text{for } i \geq 3).$$

In particular, when the number of requests follows a Poisson distribution, inter-arrival times belong to an exponential distribution with  $\overline{t_a} = 1/\lambda = T/\overline{N_A}$  and  $\sigma_{POP} = \overline{t_a} = 1/\lambda = T/\overline{N_A}$ .

**A.3** From (4) and (19):

$$\overline{N_{bi}} = \frac{1}{i} (N_{b1} + N_{b2} + \dots + N_{bi}) = \frac{1}{i} \sum_{k=1}^i N_{bk} = \frac{1}{i} \sum_{k=1}^i \left( \sum_{l=1}^k N_{Al} - k N_p \right) =$$

$$= \frac{1}{i} \left( \sum_{l=1}^i \sum_{k=l}^i N_{Al} - N_p \sum_{k=1}^i k \right) = \frac{1}{i} \left[ \sum_{l=1}^i \sum_{k=l}^i N_{Al} - \frac{i(i+1)}{2} N_p \right]$$

$$\sum_{l=1}^i N_{Al} = k \overline{N_{Al}} \cong k \overline{N_A} \pm \sigma_{POP} \cdot t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k}} \right\} \quad (\text{with } k \geq 2);$$

$$\sum_{k=1}^i \sum_{l=1}^k N_{Al} = N_{A1} + \sum_{k=2}^i \sum_{l=1}^k N_{Al} \cong N_{A1} + \sum_{k=2}^i \left( k \overline{N_A} \pm \sigma_{POP} \cdot t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k}} \right\} \sqrt{k} \right) + \overline{N_A} - \overline{N_A} =$$

$$= N_{A1} - \overline{N_A} + \sum_{k=2}^i k \overline{N_A} \pm \sum_{k=2}^i \sigma_{POP} \cdot t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k}} \right\} \sqrt{k} = N_{A1} - \overline{N_A} + \frac{i}{2} (i+1) \overline{N_A} \pm \sigma_{POP} \cdot \sum_{k=2}^i t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k}} \right\} \sqrt{k};$$

$$\overline{N_{bi}} \cong \frac{1}{2} (i+1) (\overline{N_A} - N_p) + \frac{1}{i} (N_{A1} - \overline{N_A}) \pm \frac{1}{i} \sigma_{POP} \cdot \sum_{k=2}^i t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k}} \right\} \sqrt{k}; \quad (\text{for } i \geq 2).$$

The term  $(N_{A1} - \overline{N_A})/i$  can be neglected and, assuming a Poisson arrival process,  $\sigma_{POP} = \sqrt{\overline{N_A}}$ .

**A.4** From (7) and (20):

$$\overline{\varphi_{bi}} = \frac{1}{i} (\varphi_{b1} + \varphi_{b2} + \dots + \varphi_{bi}) = \frac{1}{i} \sum_{k=1}^i \varphi_{bk} = \frac{1}{i} \sum_{k=1}^i \left( (k-1) \varphi_0 - \sum_{l=1}^{k-1} t_{al} \right) =$$

$$= \frac{1}{i} \left( t_0 \sum_{k=2}^i (k-1) - \sum_{k=2}^i \sum_{l=1}^{k-1} t_{al} \right) = \frac{1}{i} \left[ \frac{i-1}{2} t_0 - \sum_{k=2}^i \sum_{l=1}^{k-1} t_{al} \right]$$

$$\sum_{l=1}^{k-1} t_{al} \cong (k-1) \overline{t_a} \pm \sigma_{POP} \cdot t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k-1}} \right\} \sqrt{k-1} \quad (\text{with } k \geq 3);$$

$$\sum_{k=2}^i \sum_{l=1}^{k-1} t_{al} = t_{a1} + \sum_{k=2}^i \sum_{l=1}^{k-1} t_{al} \cong t_{a1} + \sum_{k=3}^i \left( (k-1) \overline{t_a} \pm \sigma_{POP} \cdot t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k-1}} \right\} \sqrt{k-1} \right) + \overline{t_a} - \overline{t_a} =$$

$$= t_{a1} - \overline{t_a} + \sum_{k=2}^i (k-1) \overline{t_a} \pm \sum_{k=3}^i \sigma_{POP} \cdot t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k-1}} \right\} \sqrt{k-1} = t_{a1} - \overline{t_a} + i \frac{i-1}{2} \overline{t_a} \pm \sigma_{POP} \cdot \sum_{k=3}^i t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k-1}} \right\} \sqrt{k-1};$$

$$\overline{\varphi_{bi}} \cong \frac{1}{2} (i-1) (t_0 - \overline{t_a}) - \frac{1}{i} (t_{a1} - \overline{t_a}) \pm \frac{1}{i} \sigma_{POP} \cdot \sum_{k=3}^i t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k-1}} \right\} \sqrt{k-1} \quad (\text{with } i \geq 3);$$

$$\overline{\varphi_{bi}} \cong \frac{1}{2} (i-1) (t_0 - \overline{t_a}) - \frac{1}{i} (t_{a1} - \overline{t_a}) \pm \frac{1}{i} \sigma_{POP} \cdot \sum_{k=3}^i t_{\text{value}} \left\{ \frac{\sigma_{POP}}{\sqrt{k-1}} \right\} \sqrt{k-1}; \quad (\text{for } i \geq 3).$$

As done similarly for the buffer expression, the term  $(t_{a1} - \overline{t_a})/i$  can be neglected and, assuming a Poisson arrival process,  $\sigma_{POP} = T/\overline{N_A}$ .

## REFERENCES

- Adan, I., Resing, J., 2002. *Queueing Theory*. Eindhoven, The Netherlands: Department of Mathematics and Computing Science – Eindhoven University of Technology.
- Boenzi, F., Mummolo, G., Rooda, J.E., 2013. Appointment scheduling for a computed tomography facility for different patient classes using simulation. *Proceedings of the 25<sup>th</sup> European Modeling and Simulation Symposium*, pp. 370-378, September 25-27, Athens, Greece
- Green, L.V., Savin, S.V., Wang, B., 2006. Managing patient service in a diagnostic medical facility. *Oper Res* 54(1): 11–25.
- Hans, E.W., Van Houdenhoven, M., Hulshof, P.J.H., 2011. *A Framework for Health Care Planning and Control*. Memorandum 1938. ISSN 1874-4850. Department of applied mathematics, University of Twente, Enschede, The Netherlands. Available from: <http://www.math.utwente.nl/publications> [Accessed 15 July 2014].
- Hassin, R., Mendel, S., 2008. Scheduling Arrivals to Queues: A Single-Server Model with No-Shows. *Management Science* 54(3): 565–572.
- Hofkamp, A.T., Rooda, J.E., 2007. *Chi 1.0 Reference Manual*. Eindhoven University of Technology. Available from: [http://seweb.se.wtb.tue.nl/sewiki/\\_media/chi/chi10\\_refman1689.pdf](http://seweb.se.wtb.tue.nl/sewiki/_media/chi/chi10_refman1689.pdf) [Accessed 15 July 2014].
- Hopp, W.J., 2008. Chapter 4: Single Server Queueing Models. In: Chhajed, D., Lowe, T.J., eds. *Building Intuition - Insights from Basic Operations Management Models and Principles*. Springer International Series in Operations Research & Management Science Volume 115, pp. 51-79.
- Huttschenreuter, A., 2004. *Waiting Patiently - An analysis of the performance aspects of outpatient scheduling in health care institutes*. BMI – Paper, Vrije Universiteit Amsterdam, The Netherlands. Available from: [http://www.few.vu.nl/en/Images/werkstuk-huttschenreuter\\_tcm39-91363.pdf](http://www.few.vu.nl/en/Images/werkstuk-huttschenreuter_tcm39-91363.pdf) [Accessed 15 July 2014].
- Kaandorop, G., Koole, G., 2007. Optimal outpatient appointment scheduling. *Health Care Manag Sci* 10: 217–229.
- Klassen, K.J., Rohleder, T.R., 1996. Scheduling outpatient appointments in a dynamic environment. *Journal of Operations Management* 14: 83–101.
- Kolisch, R., Sickinger, S., 2008. Providing radiology health care services to stochastic demand of

- different customer classes. *OR Spectrum* 30(2): 375–395.
- Kuiper, A., 2012. *Appointment scheduling in healthcare*. Master Thesis. IBIS UvA – Faculty of Economics and Business, Faculty of Science – University of Amsterdam.
- Patrick, J., 2012. A Markov decision model for determining optimal outpatient scheduling. *Health Care Manag Sci* 15: 91–102.
- Rohleder, T., Klassen, K.J., 2002. Rolling Horizon Appointment Scheduling: A Simulation Study. *Health Care Manag Sci* 5: 201–209.
- Sickinger, S., Kolisch, R., 2009. The performance of a generalized Bailey–Welch rule for outpatient appointment scheduling under inpatient and emergency demand. *Health Care Manag Sci* 12: 408–419.
- Wijewickrama, A.K.A., 2006. Simulation analysis for reducing queues in mixed-patients' outpatient department. *Int. j. simul. model.* 05 2:56-68.
- Zhu, Z. C., Heng, B. H., Teow, K. L., 2009. Simulation study of the optimal appointment number for outpatient clinics. *Int. j. simul. model.* 08 3:156-165.

# SIMULATION AND EXPERIMENTAL ANALYSIS OF LED WEATHERPROOF LUMINAIRE THERMAL PERFORMANCE

Carlos Javierre<sup>(a)</sup>, Daniel Elduque<sup>(b)</sup>, Víctor Camañes<sup>(c)</sup>, David Franch<sup>(d)</sup>

<sup>(a,b,c)</sup>Department of Mechanical Engineering, University of Zaragoza, C/ María de Luna, 3, 50018 Zaragoza, Spain  
<sup>(d)</sup>Zalux, Centro Empresarial Miralbueno, 50012 Zaragoza, Spain

<sup>(a)</sup>[carlos.javierre@unizar.es](mailto:carlos.javierre@unizar.es), <sup>(b)</sup>[daniel.elduque@gmail.com](mailto:daniel.elduque@gmail.com), <sup>(c)</sup>[victor.camanes@gmail.com](mailto:victor.camanes@gmail.com), <sup>(d)</sup>[dfranch@zalux.com](mailto:dfranch@zalux.com)

## ABSTRACT

In this paper the experimental and simulation analysis of a LED weatherproof luminaire is carried out. A CFD model that properly correlates with the real measurements has been created. After this, it is used as a tool to assess different design structures and new thermal solutions, showing that some LED luminaires designs create new thermal problems due to the coupling between the LED's and driver's heat. This problem must be solved or at least decreased in order to obtain higher LED lifetimes and efficiencies. This problem of thermal management is forcing engineers to come up with new solutions for weatherproof luminaires. The use of CFD simulation helps designers to properly understand the thermal behaviour and improve new designs reducing significantly prototyping costs.

Keywords: LED weatherproof luminaire, simulation, measurement, design

## 1. INTRODUCTION

Reducing the environmental impact of new products is key factor as environmental awareness increases in society. New techniques such as Life Cycle Assessment have been developed to assess and improve the environmental performance of products and services by using new materials (Sarasa, et al. 2009) and reducing energy and raw materials consumption (Martinez, et al. 2009) (Martinez, et al. 2010) (Grignon-Massé, et al. 2011) (Fernández, et al. 2013) (Elduque, et al. 2014) (Jiménez, et al. 2014).

Energy efficiency has become a vital factor in most electrical and electronic products (Welz, et al. 2011). For several decades fluorescent lighting has been widely used in most industrial applications. The search for higher efficiency levels is a major concern for the lighting industry. In the last decade, significant improvements have been made, making LED (Light Emitting Diode) lighting a reality.

This technology offers higher output efficiency and lowers the environmental impact of the use phase but due to its characteristics it also creates new problems that must be addressed. LED lifetime and efficiency

output is highly dependent on working temperatures, meaning that keeping the LED modules cool is a priority when developing a new luminaire.

Simulation approaches are currently used in multiples fields (Latorre, et al. 2009) (Martinez, et al. 2010) (Latorre-Biel, et al. 2012) (Latorre-Biel, et al. 2013), including LED lighting. Simulation software is also currently used to improve design results (Aisa, et al. 2006) (Javierre, et al. 2006) (Ruiz Argáiz, et al. 2008) (Jiménez, et al. 2009) (Javierre, et al. 2013) (Javierre, et al. 2014). Several authors have used simulation software to analyze the thermal behaviour of LED and improve its efficiency. Some studies have analyzed the cooling of one LED lamp, improving its heat dissipation by means of optimized heat sinks (Costa and lopes 2014) The thermal modelling of individual LED and its boundary condition have also been assessed to improve led manufacturing and its performance

Other authors have studied LED arrays and how its placement affects the thermal management (Li, et al. 2011) (Yung, et al. 2014). Also the use of new LED substrates materials has been analyzed (Yin, et al. 2010). Finally, other researchers have focused on complete luminaires

In order to solve this problem, the Spanish leading manufacturer of weatherproof luminaires and industrial battens, Zalux, has worked with the Department of Mechanical Engineering of the University of Zaragoza to develop new design solutions to improve the thermal behaviour of its luminaires.

To carry out this project both simulation and experimental measurements have been performed with the aim of developing new LED luminaires made out of different materials such as aluminium, thermoplastics or thermally conductive thermoplastics.

## 2. EXPERIMENTAL SETUP AND THERMAL MODELLING

### 2.1. Experimental Setup

LED temperatures are a key factor when developing new weatherproof luminaires. These light fittings enclose into a plastic housing all the lighting system: an ECG (Electronic Control Gear) and the LED modules,



obstructing heat dissipation and thus reducing the efficiency. This makes necessary to find new designs that guarantee appropriate heat dissipation, improving the efficiency of the luminaire. This factor is especially critical for weatherproof luminaires, as the heat transfer via convection is lower due to the enclosed housing (Figure 1).

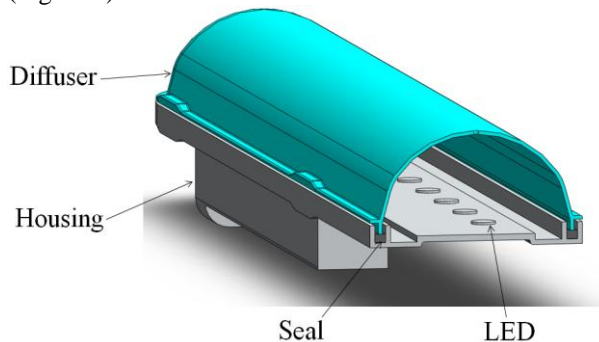


Figure 1: Weatherproof luminaire structure

A commercial LED weatherproof (Figure 2) sold by Zalux has been chosen to set the benchmark and allow a correlation between experimental data and simulation results. This luminaire has a new weatherproof design base on aluminium housing to improve LED cooling as it works as a heat sink, reducing LED temperatures. This is a fully functional design but expensive as the housing is produced by injecting aluminium into a mould.



Figure 2: Zalux Nextrema luminaire

The internal component distribution of the commercial luminaire is shown in Figure 3. The ECG is placed near end of the luminaire, where the mains cable is connected. The LED MCPCBs are placed at the centre of the luminaire.



Figure 3: Components used in Zalux Nextrema luminaire

First of all, the luminaire will be measured using a thermocouple logger and a thermo-graphic camera.

Thermal images allow us to understand how the housing cooling effect is working and which the temperatures in the external walls are (Figure 4).

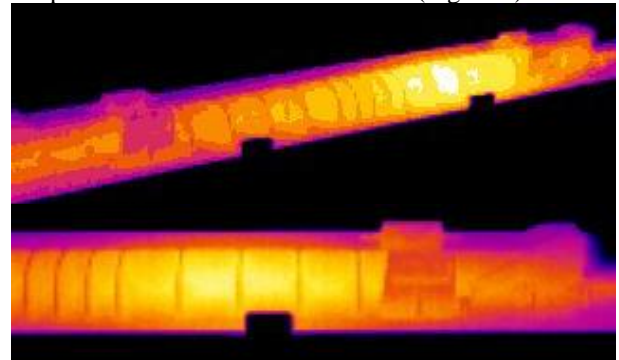


Figure 4: Thermal image

These images allow us to see that the electronic driver is an important source of concentrated heat whereas the LEDs also generate heat but it is more distributed along the length of the luminaire. This means that near the driver there is a heat coupling effect that increases the temperature in this area.

Thermocouples are used to determine the temperature of internal components, like the LEDs (Figure 5) or the electronic driver.

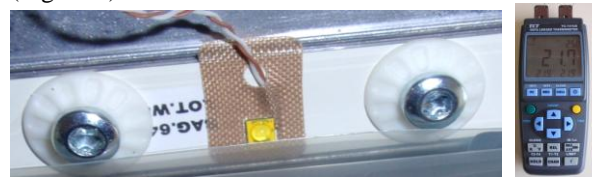


Figure 5: Thermocouple used to measure LED temperature

## 2.2. Simulation model

After obtaining the real working temperatures values, a CFD model was developed with Solid Works Flow Simulation, obtaining a simulation model that can be used to improve new designs.

Solidworks is a well-known simulation software. Its developer has validated its ability to calculate essential features of flows and to solve conjugate heat transfer problems (flow problems with heat transfer in solids). This software is currently used to solve a wide range of fluid flow and heat transfer phenomena. It numerically solves the Navier-Stokes equations (mass, momentum and energy conservation for fluid flows). These equations are supplemented by definitions of thermophysical properties and state equation for the fluids. Also radiation heat transfer is considered (SolidWorks, 2012).

To solve all the equations, a numerical solution technique is employed by Flow Simulation. It uses a cell-centered finite volume method with a rectangular computational mesh in a Cartesian coordinate system. Values of the variables are calculated at the mesh cell centres and equations are discretized with first (time) or second order (space) accuracy. Mesh cells intersecting with a solid/fluid interface are split into smaller cells. To accurately solve this interface, additional boundary

faces are used to consider boundary conditions and geometry. If needed, the mesh is refined during the calculation process to solve high-gradient regions.

Material properties (Table 1) were obtained from suppliers' data and introduced into the model. The upper wall is defined as an insulator and polycarbonate is used for the light diffuser.

Table 1: Material conductivity

| Part      | Material               | Conductivity (W/mK) |
|-----------|------------------------|---------------------|
| Wall      | Isolator (WC scenario) | 0                   |
| LED MCPCB | Aluminium              | 136.7               |
| Diffuser  | Polycarbonate          | 0.22                |
| Tray      | Steel                  | 52                  |

Thermal resistances between LED and MCPCB (Metal Core Printed Circuit Board) were also taken into account. The Solid/Fluid Interface has been improved using small solid fluid refinement, curvature and tolerance refinement and optimization of the resolution for thin walls. An example of the fluid mesh is shown in the next figure.

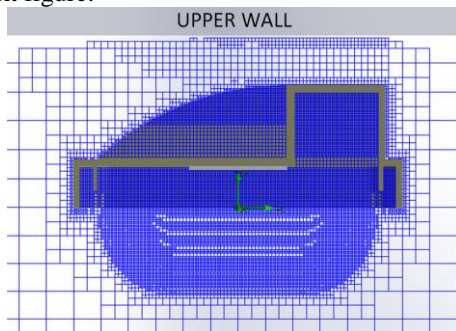


Figure 6: Example of fluid mesh definition

Also cell refinement was used, especially for solid cells so as to achieve a better modelling of the geometry; also several local meshes were used to improve the LED temperature distribution. The CFD mesh with solid and partial cells is shown below (Figures 7 and 8).

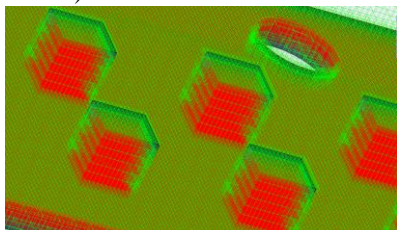


Figure 7: Example of LED mesh definition



Figure 8: Example of LED MCPCB mesh definition

Models between 4 and 8 million cells were used to obtain appropriate results with reduced computing times with a desktop PC (i7 quad-core, 32GB RAM memory).

Heat losses for LEDs and the electronic gear were obtained from the manufacturer.

After creating the CFD model, an external analysis to achieve a steady state was performed, showing that experimental and simulation results were very similar (Figure 9)

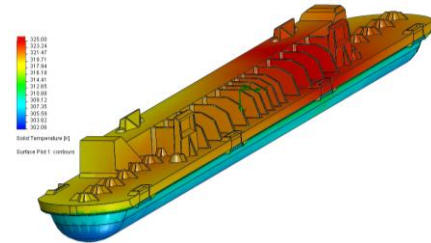


Figure 9: Temperature results of Zalux Nextrema

CFD simulations also allow us to understand how the heat sink is working and the effect of the airflow in the cooling (Figure 10).

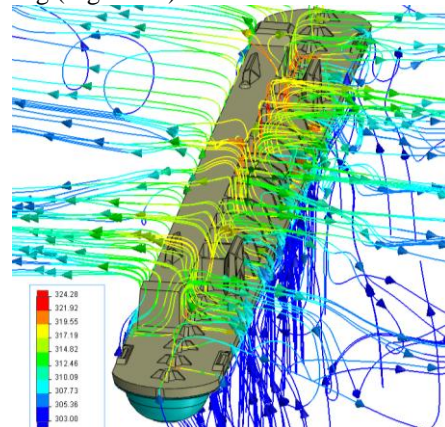


Figure 10: Airflow results of Zalux Nextrema

It also allows us to identify heat coupling and potential temperature hotspots.

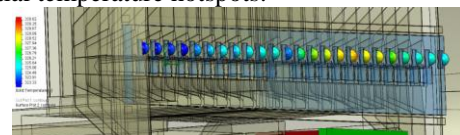


Figure 11: LED temperatures of Zalux Nextrema

### 3. DESIGN IMPROVEMENTS

These results and the simulation model are used to develop new designs in order to improve the performance and reduce temperatures. These modelling approaches were also tested with other smaller weatherproof luminaires used for emergency lighting, obtaining adequate results (Figure 12).

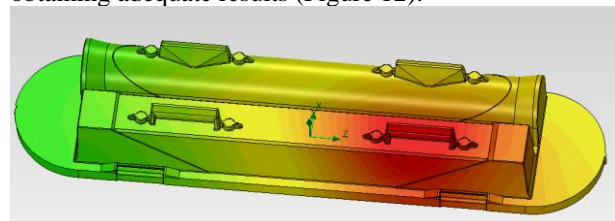


Figure 12: Temperature results of Zalux Emergency Nextrema

Also the simulation and experimental analysis are applied to propose new design configurations, like the one shown below (Figure 13), using an aluminium heatsink for the LED surrounded by a polymer housing to ensure the weatherproofness. The aluminium heatsink can be extruded, also meaning lower production costs.

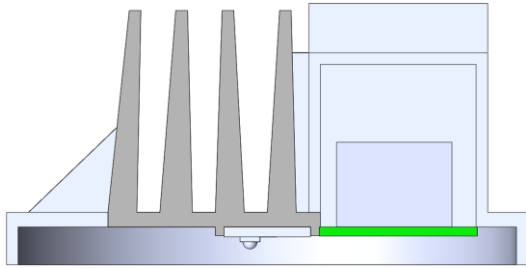


Figure 13: New design using an extruded heatsink

Instead of creating a prototype, which is expensive for plastic parts, a CFD model was carried out to assess if this solution achieves and adequate thermal behaviour (Figure 14) (Table 2).

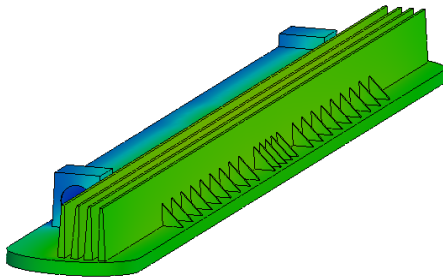


Figure 14: Temperature results of the redesigned luminaire

Table 2: LED Temperature Deltas

| Design                              | LED temperature Delta (K) |
|-------------------------------------|---------------------------|
| Nextrema Aluminium housing          | 27.8                      |
| Extruded heatsink + Polymer housing | 30.3                      |

In order to determine whether or not the use of a thermally conductive polymer housing may have a significant cooling in a standard weatherproof luminaire, three different materials were tested: production PC, and two PolyOne Thermo-Tech thermally conductive polymers (Table 3) (Figure 15)

Table 3: Housing material conductivity

| Material          | In-plane conductivity (W/mK) | Through-plane conductivity (W/mK) | Maximum temperature Delta (K) |
|-------------------|------------------------------|-----------------------------------|-------------------------------|
| Standard PC       | 0.22                         | 0.22                              | 39.2                          |
| Thermo-Tech White | 0.9-1.3                      | 0.9-1.3                           | 35.8                          |
| Thermo-Tech Grey  | 19-21                        | 4.5-5.5                           | 33.3                          |

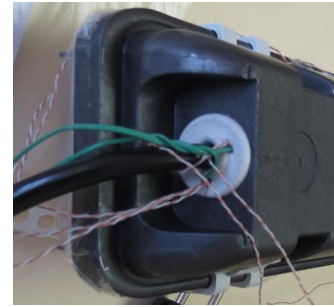


Figure 15: Temperature measurements of Thermo-Tech Grey thermally conductive polymer housing

The temperature reached by the light source shows a noteworthy reduction (Table 3), showing that the use of thermally conductive polymers is an interesting approach to improve the efficiency of the luminaire by means of reducing its operating temperature, but as this design was based on a standard luminaire structure, temperatures are higher than in designs optimized for LEDs.

Another adaptation of traditional weatherproof design solutions to use LED were also tested in this project, finding that LED technology (Light sources and the Electronic Control Gear) produce several higher problems due to heat concentration, creating hotspots that would decrease LED performance (Figure 16)

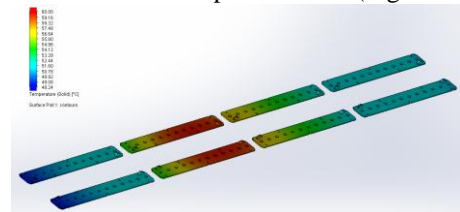


Figure 16: Temperature results showing thermal coupling

To check if the CFD simulation was producing accurate results, a prototype of this luminaire was assembled, allowing us to compare temperatures between the CFD model and the prototype (Table 4).

Table 4: Comparison between CFD and prototype

| Temperatures (°C) | Prototype values | CFD Simulation | Difference |
|-------------------|------------------|----------------|------------|
| MCPCB1            | 52.8             | 51.7           | -1.1       |
| MCPCB2            | 59.8             | 60.7           | +0.9       |
| Driver            | 65.5             | 62.6           | -2.9       |
| Tray              | 60.0             | 62.1           | +2.1       |

Differences lower than 3°C were found, meaning that the CFD was producing good results for the working conditions

The use of thermally conductive polymers was also tested for LED-specific design solution, using an anisotropic material model in the CFD simulation. Thanks to the way fins are filled in the injection process, they have a conductivity of around 20 W/mK in the proper direction to help even more with heat dissipation. A reduction of 3.8 K was obtained for the



LEDs maximum temperature when comparing with isotropic modelling.

In order to optimize the design's thermal behaviour via convection, also several simulations were performed to obtain the optimal pitch between cooling fins (Figure 17). More fins create a bigger cooling and irradiative area, but if the gap between fins is too small, the air does not flow properly due to natural convection, increasing the temperature.

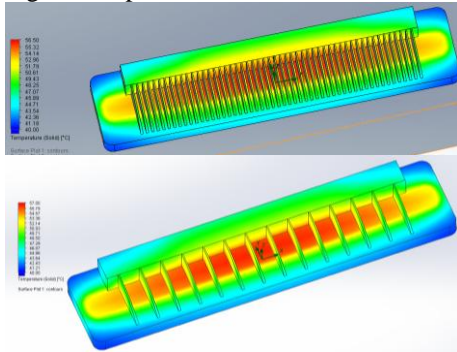


Figure 17: Temperature results for 5 mm and 30 mm fin distance

The CFD results allow us to better understand how the air flow works between the fins, as shown in Figure 18 below.

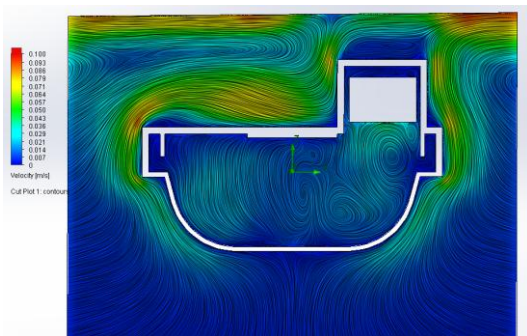


Figure 18: Air flow between fins

The optimum pitch between fins was 10mm as shown in Figure 19.

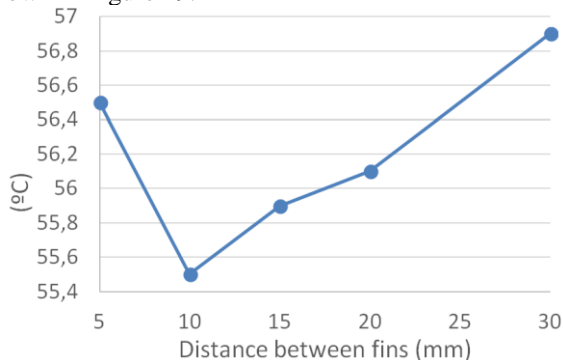


Figure 19: Maximum LED temperature depending on fin pitch

After studying the air flow, a new design was proposed, creating two banks of fins and placing the driver at the centre of the luminaire. This creates a symmetrical flow, increasing the cooling created by the

fins and reducing average temperature (Figure 20). Although, due to the thermal coupling caused by the LED and the driver the maximum temperature of the LED is increased, meaning that the lifetime and efficiency would be lower.

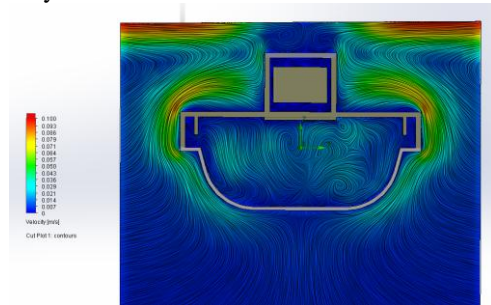


Figure 20: Temperature results showing thermal coupling

Also, in order to find other ways of improving thermal performance, the effect of the surface emissivity was studied. In natural convection flows, radiation has an important effect. Even though temperatures of a luminaire housing are not high, any way of improving the performance is useful. To do that, the effect of the material surface was studied. A comparison between a raw housing (without any surface treatment), an enamel paint (emissivity around 0.9) and black matt paint (emissivity around 0.98).



Figure 21: Black matt paint (Left) and Enamel paint (Right) inside the oven

To determine if there was any significant difference, two housing were heated at the same time in an oven (Figure 21), and the cooling time to achieve different temperatures was measured. Optimum results were obtained with the black matt pain after applying 5 layers of paint.

#### 4. CONCLUSION

In this paper the process measuring a luminaires to create a CFD model that properly correlates with the real measurements is shown. After this model is obtained, it is used as a tool to assess different design and thermal solutions, showing that LED luminaire create new thermal problems that must be solved in order to obtain higher LED lifetime and efficiency. The problem of thermal coupling between the LED and the

ECG is forcing engineers to come up with new solution for weatherproof luminaires.

The use of CFD simulation helps designers to properly understand the thermal behaviour and helps in reducing prototyping costs.

## ACKNOWLEDGMENTS

The authors would like to thank the help provided by Zalux's development team.

## 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), 15-19.
- Costa, V., & Lopes, A. (2014). Improved radial heat sink for led lamp cooling. *Applied Thermal Engineering*, 70, 131-138.
- Elduque, D. et al. 2014. Life cycle assessment of a domestic induction hob: electronic boards. *Journal of Cleaner Production*, 76, 74-84
- Fernández, A., Javierre, C., González, J. & Elduque, D., 2013. Development of thermoplastic material food packaging considering technical, economic and environmental criteria. *Journal of Biobased Materials and Bioenergy*, 7(2), 176-183.
- Grignon-Massé, L., Rivière, P. & Adnot, J., 2011. Strategies for reducing the environmental impacts of room air conditioners in Europe. *Energy Policy*, 39(4), 2152-2164.
- Javierre, C., Abad-Blasco, J., Camañes & V., F. D., 2013. Redesign of metallic parts with thermoplastic materials: Application example | Rediseño de componentes metálicos En materiales termoplásticos: Ejemplo de aplicación. *DYNA*, 88(2), 197-205.
- Javierre, C., Fernández, A., Aísa, J. & Clavería, I., 2006. Criteria on feeding system design: Conventional and sequential injection moulding. *Journal of Materials Processing Technology*, 171(3), 373-384.
- Javierre, C., Fernández, A., Clavería, I. & Elduque, D., 2014. Textile injection process characterisation by means of a spiral mould. *Measurement: Journal of the International Measurement Confederation*, 47(1), 100-109.
- Jiménez, E., Ruiz, I., Blanco, J., & Pérez, M. (2009). Design and simulation of production of injection pieces in automobile industry. *International Journal of Simulation: Systems, Science and Technology*, 10(3), 23-30.
- Jiménez, E. et al. 2014. Methodological approach towards sustainability by integration of environmental impact in production system models through life cycle analysis: Application to the Rioja wine sector. *SIMULATION*, 90(2), pp. 143-161.
- Latorre, J. I., Jiménez, E., Pérez, M., & Martínez, E. (2009). The design of a manufacturing facility. An efficient approach based on Alternatives Aggregation Petri nets. *Paper presented at the 21st European Modeling and Simulation Symposium, EMSS 2009*.
- Latorre-Biel, J. I., Jiménez-Macías, E., Blanco-Fernández, J., & Sáenz-Díez, J. C. (2012). Decision making in the Rioja wine production sector. *Paper presented at the 24th European Modeling and Simulation Symposium, EMSS 2012*.
- Latorre-Biel, J. I., Jiménez-Macías, E., Blanco-Fernández, J., & Sáenz-Díez, J. C. (2013b). Optimal design, based on simulation, of an olive oil mill. *Paper presented at the 25th European Modeling and Simulation Symposium, EMSS 2013*.
- Li, J., Yang, Q., Niu, P., Jin, L., Meng, B., Li, Y., Zhang, X. (2011). Analysis of Thermal Field on Integrated LED Light Source Based on COMSOL Multi-physics Finite Element Simulation. *Physics Procedia*, 22, 150-156.
- Martínez, E. et al. 2009. Life cycle assessment of a multi-megawatt wind turbine. *Renewable Energy*, 34(3), 667-673.
- Martínez, E., Jiménez, E., Blanco, J., & Sanz, F. (2010). LCA sensitivity analysis of a multi-megawatt wind turbine. *Applied Energy*, 87(7), 2293-2303.
- Martínez, E., Jiménez, E., Sanz, F., Pérez, M., Blanco, J., & Santamaría, J. (2010). Virtual representation of terrain through the web with VRML-Web3D and graphic libraries. *International Journal on Interactive Design and Manufacturing*, 4(2), 125-136
- Ruiz Argáiz, I., Jiménez Macías, E., Blanco Fernández, J., & Pérez de la Parte, M. (2008). Design and simulation of production of injection pieces in automobile industry. Paper presented at the *Proceedings - EMS 2008, European Modelling Symposium, 2nd UKSim European Symposium on Computer Modelling and Simulation*
- Sarasa, J., Gracia, J. & Javierre, C., 2009. Study of the biodesintegration of a bioplastic material waste. *Bioresource Technology*, 100(15), 3764-3768.
- SolidWorks, 2012. *SolidWorks Flow Simulation Technical Reference*, SolidWorks.
- Wang, J. (2014). Thermal module design and analysis of a 230 W LED illumination lamp under three incline angles. *Microelectronics Journal*, 45, 416-423.
- Welz, T., Hischer, R. & Hilty, L. M., 2011. Environmental impacts of lighting technologies — Life cycle assessment and sensitivity analysis. *Environmental Impact Assessment Review*, 31(3), 334-343.
- Yin, L., Yang, L., Yang, W., Guo, Y., Ma, K., Li, S., & Zhang, J. (2010). Thermal design and analysis of multi-chip LED module with ceramic substrate. *Solid-State Electronics*, 54, 1520-1524.
- Yung, K., Liem, H., Choy, H., & Cai, Z. (2014). Thermal investigation of a high brightness LED array package. *Applied Thermal Engineering*, 63, 105-118.

# FUPOL SIMULATORS AND ADVANCED VISUALIZATION FRAMEWORK INTEGRATION

Egils Ginters<sup>(a)</sup>, Artis Aizstrauts<sup>(b)</sup>, Mikelis Baltruks<sup>(c)</sup>, Kawa Nazemi<sup>(d)</sup>, Dirk Burkhardt<sup>(d)</sup>, Peter Sonntagbauer<sup>(e)</sup>, Susanne Sonntagbauer<sup>(e)</sup>, Jorge Martin Gutierrez<sup>(f)</sup>

<sup>(a,b,c)</sup>Sociotechnical Systems Engineering Institute of Vidzeme University of Applied Sciences, Cesu street 4, Valmiera LV-4201, Latvia

<sup>(d)</sup>Information Visualization and Visual Analytics, Fraunhofer-Institut für Graphische Datenverarbeitung, Fraunhoferstr. 5, 64283 Darmstadt, Germany

<sup>(e)</sup> Cellent AG, Ziegelofengasse 33, 1050 Vienna, Austria

<sup>(f)</sup> Dpto Expresion Grafica en Arquitectura e Ingenieria, Universidad de La Laguna, La Laguna, Spain

<sup>(a)</sup>[egils.ginters@va.lv](mailto:egils.ginters@va.lv), <sup>(b)</sup>[artis.aizstrauts@va.lv](mailto:artis.aizstrauts@va.lv), <sup>(c)</sup>[mikelis.baltruks@va.lv](mailto:mikelis.baltruks@va.lv), <sup>(d)</sup>[kawa.nazemi@igd.fraunhofer.de](mailto:kawa.nazemi@igd.fraunhofer.de), [dirk.burkhardt@igd.fraunhofer.de](mailto:dirk.burkhardt@igd.fraunhofer.de), <sup>(e)</sup>[peter.sonntagbauer@cellent.at](mailto:peter.sonntagbauer@cellent.at), [susanne.sonntagbauer@cellent.at](mailto:susanne.sonntagbauer@cellent.at), <sup>(f)</sup>[jmargu@ull.es](mailto:jmargu@ull.es)

## ABSTRACT

FP7 FUPOL project No.287119 (see [www.fupol.eu](http://www.fupol.eu)) aims at a new approach to traditional politics modeling. The FUPOL will be able to automatically collect, analyze and interpret opinions expressed on a large scale from the Internet and social networks. This will enable governments to gain a better understanding of the needs of citizens. Likewise the software will have the capabilities to simulate the effects of policies and laws and to assist governments in the whole policy design process. Basic visualization of the simulation results are supported by the simulators however visualization facilities are limited, therefore for detailed visual analysis of simulation data SemaVis environment is used.

Keywords: simulation, visualization, policy modeling, SemaVis

## 1. INTRODUCTION

The policy modeling process and lifecycle respectively is characterized by making decisions. The decision making process involves various stakeholders, that may have diverse roles in the policy making process. The heterogeneity of the stakeholders and their “way of work” is a main challenge for providing technologies for supporting the decision making as well as technologies to involve various stakeholder in the process.

Policy use case models depending on the domain could be described as discrete or continuous, and determined or stochastic systems. For simulation of the above-mentioned models different and heterogenic simulation tools could be used.

Simulation of scenarios can ask for collaboration of some separate simulation models joining for implementation of the task or policy domain use case. Therefore distributed and multi-level simulators can be designed (Aizstrauts et al. 2013). The simulation would

be ensured at two levels: micro and macro simulation levels. In the micro level the Agent-Based/ Multi-Agent Simulation (ABM/MAS) operations related to versatile and small basic components interaction and forecasting of the interaction results will be carried out. Although ABM/MAS could be used for forecasting of continuing changes the system dynamics (SD) use would be reasonable. At the macro level (if it is necessary) the SD simulation can be implemented. Moreover, the micro level would be the data source for the macro simulation model. If the scenario establish simultaneously use the set of models then distributed simulation can be realised using Easy Communication Environment (ECE) (Aizstrauts and Ginters et al. 2012).

## 2. ADVANCED VISUALIZATION

Advanced visualization techniques provide helpful instruments for the various stages of decision making and active participation of citizens in the policy creation process (Burkhardt et al. 2013b, Kohlhammer et al. 2012).

### 2.1. Visualization of Semantics

The representation of domain knowledge is a simplification and abstraction of the real world. Certain aspects and features of the real world that are deemed important or helpful to represent a domain are emphasized while numerous facets of that same domain have to be neglected. The distinction between relevant and irrelevant is represented in the domain ontology. Thus, for a computer system this domain ontology can define what entities exist and what they are called. Such representations form the semantics that guide information visualization techniques to visualize the important and hide the unimportant. The successful and efficient use of semantics for information visualization is still a major challenge in this current field of research (Keim et al. 2006, Kohlhammer 2012).



The SemaVis framework (Nazemi et al. 2014) is designed to visualize semantic information by offering effective navigation and interaction mechanisms. A special feature consists in the adequate visualization for different user groups with different preferences and background knowledge, both in terms of information to be displayed as well as in interacting with the visualizations. These users and user group orientation with customizable look and feel of principles had been developed and integrated into the SemaVis framework (Nazemi et al. 2010).

The SemaVis pipeline structures the process of semantic information processing towards an adapted visualization in three steps (see Figure 1): Semantics, Layout and Presentation. Semantics processing extracts information of the semantic data sources which are needed for the visualization of the data. The identified and extracted information are enriched in layer layout with graph and visualization layout information. At presentation level the visual presentation of data is determined (Nazemi, Stab and Kuijper 2011). At presentation level the visual variables of data is determined. This might be hue, saturation, color value, size, shape, etc. but also the 'visual behavior' of the elements, for example a highlighting of selected elements by a modified transparency value. The Presentation takes into account constraints and preferences from the user, unless the information is stored by a user model or community model and can be queried (Nazemi et al. 2011a), and even more the data characteristics (Nazemi et al. 2011b).

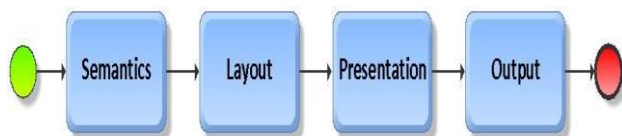


Figure 1: SemaVis Transformation Pipeline for Semantics Data Processing

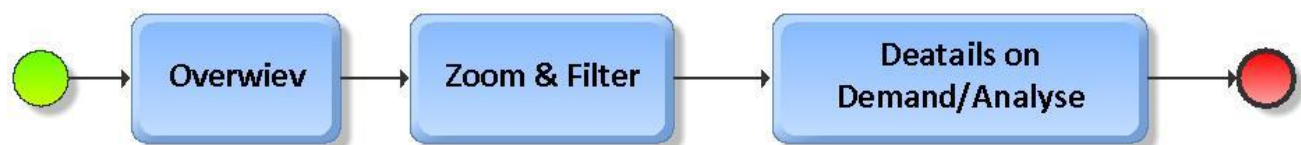


Figure 2: Overview-to-Detail Approach for the Visualization Interaction in Statistical Data to Provide an Intuitive Drill-Down Strategy in SemaVis to Find Relevant and Necessary Indicators

In a first inclusion of simulation data for visualization, we defined a similar API as many OGD portals do provide. A hierarchy for all available simulated data aspect was defined and is provided for the visualization. Even more, to each category all available simulation indicators are linked and can be chosen for the visualization. These simulation indicators are also similar to OGD bases. They hold the concrete

## 2.2. Visualization of Heterogeneous Data

Visualizing structured data is beneficial for the user, but often data-sources contain additional data too, e.g. statistics or multimedia data. These additional data provides also additional information that leads to an improved information gathering, because the user can analyse further information for her his data analysis purposes.

In particular for Open Government Data (OGD) it is essential to consider that an overview about the existing indicators and the statistical data can result in a significant better understanding about a given problem. For such kind of heterogeneous data, it is essential to consider next to a visualization pipeline that generates the visualizations, also an interaction strategy to provide users with intuitive navigation possibilities through data. Therefore, we enhanced the semantics visualization pipeline with Shneiderman's Information Seeking Mantra (Shneiderman 1996). In fact, the user starts his interaction with an overview about the hierarchy of indicators (see Figure 2). By navigating through the hierarchy and selecting his preferred indicators, he makes zooming and filtering on data level. The final step is the concrete (detailed) analysis of the statistical data.

## 2.3. Simulation Data Visualization

A specific challenge is the visualization of simulation data. On the one hand it is comparable to OGD visualization, because internally specific kinds of indicators are calculated as a future prediction. However, it is also different. OGD bases consist only about a time-based data-table. In simulation context also aggregated statistics are available in form of concepts. In fact, the visualization and the internal data-model have to consider a timely data provision (in an historical manner) or in abstracted form by category, which has a significant impact on the visualization, because some layout algorithms are not designed to illustrate data on categorical level.

data about the simulated aspect by time or category. If the user is selects a concrete indicator, the statistical data is loaded from the server and represented in the statistical visualizations.

A major advantage of the visualization is the form of an interactive dashboard (Nazemi et al. 2010 and Burkhardt et al. 2013a) that allows orchestration of the visualization in a personal preferred manner. Under this

concept, the user can select and arrange visualizations about the indicator hierarchy, the meta-information about a selected indicator, the indicator data from an OGD portal (if available) and the simulation indicators at the same time and in depend of the user's preferences.

To request the data the web-API of the Simulator and SemaVis are used (see Figure 3).

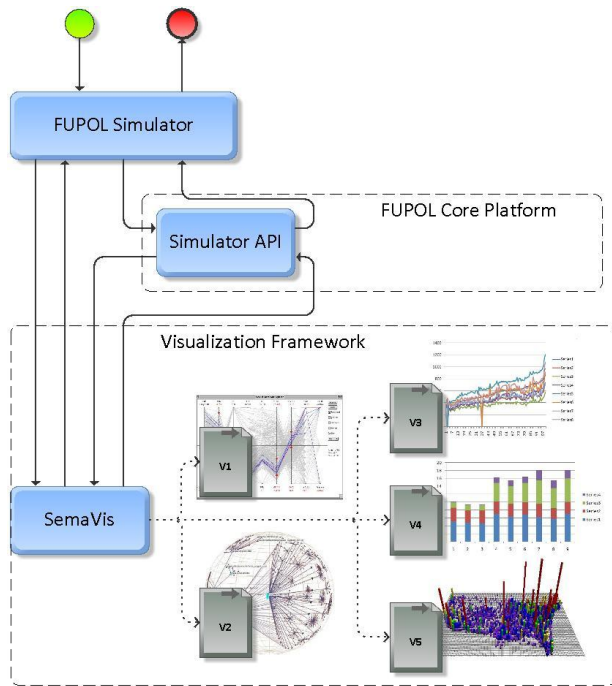


Figure 3: Visualization of Heterogeneous Data

The APIs of the simulator and SemaVis can be used directly or through the proxy at the FUPOL Core Platform.

Overall, the SemaVis API offers the control and access to internal information and states by external applications. Here the actionsript API is used which allows to trigger changes or to ask for state information at runtime. It distinguishes between internal events and actions propagated to the external system, internal state information propagated to external applications; external events and actions propagated to the SemaVis visualizations as well as external state information (which should be handled like an action within SemaVis) propagated to SemaVis visualizations.

External or third party systems can be connected to SemaVis and adapt the framework for application-specific purpose. SemaVis offers the control of and access to internal information and states by external applications.

SemaVis is a visualization framework executed on client side as Adobe Flex application within the browser. The communication between SemaVis and the Simulators is realized by the use of CSV files and

Simulator API providing two ways of integration to ensure reliability through the reservation of visualization channels (see Figure 4).

Figure 4 shows action and data flow to load and present simulation results with SemaVis framework. SemaVis is embedded as Adobe Flash application in Skopje Vodno Mountain Recreational Activities simulator which is web based application. After the simulation user is able to see the advanced visualization (SemaVis application) for particular simulation. BPMN2 flowchart (see Figure 4) has three level activities – SemaVis, FUPOL Core Platform (WP3) and Simulator. SemaVis level represents all activities related to SemaVis application, Simulator level shows all activities related to simulation data gathering. WP3 level works as data distribution, it manages all connections among all components. There are no direct connections between external component (e.g. SemaVis application) and internal system (e.g. simulator).

When the SemaVis application is loaded, simulation ID is passed to it, so that SemaVis can load data for this particular simulation. With given simulation ID SemaVis application makes HTTP request for data dictionary which explains data structure. This data dictionary is formatted as XML document (Figure 4 uses index.xml as a document name). Below see an example of such data dictionary document:

```
<xml>
<geolocation>Skopje</geolocation>
<category name="Occupancy">
category name="By day">
<indicator name="Monday"
url="/data/day1.csv?sim=1234" />
...
<indicator name="Sunday"
url="/data/day7.csv?sim=1234" />
</category >
</category >
</xml>
```

This request for data dictionary is handled by FUPOL Core Platform proxy who is used as performance booster (caching responses) and security by filtering who is allowed to use simulator. Core Platform requests simulator API for data dictionary by passing simulation ID to it. Simulator generates data dictionary (index.xml in Figure 4) for particular simulation. Data dictionary contains data structure and URLs for indicators. Data structure only describes what data is going to be visualized, what structure it has, but does not have any raw data within, the last entity in data structure is an indicator – link to raw data. Raw data are in CSV format.

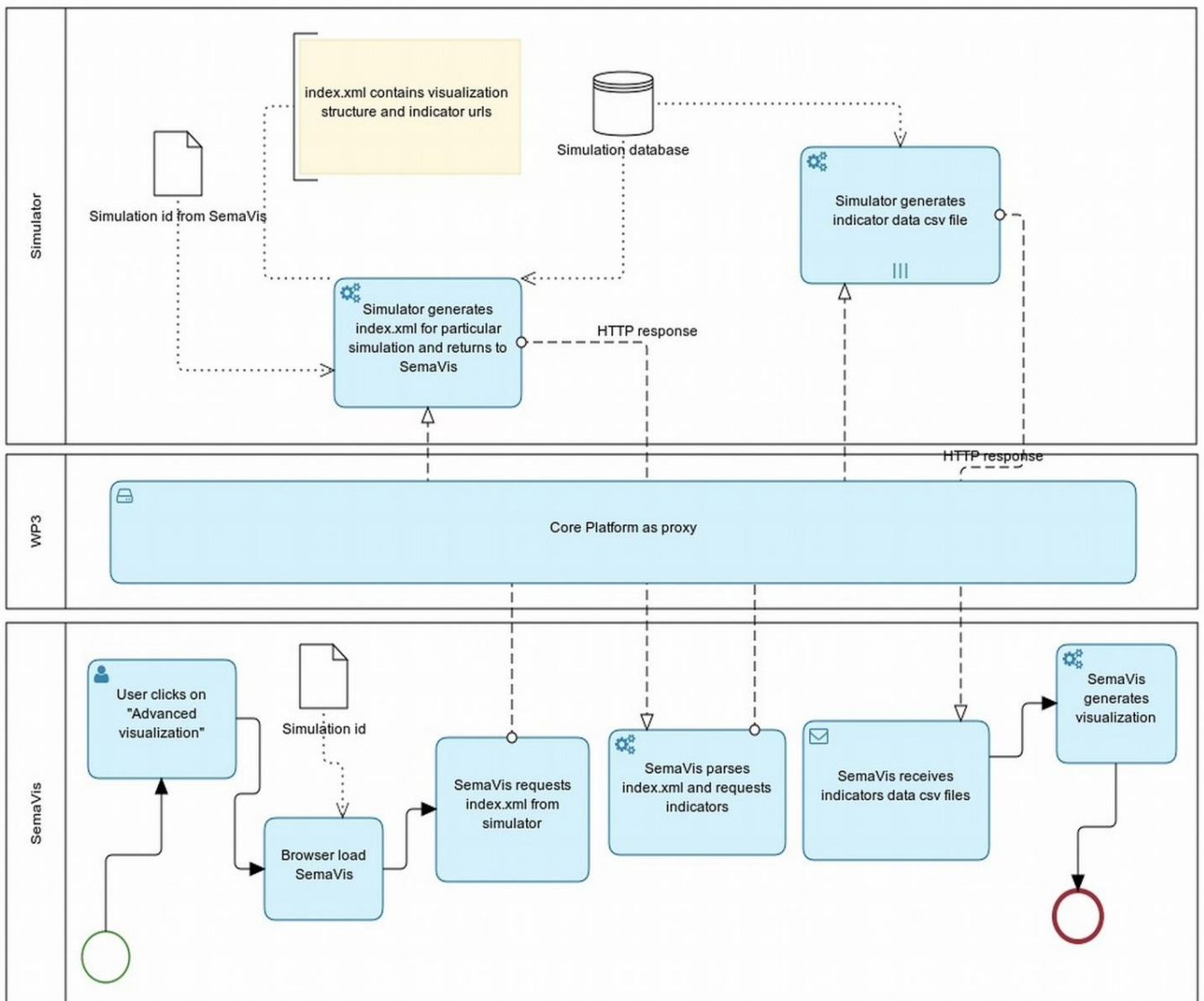


Figure 4: Simulators and SemaVis Interaction Diagram

When the simulator has built the data dictionary it returns dictionary to the SemaVis through FUPOL Core Platform and SemaVis receives data dictionary as HTTP response. SemaVis application parses data dictionary and builds structured data entities. For each indicator entity SemaVis makes a HTTP request for particular raw data. This request also goes through the FUPOL Core Platform and ends up to specific simulator API that generates raw data for each particular indicator. When the raw data are ready then the simulator returns them to the SemaVis application (through FUPOL Core Platform) and they are ready to be visualized.

### 3. VISUALIZATION OF SKOPJE VODNO MOUNTAIN RECREATIONAL ACTIVITIES SIMULATION RESULTS

The approach described above was verified designing Skopje Vodno Mountain Recreational Activities simulator (Ginters et al. 2014).

The main objective of the Vodno Mountain simulator design was elaboration the solution that offers the City of Skopje a possibility to realize a better schedule of resources and plan of activities on Vodno Mountain, which is located to the southwest of the capital city Skopje. The simulator offers to Skopje citizens the opportunity to forecast the occupancy of the recreational resources on Vodno Mountain and suggest new schedule, new ideas to the administration of City of Skopje. The system would help the administration of City of Skopje in improving the scheduling and resource planning, initiation and creating new projects involving the recreational area at Vodno Mountain. The

citizens of City of Skopje also are involved in the decision making process by constant communication and expressing their opinion to the authorities, making the whole process more transparent and efficient.

The simulation is realized by using a simulation agents that simulates a typical behaviour of a person from a given user group. The agent's behaviour depends on the type of the simulation, with a specific input and output parameters. There are two types of simulations called Simulation 1 and Simulation 2. The Simulation 1 spans through the period of one year. Only hourly average data obtained by simulation are recorded and available for visualization. Simulation runs with random selection of input parameters within a defined range, such as, simulation of a weather condition, or agent

behaviour. The input of this simulation is a complete simulation configuration. All parameters are simulated according to their behaviour and range of defined values. Citizens can initiate the Simulation 2. This simulation spans through a period of one week within a selected month. Users can change several parameters, such as desired weather, month selection, or number of persons in a certain user group. After simulation, the result data is available for visualization and users can see the effects of the changes.

The main kind of embedded visualization is map of Vodno Mountain recreational area allowing showing the routes and occupancy the resources using pie charts (see Figure 5).

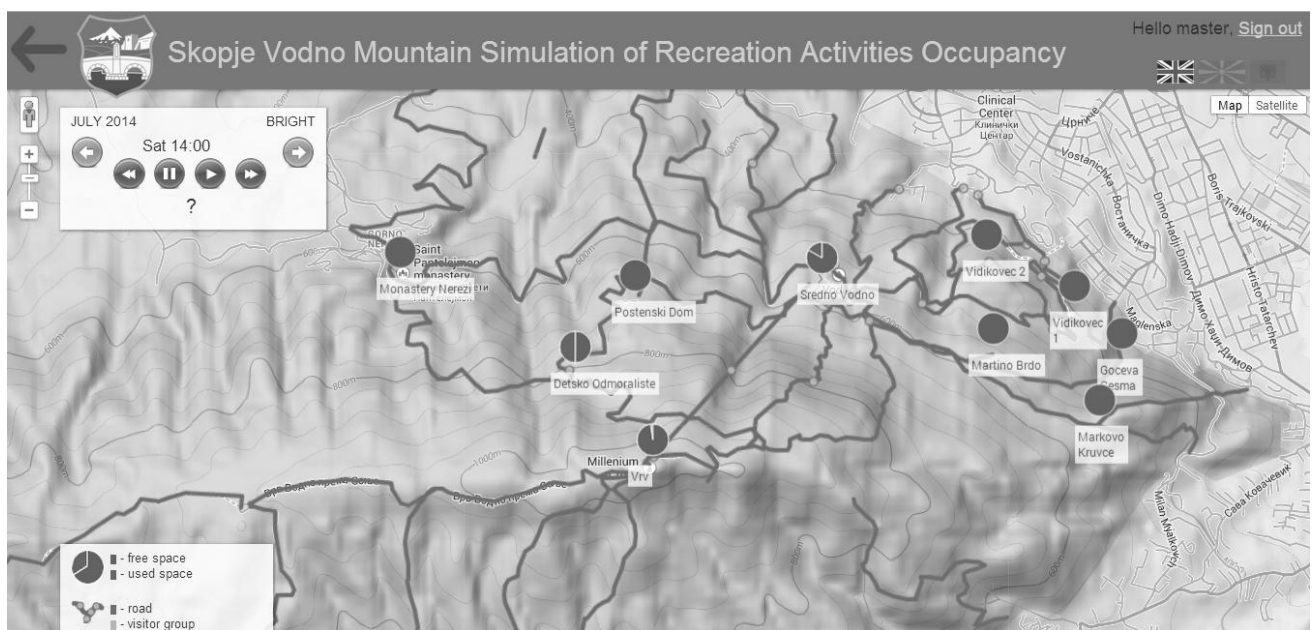


Figure 5: Resources Occupancy Map of Vodno Mountain Area

Figure 5 represents the situation in Vodno Mountain in particular timeslot. Pie charts represent occupancy of the resource (e.g. hotel, restaurant, etc.), black lines represent trails and white icons represent activity for each particular user group in this current timeslot. In this example there are three user groups with the bicycle activity and four hiker groups.

However, detailed analysis of simulation results seeks/requires for more versatile visualization that is ensured by integration with SemaVis (see Figure 6).

Figure 6 shows an example of SemaVis application integrated with Skopje Vodno Mountain Recreational

Activities simulator. SemaVis application screen consists of several (manageable) blocks. One of the blocks contains an integrated simulator with simulation results; other blocks around it contains different visualization for this particular simulation. User can add new visualization approaches, choosing from the right side menu. In this example (see Figure 6) there are four data representation blocks, besides simulator block, data structure block, data structure as graph block, raw data table block, line chart block and candlestick chart. All of these blocks are related to simulator block.



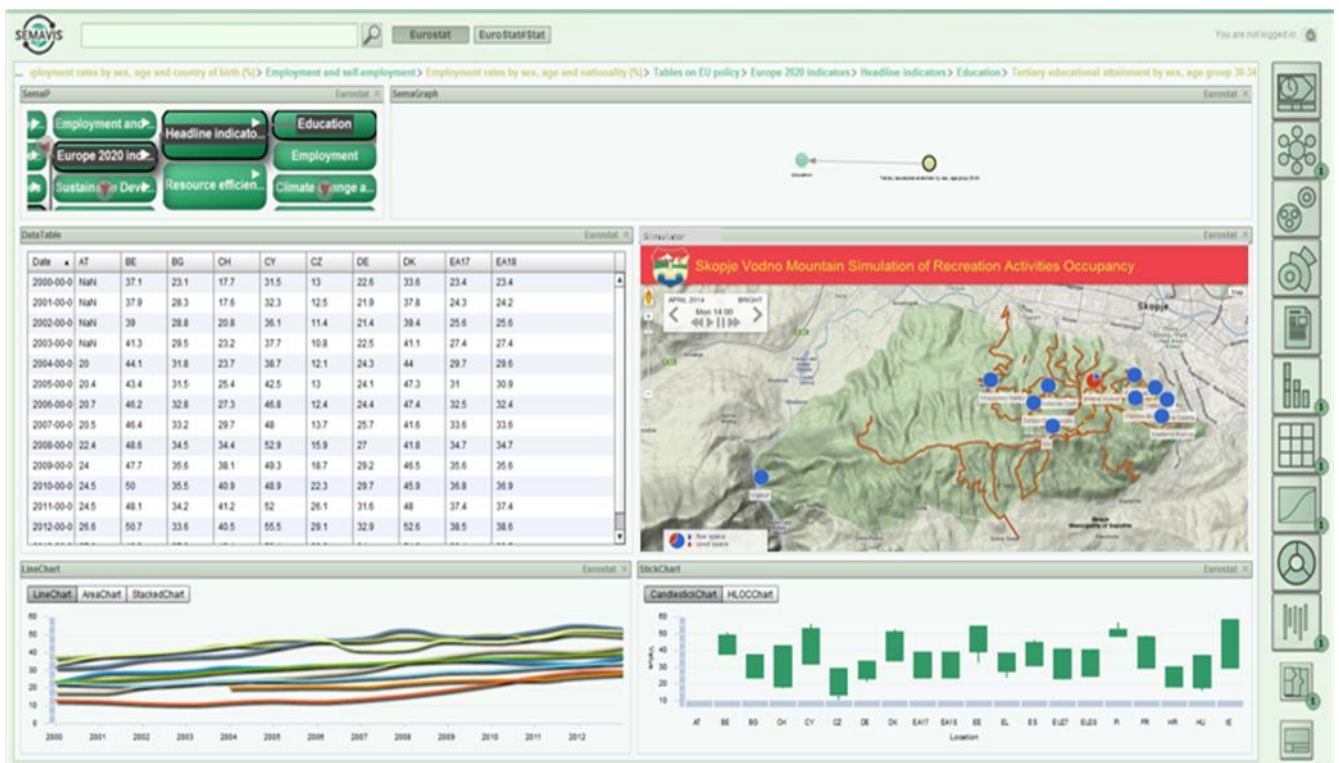


Figure 6: Vodno Mountain Simulator Integration with SemaVis

The approach mentioned above ensures interactive visualization of simulation results as statistical data. The visual analysis of the results can be performed: drilling-down navigation to select the relevant indicators; identifying of relevant influencing factors or impacts; interlinking of various data-sources for an advanced information acquisition and context comprehension; combining of an analysis cockpit to see the data from “different perspectives”.

#### 4. CONCLUSIONS

In situation of ambiguous potential solutions and significant amount of important influencing factors the simulation is right tool for possible trends and features interactive identification.

Even the best solution can remain misunderstood if it will not be displayed in the language and form clear for the user.

One of more traditionally efficient form of perception is visual. However this is necessary requirement only, because difficult and multilayer results must be shown in understandable form in conformity with the user’s requests.

The visualization functions built-in in simulators are limited, therefore for visualization of versatile and intelligent data the use of specific visualizations tools is reasonable, for example, SemaVis in FUPOL case.

If the simulator is Open Source Software (OSS) tool, but visualization environment belongs to commercial products then for the integration CSV formats can be used that does not break the rules of commercial license.

Respecting development of Virtual and Augmented Reality (VR/AR) solutions the next step could be use of VR/AR authoring platforms for simulation results visualization.

Nevertheless more or less standardization activities must be carried out before related with simulation tools connection alignment with VR/AR environments. Unfortunately this is the future of research which must be implemented together with Future Internet standardization.

#### ACKNOWLEDGEMENTS

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

#### REFERENCES

- Aizstrauts, A., Ginters, E., Aizstrauta, D. and Sonntagbauer, P., 2012. Easy Communication Environment on the Cloud as Distributed Simulation Infrastructure. *Proceedings of the 5th WSEAS World Congress on Applied Computing Conference (ACC '12), Recent Advances in Computing Engineering Series 2*. ISBN: 978-1-61804-089-3, ISSN: 1790-5109, 2-4 May, Faro, Portugal, pp. 173-179.
- Aizstrauts, A., Ginters, E., Lauberte, I., Piera, M-A., 2013. Multi-level Architecture on Web Services Based Policy Domain Use Cases Simulator. In J. Barjis, A. Gupta, A. Meshkat, eds. *Enterprise & Organizational Modeling and Simulation (EOMAS 2013)*. Lecture Notes in Business Information

- Processing (LNBIP), Vol. 153, pp.130-146, ISSN 1865-1348, Springer, Heidelberg.
- Burkhardt, D., Nazemi, K., Stab, C., Steiger, M., Kuijper, A., Kohlhammer, J., 2013. Visual Statistics Cockpits for Information Gathering in the Policy-Making Process. *In Advances in Visual Computing*. 9th International Symposium, ISVC 2013, Part II. Berlin, Heidelberg, New York : Springer, pp. 86-97, LNCS 8034.
- Burkhardt, D., Nazemi, K., Sonntagbauer, P., Kohlhammer, J., 2013. Interactive Visualizations in the Process of Policy Modeling. In M. A. Wimmer, ed. et al. Electronic Government and Electronic Participation. *Joint Proceedings of Ongoing Research of IFIP EGOV and IFIP ePart 2013*. (GI-Edition - Lecture Notes in Informatics (LNI) P-221), pp. 104-115. Bonn: Köllen.
- Ginters, E., Aizstrauts, A., Dreija, G., Ablazevica, M., Stepucev, S., Sakne, I., Baltruks, M., Piera, M-A., Buil, R., Gusev, M., Velkoski, G. 2014. Deliverable 4.5 – FUPOL Simulator Software prototype (Vodno Mountain Recreation Activities), 225.
- Keim, D. A., Kohlhammer, J., May, T., Thomas, J. J., 2006. Event Summary of the Workshop on Visual Analytics. *Computers&Graphics*, Vol. 30:2, Elsevier.
- Kohlhammer, J., Nazemi, K., Ruppert, T., Burkhardt, D., 2012. Toward Visualization in Policy Modeling. *IEEE Computer Graphics and Applications*, Vol.32, 5, pp. 84-89.
- Nazemi, K., Breyer, M., Burkhardt, D.; Stab, C. and Kohlhammer, J., 2014. SemaVis - A New Approach for Visualizing Semantic Information. *In Towards the Internet of Services: The Theseus Program*, Springer.
- Nazemi, K., Burkhardt, D., Breyer, M., Stab, C., Fellner, D. W., 2010. Semantic Visualization Cockpit: Adaptable Composition of Semantics-Visualization Techniques for Knowledge-Exploration. *In ICL 2010 Proceedings: International Conference Interactive Computer Aided Learning*. Kassel: University Press, pp. 163-173.
- Nazemi, K., Burkhardt, D., Breyer, M., Kuijper, A., 2011. Modeling Users for Adaptive Semantics Visualizations. *In Universal Access in Human-Computer Interaction: Part II: Users Diversity*. Berlin, Heidelberg, New York: Springer, pp. 88-97, LNCS 6766.
- Nazemi, K., Burkhardt, D., Praetorius, A., Breyer, M., Kuijper, A., 2011. Adapting User Interfaces by Analyzing Data Characteristics for Determining Adequate Visualizations. *In Human Centered Design: Second International Conference: HCD 2011*. Berlin, Heidelberg, New York: Springer, pp. 566-575, LNCS 6776.
- Nazemi, K., Stab, C., Kuijper, A., 2011. A Reference Model for Adaptive Visualization Systems In: Jacko, Julie A. (Ed.): *Human-Computer Interaction: Part I: Design and Development Approaches*. Berlin, Heidelberg, New York: Springer, 2011, pp. 480-489. (Lecture Notes in Computer Science (LNCS) 6761).- DOI 10.1007/978-3-642-21602-2\_52.
- Shneiderman, B., 1996. The Eyes Have It: A Task By Data Type Taxonomy For Information Visualization, *IEEE Visual Languages*, pp. 336-343, College Park, Maryland, US.

#### AUTHORS BIOGRAPHY

**Egils Ginters** is director of Socio-technical Systems Engineering Institute. He is full time Professor of Information Technologies in the Systems Modelling Department at the Vidzeme University of Applied Sciences. He is a Senior member of the Institute of Electrical and Electronics Engineers (IEEE), member of European Social Simulation Association (ESSA) and Latvian Simulation Society. His main field of interests involves: systems simulation technologies, logistics information systems, and technology acceptance and sustainability assessment. He has more than 140 scientific articles related with the research fields.

**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.

**Mikelis Baltruks** graduated in Information Technology at the Vidzeme University of Applied Sciences in 2014. Henceforth he worked as programmer at Socio-technical Systems Engineering Institute. His research interests are software designing and simulation.

**Kawa Nazemi** graduated in Media and Computer Science at the University of Applied Sciences Giessen-Friedberg in 2006. Henceforth he worked as researcher at the Fraunhofer IGD and was involved in several national and European projects, where he developed user-centred systems for knowledge representations. Since 2011 he leads the research group Semantics Visualization and is the deputy head of the associated department. His research interests are focused on technologies and concepts for user-, context-, and content-adaptive knowledge representation and intelligent visualizations.

**Dirk Burkhardt** holds a diploma in Computer Sciences from the University of Applied Sciences Zittau/Görlitz. His research interests are focused on concepts and technologies for user-centred process modeling and visualization technologies. He also researches in gesture-based multimodal interactions for intuitive interactions and navigations within graphical user-interfaces.



# VISUALIZATION TECHNIQUES FOR SIMULATION MODELS IN VIRTUAL REALITY

Arnis Cirulis<sup>(a)</sup>, Egils Ginters<sup>(b)</sup>

<sup>(a)</sup> Sociotechnical Systems Engineering Institute, Valmiera, Latvia

<sup>(b)</sup> Sociotechnical Systems Engineering Institute, Valmiera, Latvia

<sup>(a)</sup>[arnis.cirulis@va.lv](mailto:arnis.cirulis@va.lv), <sup>(b)</sup>[egils.ginters@va.lv](mailto:egils.ginters@va.lv)

## ABSTRACT

The aim of this paper is to merge simulation modelling and virtual reality technologies more deeply together. There are lot of platforms in field of simulation modelling which provide acceptable three dimensional capabilities and more flexible visualization properties to ease the understanding of the results of simulation model. However virtual and augmented reality technologies must be considered as well, because their development in last decade is strongly considerable and could improve perception. To hand the results or the process of simulation model to people, students or trainees in easy understandable and interactive way could be a great challenge. Thereby the researcher team of Virtual and Augmented Reality laboratory of Vidzeme University of Applied sciences develop new methods for basic 3D data depiction in virtual and augmented environment. Petri nets based modelling and notation could be appropriate and most comparable to functional block diagrams and parallel processes which are organized in scenarios of virtual reality systems. If transformation between these two technologies could be achieved, then successful approbation in different scenarios could open new possibilities in education and wide use of simulation modelling.

Keywords: virtual reality, augmented reality, Petri Nets, NetLogo, simulation modelling, visualization

## 1. INTRODUCTION

The history of necessity to merge real and virtual environment is longstanding, however definitions are formulated just in last decades. If the first approaches of mixing real and virtual appeared in cinematography in 1950s, then definition appeared only in 1997 by Azuma. History facts are summarized by Kent [2011], and the first entry devoted to cinematographer Morton Heiling, who created Sensorama simulator in 1957, this simulator was patented and provided visuals, sound, vibrations and smell. The AR term appeared only in 1990 by Tom Caudell whose work was related to aircraft maintenance works. Azuma's definition [1997] in 1997 prescribed AR as combination of real and virtual worlds in 3D with real-time interaction capabilities. Despite this definition, most of nowadays AR applications still offer real worlds augmentation

with two dimensional data. Most of applications from Google Play and Apple's App Store for Android and iPhoneOS based smartphones and tablets provide two dimensional information. For example, text or image based descriptions or instructions, navigation arrows and instructions, or poorly interactive marker solutions with 3D object integration. Rapid growth of mobile devices usage and also technological leap of supported features makes professionals to think over new solutions in different economic sectors. Modern mobile and entertainment AR systems use one or more of the following tracking technologies: digital cameras and/or optical sensors (Nintendo Wii, Sony PlayStation Move, Microsoft Kinect), accelerometers, GPS, gyroscopes, solid state compasses, RFID and wireless sensors. These technologies offer varying levels of accuracy and precision. By Kent's literature analysis [2011] most important is the position and orientation of the user's head. Tracking the user's hand(s) or handheld input device can provide 6DoF interaction technique. In Furht's turn [2011] performance is relevant consideration. In spite of rapid advances in industry of mobile devices, their computing platform for real time imaging applications is still rather limited if done using cell phone's platform. As a result, many applications send data to a remote computer that does the computation and sends the result back to the mobile device, but this approach is not well adapted to AR due to limited bandwidth. Nevertheless, considering the rapid development in mobile devices computing power, it can be considered feasible to develop real-time AR applications locally processed. This Furht's conclusion [2011] confirmed a choice for AR pilot product use the tablet PC with camera and already integrated sensors. More immersive solution with Vuzix AR HMD will be developed after development of theoretical model and working portable tablet solution. The usefulness of mobile devices use is also affirmed by Lee [2009]. Mobile computing devices have been widely deployed to customers; the interest in mobile AR is increasing. As the need for mobility is growing, computing devices shrink in size and gain acceptance beyond an audience of tech-savvy specialists. With the help of enhanced computing power, small devices have sufficient computational capability to process complex visual computations. For example, a cell phone with a camera

and wide screen can serve as an inexpensive AR device for the general public. Emerging wearable computers are also excellent platforms for enabling AR anywhere users may go.

## 2. PROBLEM AREA AND GLOBAL PRACTICE IN CONTENTS VISUALIZATION

Today, image processing and computer vision technologies have been progressed to a stage that allows us to infer the 3D information of the world directly from the images. Because of the success of these technologies, more and more vision-based AR applications are emerged. Augmented reality refers to the combination of virtual objects and real-world environment, so that users can experience a realistic illusion when using the interactive virtual object to explore the real-world environment (Kan et al., 2011). Several technologies and approaches of augmented reality implementations are available today. These technologies can be divided in two basic groups: marker-based systems and marker-less systems. Kan et al. [2011] explains marker-based systems by use of specific marker for 3D tracking and positioning. This marker is employed to identify the corresponding virtual object that is to be placed in the real-world environment. When the marker is used as a tracking target, it has to be registered in the system in advance, as well as the virtual objects it associates with. However, since the registered information is independent for different AR systems, markers used in one system may not be applied in another system, unless an additional registration procedure is applied. For that reason in effective way RFID technologies could be used. Ginters et al. [2011] introduced *AR-RFID* low cost visualization solution for logistics. To visualize the items in the storehouse which must be moved to assembly bay the *AR-RFID* solution is used ensuring additional possibilities of checking. That allows reducing the amount of potential errors and diminishing possible losses. RFID scanner reads the component code and 3D model of the item is visualized on the screen. *AR-RFID* by Ginters et al. [2011] eliminates the problem of generating an equally stable and recognizable fiducial marker, because single marker is used only for tracking, not to recognize the 3D model. This makes it possible to use single marker with the highest stability. To identify the 3D model, an identification code, which is read from the RFID tag, is used. The 3D model is found using the identification code and displayed on the screen of the tablet PC. As well considerable marker-based version is described by Kan et al. [2011] how traditional fiducial markers are replaced by QR codes, allowing to use AR in public domain systems. The QR code can be easily generated by any user and the AR system can track it no matter what the information it embedded. Thanks to *FullHD* video cameras and increase of computing performance, it is easier to pull down traditional marker restrictions, Seac02 company offers in its authoring platform LinceoVR to use self-taken photos as markers,

providing limitless object-marker alignment options also solving the problem with fiducial marker hiding, which sometimes influences immersion level, if it is visible.

The second AR technologies group is based on marker-less solutions, meaning that marker-less pose trackers mostly rely on natural feature points which are visible in the users environment. Four requirements are summarized by Herling and Broll [2011] to allow accurate pose determination from natural points. These requirements are as follows: fast computation time, robustness with respect to changing lighting conditions and image blurring, robustness against observation from different viewing angles, scale invariance or changing viewing distance. Numerous feature detectors have been proposed providing different properties concerning detection robustness or detection speed. Herling and Broll [2011] divides exiting detectors in two classes: corner detectors spotting corner-like feature, blob detectors not spotting unique corner positions but image regions covering blob-like structures with a certain size.

If nowadays visualization technologies get more and more advanced and realistic in different environments, then also importance of visualized contents must be considered as well. It is coming more and more difficult to construct intelligent environment with highly intelligent contents. Despite complex scenario trees and operation matrixes, simulation modelling is a key to a truly intellectual and realistic content of the virtual environment. Lots of nowadays simulations modelling tools already evaluate three dimensional graphics to visualize results of the model and to make them more perceivable also to unprofessional auditory.

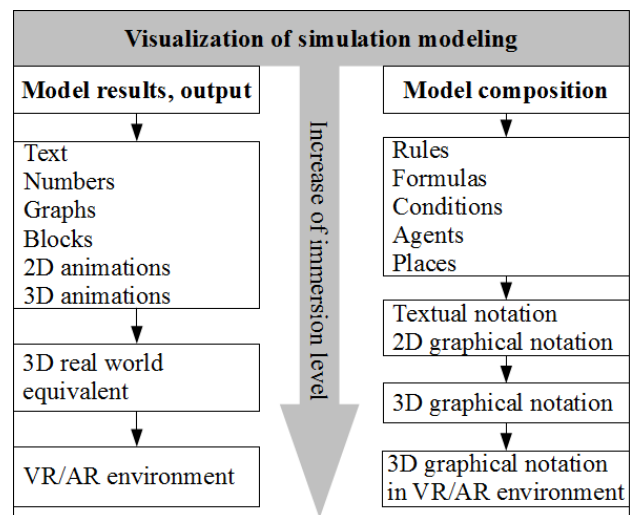


Figure 1: Visualization domains for simulation modelling

In figure 1 virtual and augmented reality is presented as even more immersive environment comparing to typically animated 2D and 3D objects. Thus brings the idea to next step not only to visualize results of the simulation model but also make more

immersive the composition process of simulation models (see fig. 1) with 3D graphical notation carried out in VR/AR environment. In this paper NetLogo and Petri nets experience is used to analyse theoretical aspects and chances in usage of VRAR technologies in field of simulation modelling.

### 3. CHARACTER RECOGNITION IN VR/AR ENVIRONMENT

As the test platform TemPerMod simulation model will be used. The model TemPerMod is created in the NetLogo environment. NetLogo is a multi-agent programmable modelling environment for simulating natural and social phenomena. The agent-based model TemPerMod consists from three levels, where first level is that the specialist could change corresponding attributes to every temperament type. That is, he or she can adjust such attributes: figure, colour and activity from predefined list. As depicted in top of figure 2, the modelling desktop has the form of a square and it is divided in four frames. In each frame has equal amount of the agents, which have the same form and colour. At the beginning and the end of simulation session applicant has to answer some questions, which are provided using a popup. This type of simulation is very useful for identifying person's temperament, because it takes less time than executing a long traditional type test. Results from the NetLogo model "TemPerMod" is matching person's temperament type better than results from traditional tests (Lauberte, Ginters and Cirulis, 2010).

In bottom of figure 2 is depicted future evaluation of model which will improve test results and solve problems with users' patience during the test. For interactive final questions data glove or motion recognition for object selection will be used. Depending on runtime conditions Vuzix head mounted display can be used in virtual reality (VR) mode or in augmented reality (AR) mode. If surrounding is too disturbing then VR mode with predefined virtual environments can be loaded, otherwise AR mode can be more entertaining and immersive but effect on test results in both modes needs to be analysed.

For present TemPerMod model during the modelling session every type of agents (geometrical objects) moves with different speed of motion and step by step they eventually bounce into inner quadrant. There they take a random spot and stop till beginning of next cycle. The length of the modelling session is over 100 cycles, where one cycle continues eight to ten seconds. Agents change shape, colour, behaviour and position in every cycle. During the modelling session also ten phrases in written form are announced. The phrases change on screen at every tenth cycle. Full modelling session takes approximately 20 minutes. At the start and at the end of modelling session, applicant is prompted some questions and is asked to input some information about himself, but the end of modelling session applicant has to answer to questions about

preferred agents, colour, activity and phrase on the screen (Lauberte, Ginters and Cirulis, 2010).

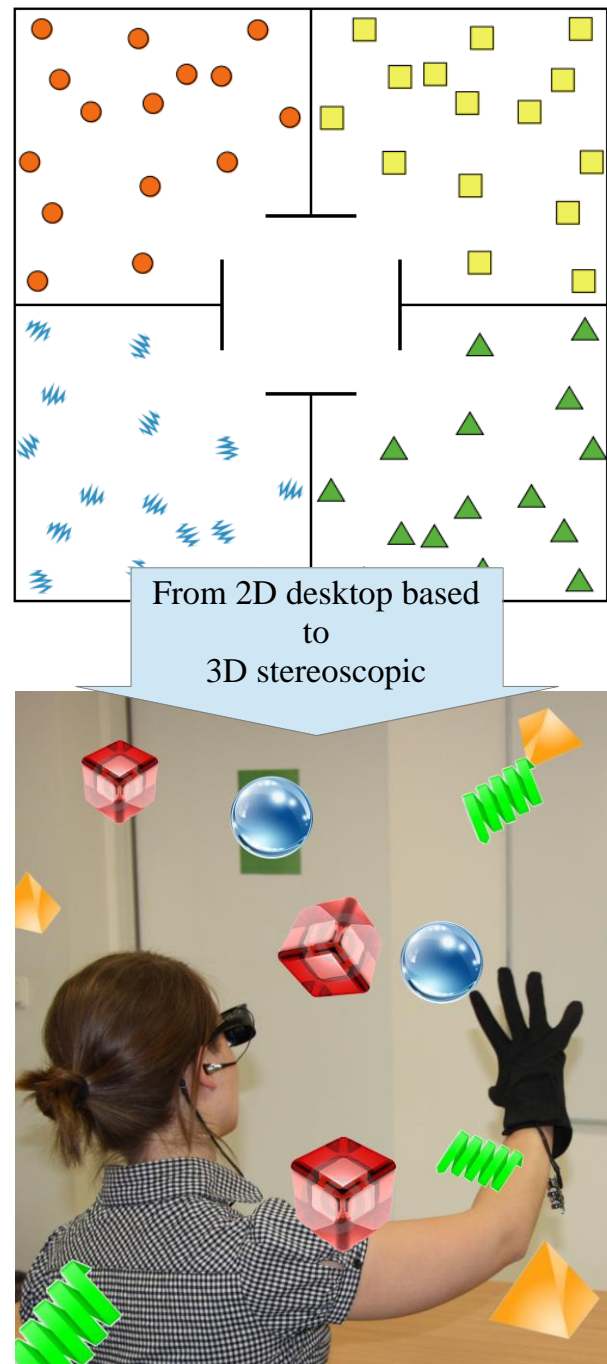


Figure 2: TemPerMod in 3D augmented and stereoscopic environment

To achieve working environment in this paper Virtools VR/AR platform is considered as effective platform which integrates support for many VR/AR devices. As depicted in figure 3 estimated flowchart data from NetLogo environment should be collected by development of TemPerMod procedure converter software module for Virtools platform. Also NetLogo exports can be used to collect data in temporary files during simulation runtime. Depending on variables used in TemPerMod and users FOV (field of view) 3D space

is constructed in form of global and local coordinate system. Then comes 3D object generation phase where count, size and random location is defined. Colour is assigned with glass type shaders or typical colour codes or predefined coloured textures can be used. When the virtual world is designed the simulation must be run for predefined amount of time. 3D objects transforms, rotates in provided space and changes their colour during predefined phases. Such hypothesis is also considered that in VR/AR environment the necessary simulation time can be decreased to get the successful results.

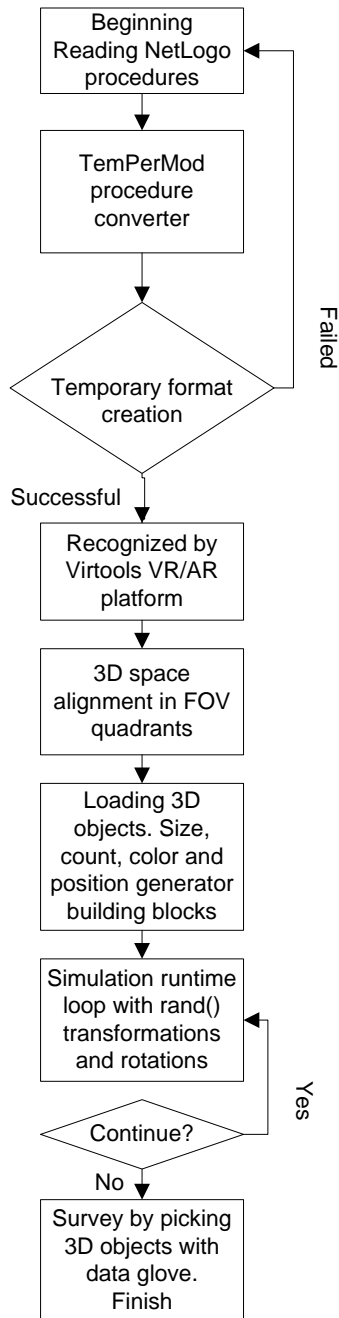


Figure 3: Flowchart for 3D stereoscopic TemPerMod

The final survey form is improved with interaction with 3D objects and audial questions and voice recognition for answer analysis. The following section will introduce visualization for Petri nets which are also used in simulation modelling.

#### 4. PETRI NETS IMPLEMENTATION IN VR/AR ENVIRONMENT

Petri nets provide a mathematically defined approach to specification, project, analysis, and verification, performance assessment for systems of parallel and distributed processes. They offer not only accurate semantics and a theoretical foundation but also a graphic form that facilitates the understanding of information and management flow within the same formalism. Petri nets are also utilised in various applications because of the intuitively attractive graphic form in the depiction of different diagrams. Data transmission networks and other communications networks – from telecommunications to internet-based services – are characterised by distributed operation and form systems with concurrent components (WoPeD Petri Nets, 2012).

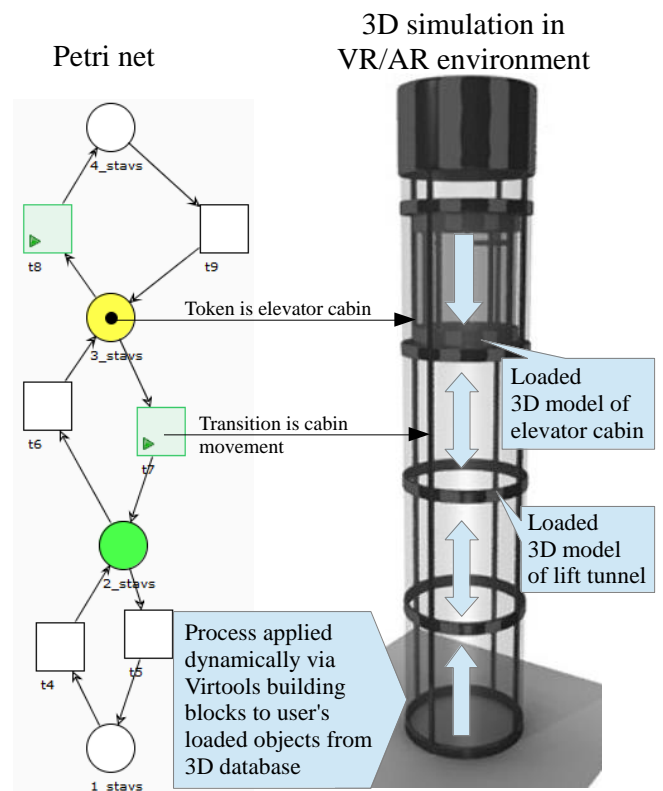


Figure 4: Elevator Petri net look in VR/AR environment

In VR/AR laboratory Petri nets were used to model the operating framework of the method, which is based on dynamic selection of device sets during the training process, using the trainee's individual personality type as the foundation. Petri net algorithm was developed for the enhanced training method and for the location of the most suitable devices, as well as suitability index



matrices of devices and device sets, based on which the selection of devices takes place (Cirulis, 2011).

The author used Petri net modelling to explore the nature of the method so that it could be implemented in a real, existing VR/AR authoring platform. There is a possibility of also using UML (Unified Modelling Language), yet UML notation describes aspects of the programming system (from requirements to implementation) from various points of view, whereas Petri nets are fundamentally intended for modelling dynamic system behaviour, but they are also to some extent comparable to UML state and activity diagrams.

In this case classical Petri net for elevator operation is used to explain the idea of result depiction in VR/AR environment (see figure 4). As summarized by Kindler [2004] there are several solutions and tools which provide visualization for Petri net structures. Tool ExSpect uses the concept of dashboards in order to visualize the dynamic behaviour of a system in a way that is familiar to the experts in the application area. Mimic library of Design/CPN allows a Design/CPN simulation to manipulate graphical objects, and the user can interact with the simulation via these graphical elements. GenGED tool for visualizing Petri nets based on graph transformations and their animations. PEP is one of the first Petri Net tools that came up with a 3D visualization of Petri net models where SimPep triggers animations in a VRML model while simulating the underlying Petri net (Kindler, 2004). Kindler successfully visualized Petri net of toy train system in 3D. Lots of principles offered by Kindler for behaviour animation and geometry functions will be used in VR/AR approach for Petri nets (see figure 5). Instead of Petri net kernel, XML and VRML models WoPeD tool's source code is used to develop daemon process for data delivery from token game to Virtools. WoPeD's coverability graph is also exported to prepare matrix of 3D objects' values for determined scenario. These values include information on 3D objects' sizes, scaling, transformations and rotations. For Virtools platform two building blocks are developed by SDK to achieve functionality of data receive from daemon process and adjustment of 3D objects behaviour values which are necessary during the simulation session. As this is the very beginning of the project, more detailed workflows will be presented after development of this approach.

By development of 3D visualization for Petri net with technologies mentioned above, it gives varies of ways and looks to run the simulation. Virtual reality gives opportunity of immersion, e.g. instead of watching 3D simulation on a monitor you can be in the virtual elevator cabin and see the world from other perspective. This brings us to a reverse approach by which the token game is driven by actions in virtual world. In elevator case the VR participant presses buttons in cabin and the model professional receives information through a model (figure 5). In the beginning it could be easy and interactive way for editing exiting models rather than for development of new ones.

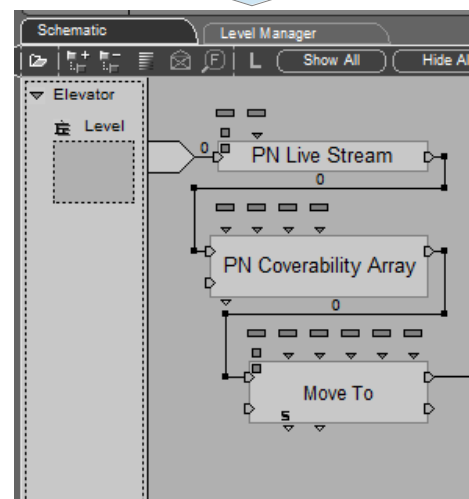
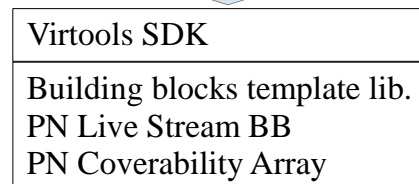
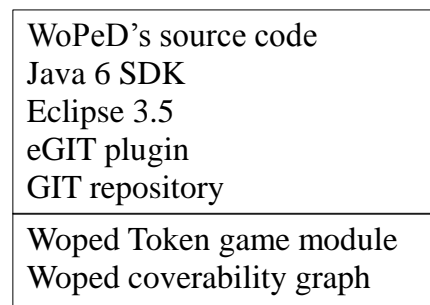


Figure 5: Technological approach for Petri net VR/AR visualization

## 5. CONCLUSIONS

Logics of visualization for different simulation modelling tools are different but technical realization and tools can be similar. As VR/AR platforms are complicated and contain plenty of other tools, then best and cost effective way is to use VR/AR platform with software module development capabilities. By provision of VR/AR environment for simulation models also non experts of the field can perceive information in understandable way. In these paper ideas and possible realization ways were introduced for NetLogo simulation model and Petri net model.

Interesting aspect was brought also in figure 1 and figure 5 offering simulation in interactive way by use of VR/AR technologies. 3D graphical notation for Petri nets was introduced by Dance already in 2001. Also reverse approach of simulation modelling or simulation model editing must be seriously considered in future research. Dance [2001] also confirms that it is more effective and easier way to convey the structure of a physical object in 3D space using 3D objects than it is

to do so with a 2D symbolic representation, or, a text description, or, a mathematical representation.

## REFERENCES

- Azuma, R., 1997. A survey of augmented reality, <http://www.ronaldazuma.com/papers/ARpresence.pdf>, Presence: Teleoperators and Virtual Environments, pp. 355-385.
- Cirulis A., Ginters E., 2011. Training Scenario Operations Realization In Virtual Reality. Proceedings of the 13th WSEAS International Conference on Automatic Control, Modelling and Simulation (ACMOS 11), Lanzarote, Spain, May 27-29, 2011, ISBN 978-1-61804-004-6, pp.39-44.
- Dance, L. K., Fishwick, P. A., 2001. Methodology for the 3D modeling and visualization of concurrency networks. In Aerospace/Defense Sensing, Simulation, and Controls (pp. 152-163). International Society for Optics and Photonics.
- Furht, B., 2011. Handbook of Augmented Reality. Springer. Florida Atlantic University, USA, ISBN 978-1-4614-0063-9, 746 p.
- Ginters, E., Martin-Gutierrez, J., 2011. Low cost augmented reality and RFID application for logistics items visualisation. Annual Proceedings of Vidzeme of Applied Sciences "ICTE in Regional Development".
- Herling, J., Beoll, W.: Markerless tracking for augmented reality. 255 p. Handbook of Virtual Reality. Springer (2011)
- Kan, T., 2011. QR code based augmented reality applications. 339 p. Handbook of Virtual Reality. Springer.
- Kent, J., 2011. The Augmented reality Handbook - Everything you need to know about Augmented reality.
- Kindler, E., Páles, C., 2004. 3D-visualization of Petri net models: Concept and realization. In Applications and Theory of Petri Nets 2004 (pp. 464-473). Springer Berlin Heidelberg.
- Lauberte, I., Ginters, E., & Cirulis, A., 2010. Agent-based simulation use in multi-step training systems based on applicant's character recognition. In Advances in Databases and Information Systems (pp. 16-22). Springer Berlin Heidelberg.
- Lee, T., 2009. Multithreaded Hybrid Feature Tracking for Markerless Augmented Reality. Visualization and Computer Graphics, IEEE Transactions. Comput. Sci. Dept., Univ. of California, Los Angeles, pp. 355--368.
- WoPeD Petri Nets, 2010. Available from: <http://193.196.7.195:8080/woped/PetriNets> [accessed 10 April 2012]

## AUTHORS BIOGRAPHY

**Dr. Arnis Cirulis** is a researcher at Sociotechnical Systems Engineering institute of Vidzeme University of Applied Sciences and is responsible for research in field

of VR/AR (virtual and augmented reality) technologies. He holds Dr.sc.ing in 2012 and he is also an assistant professor in Faculty of Engineering.

**Prof. Egils Ginters** is director of Sociotechnical Systems Engineering Institute. He is Full time Professor of Information Technologies in the Systems Modelling Department at the Vidzeme University of Applied Sciences. Graduated with excellence from Riga Technical University Department of Automation and Telemechanics in 1984 and holds Dr.Sc.Eng. in 1996. He is a member of Institute of Electrical and Electronics Engineers (IEEE), European Social Simulation Association (ESSA) and Latvian Simulation Society.



# A COLOURED PETRI NET MODEL FOR THE MAS SIMULATION OF URBAN ECONOMICS IN YANTAI

Roman Buil<sup>(a)</sup>, Miquel Angel Piera<sup>(b)</sup>, Egils Ginters<sup>(c)</sup>, Artis Aizstrauts<sup>(d)</sup>

<sup>(a,b)</sup> Universitat Autònoma de Barcelona, Department of Telecommunications and Systems Engineering, Unit of Logistics and Aeronautics, C. dels Emprius, 2, 08202, Sabadell, Barcelona, Spain

<sup>(c,d)</sup> Sociotechnical Systems Engineering Institute of Vidzeme University of Applied Sciences, Cesu street 4, Valmiera LV-4201, Latvia

<sup>(a)</sup>[miquelangel.piera@uab.cat](mailto:miquelangel.piera@uab.cat), <sup>(b)</sup>[roman.buil@uab.cat](mailto:roman.buil@uab.cat), <sup>(c)</sup>[egils.ginters@va.lv](mailto:egils.ginters@va.lv), <sup>(d)</sup>[artis.aizstrauts@va.lv](mailto:artis.aizstrauts@va.lv)

## ABSTRACT

E-governance is one area in which the use of simulation could help the decision maker to take better decisions with positive impact in the society. The Multi Agent Systems (MAS) models were already used to foster e-participation in which citizens could be involved in the design of urban policies that affects their habitat environment. Coloured Petri Net (CPN) models have been already used as a base to develop MAS models in Repast Simphony.

This paper presents the use case developed in FUPOL for Yantai, China. The model will be used to decide which policy to use in order to establish a criteria to force industries to upgrade (to reduce consumptions and pollution) or to close. Results will show how different are the closed industries during a simulation depending on the indicators weights in the cost function.

Keywords: Urban policy, urban economics, petri net, MAS simulation, e-participation

## 1. INTRODUCTION

Digital Simulation techniques can contribute to increase the knowledge about the possible consequences of taking some decisions or others. E-governance is one area in which the use of simulation could help the decision maker to take better decisions with positive impact in the society. A new methodology based on the transformation of Coloured Petri Net (CPN) models into Multi Agent System (MAS) models was already presented in Buil and Piera (2013) and Piera et al (2014) as part of their work in the FP7 project FUPOL.

Buil and Piera (2013) presented a causal modelling approach, which takes into consideration that actors have different skills and pre-knowledge of the complex urban system. By incorporating users' heterogeneity in non-scientific knowledge (subjective values), interests and preferences, FUPOL modelling approach contributed to overcome one of the main shortages of present policy models: citizen e-participation.

E-participation in urban decision-making means the use of ICT for enabling and strengthening citizen participation in democratic decision-making processes.

The use of ICT in e-Participation process consists on motivation and engagement of a large number of citizens through diverse modes of technical and communicative skills to ensure broader participation in the policy process, real-time qualitative and accessible information, and transparent and accountable governance.

Piera et al (2013) consider some of the difficulties in establishing verification and validation of agent based models, and proposes the use of CPN formalism to specify agent behaviour in order to check if the model looks logical and the model behaves logical. Model plausibility is used to express the conformity of the model with a priori knowledge about the process. A proof-of-concept is presented by means of a case study for testing the robustness of emergent patterns through sensitivity analyses and can be used for model calibration.

This paper presents a special use case in urban economics. If the first approach was based on citizens' behaviour related to some specific system; in this case, due to cultural aspects, the model will be based on industries' behaviour. The pilot city interested in this model, and participating in its development, is Yantai, China. The simulation model will be used to help a decision making process, in which, the industrial park of the city must be renovated in order to deal with the problem of high energy consumption together with the high generation of pollution.

Section 2 and 3 present the modeling approaches used during the model development: CPN and MAS. The case study is presented in Section 4; and the conclusions in Section 5.

## 2. COLOURED PETRI NETS

Coloured Petri Nets (CPN) is a simple yet powerful modelling formalism, which allows to properly modelling discrete-event dynamic systems that present a concurrent, asynchronous and parallel behaviour (Moore et al. 1996, Jensen 1997, Christensen et al. 2001). CPN can be graphically represented as a bipartite graph, which is composed of two types of nodes: the place nodes and the transition nodes. Place nodes are commonly used to model system resources or logic

conditions, and transition nodes are associated to activities of the real system. The entities that flow in the model are known as tokens and they have attributes known as colours. The use of colours allows modelling not only the dynamic behaviour of systems but also the information flow, which is a key attribute in decision making (Mujica and Piera 2011).

The formalism can be graphically represented by a bipartite graph where the place nodes are represented by circles and the transition nodes by rectangles or solid lines. Figure 2 illustrates a graphical representation of a CPN model.

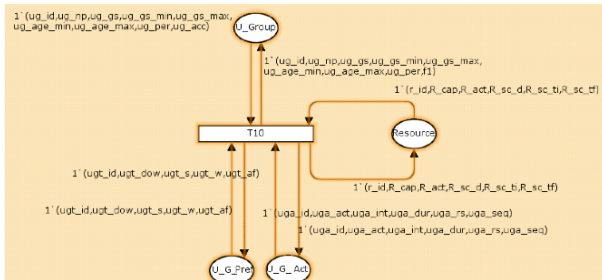


Figure 1: CPN Model Example

### 3. MULTI AGENT SYSTEMS

Multi-agent systems (MAS) have been applied in various fields related to Human Sciences, such as Political Sciences, Economics, and Social Sciences (Segal-Halevi (2012), Wilensky, 1999), in which an agent can be seen as an actor that has the power of “Agency”. The concept of agency is paramount since it extends beyond a single human being to organizations and social systems, where some have “agency” power with the ability to make decisions, while others do not have such power. These agents communicate, collaborate and negotiate among each other in order to meet their design objectives.

During the developing phase of MAS, it is necessary to simulate the system feasibility before the formal implementation. One way to achieve this objective is to develop a model to represent the behaviours of the MAS and then to simulate the model performance. In practice, a lot of time might be spent on modelling MAS behaviours while the resulting process models are typically still not compatible with the original system due to the lack of information about actual behaviours of the MAS. To avoid this difficulty, CPN models can be used to describe agents’ behaviour and predict the performance by means of state space analysis tools.

In FUPOL several MAS platforms has been analysed and, among them, Repast Symphony has been chosen as a simulation environment.

Repast Symphony is a widely used, free, and open source environment for agent-based modelling of complex adaptive system. The most recent version, 2.0, was released on March 5, 2012. Repast Symphony is a second-generation environment that builds upon the previous Repast 3 library (North et al. 2013).

Due to its flexibility it is suited to fully integrate the semantic rules present in CPN, and their implementation allow to govern the agents that interact within the environment in a more transparent way which is useful to understand the emergent dynamics caused by the agent interaction.

### 4. URBAN ECONOMICS - YANTAI

The municipality of Yantai needs to reduce the energy (electricity, gas, water, oil) consumption in the city and to reduce the generated pollution. There is a particular problem with the level of electricity consumption.

The initial idea was to start closing industries with high levels of electricity consumption year by year and let the establishment of other industries with reduced electricity consumption; however, there are other factors that they would like to consider. For that reason, the officers of Yantai municipality are planning to face the decision making process of closing industries, by considering several indicators and studying the effects of the decisions in some years time. This is the reason for the use of simulation to define the best conditions (indicators to consider) that will determine which industry must be closed.

The process can be seen as an industrial renovation process, in which, industries upgrade their infrastructure and/or organization in order to reduce consumptions and pollution emissions (to decrement the cost function value). Even they upgrade, companies with worse cost function value (higher cost function) will be closed every year.

#### 4.1. Model Interpretation

As mentioned above, the model will be based on industries’ behavior. However, it is easy to understand that the behavior of an industry is determined by its personnel (director, president, etc.).

The model consists in a set of industries considered as candidates to be closed, and a set of industries’ profiles to be considered as potential established industries in Yantai. They are categorized by a set of attributes:

- Electricity Consumption
- Water Consumption
- Gas Consumption
- Oil Consumption
- Pollution generated
- Waste Indirect Jobs
- Direct jobs
- Exports Trade
- Imports Trade
- Revenue
- Local Government Taxes
- National Government Taxes
- Suppliers in the City
- Clients in the City
- GDP Closure Affectation

There is a cost function depending on these attributes, and the final user will be able to choose the attributes to consider in this cost function, and to assign a weight to each one depending on whether they should be more or less important. For example, it can be tested what would happen if it is taken into account only the consumption of gas, electricity and water, or if it is also taken into account direct jobs and taxes.

Depending on the ranking generated considering the cost function, the decision maker will decide which industries must be closed and which new profiles should be considered.

The simulation model lists the industries and the profiles in descendent order, from the one with more options to be closed, to the one in better conditions depending on the cost function. After one year, the city council will decide which industry must be closed. Therefore, the industries must react to try to modify the order of that list before the decision is taken.

An industry can decide whether to let time pass and see what happen, or to try to upgrade what is needed in order to change the order of the list, moving itself down. The economic capabilities of the industries will affect the possible improvements; moreover, according to which industry concerned and which sector, it will have more or less room for improvement depending on the used technology. Therefore, the final possible improvement will depend on two factors: Technical and Financial.

Considering that not always the expectations are satisfied, the model includes a random variable indicating the percentage of fulfillment for all the attributes involved in the cost function, because industries do not always accomplish with the 100% promised. The percentage of compliance heavily depends on the industry sector and culture. Initially, due to a lack of information about this parameter, percentages of compliance are randomly decided, and they are:

- More than 75% compliance in 35% of cases.
- More than 50% and less than 75% compliance in 25% of cases.
- Between 0% and 50% compliance in 25% of cases.
- And 0% of compliance in 15% of cases.

Final results will give an idea about a future picture of the industrial park in Yantai depending on the attributes used in the cost function.

#### 4.2. Yantai CPN Model

The CPN model is used to analyze the behavior of the industries and the use of the cost function. It helps to determine how to model the behavior of the MAS model agents. The CPN model is not exactly replied in the MAS model due to after its analysis and presentation in Yantai, it was realized that some aspects must be changed, and there was no need to modify the CPN to be able to build the MAS model.

The CPN model has been developed using CPN Tools (<http://cpntools.org/>) and it is presented in the following subsections. However, due to the size of the model, it is not possible to include all the CPN in this paper. Therefore, the list of places types, the list of colors, and the list of transitions, together with one submodel of the CPN, are presented in the following subsections.

##### 4.2.1. Places Types

The types of places are defined using the color assigned to them, and are the following:

- **Improvements:** It contains information regarding the industries upgrade. The first colour indicates the identifier of the industry, while the rest correspond to the percentage of improvement that applies to different indicators.
- **Fulfillment:** It indicates the percentage of satisfied upgrading.
- **Count, Actrans:** They are used as a controller to enable and disable transitions.
- **Or:** It is used to make comparisons.
- **LL:** It indicates the cost of each industry.
- **Result:** It indicates information about the results: industries position in the list and associated cost.
- **Emp:** List of industries to consider.
- **Weight:** It determines the importance of the attributes in the cost function. One value for each industry attribute.
- **FinalImp:** Place with the improvements to be applied.

The places are named  $N_{xx}$ , where  $xx$  is a number and they are of one of the types described above. Table 1 includes the places for each place type. Notice that there are 57 places; however, due to the CPN software, which do not allow to duplicate places, some of the places are between parenthesis, indicating they represent the same as the place in front.

Table 1: Places by type (depending on the color)

Place Type	Places with that color
Improvements	N39, N42 (N56)
Fulfilment	N40
Count	N12 (N15), N19 (N23, N26), N33 (N57)
Actrans	N2, N10 (N58), N16 (N20), N18, N27 (N48, N51), N32 (N52)
Or	N13, N24 (N47)
LL	N8 (N9, N22), N11, N14, N17, N25 (N49), N28 (N29), N34 (N38), N36 (N43, N50), N41 (N55)
Result	N21 (N54), N53
Emp	N4, N5, N6 (N31), N7 (N30), N35 (N37), N45
Weight	N1, N3 (N46)
FinalImp	N44

##### 4.2.2. Colors

There is a color for each place type. Table 2 to Table 11 present the set of colors for each place type color. Left

column is the type of variable and right column, the meaning of the color.

Table 2: Colors at Improvements place color

INT	ID of the company
REAL	Percentage of improvement in electricity consumption
REAL	Percentage improvement in gas consumption
REAL	Percentage improvement in CO2 emissions
REAL	Percentage of improvement in water consumption

Table 3: Colors at Fulfillment place color

REAL	Possible satisfaction of the proposed upgrade
------	-----------------------------------------------

Table 4: Colors at Count place color

INT	Control variable to enable or disable transitions
INT	Control variable to enable or disable transitions

Table 5: Colors at Actrans place color

INT	Control variable to enable or disable transitions
-----	---------------------------------------------------

Table 6: Colors at Or place color

INT	ID of the industry
REAL	Weight associated with the industry
INT	Control variable to enable or disable transitions

Table 7: Colors at LL place color

INT	ID of the industry
REAL	Weight associated with the industry

Table 8: Colors at Results place color

INT	Position in the list
INT	ID of the industry
REAL	Weight associated with the industry

Table 9: Colors at Emp place color

INT	Company ID
REAL	Electricity consumption.
REAL	Gas consumption.
REAL	CO2 emissions.
REAL	Water consumption.
REAL	Number of jobs.
REAL	Number of suppliers in the area.
REAL	Number of customers in the area.
REAL	Local government taxes.
REAL	Surface used.
REAL	Transportation (private).
REAL	Fleet company.

Table 10: Colors at Weight place color

INT	Weight associated with the consumption of electricity
REAL	Weight associated with the consumption of gas
REAL	Weight associated CO2 emissions
REAL	Weight associated with the consumption of water
REAL	Weight associated with the number of jobs
REAL	Weight associated with the number of suppliers in the city
REAL	Weight associated with the number of clients in the city
REAL	Weight associated with local government taxes
REAL	Weight associated with the surface used
REAL	Weight associated with transport (private)
REAL	Weight associated with the company's fleet

REAL	Weight associated with the consumption of electricity
------	-------------------------------------------------------

Table 11: Colors at FinalImp place color

INT	Percentage applied to improve power consumption.
REAL	Percentage applied to improve gas consumption.
REAL	Percentage applied to improve CO2 emissions.
REAL	Percentage applied to improve water consumption.

#### 4.2.3. Transitions

Table 12 presents the list of transitions: Id, input places, output places and description.

Table 12: List of Transitions

ID	Input	Output	Description
T1	N1, N2	N3	Choosing what will be chosen to perform the simulation heuristic.
T2	N3, N4	N7, N8, N57, N58	Calculating the weight associated with each industry already established.
T3	N3, N5	N6, N8	Calculating the weight associated with each industry to be established.
T4	N9, N10	N9, N11	Selection of industries already established in the whole business.
T5	N11, N12, N13	N12, N13, N14	Comparison of the weights associated to find the smallest.
T6	N13, N14, N15, N16	N13, N15, N16, N17	Selection of the industry associated with the lowest weight.
T7	N17, N18, N19, N20	N18, N19, N20, N21	Formation of a list of established industries ordered more associated with less weight.
T8	N12, N14, N16	N11, N12, N16	Return the set of tokens industries are being sorted.
T9	N22, N23, N24	N23, N24, N25	Comparison of the weights associated to find the smallest.
T10	N24, N25, N26, N27	N24, N26, N27, N28	Selection of the industry associated with the lowest weight.
T11	N23, N25, N27	N22, N23, N27	Return the set of tokens industries are being sorted.
T12	N29, N30, N32, N33	N30, N32, N33, N34, N35, N36	Select established industries and decide if they have to raise their consumption or not.
T13	N29, N31	N31, N36	Select industries that want to send them down to the final result.
T14	N37, N38, N39, N40	N40, N41, N42, N43, N44, N45	Select the percentage improvements and compliance with these associated industries. Calculate improve overall application.
T15	N44, N45, N46, N47, N48	N46, N47, N48, N49	Calculating the weight associated with the new industry with the improvements implemented.
T16	N50, N51, N52	N51, N52, N53	Make a list of the final result.

#### 4.2.4. Submodel

Figure 2 presents part of the CPN model, which is used to determine the order of the industries depending on the cost function.

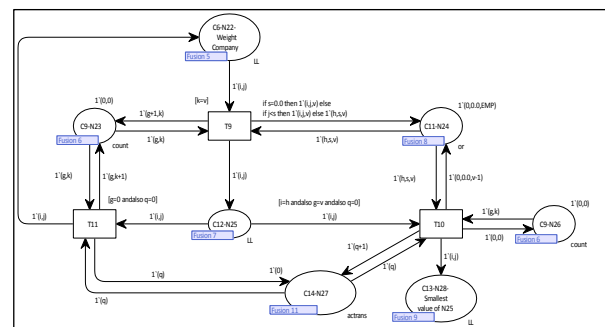


Figure 2: Part of the developed CPN model

### 4.3. Yantai MAS Model

The introductory part of this Section 4 and Section 4.1 already introduced how the MAS model works. The CPN model is the base of the MAS model; however, in that case, there are some modifications due to the results of the simulation of the CPN and the feedback obtained from the responsables of Yantai Municipality.

The main difference between both models is that in the CPN model, the decisions to make at the end of the year were two:

1. Which industry or industries must be closed
2. Which industry or industries must be forced to upgrade

However, in the final MAS model, industries decide if they upgrade by themselves depending on their possibilities, their situation in the list, and considering if the industries of the same sector are upgrading or not.

The cost function includes some of the indicators presented in Section 4.1 multiplied by a weight that can be set to 0. The considered indicators with their sign in the cost function are:

- Electricity Consumption (+)
- Water Consumption (+)
- Gas Consumption (+)
- Oil Consumption (+)
- Pollution generated (+)
- Direct jobs (-)
- Local Government Taxes (-)
- National Government Taxes (-)
- Suppliers in the City (-)
- Clients in the City (-)

### 4.4. Test Results

Due to the partial lack of data, tests have been performed using real data (not complete) and randomly generated data. There is data from 8 real industries containing electricity, water and gas consumption, direct jobs, local taxes (partially), and suppliers in the city. It is considered that 2 industries are upgrading from the beginning (I1 and I2). Four new industry profiles are randomly generated (I9, I10, I11, I12), the simulation time is 5 years with periods of 1 year in which one industry is closed and a new one is established in the city. Table 13 presents the industries' data. If the value is N/A, it is considered as 0.

Table 13: Industries Input Data

Id	Electricity (kw)	Water (tons)	Gas (tons)	Direct Jobs	Loc Tax (bilions)	Suppliers in city	Upgrading
I1	866325	10220000	1800	700	90	2	1
I2	970132	13212000	2100	650	2.1	3	1
I3	755327	9800000	20	150	N/A	5	0
I4	1200	3000	0	12	N/A	0	0
I5	1500	2100	0	15	N/A	0	0
I6	872121	120000	0	100	N/A	6	0
I7	1013222	17000000	2300	200	70	8	0
I8	731121	2100000	0	120	3.95	55	0
I9	20000	10000	0	20	0.5	4	0
I10	10000	8000	1500	15	0.5	6	0
I11	15000	20000	500	10	0.5	5	0
I12	5000	7000	100	25	0.5	10	0

Three different tests have been performed, depending on the cost function weights:

- Test 1: Just the weight for electricity consumption is set to 1; all the others to 0.
- Test 2: Weights for electricity, water and gas consumption set to 1; all the others to 0.
- Test 3: All weights set to 1.
- Test 4: Weights for direct jobs, local taxes and number of suppliers set to 1; all the others to 0.

The initial list of ordered industries (first row) and ordered industry profiles (second row) depending on the cost function, for each test, are:

- Test 1: I7, I2, I6, I1, I3, I8, I5, I4; I12, I10, I11, I9
- Test 2: I7, I2, I1, I3, I8, I6, I4, I5; I12, I10, I9, I11
- Test 3: I7, I2, I1, I3, I8, I6, I4, I5; I12, I10, I9, I11
- Test 4: I4, I5, I6, I3, I8, I7, I2, I1; I12, I9, I10, I11

And the final lists of ordered industries (first row) and the list of closed industries (second row), by order, are:

- Test 1: I1, I8, I9, I11, I10, I12, I4, I5; I7, I2, I6, I3
- Test 2: I8, I6, I11, I9, I10, I12, I5, I4; I7, I2, I3, I1
- Test 3: I8, I6, I11, I9, I10, I12, I5, I4; I7, I2, I3, I1
- Test 4: I11, I10, I9, I12, I8, I7, I2, I1; I4, I5, I6, I3

The total of the indicators considering all the industries are usually reduced irregularly depending on the indicators' weights.

Table 14: Total indicators' values

Test	Electricity (kw)	Water (tons)	Gas (tons)	Direct Jobs	Loc Tax (bilions)	Suppliers in city
0	5210948	52457100	6220	1947	166.05	79
1	1259338	9135528	3397	952	95.95	82
2	1353646	1807049	2100	363	5.95	86
3	1353646	1807049	2100	363	5.95	86
4	3545791	41150998	8090	1786	595.06	93

Table 14 shows the difference between the performed tests. When giving importance just to electricity (Test 1), its consumption decrease a lot after 5 years; however, the number of direct jobs decrease too much. When giving importance to all the energy consumptions (Test 2 and 3), the consumptions decrease together with the direct jobs, which are dramatically reduced to around 20% of the initial number of direct jobs. Notice that the negative indicators do not affect the results when the consumptions are also considered. This is due to the units used and the weights. It will be necessary to perform a high number of tests in order to determine which kind of units to use or which proportions for the weights.



Test 4 gives importance to the negative indicators (direct jobs, local taxes and suppliers) and the results confirm that these indicators are more or less kept or increased; however, the consumptions are not reduced as they are when considering them in the cost function.

## 5. CONCLUSIONS

Under the framework of the FUPOL project one challenging task is the policy modelling and analysis. The proposed model allow the study and analysis of a urban economic model using MAS modelling and simulation, in which, the agents represent industries instead of persons, as made in the rest of the simulation models developed for the FUPOL project.

Results show how the model can be run using different cost functions, which will allow the analysis of a large number of alternatives. The Yantal City Council has no idea about which indicators are the best ones in order to make the correct decisions. Using this model they will be able to test what would happen depending on which indicators and weights they select for the cost function.

## ACKNOWLEDGMENTS

This work is funded by Future Policy Modelling Project (FUPOL), FP7-ICT-2011-7, ref. 287119 (www.fupol.eu).

## REFERENCES

- Buil, R. and Piera, M.A. (2013). A MAS Approach Based on Coloured Petri Net Formalism for Urban Policy Design, *25th European Modeling and Simulation Symposium, EMSS 2013*, pp. 510.
- Christensen, S., Jensen, K., Mailund, T., Kristensen, L.M., 2001. State Space Methods for Timed Coloured Petri Nets. Proc. of 2nd International Colloquium on Petri Net Technologies for Modelling Communication Based Systems, 33-42, Berlin.
- Islam M. S., 2008, Towards a sustainable e-Participation implementation model, *European Journal of ePractice*, www.epracticejournal.eu 1 N° 5, October 2008, ISSN: 1988-625X
- Jensen, K., 1997. Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use. 1 Springer-Verlag, Berlin.
- Moore, K.E., Gupta, S.M., 1996. Petri Net Models of Flexible and Automated Manufacturing Systems: A Survey. *International Journal of Production Research*, 34(11), 3001-3035.
- Mujica, M.A., Piera, M.A., 2011. A Compact Timed State Approach for the Analysis of Manufacturing Systems: Key Algorithmic Improvements, *International Journal of Computer Integrated Manufacturing, Vol.24 (2)*, February 2011.
- North, M.J., N.T. Collier, J. Ozik, E. Tatar, M. Altaweel, C.M. Macal, M. Bragen, and P. Sydelko, 2013. Complex Adaptive Systems Modeling with Repast Symphony, *Complex Adaptive Systems Modeling, Springer, Heidelberg, FRG (2013)*.
- Piera, M.A., Buil, R. and Ginters, E. (2013). "Validation of Agent-Based Urban Policy Models by means of State Space Analysis", 8th EUROSIM Congress on Modelling and Simulation, pp. 403-408.
- Piera, M.A., Buil, R. and Mújica, M. (2014). "Specification Of CPN Models Into MAS Platform For The Modelling Of Social Policy Issues: FUPOL Project", Accepted for International Journal of Simulation and Process Modelling (IJSPM).
- Segal-Halevi, Erel, 2012, NetLogo Land-Random model.
- Wilensky, U., 1999, NetLogo, <http://ccl.northwestern.edu/netlogo/>, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL

## AUTHORS BIOGRAPHY

**Roman Buil** received the B.S. degree in Mathematics from Universitat Autònoma de Barcelona, Barcelona, Spain, in 2002, the M.S. degree in industrial engineering - advanced production techniques from Universitat Autònoma de Barcelona, in 2004. He is currently working towards the Ph.D. degree in industrial engineering - advanced production techniques at the same University. He is research scientist, assistant teacher and project manager at the Logistics and Aeronautical Unit of the Telecommunications and Systems Engineering Department of Universitat Autònoma de Barcelona. His research interests include modelling and simulation methodologies, optimization techniques, policy modelling, production planning and decision making for production planning and logistics. He is member of LogiSim, a recognized research group on Modelling and Simulation of Complex Systems and he has been involved in industrial projects working as consultant of different companies. He participates in FP7 FUPOL Project No.287119.

**Miquel Angel Piera Eroles** is the delegate for Technical Innovation Cluster, and director of LogiSim, a recognized research group on Modeling and Simulation of Complex Systems. He is Full time Professor in the System Engineering Department at Universitat Autònoma de Barcelona (UAB). Graduated with excellence from UAB in Computer Engineering (1988), Msc from University of Manchester Institute of Science and Technology in Control Engineering (1991), and he got his Phd in 1993. He is member of the Editorial board of 3 international journals and Editor in Chief of IJBRM. Dr Piera has been general chair and invited speaker of many International conferences. He has been nominated as Deputy Director of the UAB Engineering School and Coordinator of the Spanish CEA-IFAC research team on Modeling and Simulation. He has coordinated many research and industrial projects, he has also participated in some EC funded research and academic projects, such as: LOGIS



MOBILE LV/B/F/PP-172.001 (2004-2006), Curriculum Development on Logistics and Supply Chain Management. 134522-LLP-1-2007-1-ESERASMUS-ECDSP and FP7 FUPOL No.287119.

**Egils Ginters** is director of Socio-technical Systems Engineering Institute. He is full time Professor of Information Technologies in the Systems Modelling Department at the Vidzeme University of Applied Sciences. He is a Senior member of the Institute of Electrical and Electronics Engineers (IEEE), member of European Social Simulation Association (ESSA) and Latvian Simulation Society. He participated and/or coordinated some of EC funded research and academic projects: FP7 FUPOL project No. 287119 (2011-2014), FP7-ICT-2009-5 CHOREOS project No. 257178 (2010-2013), e-LOGMAR-M No.511285 (2004-2006), SocSimNet LV/B/F/PP-172.000 (2004-2006), LOGIS MOBILE LV/B/F/PP-172.001 (2004-2006), IST BALTPORTS-IT (2000-2003), LOGIS LV-PP-138.003 (2000-2002), European INCO Copernicus DAMAC-HP PL976012 (1998-2000), INCO Copernicus Project AMCAI 0312 (1994-1997). His main field of interests involves: systems simulation technologies, logistics information systems, and technology acceptance and sustainability assessment. He has more than 140 scientific articles related with the research fields.

**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.

# SINGLE SOURCING VS. DOUBLE SOURCING: A SIMULATION APPROACH TO SUPPLIER SELECTION

Mattia Armenzoni <sup>(a)</sup>, Marta Rinaldi <sup>(b)</sup>, Roberto Montanari <sup>(c)</sup>, Eleonora Bottani <sup>(d)</sup> and Federico Solari <sup>(e)</sup>

<sup>(a)</sup> Interdepartmental Centre SITEIA.PARMA, c/o Department of Industrial Engineering, University of Parma, Parco Area delle Scienze, 181/A – 43124 Parma (ITALY)

<sup>(b)</sup> Interdepartmental Centre CIPACK, c/o Department of Industrial Engineering, University of Parma, Parco Area delle Scienze, 181/A – 43124 Parma (ITALY)

<sup>(c), (d), (e)</sup> Department of Industrial Engineering, University of Parma, Parco Area delle Scienze 181/A - 43124 Parma (ITALY)

<sup>(a)</sup>[mattia.armenzoni@unipr.it](mailto:mattia.armenzoni@unipr.it), <sup>(b)</sup>[marta.rinaldi@unipr.it](mailto:marta.rinaldi@unipr.it), <sup>(c)</sup>[roberto.montanari@unipr.it](mailto:roberto.montanari@unipr.it), <sup>(d)</sup>[eleonora.bottani@unipr.it](mailto:eleonora.bottani@unipr.it), <sup>(e)</sup>[federico.solari@unipr.it](mailto:federico.solari@unipr.it)

## ABSTRACT

The paper presents a simulation approach to the supplier selection problem. Specifically, the aim of the model is to compare a scenario where a company exploits only one supplier (single sourcing) with a double sourcing option. In the case of double sourcing, the second supplier has a higher reliability compared to the first one, meaning that it is always able to deliver the product required within a defined lead time; however, products are supplied at a higher price, thus generating higher costs for the company. Under both scenarios, it is hypothesized that the company adopts an Economic Order Interval (EOI) reorder policy. Overall, the study is articulated into two steps and has the final aim to compare the single sourcing and double sourcing strategies, to assess the economic profitability of those solutions depending on the operating conditions of the company. Related results will provide companies with some economic benchmark for pondering purchasing options.

Keywords: supplier selection, single sourcing, multiple sourcing, simulation, economic analysis, benchmark.

## 1. INTRODUCTION

A supply chain is a network of suppliers, manufacturers, distributors and retailers, connected by transport infrastructures and information sharing mechanisms. Currently, supply chains include many potentially suppliers, serving one or more manufacturers and/or assemblers, who often use various distributors to send products to the different retailers. A supply chain must ensure an adequate service level to the customers, at the same time minimizing its total cost, including, among others, stock-out cost, cost of holding stocks, order, transport and purchasing.

Purchasing decisions play a main role both to ensure the quality of the final product and to control the

cost of procurement (Monczka et al., 2011). Traditionally, purchasing decisions have mainly been made based on cost considerations; currently, many other characteristics are considered for the selection of suppliers, including quality, reliability, flexibility and so on. Weber et al. (1991) review 47 articles, where more than one criterion was exploited for supplier selection. Many other authors (e.g., Roa and Kise 1980, Ellram 1990, Stamm and Golhar 1993, Bottani and Rizzi, 2005) propose detailed lists of criteria for the selection of suppliers.

Besides those selection criteria, some particular procurement strategies also require specific selection procedures for suppliers. For instance, just in time (JIT) strategies ground on the principles of improving quality of the final product, flexibility of the supply chain and service level delivered to customers. A practical way to achieve the objectives above is to decrease the lot size, thus reducing the amount of waste, the cost of inspections and quality checking. With the decrease of the lot size, the number of suppliers decreases too.

The most common procurement strategies are single sourcing, dual sourcing and multiple-sourcing (Yu et al., 2009). Single sourcing means that a company selects one single supplier, starting from a set of suppliers and evaluating them based on relevant selection criteria (Newman, 1989). The second strategy, i.e. the dual sourcing, refers to the case where the buyer uses two suppliers, with different characteristics in terms of market share, price, reliability, or other aspects (Tomlin and Wang, 2005). In the last strategy, i.e. the multiple-sourcing, the buyer works with several suppliers and encourages competition among them, to keep advantage of the better procurement conditions.

The strategic importance of selecting among single sourcing or dual sourcing has been widely debated in literature. A wide branch of studies discusses the advantages and disadvantages of adopting single

sourcing vs. double/multiple sourcing. For instance, Costantino and Pellegrino (2010) summarize the main pros and cons of multiple sourcing against single sourcing, as follows:

- Single sourcing allows establishing a long-term partnership between company and supplier. On the other hand, however, the company is more dependent on the supplier, and, therefore, the supply channel is more vulnerable;
- Multiple sourcing allows reducing the risk of unexpected interruption of furniture, and, at the same time, it increases competition among suppliers. Hence, lower price of the products can be obtained. However, the overall cost of the purchasing process could be higher than in the case of single sourcing, because of the need for managing more suppliers.

On the basis of the above considerations, many researchers have tried to define the optimal procurement strategy of the company, by identifying the optimal number of suppliers (e.g., Agrawal and Nahmias 1997; Zeng 2001; Burke et al. 2007). Others studies propose the evaluation of supply contracts as a function of the level of uncertainty where the company operates (Kamrad and Siddique 2001). Costantino and Pellegrino (2010) propose a Real Options approach for evaluating, through a probabilistic model, the benefits of multiple sourcing in managing the supplier default risk, compared to the single sourcing strategy. Sajadieh and Thorstenson (2014) investigate four different sourcing models, including sole sourcing and dual sourcing, in a two-echelon supply chain. The sourcing strategies are compared with respect to the total cost they generate, including inventory holding cost, backorder cost, order cost and setup costs for the vendors. The authors found that there is not a strategy that dominates the remaining ones; therefore, they conclude that the selection of the sourcing strategy needs a careful evaluation. A similar work is carried out by Chung et al. (2010). These authors study a decentralized supply chain composed of two suppliers and one buyer. One supplier has more flexibility in terms of quantity, while the second one offers lower purchasing price of items. The purchasing problem is faced from the buyer's perspective, to identify the optimal decision. Yu et al. (2009) evaluate the single sourcing vs. dual sourcing methods in presence of supply disruption risks. They examine a two-stage supply chain with a non-stationary and price-sensitive demand, and derive the expected profit functions of the two sourcing modes for the supply chain. Similarly, Sawik (2014) presents a stochastic mixed integer programming model to selecting suppliers in the presence of supply chain disruption risks, both in the case of single and dual sourcing strategies. Glock (2012) examines the case of a buyer who has to decide between single or dual sourcing for a homogeneous product. In the scenario examined, the production process of the suppliers is subject to learning effects, meaning that the production cost can

progressively decrease and the production capacities of the suppliers can increase correspondingly.

The brief discussion of the literature proposed above suggests that, in general, researchers compare single sourcing and dual sourcing policy with the purpose of identifying the optimal purchasing strategy, in terms of total cost. Another finding from the literature is that dual sourcing is a preferred strategy when there is the risk of interruptions of furniture.

Starting from the considerations above, in this paper we compare the single sourcing strategy and the dual sourcing one, in terms of their total cost. A simulation model is developed to reproduce the sourcing strategy of a buyer, under two different scenarios. For each scenario, the total cost of sourcing is evaluated, with the purpose of identifying the optimal purchasing strategy of the buyer.

The remainder of the paper is organized as follows. Section 2 describes the simulation model developed and the scenarios examined. Section 3 provides the results of the simulation runs. Section 4 discusses the main findings of the study, the related limitations and implications, and highlights future research directions.

## 2. MODELLING FRAMEWORK

### 2.1. The scenarios examined

As mentioned, a simulation approach has been chosen in this paper, to study the behavior of single sourcing and dual sourcing systems.

Two different scenarios have been modelled and evaluated. The first one reflects the single sourcing strategy, and considers only one supplier who is characterized by a defined level of reliability. Several simulation runs have been launched varying the degree of reliability, from 100% (i.e., the supplier always delivers the product on the right quantity and on time) to 80% (reflecting the percentage of times the supplier is able to satisfy the customer's request). Obviously, in general a more reliable seller is preferred. Nonetheless, a company could accept a lower level of reliability in the case this corresponds to a lower cost of purchasing. In this respect, simulation is a valuable tool to evaluate the trade-off between cost and reliability, and to assess the impact of reliability on the total cost of the system. By means of simulation, for each level of reliability (from 100% to 80%, step 1%), the optimal combination of the operating leverage of the reorder policy (i.e.,  $\Delta T$  and OUTL) is determined. By optimal, we mean the combination that minimizes the total cost of the system, as described in section 3.

The second scenario reflects the dual sourcing option; therefore, two different suppliers are considered. The first one has the same characteristics of the supplier described in the previous scenario. The second one, instead, is characterized by 100% reliability; however, compared to the first vendor, it sells its product at a higher price. For this reason, the buyer will place an order to the second vendor only when the first one is not able to deliver the right quantity of product in the required time. Again, several scenarios are considered,

with the reliability of the first vendor varying in a defined interval, as per the first scenario.

The aim of this work is to determine the optimal (in terms of total cost) level of reliability for the scenarios described. Moreover, the single sourcing and dual sourcing options are also compared, to identify the positive and negative aspects of both solutions.

## 2.2. Software implementation under MS Excel™

The simulation model consists of a MS Excel™ file which reproduces the flow of orders, under an EOI reorder policy. The simulation duration was set at 100,000 days.

The customer's demand is generated as a sequence of random numbers, whose statistical distribution is uniform between 0 and 200 items/day.

According to the EOI policy, the buyer places an order to the vendor at fixed periodic interval (designated as  $\Delta T$ ), regardless of the current inventory position. The amount of product ordered can vary, as it should allow raising the current stock to the order-up-to level (OUTL) threshold. This latter reflects the amount of stock to recover. Therefore, the reorder policy simulated is characterized by 2 operating leverages, such as  $\Delta T$  and OUTL.

During the simulation, the model was exploited to examine different settings of the operating leverages, with the purpose of identifying the combination of  $\Delta T$  and OUTL which generates the minimum total cost. To this extent,  $\Delta T$  was varied from 20 and 40 days while OUTL from 3000 to 4000 items.

The total cost, that the simulation model minimizes, includes the inventory holding cost ( $C_h$ ), the stock-out cost ( $C_{so}$ ) and order cost ( $C_o$ ). All contributions are expressed in [€/day], and their amount is derived as the average cost over the simulation duration.

The holding cost considers only one variable, i.e. the level of the inventory at the buyer's site. The stock out cost results from the number of items the buyer is not able to deliver to the customer (which is a direct outcome of the simulation). The economic loss for those items is assumed to be very high, accounting for 100 times the unitary cost of holding stocks. Such a value should highlight the need for avoiding out-of-stock situations. Finally, the order cost covers the administrative aspects that a company has to consider when it places an order (e.g., the cost of personnel). Its unitary amount is expressed as a fixed cost per order, i.e. [€/order]

On the basis of the description above, it can be easily deduced that, in scenario 1, the total cost  $C_{tot}$  is computed as:

$$C_{tot} = C_h + C_{so} + C_o \quad [€/day] \quad (1)$$

In scenario 2, two additional cost components are included in the total cost, i.e.:

- a difference in the purchasing cost ( $\Delta C_p$ ) between vendor 1 and 2. This contribution reflects the fact that vendor 2 delivers the product at a higher price;
- a fixed additional cost component, which consider the fixed cost of managing the order (and, more in general, the relationship) with the second supplier ( $\Delta C_m$ ). This cost contribution is computed each day.

Moreover, as already mentioned, in scenario 2 the order cost is different for seller 1 and 2. Overall, the total cost in scenario 2 can be calculated as:

$$C_{tot} = C_h + C_{so} + C_{o1} + C_{o2} + \Delta C_p + \Delta C_m \quad [€/day] \quad (2)$$

Table 1 shows the input data set in the simulation for the 2 different sellers. As already explained, scenario 1 considers only one supplier, with not perfect reliability (vendor 1). Conversely, scenario 2 includes 2 suppliers, i.e. vendor 1, with the same characteristics described above, and vendor 2, this latter being more expensive but perfectly reliable.

Table 1: Input data of the simulation

	Vendor 1	Vendor 2
Procurement lead time [days]	3	3
Cost of holding stock [€/unit/day]	0,001	0,001
Stock-out cost [€/unit/day]	0,1	0,1
Order cost [€/order]	50	100
$\Delta$ purchase cost [€/unit]	-	0,05
Seller management [€/unit]	-	1,8
Reliability	80-100%	100%

## 3. RESULTS

At the end of the simulation process, the results have been collected, to study the behavior of the system as a function of the different operating conditions. Results are discussed below for the two scenarios simulated.

### 3.1. Results under scenario 1

The optimal combination of  $\Delta T$  and OUTL (i.e. the combination which minimize the total cost), as a function of the reliability of the supplier, has been first investigated. Results are reported in Table 2.

Table 2: Scenario 1 – total cost,  $\Delta T$  and OUTL as a function of the supplier's reliability.

Reliability	Total cost	$\Delta T$	OUTL
99%	4.53	23	3050
98%	5.03	19	3125
97%	5.24	19	3475
96%	5.39	19	3650
95%	5.5	18	3500
94%	5.61	17	3575
93%	5.74	16	3375
92%	5.84	16	3550
91%	5.96	15	3400
90%	6.08	15	3625

89%	6.18	15	3675
88%	6.27	14	3550
87%	6.36	14	3775
86%	6.41	14	3900
85%	6.49	13	3700
84%	6.57	13	3775
83%	6.62	13	3850
82%	6.71	12	3900
81%	6.77	12	3700
80%	6.8	12	3957

The trend of  $\Delta T$  as a function of the reliability level for this scenario is shown in Figure 1.

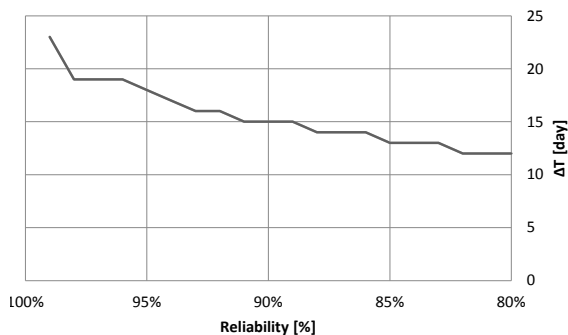


Figure 1: Scenario 1 – trend of  $\Delta T$  as a function of the supplier's reliability.

From Figure 1, it is easy to see that there is a significant decrease in the optimal  $\Delta T$  when the supplier's reliability decreases. This means that, if the supplier is less reliable, the company will increase the order frequency (i.e., will decrease the reorder interval), to protect itself from possible delays and avoid stock-out situations.

Conversely, the trend of the OUTL as a function of the supplier's reliability is opposite (Figure 2): this parameter increases when the reliability decreases. Indeed, a company should set a high safety stock level in order to face a potential lack of product ordered.

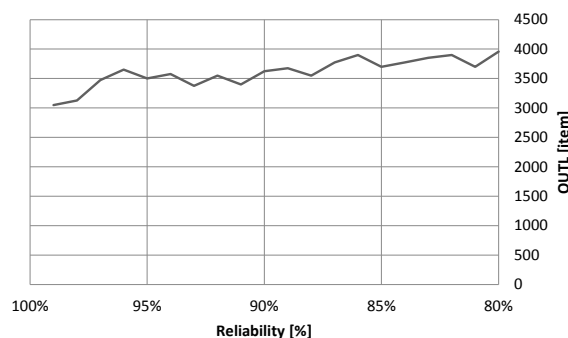


Figure 2: Scenario 1 – trend of OUTL as a function of the supplier's reliability.

With respect to the total cost of this scenario, it increases with the supplier's reliability, with a logarithmic trend (Figure 3). Indeed, if a supplier is 100% reliable, the company could work at optimal level of  $\Delta T$  and OUTL minimizing the total cost. Otherwise,

if the supplier's reliability decreases (meaning that it is not always able to deliver the product), the company should adapt the parameters of its reorder policy to the supplier's reliability, resulting in an increased total cost.

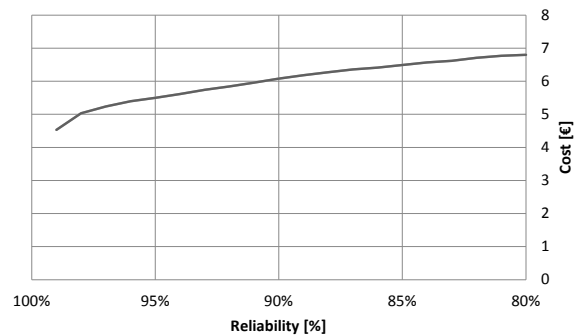


Figure 3: Scenario 1 – trend of the total cost as a function of the supplier's reliability.

### 3.2. Results under scenario 2

As per the previous case, the trend of the two main parameters was analyzed also in scenario 2. In this case, however, the reliability of the first supplier was varied from 80% to 100%, while the second vendor owns a perfect (100%) reliability.

The optimal combination of  $\Delta T$  and OUTL, as a function of the reliability of the first supplier, is reported in Table 3.

Table 3: Scenario 2 – total cost,  $\Delta T$  and OUTL as a function of the supplier's reliability.

Reliability	Total cost	$\Delta T$	OUTL
0.99	5.47	30	3550
0.98	5.54	30	3575
0.97	5.61	30	3525
0.96	5.67	30	3525
0.95	5.74	30	3575
0.94	5.8	31	3650
0.93	5.87	31	3675
0.92	5.93	31	3650
0.91	6.00	32	3775
0.90	6.07	32	3775
0.89	6.13	32	3700
0.88	6.19	32	3775
0.87	6.26	33	3875
0.86	6.32	33	3825
0.85	6.39	33	3825
0.84	6.46	33	3825
0.83	6.52	33	3875
0.82	6.59	33	3850
0.81	6.65	33	3875
0.80	6.71	34	3975

Compared to scenario 1, the trend of  $\Delta T$  as a function of the supplier's reliability is opposite (Figure 4). Specifically, the optimal  $\Delta T$  increases slightly, and the optimal value varies in a range of only 4 days. This result is justified by the fact that, if the first supplier is not able to deliver the product, the company will place

an order to the second one, whose order cost, however, is higher. For this reason, less frequent orders are preferable.

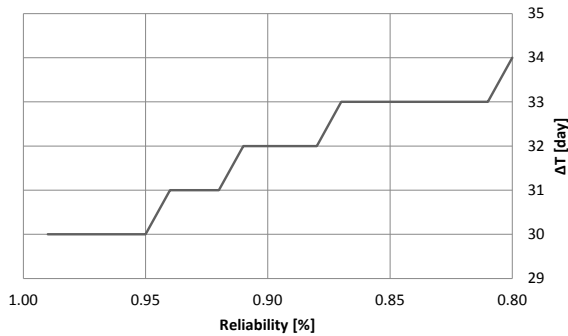


Figure 4: Scenario 2 – trend of  $\Delta T$  as a function of the reliability of the first supplier.

Conversely, from Figure 5 one can observe that the OUTL trend is the same as that of the previous scenario, even though the increase is less appreciable, as can be seen from the lower slope of the curve.

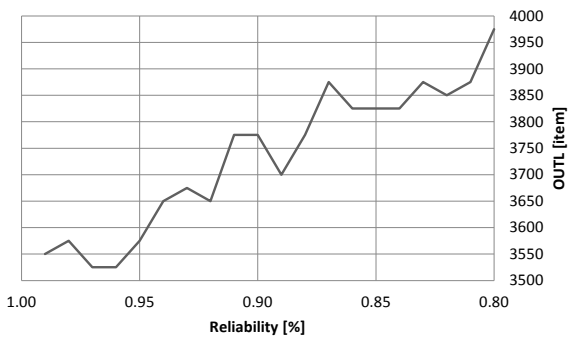


Figure 5: Scenario 2 – trend of OUTL as a function of the reliability of the first supplier.

As already mentioned, the total cost of scenario 2 includes more cost components than scenario 1. Nonetheless, even in this scenario, the cost increases if the reliability decreases, but with a different (linear) trend.

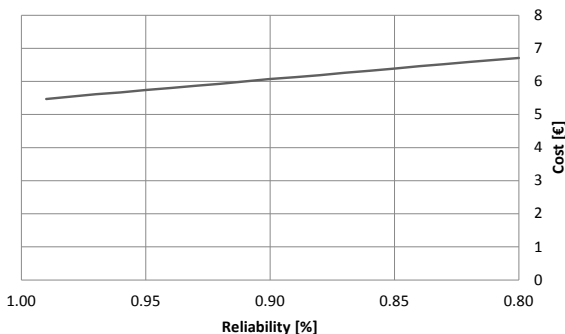


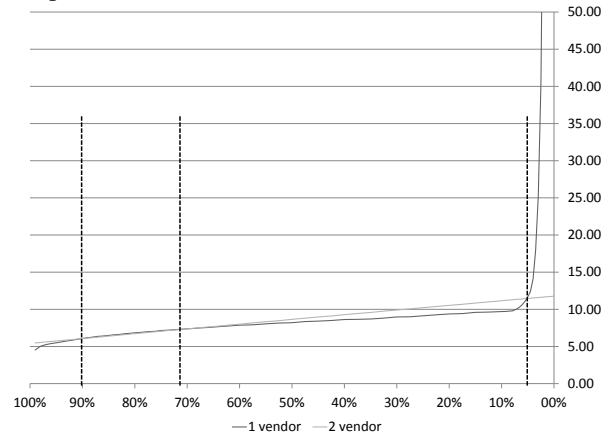
Figure 6: Scenario 2 – trend of the total cost as a function of the reliability of the first supplier.

### 3.3. Cost comparison

On the basis of the outcomes of the scenarios 1 and 2, a trade-off between cost and supplier's reliability can be

investigated, by varying the reliability of the first vendor between 0% and 100%. The corresponding outcomes are reported in Figure 7.

Figure 7: scenario 1 vs. scenario 2 - total cost comparison.



The comparison of cost shows that the economic profitability of single sourcing vs. dual sourcing chiefly depends on the reliability of the first vendor. Specifically:

- in the case vendor 1 has a high reliability (approx. ranging from 100% to 90%), there is actually no economic profitability in exploiting the dual sourcing strategy. Indeed, under that circumstance the stock-out experienced by the buyer is limited, and the related cost is not balanced by the additional cost of purchasing items from vendor 2;
- in the case the reliability of vendor 1 is lower (approx. from 90% to 70%), dual sourcing becomes the most profitable strategy. This suggests that, under this scenario, the additional cost of sourcing at vendor 2 is counterbalanced by the avoidance of out-of-stock situations, resulting in a lower total cost;
- if the reliability of vendor 1 decreases (from 70% to 7% approximately), dual sourcing is no longer profitable. Indeed, despite the low reliability, the optimal cost configuration is obtained with single sourcing at vendor 1. The rationale for this outcome, that could sound strange, is that, under this scenario, if multiple sourcing was allowed, the buyer would always place order to vendor 2, generating higher total cost;
- finally, in the case the reliability of vendor 1 is lower than 7%, the optimal sourcing strategy becomes single sourcing with vendor 2. This means that, with this low reliability, placing orders to vendor 1 would always generate out-of-stock situations. Therefore, the optimal situation is sourcing only at vendor 2, to avoid out-of-stock occurrence.



#### 4. DISCUSSION AND CONCLUSIONS

This paper has proposed a simulation approach to a well-known problem of supplier selection literature, i.e. the choice between single sourcing and dual sourcing. A simulation model has been built under MS Excel™ to reproduce the order process of a buyer, under two scenarios, i.e.:

- The single sourcing scenario, where the buyer can place orders only to one vendor;
- The dual sourcing scenario, where the buyer can exceptionally place orders to a second vendor. This latter has higher reliability compared to the first vendor, and therefore is always able to send products to the buyer, but at higher price.

The economic profitability of the two scenarios has been analyzed as a function of the reliability of the first supplier, as well as the trend of the reorder policy (i.e., EOI) parameters, namely the reorder interval and the order-up-to level.

Although the numerical outcomes of our study can be affected by the specific input data we set, the considerations related to the economic profitability of the two scenarios can be useful to a company wishing to ponder different purchasing decisions. This is an interesting practical contribution of this study.

From the theoretical perspective, we have mentioned that single sourcing and dual sourcing are quite debated topics of supply chain management literature. Nonetheless, to our knowledge, there are very few studies that propose a simulation approach to solve this problem. This paper, therefore, demonstrates the usefulness of a simulation approach to this kind of managerial decisions. Simulation is also useful to vary the problem settings (as done also in this paper), in terms, for instance, of the EOI parameters, the supplier's reliability or the economic parameters.

Starting from this paper, future research activities can be carried out to investigate in further detail the sensitivity of the results to the input data set. A formal design of experiments (DOE) procedure could be useful to this extent. Moreover, applying the approach developed to a real case example could be another suitable research to be undertaken in the future.

#### REFERENCES

- Agrawal, N., Nahmias, S., 1997. Rationalization of the supplier base in the presence of yield uncertainty. *Production and Operations Management*, 6(3), 291–308.
- Bottani, E., Rizzi, A., 2005. A fuzzy multi-attribute framework for supplier selection in an E-procurement environment. *International Journal of Logistics: Research and Applications*, 8(3), 249–266.
- Burke, G.J., Carrillo, J.E., Vakharia, A.J., 2007. Single versus multiple supplier sourcing strategies. *European Journal of Operational Research*, 182, 95–112.
- Chung, W., Talluri, S., Narasimhan, R., 2010. Flexibility or Cost Saving? Sourcing Decisions with Two Suppliers. *Decision Sciences*, 41(3), 623–650.
- Costantino, N., Pellegrino, R., 2010. Choosing between single and multiple sourcing based on supplier default risk: A real options approach. *Journal of Purchasing & Supply Management*, 16, 27–40.
- Ellram, L., 1990. The supplier selection decision in strategic partnerships. *Journal of Purchasing Material Management*, 26(4), 8–14.
- Glock, C.H., 2012. Single sourcing versus dual sourcing under conditions of learning. *Computers & Industrial Engineering*, 62, 318–328.
- Kamrad, B., Siddique, A., 2001. Risk sharing and supplier switching contracts. *Proceedings of the Fifth Annual International Conference on Real Options*, UCLA, Los Angeles.
- Monczka, R.M., Handfield, R., Gunipiero, L.C., Patterson, J.L., 2011. *Purchasing and Supply Chain Management (5<sup>th</sup> edition)*. South-Western, Cengage Learning.
- Newman, R.G., 1989. Single Sourcing: Short-term savings versus long-term problems. *Journal of Purchasing and Materials Management*, 20–5.
- Roa, C.P., Kiser, G.E., 1980. Educational buyer's perception of vendor attributes. *Journal of Purchasing Material Management*, 16, 25–30.
- Sajadieh, M.S., Thorstenson, A., 2014. Comparing sourcing strategies in two-echelon supply chains. *Computers & Operations Research*, 45, 108–115.
- Sawik, T., 2014. Joint supplier selection and scheduling of customer orders under disruption risks: Single vs. dual sourcing. *Omega*, 43, 83–95.
- Stamm, C.L., Golhar, D.Y., 1993. JIT purchasing attribute classification and literature review. *Production Planning & Control*, 4(3), 273–282.
- Tomlin, B., Wang, Y., 2005. On the value of mix flexibility and dual sourcing in unreliable newsvendor networks. *Manufacturing & Service Operations Management*, 7(1), 37–57.
- Weber, C.A., Current, J.R., Benton, W.C., 1991. Vendor selection criteria and methods. *European Journal of Operational Research*, 50(1), 2–18.
- Yu, H., Zeng, A.Z., Zhao, L., 2009. Single or dual sourcing: decision-making in the presence of supply chain disruption risks. *Omega*, 37, 788–800.
- Zeng, A.Z., 2001. Single or multiple sourcing: an integrated optimization framework for sustaining time-based competitiveness. *Journal of Marketing*, 6(4), 10–25.

#### AUTHORS' BIOGRAPHY

**Mattia Armenzoni** is scholarship holder in Industrial Engineering at the Interdepartmental Center SITEIA.PARMA of the University of Parma (Italy). He got a Master degree in Mechanical Engineering of the Food Industry at the same University, discussing a thesis related to the design of a static dryer for pasta with simulation tools. He attended numerous conferences on the topics of food processing, food

plants, modelling and simulation, and published several papers on the same topics in international conferences.

**Marta Rinaldi** is Ph.D. student at the University of Parma (Italy) and scholarship holder in Industrial Engineering at the Interdepartmental Center for Packaging (CIPACK). She graduated (with distinction) in Industrial Engineering and Management in 2011, with a thesis titled 'Analysis and evaluation of energy efficiency of a shrinkwrap-packer'. She is author of approx. 15 papers in the field of logistics, supply chain management, inventory management, discrete event simulation and its application to industrial plants and manufacturing systems.

**Roberto Montanari** graduated (with distinction) in 1999 in Mechanical Engineering at the University of Parma (Italy), where he is currently employed as Full Professor of Industrial Plants. His research activities concern equipment maintenance, power plants, food plants, logistics, supply chain management, supply chain modelling and simulation, inventory management. He has published his research papers in several qualified international journals, as well as in national and international conferences. He acts as a referee for numerous scientific journals and is editorial board member of two international scientific journals.

**Eleonora Bottani** graduated (with distinction) in 2002 in Industrial Engineering and Management at the University of Parma (Italy), where she is Lecturer (with tenure) in Mechanical Industrial Plants since 2005. In March 2006, she got her PhD in Industrial Engineering at the same university. She authored or co-authored more than 100 papers in the fields of logistics, supply chain management, supply chain modelling, performance analysis, industrial plants safety and optimisation. She acts as a referee for about 60 international scientific journals and for several international conferences. She also owns editorial responsibilities for several international journals.

**Federico Solari** is research fellow at the Department of Industrial Engineering of the University of Parma (Italy). He got a Master degree in Mechanical Engineering of the Food Industry at the same university, discussing a thesis related to the design of a plant for the extraction of volatile compounds. He received his Ph.D. with a thesis entitled "Advanced approach to the design of industrial plants by means of computational fluid dynamics". He attended several conferences related to food process and plants, and published several papers on the same topics in international conferences.

# A CONJECTURE FROM LEARNING SIMULATIONS OF SERIES AND PARALLEL CONNECTIONS OF COMPONENTS

Alexandre Muzy<sup>(a)</sup>, Bernard P. Zeigler<sup>(b)</sup>

<sup>(a)</sup>IBS UMR CNRS 7271, Bio-info, CS 40121 - 06903 Sophia-Antipolis Cedex, France, Email: alexandre.muzy@cnrs.fr.

<sup>(b)</sup>RTSync Corp. and Arizona Center for Integrative Modeling and Simulation, AZ, United-States of America, Email :zeigler@rtsync.com

## ABSTRACT

Abstract sequential machines can be realized by series and parallel connections of components. These connections can be automatically learned by simulation, rating and selecting component models. Our goal here is threefold: (i) discuss implications for model reconstruction and reconfiguration in model engineering, (ii) present mathematical analyses of search simulation results, and (iii) conjecture that the whole search space of achievement components seems to be well ordered. This conjecture opens new perspectives for studying parallel and series connections of components upon a simulation-based framework. More generally, the automatic learning demonstrated here constitutes a formal basis for model reconstruction and reconfiguration in model engineering.

Keywords: Sequential machines, activity, discrete-events, simulation, learning.

## 1 INTRODUCTION

*Linear sequential machines* can be decomposed into parallel and series connections [1, 2]. In sequential machine theory [3], logical and functional functions in machines are studied describing possible realizations independently from the implementation of physical components.

On the other hand, *activity* can be used to link the energy of physical components to information [4] and activity-based credit assignment (ACA) can be used for automatically rating component systems and composing them according to the experimental frames in which they are placed [5, 6]. This includes both arbitrarily connected coupled models, i.e., having series, parallel, or feedback loop couplings in any combination. Here we focus on series and parallel coupled models where mathematical analysis of the ACA's behavior is possible, as we shall show.

From a systems science perspective, *systems problem solving* [7] and *system design and engineering* [8] are the approaches closest to the approach presented here. However, these approaches remain at modeling level and do not use activity at component level for building systems at network level. Concerning other credit assignment methods, a comparison is presented in [5].

This manuscript aims at using simulation for inductively modeling the automatic search process of parallel and series connections. The difficulty in such approaches is to find a metrics being generic enough to allow mathematical analysis. The metrics presented here takes advantage of both activity measure and connections topology. This metrics proved to be consistent in both modeling (mathematical) and simulation (computational) worlds. This allows, for parallel and series connections, to induce a new conjecture. As far as we know, it is the first attempt of a method for the mathematical analysis of the learning simulation of basic structures (parallel and series) of components, at system level.

The manuscript is organized as follows. Section 2 presents the mathematical structures and activity-based metrics used. Section 3 situates the approach presented in the context of model engineering. Section 4 describes the simulation and mathematical results obtained for both parallel and series connections. Finally, a conclusion presents the new conjecture and draws the lines of the corresponding formal perspectives.

## 2 TOOLS

Discrete event system specification allows specifying locuses of activity (components) [9] and thus storing the amount of activity of components. The activity (efficiency) of components can be combined with the behavior

(performance) of compositions to find target solutions.

## 2.1 Discrete Event System Specification (DEVS)

The structure of both network and basic discrete event systems is presented here.

**Definition 1.** A basic Discrete Event System Specification (DEVS) is a structure:

$$DEVS = (X, Y, S, \delta_{ext}, \delta_{int}, \lambda, ta)$$

Where,  $X$  is the set of input events,  $Y$  is the set of output events,  $S$  is the set of partial states,  $\delta_{ext} : Q \times X \rightarrow S$  is the external transition function with  $Q = \{(s, e) \mid s \in S, 0 \leq e \leq ta(s)\}$  the set of total states,  $\delta_{int} : S \rightarrow S$  is the internal transition function,  $\lambda : S \rightarrow Y$  is the output function, and  $ta : S \rightarrow \mathbb{R}^{0,+}$  is the time advance function.

**Definition 2.** A DEVS network is a structure:

$$N = (X, Y, D, \{M_d\}, \{I_d\}, \{Z_{i,d}\}, Select)$$

Where  $X$  is the set of input events,  $Y$  is the set of output events,  $D$  is the set of component names, for each  $d \in D$ ,  $M_d$  is a basic model (whose structure differs from one DEVS-based formalism to another), for each  $d \in D \cup \{N\}$ ,  $I_d$  is the set of influencers of  $d$  such that  $I_d \subseteq D \cup \{N\}$ ,  $d \notin I_d$  and for each  $i \in I_d$ :  $Z_{i,d}$  is a coupling function, the  $i$ - to  $-d$  output translation, defined for: (i) external input couplings:  $Z_{self,d} : X_{self} \rightarrow X_d$ , with  $self$  the network name, (ii) internal couplings:  $Z_{i,j} : Y_i \rightarrow X_j$ , and (iii) external output couplings:  $Z_{d,self} : Y_d \rightarrow Y_{self}$ , and  $Select : 2^D - \{\emptyset\} \rightarrow D \cup \{\emptyset\}$  is the sequential select function (to select one component to execute its transition/output functions, among imminent components). Considering a set of components  $C$  candidate for internal transition, the sequential select function has constraint  $Select(C) \in C \cup \{\emptyset\}$ , i.e., only one component or no components can be selected among candidates.

## 2.2 Activity credit assignment (ACA)

**Definition 3.** Event-based activity  $A_\xi(t' - t)$  [9] in an event set  $\xi$  consists of:

$$A_\xi(t' - t) = |\{ev_i = (t_i, v_i) \in \xi \mid t \leq t_i < t'\}|$$

Where  $t, t_i, t'$  are time-stamps and  $v_i \in V$  is an event value.

Average event-based activity consists then of  $\overline{A_\xi(t' - t)} = \frac{A_\xi(t' - t)}{t' - t}$ .

For example, assuming the event trajectory depicted in Figure 1, the average event-based activity of the system corresponds to the following values for different time periods:  $\overline{A_\xi(10)} = 0.3$ ,  $\overline{A_\xi(20)} = 0.15$ ,  $\overline{A_\xi(30)} \simeq 0.133$ ,  $\overline{A_\xi(40)} = 0.175$ .

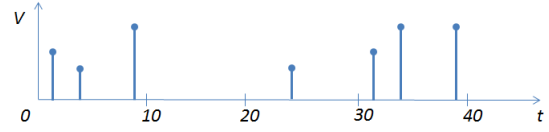


Figure 1: An example of event trajectory.

**Definition 4.** Average external activity  $\overline{A_{ext}}$ , related to the counting  $n_{ext}$  of external transitions  $\delta_{ext}(s, e, x)$ , over a time period  $[t, t']$  consists of:

$$\begin{cases} s' \leftarrow \delta_{ext}(s, e, x) \Rightarrow n'_{ext} \leftarrow n_{ext} + 1 \\ \overline{A_{ext}(t' - t)} = \frac{n_{ext}}{t' - t} \end{cases}$$

**Definition 5.** Average internal activity  $\overline{A_{int}}$ , related to the counting  $n_{int}$  of internal transitions  $\delta_{int}(s)$ , over a time period  $[t, t']$  consists of:

$$\begin{cases} s' \leftarrow \delta_{int}(s, e) \Rightarrow n'_{int} \leftarrow n_{int} + 1 \\ \overline{A_{int}(t' - t)} = \frac{n_{int}}{t' - t} \end{cases}$$

**Definition 6.** Total average simulation activity  $\overline{A_s(t' - t)}$  is equal to:

$$\overline{A_s(t' - t)} = \overline{A_{ext}(t' - t)} + \overline{A_{int}(t' - t)}$$

$e_v : \Omega \times P \times P^* \rightarrow \mathbb{R}$  is the evaluation function, where  $\rho^* \in P^*$  is the target output segment (a partial function  $\rho^* : [t, t'] \rightarrow Y^*$ ) of the target network name (noted target network for short after)  $k^* \in K$ ,  $\rho \in P$  is the output segment (a partial function  $\rho : [t, t'] \rightarrow Y$ ) of a candidate network  $k \in K$ , and  $\omega \in \Omega$  is the input segment (a partial function  $\omega : [t, t'] \rightarrow X$ ) of both networks. For each input segment  $\omega \in \Omega$ , the evaluation function  $e_v$  computes the distance between each output segment  $\rho \in P$  (obtained by simulation) and the target output segment  $\rho^* \in P^*$ . Evaluation function compares simulation results between a candidate network  $k \in K$  and a target network  $k^* \in K$ , i.e.,  $e_v(k, k^*)$  (noted  $e_v(k)$  for short hereafter).

**Definition 7.** A local optimum  $\hat{k} \in K$  consists of  $e_v(\hat{k}) < e_v(k^*)$ , where  $k^* \in K$  is the global optimum (target network) and  $e_v(k^*) = v^*$  with  $v^*$  the maximum evaluation value.

**Definition 8.** At trial  $0 < r \leq R$  (with  $r \in \mathbb{N}$ ), the simulation credit (achievement) of each component  $i \in D$ , over a simulation duration  $[t, t']$ , consists of  $c_{i,r}(t' - t) = \overline{A_s(t' - t)} \times e_v(k)$ , with  $k \in K$ .

**Definition 9.** Over a number of trials  $R \in \mathbb{N}$ , the *accumulated simulation credit* of a component  $i \in D$  is the sum of accumulated credits at each *trial*  $0 < r \leq R$ :  $c_{i,R}(t' - t) = \sum_{r=1}^R c_r(t' - t)$ .

**Definition 10.** The *accumulated simulation credit ratio* of a component  $i \in D$  consists of  $\hat{c}_{s,i} = \frac{C_{s,i}}{C_{s,n}}$  where  $n$  is the index of the *last component* of a sequence of components  $(DEVS_i)_{i=1}^n$  with  $i \in D$  and  $D \subset \mathbb{N}$ .

**Definition 11.** The *accumulated predicted credit* of a component  $i \in D$  consists of  $c_{p,i} = \sum_{i \in D} P(i)ev(k)$ , with  $P(i)$  the *probability of activation of component*  $i \in D$  and  $k \in K$  a candidate network.

**Definition 12.** The *accumulated predicted credit ratio* of a component  $i \in D$  consists of  $\hat{c}_{p,i} = \frac{c_{p,i}}{c_{p,n}}$ .

### 3 MODEL ENGINEERING

In [10], authors state that model reconstruction and re-configuration is one of the important problems in model engineering: “According to the performance requirement change of model function caused by the diversity of demand and the environmental uncertainty, the model needs to be reconstructed and configured quickly. The model reconstruction is to adjust its internal structure without changing the main external features, in order to optimize the performance and make it easier to understand, maintain and transplant. The model configuration is to adjust and optimize the internal components and parameters without changing the basic structure of the model, so as to adapt the different requirements or changes in function and performance of the model. For the model engineering of the complex system, the model reconstruction and configuration management is a very important and challenging research.”

Muzy and Zeigler [5] developed a framework for automatically rating component models and composing them according to the experimental frames in which they are placed. Components are assigned credit by correlating measures of their participation (activity) in simulation runs with run outcomes. These ratings are employed to bias component selection in subsequent compositions. Figure 2 sketches the position of the new decision and search process layers on the top of usual modeling and simulation processes. Model construction is biased via the synthesis from achievement components stored in repositories (or model-base).

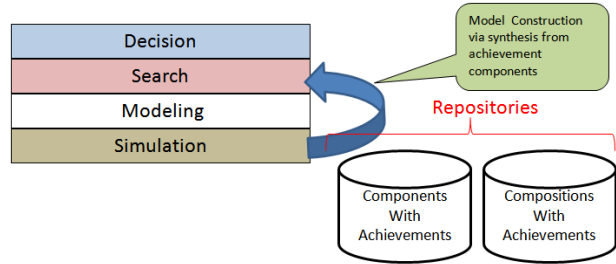


Figure 2: Decision and search in modeling and simulation.

Figure 3 presents the basic modeling and simulation entities used in activity-based credit assignment. A correlator entity is in charge of correlating performance evaluations obtained in an experimental frame and simulation results for credit assignment at component-based level.

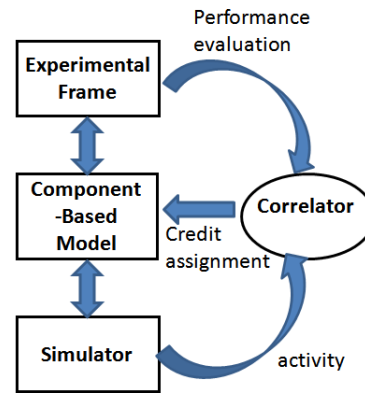


Figure 3: Basic modeling and simulation entities in activity-based credit assignment.

This activity-based approach to model construction also applies to model reconstruction and reconfiguration as defined by [10]. This is because the experimental frame represents the requirements of the model and when the latter changes, so should the frame representing them. A change in frame can then automatically induce a new search for composition of components to best meet the new requirements.

### 4 SERIES AND PARALLEL CONNECTIONS

In this set-up, the requirement for ACA to demonstrate is that it must converge to a stable assignment of credits to all steps where each right step has a credit that exceeds that of the corresponding wrong step. In this way, a composition strategy that selects from high valued component alternatives will select the correct composition. Assuming uniform independent random selection of alternatives.

We now proceed to show that this requirement holds using a stochastic analysis approach. We also verify the probability estimates obtained using direct simulation of the ACA as described in [6].

The *main assumptions* of the simulations are:

- At a given trial, for a right component, to be activated, all the precursors that are right have to be activated.
- At each step, a uniform random selection of components is achieved.
- Evaluations are deterministic. Each composition is evaluated once.
- Components contribute additively to the total outcome measure.
- Each wrong component has the same activity, when activated.
- Each right component has the same activity, when activated.

#### 4.1 Series connections

Figure 4 represents the component-based implementation of a series connection, for four right steps. A *skill* is a learned sequence of actions. The problem here is to learn a skill that requires a sequence of actions. Assume that the set of base slots for actions and their coupling is known, the actions exist as alternative components for selection – the right component has to be selected for each slot. The sequence of activations is assumed to be known, and the actions exist as alternative components for selection – the right component has to be selected for each step. For simplicity, there are two actions, right and wrong, for each step. For example, “move an object from one place to another” consists of *1<sup>st</sup> step*: “move grabber to right place” vs. “move to another place”; *2<sup>nd</sup> step*: “lift the object” vs. “drop the object”; *3<sup>rd</sup> step*: “move the grabber to the target location”, etc. Each step must be right before learning the next. For example, a skill has 4 steps. A coupled model is shown in Figure 4 that represents the components and couplings, where each component can be either *right* or *wrong*. Step 1 starts and if it is right, triggers step 2; if step 1 is wrong, then the trial is over and a new candidate is generated. If step 2 is right, it triggers step 3, etc.

Each right step sends an output of value 1 to be summed in the *Sum* component (*experimental frame*). Maximum score is then 4 if all steps are completed. A partial score of  $i$  is thus obtained for  $i$  steps completed

Figure 5 describes the accumulated predicted credits, the score, and the activation probability of right step components. Each right step accumulates its own score plus the sum of the *downstream* right step scores (since it participates in them). For example, for 2 right steps, with  $\frac{1}{8}$  probability, the activation ends after right step 2, and the score is then 2.

Table 1 presents both predicted and simulation accumulated credits of each right step component. Looking at the table, it can be seen that for long enough runs (including many trials), for each right step component  $i \in D$ , the accumulated predicted credit ratio is equal to the accumulated simulation credit ratio  $\hat{c}_{p,i} = \hat{c}_{s,i}$ .

Right step	$\hat{c}_{p,i}$	$c_{s,i}$	$\hat{c}_{s,i}$
1	3.75	0.30	3.75
2	2.75	0.22	2.75
3	1.75	0.14	1.75
4	1	0.08	1

Table 1: Analysis of accumulated credit for each right step.

Figure 6 describes the *accumulated credits*, the *score*, and the *activation probability* of *wrong step* components. Each wrong step accumulates the previous step score. For example, for wrong step 3, with  $\frac{1}{8}$  probability, the activation ends after *right step 2*, and the *score* is then 2.

As for right steps, Table 2 represents the *predicted* and *simulated accumulated credits* of each *wrong step* component. This table tells us that for each wrong step component  $i \in D$ , the accumulated predicted credit ratio is equal to the accumulated simulation credit ratio  $\hat{c}_{p,i} = \hat{c}_{s,i}$ . Also, when comparing Tables 1 and 2, it can be observed that the wrong step component with the highest accumulated credit  $c_{s,i}$  still gets a lower accumulated credit than the right step component with the lowest accumulated credit. This should ensure a good ranking of all the components of the candidate set.

Wrong step	$\hat{c}_{p,i}$	$c_{s,i}$	$\hat{c}_{s,i}$
1	0	0	0
2	1.33	0.04	1.33
3	1.33	0.04	1.33
4	1	0.03	1

Table 2: Analysis of accumulated credit for each wrong step, with  $\hat{c}_{s,i} = \hat{c}_{p,i}$ .

The stochastic analysis generalizes to a series coupling of any length. For convenience let’s consider such a connection with infinite length. Then at any step number,  $N$ , a wrong step gets credit for the previous  $N - 1$  steps and



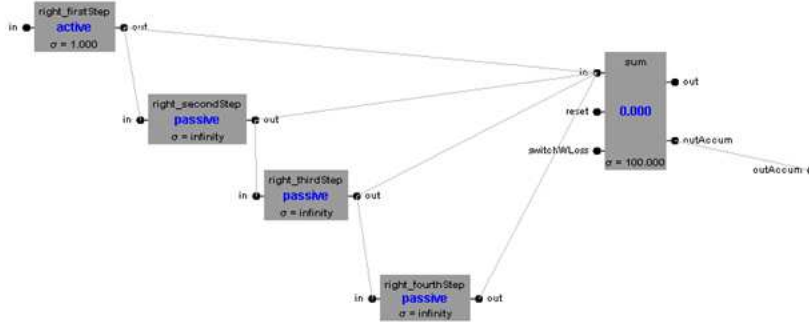


Figure 4: Series connection. *Step 1* initiates action – it needs to be right to activate *step 2* and to add 1 to the sum. *Step 2* continues action – it needs to be right to activate *step 3* and to add 1 to the sum – and so on.

this occurs with probability  $P_F(N)$  given by a geometric (distribution) first failure process. A right step receives this credit plus credit for any subsequent right steps until a wrong step, for which the probability distribution is a geometric first failure process. Thus each right step has a credit that exceeds that of the corresponding wrong step by an expected amount given by the mean of the geometric first failure distribution.

## 4.2 Parallel connections

Here the simulation consists of learning a skill that requires concurrent execution of actions. Assume that the set of base slots for actions is known, and the actions exist as alternative components for selection – the right component has to be selected for each slot. For simplicity, assume there are two actions, right and wrong, for each slot. For example: Move right arm forward and left leg backward simultaneously. You must get each slot right to get the whole skill right, but you get partial reward for getting one slot right while the other is wrong.

Figure 7 represents the component-based implementation of the parallel case, for four right steps. The skill has 4 slots. A coupled model is shown next that represents the components and coupling, where each component can be either right or wrong. Step 1 is set to be right. When it starts triggering the other slots. Each right slot sends an output to the Sum (experimental frame).

In such a parallel composition, with every component uniformly and independently randomly selected as either right or wrong, a Bernoulli distribution determines the number of right and wrong selected components at each trial. Focusing on one component, the distribution for other components is the same whether it is selected as right or wrong. The only difference being that a right

selection is active and therefore receives one more unit of credit than the corresponding wrong alternative.

Table 3 represents the *accumulated predicted credit* of each *right step*  $r \in D$ ,  $c_{p,r}$ , the *accumulated predicted credits* of each *wrong step*  $w \in D$ ,  $c_{p,w}$ , and the ratios of total accumulated credits of right slots to wrong slots defined as predicted  $\tilde{c}_p = \frac{\sum_{r \in D} c_{p,r}}{\sum_{w \in D} c_{p,w}}$  and simulated  $\tilde{c}_s = \frac{\sum_{r \in D} c_{s,r}}{\sum_{w \in D} c_{s,w}}$ .

# right	$P(i)$	$c_{p,w}$	$c_{p,r}$
0	.25	1	2
1	.5	1.5	2.5
2	.25	2	3
Total		4.5	7.5
$\tilde{c}_p$			7.5/4.5=1.6
$\tilde{c}_s$			12/7.0=1.7

Table 3: Analysis of accumulated credits for each right and wrong step.

Table 3 suggests that both predicted and simulated ratios of total accumulated credits of right slots to wrong slots are approximately equal  $\tilde{c}_p \simeq \tilde{c}_s$ . This should ensure a good ranking of right and wrong components for simulation-based learning of target parallel connections.

## 5 CONCLUSION AND PERSPECTIVES

In the context of sequential machines [3], for series and parallel connections, accumulated credit seems to be the right metrics of the set of candidate network solutions.

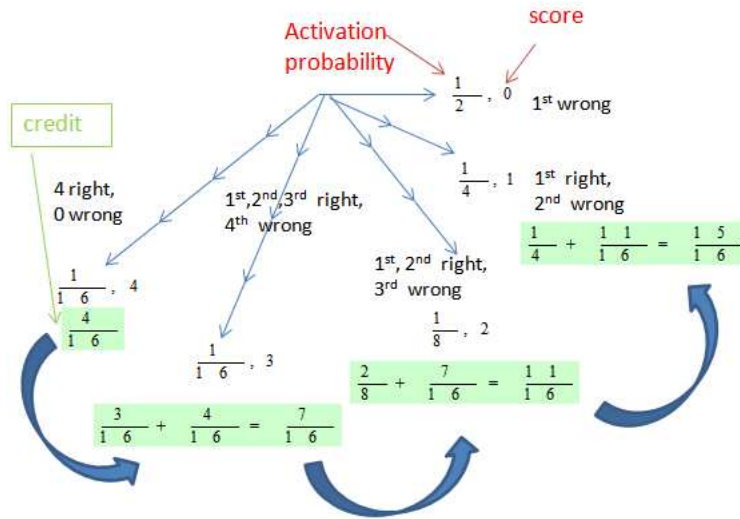


Figure 5: Analysis of accumulated credit for each right step – component in the target set.

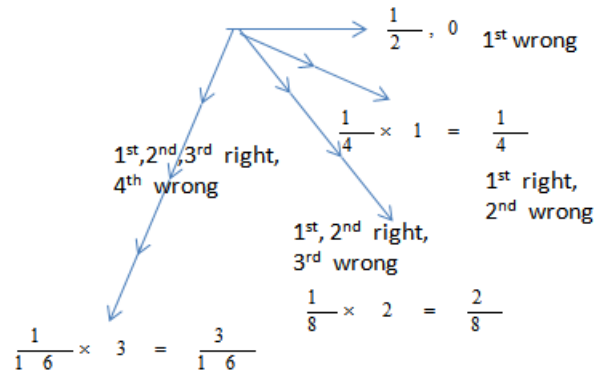


Figure 6: Analysis of accumulated credit for each wrong step – component not in the target set.

Based on the results obtained here, the following conjecture can be derived:

**Conjecture 1.** Consider any composition whether series, parallel, or feedback in any combination. Let each component have right and wrong alternatives and let the outcome be measured by the number of right selections. If the alternatives are selected uniformly and independently at random in every trial, then the expected credit assigned by ACA to each right alternative always exceeds the expected credit assigned to the corresponding wrong alternative.

Once this conjecture has been proved, i.e., that connection candidate sets are well ordered according to their achievements, it would be possible to correlate learning times (to find components) to the amount of activity required. It should be that the more activity invested in

the learning process the smaller the learning time. More generally, the automatic learning of system structures, as demonstrated here through parallel and series connections, constitutes a formal basis for model reconstruction and reconfiguration in model engineering as defined by references [10, 5]. In [5], ACA proved to be more efficient than a random model-base approach (selecting randomly components in a repository), for a model embedding both parallel and series chains of components and possibly feed-back loops. Testing the conjecture obtained here in the framework presented in [5], constitutes a challenging perspective for ACA mathematical analysis of more complex systems.

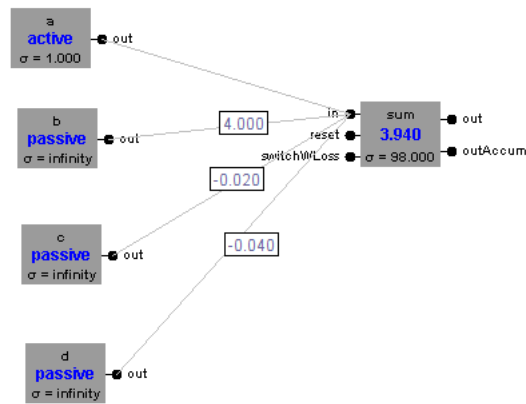


Figure 7: Parallel connections.

## REFERENCES

- [1] Hervé Gallaire and Michael A. Harrison. Decomposition of linear sequential machines. *Mathematical Systems Theory*, 3(3):246–287, 1969.
- [2] Hervé Gallaire. Decomposition of linear sequential machines. ii. *Mathematical Systems Theory*, 4(2):168–190, 1970.
- [3] Juris Hartmanis. *Algebraic Structure Theory of Sequential Machines (Prentice-Hall International Series in Applied Mathematics)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1966.
- [4] Xiaolin Hu and Bernard P. Zeigler. Linking information and energy - activity-based energy-aware information processing. *Simulation*, 89(4):435–450, 2013.
- [5] Alexandre Muzy and Bernard P. Zeigler. Activity-based credit assignment (aca) heuristic for simulation-based stochastic search in a hierarchical model-base of systems. *Accepted for publication in IEEE Systems Journal*, 2014.
- [6] Alexandre Muzy and Bernard P. Zeigler. Activity-based credit assignment (aca) in hierarchical simulation. In *SpringSim (TMS-DEVS)*, page 5, 2012.
- [7] George J. Klir and Doug Elias. *Architecture of Systems Problem Solving*. Springer, 2003.
- [8] A.W. Wymore. *Model-Based Systems Engineering*. Systems Engineering. Taylor & Francis, 1993.
- [9] Alexandre Muzy, Luc Touraille, Hans Vangheluwe, Olivier Michel, Mamadou Kaba Traoré, and David R. C. Hill. Activity regions for the specification of discrete event systems. In *Spring Simulation Multi-Conference Symposium On Theory of Modeling and Simulation (DEVS)*, pages 176–182, 2010.
- [10] Lin Zhang, Yuewei Shen, Xuesong Zhang, Xiao Song, Fei Tao, and Ying Liu. The model engineering for complex system simulation. In *26th European Modeling and Simulation Symposium (EMSS) - Model Engineering Workshop, accepted for publication*, 2014.

# SKOPJE BICYCLE INTER-MODALITY SIMULATOR – E-INVOLVEMENT THROUGH SIMULATION AND TICKETING

Egils Ginters<sup>(a)</sup>, Artis Aizstrauts<sup>(b)</sup>, Girts Dreija<sup>(c)</sup>, Maija Ablazevica<sup>(d)</sup>, Sergey Stepucev<sup>(e)</sup>, Inita Sakne<sup>(f)</sup>, Mikelis Baltruks<sup>(g)</sup>, Miquel Angel Piera Eroles<sup>(h)</sup>, Roman Buil<sup>(i)</sup>, Marjan Gusev<sup>(j)</sup>, Goran Velkoski<sup>(k)</sup>

<sup>(a,b,c,d,e,f,g)</sup>Sociotechnical Systems Engineering Institute of Vidzeme University of Applied Sciences, Cesu street 4, Valmiera LV-4201, Latvia

<sup>(h,i)</sup>Dept. de Telecomunicació i Enginyeria de Sistemes, Universitat Autònoma de Barcelona, Carrer dels Emprius, 2 08202 Sabadell-Barcelona, Spain

<sup>(j,k)</sup>Faculty of Information Sciences & Computer Engineering, University Sts Cyril and Methodius, Arhimedova bb, p. o. box 162, 1000, Skopje, Macedonia

<sup>(a)</sup>egils.ginters@va.lv, <sup>(b)</sup>artis.aizstrauts@va.lv, <sup>(c)</sup>girts.dreija@va.lv, <sup>(d)</sup>maija.ablazevica@va.lv, <sup>(e)</sup>sergejs.stepucevs@va.lv, <sup>(f)</sup>inita.sakne@va.lv, <sup>(g)</sup>mikelis.baltruks@va.lv, <sup>(h)</sup>[miquelangel.piera@uab.cat](mailto:miquelangel.piera@uab.cat), <sup>(i)</sup>roman.buil@uab.cat, <sup>(j,k)</sup>marjangusev@gmail.com

## ABSTRACT

For successful crafting political decisions it would be reasonable to verify them theoretically through modelling of potential effect. Therefore, the set of adequate models must be designed demonstrating reaction on policy maker activities. Also is necessary to understand the reaction of potential audience and the new policy accepted by citizens.

The article dealt with simulation and ticketing use in policy decision making based on Skopje Bicycle Inter-modality Simulation case.

Keywords: simulation, e-involvement, ticketing, agent-based models

## 1. INTRODUCTION

Despite of it that many well established policy decision making methodologies exist IT has not played a significant role in most of these efforts until now (Osimo et. al. 2010). This is especially true for urban policies, which are important on a worldwide scale since the majority of the world's population is living in urban areas. In Europe currently more than 79 % of the population is urbanized and it is expected that it will further increase and reach 85 % by 2030. It means the vast majority of the population in Europe is affected by urban policies in their daily lives.

The Major Cities of Europe Association (MCE) has conducted a study on the "Citizen Web Empowerment in a network of European Municipalities: "value for citizens" in Web 2.0 projects" (Buccoliero and Bellio 2010). It studies the growing demand of citizen empowerment and benchmarks the degree of citizen empowerment across the network of European municipalities in four areas e-information, Web 2.0, e-consultation and e-decision. The major outcome is that e-information (74 of 100) is sufficiently addressed, while Web 2.0 (23.2 of 100), e-consultation (32 of 100)

and e-decision (8.3 of 100) are not developed yet. It means there is a huge demand in Europe to benefit from this hidden potential, which has not been exploited yet.

For successful introduction of political decisions those must be verified by modelling of potential effect. Second, it must be checked would be the new policy accepted by citizens.

Therefore, at least the social networks must be monitored. And, third, audience must be informed in time about activities expected to make up the citizens for policy introduction. Cohen (Cohen 2012) provides at least eleven tips for crafting a policy, but mentioned tasks are important part of them.

Typical sample of putting into practice of the ideas mentioned above is Skopje Bicycle Inter-modality Simulator.

## 2. SKOPJE BICYCLE INTER-MODALITY SIMULATOR

Skopje Bicycle Inter-modality Simulator (Ginters et. al. 2014) is created to find useable solution for bicycle station and track building location in the City of Skopje (see Figure 1).

To encourage the intermodal transport, the citizens are also involved in the collection of ideas on how bicycle inter-modality can be fostered. The citizen participation is supported including the use of a simulation tool and ticketing. Public simulation is accessible at <http://prod.fupol.lv:8080/skopjebicycle>.

### 2.1. Simulation model

The relevant elements of the Skopje Bicycle Inter-Modality model (Buil, Piera and Gusev 2014; Buil and Piera et.al. 2014) are the profiles of the people, the stations along the city, the tracks between stations, and the bicycle users. Following subsections present the data tables with all the necessary columns to correctly define these elements.

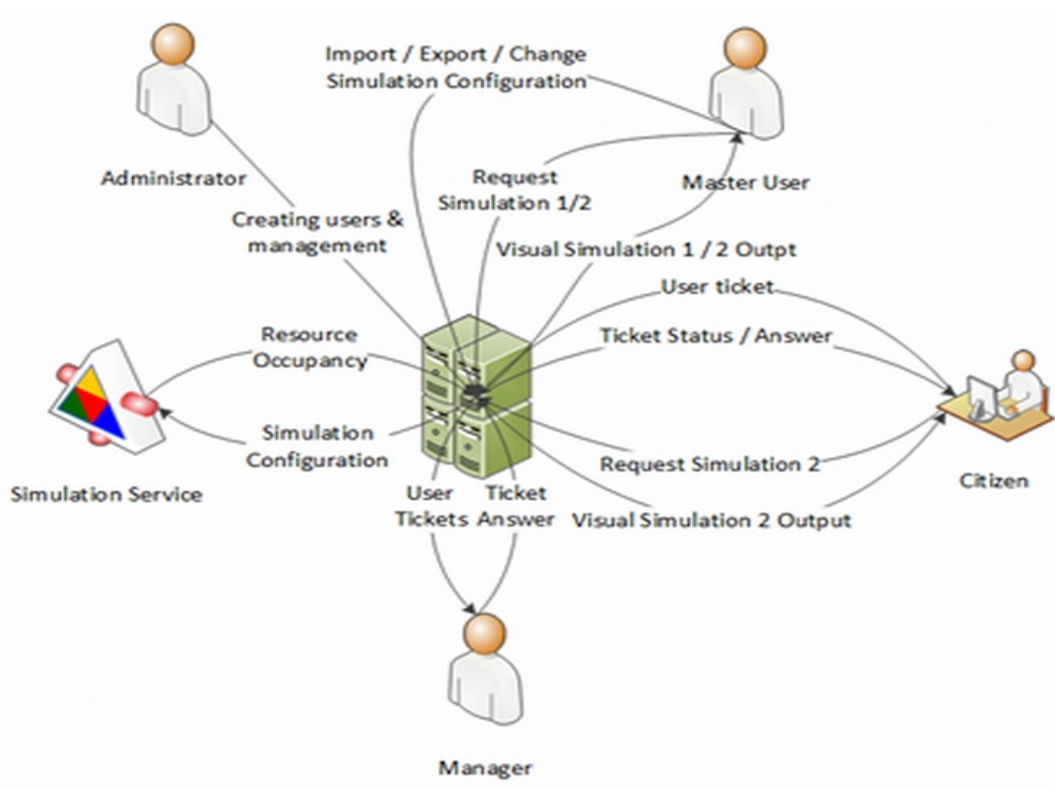


Figure 1: Functional Diagram of Skopje Bicycle Inter-modality Simulator

The Bicycle Inter-modality input has seven tables, and they are named:

- Municipalities;
- Stations;
- Tracks;
- Profiles;
- UsersData1;
- UsersData2;
- Weather Conditions.

The agents defined for the multi-agent systems (MAS) model are:

- Station;
- Track;
- User
- Observer.

Each agent has its attributes and methods.

The attributes defined to describe a station are the followings:

- id: it is the identifier of the station;
- description: name or small description of the station;
- code: it is the area code of the station;
- latitude: latitude of the station;
- longitude: longitude of the station;
- bikepath: indicates if there is a bike starting, ending or going through this station;

- bus, train and point: different variables indicating if the station is a bus station and/or a train station and/or a point of interest;
- rentingStatus: it indicates the status of the renting station (Existing, planned, possible, notPossible);
- capacityRenting: if it is a renting station, it indicates its capacity; parkingStatus: it indicates the status of the parking (Existing, planned, possible, notPossible);
- capacityPrivate: if it is a parking, it indicates its capacity;
- occupancyRenting: number of renting bikes in the station;
- occupancyParking: number of private bikes in the station;
- occupancyExtra: number of private bikes parked outside station due to the lack of parking slots;
- municipality: municipality it belongs;
- oSatisfaction: satisfaction of persons living around this station. Summation of the satisfactions of each one;
- bikeUsers: number of possible users around.

There are not relevant methods; therefore, just attributes are listed:

- Origin: the origin of the track;

- Destination: the end of the track;
- length: length of the track;
- occupancy: the amount of people in the track;
- bikePathStatus: it indicates if the bike path exists in the track, if it is planned, if it is possible or if it is not;
- bikePathQuality: it indicates the status of the track. 0 means it is not built, 1 means it is a bike path in perfect conditions, and any value in between indicate the level of conditioning.

This agent represents the users of bikes. The considered attributes for User are:

- id: it is the agent id, automatically generated;
- groupID: it is the group Id taken from the input tables;
- profile: indicates the type of user;
- description: indicates an area or neighbourhood, or indicates that it has an exact location for the house;
- age: it indicates the age of the agent;
- destination: it indicates the Municipality destination (code);
- destinationID: it indicates the exact ID of the destination station;
- originID: it indicates the exact ID of the origin station;
- tripStations: set of ordered stations from origin to destination;
- duration: duration of the trip between stations;
- startTime: time to start the first trip (going to destination);
- backTime: time to start the second trip (going back to origin);
- durationTotal: duration of the trip;
- distance: length of the trip;
- personality: level of personality of the agent;
- satisfaction: level of satisfaction about the bicycle infrastructure;
- tripStatus: variable in which the status of the tracks are added in order to generate a global status;
- trackID: id of the performed trip;
- stationID: the station where the user is parking or taking a bike;
- pDay: day parameter to determine if the bike is used;
- WEcomp: parameter to determine if the pDay value for weekend is complementary or not, which means that bike must be used one of the two days;
- weVisit: indicates if a complementary weekend, the bike has been already used;
- pMonth: month parameter to determine if the bike is used;
- pWeather: weather parameter to determine if user uses the bike;

- park: indicates if the user bike is parked in a bike parking slot.

The profiles will be useful to generate the correct amount of agents. They indicate a population percentage depending on the type and the age. Transport means indicates the percentage of times the profiles use bus, car, taxi, foot or motorbike.

There relevant methods for the agent User are:

- nextDay: it calculates the next day that the user takes the bike and schedule next scheduleDayDecision, or schedule the day calculation (same method) for the next week if scheduleDayDecision is not randomly scheduled at any day of this week;
- nextTrip: scheduling the next trip of the user;
- parameterMonthWeather: it calculates the decision parameter using the month and weather input parameters;
- scheduleDayDecision: deciding if the user will use the bike depending on the weather parameter and scheduling next use of the bike if the user finally does not use it;
- updateSatisf\_endDay: it updates the satisfaction at the end of the trip using tracks and stations status information;
- updateSatisf\_location: it updates the agent satisfaction depending on the citizens around the agent origin;
- updateSatisf\_Occupancy: it updates the satisfaction depending on the occupancy of the stations;
- updateSatisf\_track\_by\_track: it updates the satisfaction at the end of each track.
- The observer agent, in that case, has been used to manage the data used in the software interface, which can be input and output data. The input attributes are:
  - simulationENDtime: simulation time;
  - fixweather: indicates if the weather is fixed;
  - fixMonth: indicates if the month is fixed;
  - tracksFile, userFile, user2File, weatherFile, stationsFile: names of the files used as input data.

The output attributes are:

- numConflicts: number of conflicts during the simulation;
- conflictList: list of conflicts including Station id, user profile and time;
- currentBikeUsersPercentage: percentage of bike users at any time;
- occupancy: structure with the occupancy of the stations and tracks.

Other attributes are:

- distMatrix: matrix of the distances between stations. Used to calculate shortest routes;



- stationIndex: list of stations index used to calculate the shortest route applying the Dijkstra algorithm;
- municipalityStations: list of stations of each municipality. Used in order to generate users origin station;
- cycling\_velocity: average cycling velocity. Used to calculate the time of the trips depending on the paths length.

The Observer agent includes methods to initialize the model, the linking methods to reflect the interactions between agents, and other methods.

The relevant methods of the observer are:

- generateStations: it is in charge of generating the stations to consider;
- generateTracks: it is in charge of generating the tracks to consider;
- generateTripStations: it is in charge of calculating the intermediate stations of the trips;
- readUsersData: read the Users data and schedule the first agents generation;
- readWeatherConditions: reading weather conditions from an input file and saving in the weather conditions structure;
- updateRealTime: method used to calculate the real time and update weather conditions, if necessary.

## 2.2. Simulator Graphic User Interface

Simulator provides three language choices - English (default), Macedonian and Albanian. Each of simulation settings has multiple data sets that contain data about users, stations, tracks and weather. Each registered user can have one or more of three user roles (each role describes different permissions). User roles are - administrator, manager and master. Simulation configuration is section for system high-level administration. It contains such system data as e-mail link for ticket senders, and multiple simulation model settings. Each user can send to administration information about simulation results. Each simulation ticket has multiple informative fields - Status (Sent, when ticket is sent; Open - when ticket is first time opened by Administration; Answered - when ticket has received a reply from Administration; Closed - when sender has closed ticket), Close status (Like - if sender has approved his ticket; Dislike - if sender has disapproved his ticket) and Suggest status (Suggested - if manager has suggested ticket; Rejected - if master has rejected the simulation settings for representing ticket in suggested simulation's visual results block; Approved - if master has approved the settings). Users with master role can access Simulation 1, which is set of multiple simulations that are run for full year. Each of them is run for different simulation configurations. Simulation 2 runs for one week by master and regular users from public side (see Figure 2).



Figure 2: Running Public Simulation View

Each simulation has status, ID that links to simulation visual results, progress percentage, description, chosen simulation settings, bike user percentage and date created.

All simulations have similar visual result page with minor differences between Simulation 1 and Simulation 2. Clicking on each station and track opens its data window, which shows park station status and park station capacity. For tracks there are track quality rate and track status. Results block can be opened by clicking on icon on the results page.

Simulator is created to offer the City of Skopje a possibility to realize a better schedule of track and stations and plan of occupancy. To its citizens it offers

the opportunity to simulate the occupancy of the bicycle tracks and stations in Skopje city and suggests new ideas to the administration of City of Skopje. The system helps the Administration of City of Skopje to improve the scheduling and resource planning, initiation and creating new projects involving the bicycle area at Skopje city. Citizens of Skopje can also get involved in decision making process by communication and opinion expression to the authorities throughout the system, making the whole process more transparent and efficient.

In current sample double cycled policy crafting model is realized (see Figure 3).

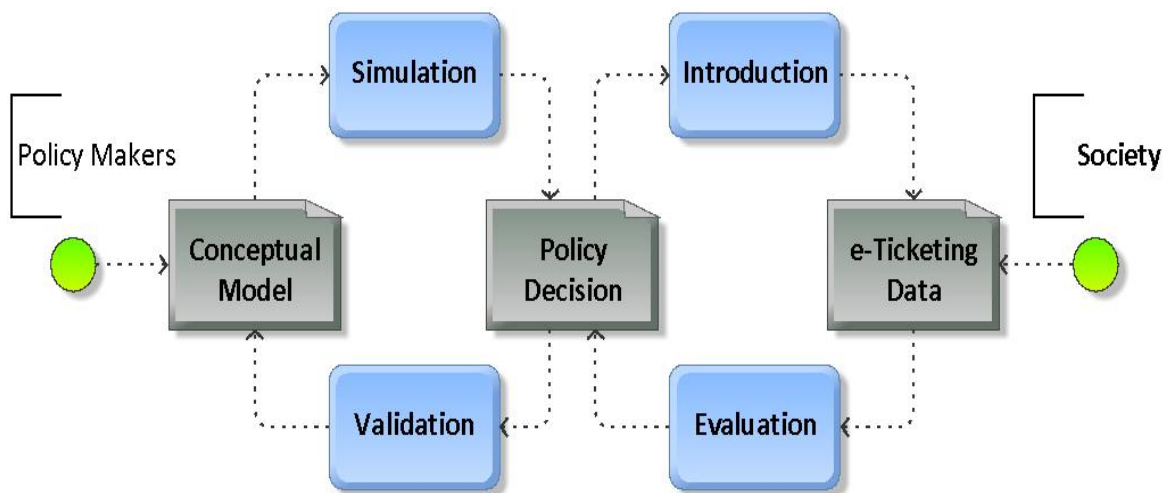


Figure 3: Double Cycled Policy Crafting Model

First Conceptual model is simulated to validate and assess the trend and only after that the policy decision could be introduced. However all the time opinions of the society is monitored through semantic search in social networks and e-ticketing to be sure that forecasts are right and justified.

### 3. CONCLUSIONS

Before introduction of policy decisions it is reasonable to validate the forecasted results. Also it is recommended to ascertain opinions of the society about potential changes.

Skopje Bicycle Inter-modality Simulator ensures validation of potential activities of Skopje Municipality using multi-agent systems (MAS) simulation in Repast Symphony environment. MAS model is based on statistical data provided by Skopje municipality. However society opinions are collected by ticketing.

Skopje Bicycle Inter-modality Simulator provides possibility for citizens to participate directly in policy

planning and crafting. On other hand through the ticketing municipality has the feedback and can be assured about adequacy of their decisions to requirements of the citizens. Inter-modality is actuality of big cities therefore right solution can be interesting for widespread set of policy decision makers.

Further activities of Skopje Bicycle Inter-modality Simulator designing will be related with enhancing and statistical validation of simulation model.

### ACKNOWLEDGEMENTS

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

### REFERENCES

Buccoliero, L., Bellio, E., 2010. Citizens Web Empowerment in European Municipalities. *Journal of E-Governance*. Volume 33, Issue 4, pp. 225-236.

- Buil, R., Piera, Miquel A., Gusev, M., 2014. *Deliverable 2.9 FUPOL Model Parameterization. Advanced and Verified Version*. 20 p.
- Buil, R., Piera, Miquel A. et.al., 2014. *Deliverable 2.11 FUPOL Cognitive and Causal Models. Validated Version*. 73 p.
- Cohen, A., 2012. *11 tips for crafting a social media policy*. Available from: <http://www.insidecounsel.com/2012/05/01/11-tips-for-crafting-a-social-media-policy?page=2> [accessed 29 April 2014].
- Ginters, E., Aizstrauts, A., Dreija, G., Ablazeveica, M., Stepucev, S., Sakne, I., Baltruks, M., Piera, M-A., Buil, R., Gusev, M., Velkoski, G., 2014. *Deliverable 4.6 – Intermediate Release FUPOL Simulator Software advanced prototype*. Skopje Bicycle Inter-modality Simulator. 224 p.
- Osimo, D. et. al., 2010. *Project Crossroads*. Deliverable D4.1 First Roadmap Draft.

## AUTHORS BIOGRAPHY

**Egils Ginters** is director of Socio-technical Systems Engineering Institute. He is full time Professor of Information Technologies in the Systems Modelling Department at the Vidzeme University of Applied Sciences. He is a Senior member of the Institute of Electrical and Electronics Engineers (IEEE), member of European Social Simulation Association (ESSA) and Latvian Simulation Society. He participated and/or coordinated some of EC funded research and academic projects: FP7 FUPOL project No. 287119 (2011-2014), FP7-ICT-2009-5 CHOREOS project No. 257178 (2010-2013), e-LOGMAR-M No.511285 (2004-2006), SocSimNet LV/B/F/PP-172.000 (2004-2006), LOGIS MOBILE LV/B/F/PP-172.001 (2004-2006), IST BALTPORTS-IT (2000-2003), LOGIS LV-PP-138.003 (2000-2002), European INCO Copernicus DAMAC-HP PL976012 (1998-2000), INCO Copernicus Project AMCAI 0312 (1994-1997). His main field of interests involves: systems simulation technologies, logistics information systems, and technology acceptance and sustainability assessment. He has more than 140 scientific articles related with the research fields.

**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.

**Mikelis Baltruks** graduated in Information Technology at the Vidzeme University of Applied Sciences in 2014. Henceforth he worked as programmer at Socio-technical Systems Engineering Institute. His research interests are software designing and simulation.

**Miquel Angel Piera Eroles** is the delegate for Technical Innovation Cluster, and director of LogiSim,

a recognized research group on Modeling and Simulation of Complex Systems. He is Full time Professor in the System Engineering Department at Universitat Autònoma de Barcelona (UAB). Graduated with excellence from UAB in Computer Engineering (1988), Msc from University of Manchester Institute of Science and Technology in Control Engineering (1991), and he got his Phd in 1993. He is member of the Editorial board of 3 international journals and Editor in Chief of IJBRM. Dr Piera has been general chair and invited speaker of many International conferences. He has been nominated as Deputy Director of the UAB Engineering School and Coordinator of the Spanish CEA-IFAC research team on Modeling and Simulation. He has coordinated many research and industrial projects, he has also participated in some EC funded research and academic projects, such as: LOGIS MOBILE LV/B/F/PP-172.001 (2004-2006), Curriculum Development on Logistics and Supply Chain Management 134522-LLP-1-2007-1-ESERASMUS-ECDSP and FP7 FUPOL No. 287119.

**Roman Buil** received the B.S. degree in Mathematics from Universitat Autònoma de Barcelona (UAB) in 2002, the M.S. degree in industrial engineering - advanced production techniques from UAB in 2004. He is currently working towards the Ph.D. degree in industrial engineering - advanced production techniques at the same University. He is research scientist, assistant teacher and project manager at the Logistics and Aeronautical Unit of the Telecommunications and Systems Engineering Department of UAB. His research interests include modelling and simulation methodologies, optimization techniques, policy modelling, production planning and decision making for production planning and logistics. He is member of LogiSim, a recognized research group on Modelling and Simulation of Complex Systems and he has been involved in industrial projects working as consultant of different companies. He participates in FP7 FUPOL Project No. 287119.

**Marjan Gusev** (Professor, Ph.D in Informatics) has been expert and consultant of various IT initiatives and projects in the former Yugoslav Republic of Macedonia and broader. He has been member of national committee for eMacedonia since 1999 and president of National IT experts group (Macedonian Association of IT companies). For his activities in e-Government area he has been elected as president of Task Force for eGov in the Macedonian IT Cluster, and was member of team that produced National Information Strategy in Macedonia responsible for e-Government part. He has participated in, coordinated and evaluated many national IT projects and participated and coordinated several TEMPUS projects of the European Commission.

# DEVS Models Design and Test using AGILE-based Methods with DEVSimPy

Timothée Ville, Laurent Capocchi, Jean-François Santucci

UMR SPE University of Corsica

ville@univ-corse.fr, capocchi@univ-corse.fr, santucci@univ-corse.fr

**Keywords:** Discrete event simulation; DEVS; behavior test; AGILE

## Abstract

Validation and test of DEVS models at the early phases of the Design process is a crucial topic when dealing with complex DEVS models. Based on Software Engineering test methods, we present in this paper a new approach which integrates Agile methods in the process of the simulation in order to design and test DEVS models. We propose an implementation in Python language based on the use of aspect programming concept (patch, mocking objects and decorators). This implementation is performed in the framework of the DEVSimPy environment with the definition of a plug-in dedicated to the automatic generation and execution of test scenario. Two pedagogical example have been used in order to point out the feasibility of the approach.

## 1. INTRODUCTION

This paper deals with validation and test of DEVS models at the early phase of the design process. DEVS (Discrete Event System specification) is a widely used formalism in the framework of simulation of complex systems. The validation of models is traditionally a step which is relegated at the end of the design process: once the models have been defined and coded, experiments are conducted in order to validate them using simulation. However this traditional way to perform validation of models is often an expensive and time-consuming activity and the resulting quality of the models is still poor. Consequently, new approaches for coping with these challenges are necessary. The same remarks can be formulated when dealing with Software Testing. Considering software testing, one emerging trend is stronger integration of testing as early as possible in the design process of a program. For that reason software engineering has proposed new design and test as Agile methods which include Test Driven Development (TDD) [Fraser et al., 2003] and Behavioral Driven Development (BDD) [Solis and Wang, 2011] methods.

In order to go on with the analogy between modeling and simulation and Software Engineering, one can imagine applying BDD and TDD Agile methods to the Design and Test of DEVS models. Our main objective is to develop an approach that is able to use different Software Testing techniques stemming from Software Engineering (Agile methods to be more

specific) that are applied to DEVS models design in order to improve quality assurance of the resulting DEVS models while proposing inexpensive and no time-consuming activity.

Our approach consists in applying the BDD method for the design of DEVS models. In order to achieve this goal, we have carefully performed a correspondence between the BDD method when applied to software and the BDD method which has to be applied to DEVS models design.

The problem is to perform a BDD method when defining DEVS models. Defining such a method in the DEVS Modeling and Simulation context requires the resolution of the following basic problems: (i) to define a semi-formal format for the behavioral specification of the test for any atomic models involved in a DEVS model, (ii) to define how to generate parameters for a test from a specification document of the tests of DEVS models, (iii) to define how to perform the previously defined tests using simulations. To solve these problems we proposed to: (i) to define a semi-formal format from the natural specifications of FDDEVS as proposed by B.P. Zeigler [Zeigler, 1976], (ii) to use the specificity of the Python Language to generate a test (by means of the decorator programming concept), (iii) the tests are executed by combining software engineering programming concepts such as decorators, patch, mocking objects in order to perform the tests using DEVS simulations in the framework of the DEVSimPy [Capocchi et al., June] environment. DEVSimPy is a collaborative general user interface implemented in Python language allowing us to experiment new approaches inside the DEVS formalism. To ensure this, we use DEVSimPy plug-ins in order to be more generic.

The next part of the paper gives the background of the paper: the DEVS formalism and the DEVSimPy framework are briefly introduced before the presentation of the main notions involved in Agile methods. In section 3. an overview of the proposed approach is briefly presented. Section 4.1. is devoted to the definition of the semi-formal format chosen for the user to write the behavioral specification of the tests as required in a BDD method. In section 4.2. we describe how we have been able to generate the test parameters which will be used to validate the models. Section 4.3. presents how the tests are performed using simulations within the DEVSimPy framework by integrating decorators, patches and mocking objects into DEVS models. The last part will permit to conclude and to give a brief overview of future work we envision.



## 2. BACKGROUND

### 2.1. DEVS Formalism and Tools

The Discrete Event system Specification (DEVS) formalism introduced by Zeigler [Zeigler, 1976] provides a means of specifying a mathematical object called a system. Basically, a system has a time base, inputs, states, outputs, and functions for determining next states and outputs given current states and inputs. The DEVS formalism is a simple way in order to characterize how discrete-event simulation languages may specify discrete-event system parameters. It is more than just a means of constructing simulation models. It provides a formal representation discrete-event systems capable of mathematical manipulation just as differential equations serve this role. Furthermore by allowing an explicit separation between the modeling phase and simulation phase, the DEVS formalism is one of the best ways to perform an simulation of systems using a computer. In the DEVS formalism, one must specify: (i) basic models from which larger ones are built, and (ii) how these models are connected together in hierarchical fashion. An atomic model allows specifying the behavior of a basic element of a given system. Connections between different atomic models can be performed by a coupled model. A coupled model, tells how to couple (connect) several component models together to form a new model. This latter model can itself be employed as a component in a larger coupled model, thus giving rise to hierarchical construction. An atomic DEVS model can be considered as an automaton with a set of states and transition functions allowing the state change when an event occur or not. When no events occurs, the state of the atomic model can be changed by an internal transition function called  $\delta_{int}$ . When an external event occurs, the atomic model can intercept it and change its state by applying an external transition function called  $\delta_{ext}$ . The life time of a state is determined by a time advance function called  $t_a$ . Each state change can produce output message via an output function called  $\lambda$ . A simulator is associated with the DEVS formalism in order to exercise instructions of coupled model to actually generate its behavior. The architecture of a DEVS simulation system is derived from the abstract simulator concepts [Zeigler, 1976] associated with the hierarchical and modular DEVS formalism.

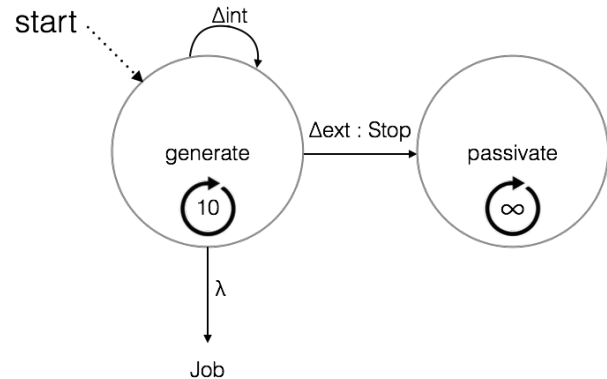
Concerning the natural language specifications FDDEVS proposed by Zeigler in [Zeigler and Sarjoughian, 2012]. FDDEVS gives us many advantages; in particular it offers a support for XML translation or graphic representation. Here is a basic example of a generator DEVS model specification.

```

to start hold in generate for time 10!
after generate output Job!
from generate go to generate!
when in generate and receive Stop then go to
 passive!
passivate in passive!
```

**Listing 1.** DEVS specification of the Generator model.

The Figure 4.1. show what are information are involved by the specification of Listing 1.



**Figure 1.** State automaton of model from Listing 1.

There are many tools which provide a user interface dedicated to help the user to define DEVS models and to perform simulations. A non exhaustive list can be done: PowerDEVS, DEVSIM++, DEVJSJAVA, VLE, DEVSIMPy, CD++Builder, ATOM3, MS4Me, etc. Special attention will be given to DEVSimPy (stand for DEVS simulator in Python language) which is a collaborative modeling and simulation (M&S) software.

DEVSimPy (Python Simulator for DEVS models) [Capocchi et al., June] is a user-friendly interface for collaborative modeling and simulation of DEVS systems implemented in Python language. Python is a programming language known for its simple syntax and its capacity to allow modelers to implement quickly their ideas [Langtangen, 2005]. The DEVSimPy project used the Python language and provides a GUI based on PyDEVS [Bolduc and Vangheluwe, juin 2001] API in order to facilitate both the coupling and the re-usability of PyDEVS models. This API is used in the excellent multi-modeling GUI software named ATOM3 [de Lara and Vangheluwe, 2002] which allows to use several formalisms without focusing on DEVS. DEVSimPy is an open source project under GPL V3 license and its development is supported by the University of Corsica Computer Science research team. It uses the wxPython graphic library which is a wrapper of the most popular WxWidgets C library. DEVSimPy can be downloaded at <http://code.google.com/p/devsimpy>.

The main goal of this environment is to facilitate the modeling of DEVS systems using the GUI dynamic libraries and the drag and drop functionality. With DEVSimPy, models can be stored in a dynamic library in order to be reused and shared. The creation of dynamic libraries composed with DEVS components is easy since the user is coached by dialogs and wizard during the building process. With DEVSimPy, complex system can be modeled by a coupling of

DEVS models and the simulation is performed in a automatic way. Moreover, DEVSImPy allows the extension (or the overwrite) of their functionality in using special plug-ins managed in a modular way. The user can enabled/disabled a plug-in using a simple dialog window.

## 2.2. Agile Test Methods

Agile methods [Cockburn, 2002] is the increasingly common practice throughout the lifecycle to develop a software iteratively. These methods may include: the Test Driven Development (TDD) [Fraser et al., 2003] method and its extension/revision, the Behavior Driven Development (BDD) [Solis and Wang, 2011] method. TDD is a software development methodology which essentially states that for each unit of software, a software developer must: (i) define a test set for the unit first, (ii) then implement the unit, (iii) finally verify that the implementation of the unit makes the tests succeed. BDD is a specialized version of TDD which focuses on behavioral specification of software units. It is based on: (i) the use of examples to describe the behavior of the application, or code units; (ii) automating those examples to provide quick feedback and regression testing; (iii) finally, the use of “Mocks ” [moc] replacing the code modules which have not yet been written.

The main steps of the BDD method can be summarized by the following points:

1. Behavior-driven development specifies that tests of any unit of software should be specified in terms of the desired behavior of the unit; usually the desired behavior should be specified using a semi-formal format for behavioral specification.
2. Then a specification document has to be read and each scenario of the document is breaking up into meaningful clauses. Each individual clause in a scenario is transformed into some sort of parameter for a test.
3. The framework then executes the test for each scenario, with the parameters from that scenario.

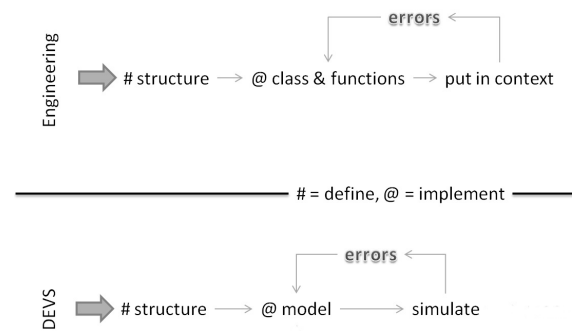
In this paper, we present a set of main BDD characteristics which are used to implement the test of DEVS model.

## 3. PROBLEM DESCRIPTION

The context of the proposed work relates to the cycle of software development. Traditionally software development corresponds to a logical and intuitive approach described in Figure 2.

The basic approach of such traditional cycle is: we code first and then we perform the tests.

As explained in the introduction, we have developed an analogy between Software Engineering design process and DEVS



**Figure 2.** Traditional cycle of model development.

models design process. From Software Engineering point of view, the structure of the software is first defined, the implementation of the related classes and functions is performed and finally the implementation is executed in order to find errors (as it can be seen in the engineering part in Figure 2). The same kind of development cycle is used for DEVS models development: the structure of models is first defined, models are then implemented and the obtained implementation is simulated in order to find potentials errors (as it can be seen in the DEVS part in Figure 2). If errors are raised during simulation, they have to fixed one by one and verified each time by simulation. We also can deduce numerous problems from a validity and productivity points of view from the Figure 2:

- The models cannot be prepared for all the situations and the predictions have a reliability decreased because of the visible uncertainty of the produced results.
- The behavior of the developed models has big chances to be erroneous and it will certainly be necessary to adjust certain parameters or certain functions.
- The maintenance, the re-factorization or the evolution of a model guaranteed not the preservation of the behavior.
- The same model implemented in various environments is completely different. It can raise problems in the optics of a standardization.

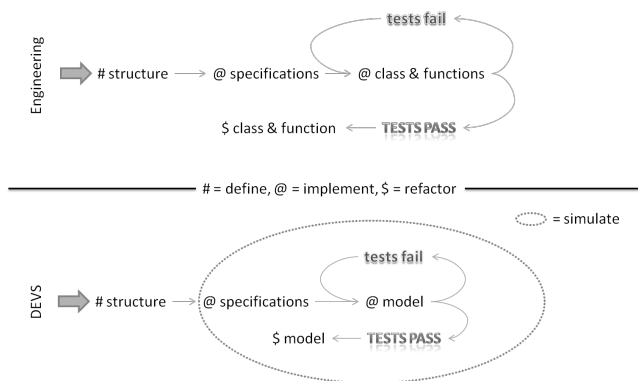
There are many manners to test a DEVS model (atomic or coupled model). In [Li et al., 2011] the authors focus on the validation of DEVS formalism implementation. The approach described in [Byun et al., 2009] concerns the correctness of a given simulation. It proposes a framework allowing to check all the possible paths involved in a given simulation model. In [Hu et al., 2007], test agents are presented. This paper proposes a definition of a test agent which is connected to the I/O of a given model and is used to point out the behavioral information concerning the model. The test is performed using this information. However the proposed approach is not a generic one since test agent should be specific to a given model. In the same idea, we have developed a DEVSImPy plug-in to manage an universal test agent using the Behave Framework tool [beh, Ville, 2013]. However an im-



portant drawback was the fact that the simulation has to be interrupted in order to manually performed the tests. In the next sub-section we give an overview of the developed solution in order to propose an embedded mechanism allowing to automatically take into account the test part at the beginning phase of the design process of DEVS models.

#### 4. PROPOSED SOLUTION

As described before, the traditional cycle for DEVS model design raises numerous problems. The proposed approach is summarized in Figure 3.



**Figure 3.** Proposed behavior driven development cycle.

We have developed a solution by analogy with the BDD approach defined in the Software Engineering domain. As shown in Figure 3 (engineering part) this approach consists in firstly define the structure of the software to be designed and immediately after to write the test specifications. Then the designer can implement and at the same time test the class and functions of the software to be designed. The same kind of development cycle is proposed for the design of DEVS models as it is shown on the DEVS part of Figure 3. The only difference is that the specifications writing and the model implementation are performed using a modeling and simulation framework. The simulation engine is used in order to perform the tests.

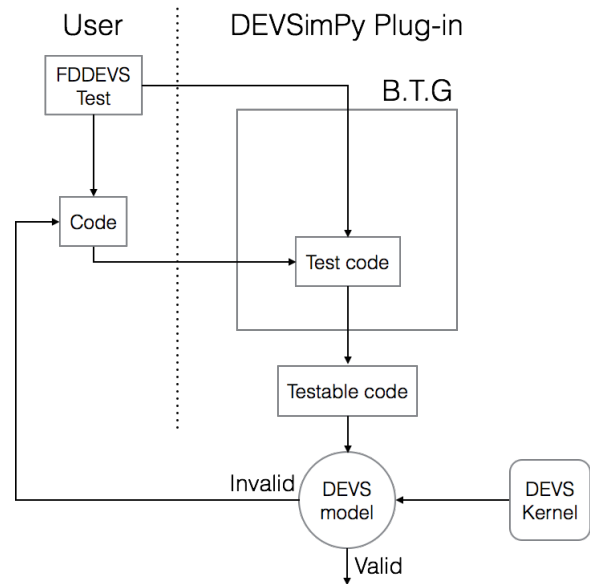
Implemented behavior in a DEVS model are directly tested. This brings numerous advantages:

1. Produced code is reliable.
2. The basic elements of the DEVS model can be tested one after the other.
3. Even if there is an evolution of the implementation of the DEVS model, the behavior remain the same.

The previously presented approach has been implemented in the framework of the DEVSimPy environment. The design and test process has been introduced in this Python Language oriented simulation environment. In order to propose

a generic implementation we choose to define a DEVSimPy plug-in dedicated to: (i) the automatic generation of test scenario and (ii) the execution of the test scenario using the simulation kernel.

Figure 4 describes explicitly how a user is going to develop the Design and Test proposed methodology.



**Figure 4.** DEVS design and test proposed methodology.

The user has first to write the FDDEVS test specifications. Then he has to code part of (or completely) the DEVS model corresponding to the previously written specifications. The DEVSimPy plug-in will allow to:

- Generate testable code from the FDDEVS specification (own to the Behavior Test Generation).
- Integrate the testable code into the DEVS model already defined by the user.
- Simulate the integration of the testable code and the already defined DEVS model part.

If the result of the simulation points out an invalid DEVS model definition, the user has to rewrite the DEVS model and again execute the plug-in.

#### 4.1. Test Scenario Specification

Specification is an important point when dealing with a BDD approach for DEVS formalism. It gives a manner to describe the test patterns associated with the behavior of a DEVS model which has to be defined and implemented. Instead of defining a new language we choose to select an already defined language which allows to describe DEVS modeling scheme under a semi-formal natural language. The pseudo-natural language proposed by Prof. Zeigler which is inspired from the grammar detailed in [Hong and Kim, 2006] has been adapted in order to offer a language for specify-



```

1 #original function : Empty
2 model.extTransition()
3
4 # {(received messages) : prepared message}
5 criticals_data = {(1, 2): 1, (2, 3): 2}
6 def criticalsData(*args):
7 return criticals_data[args]
8
9 # Patch of the extTransition method with side effect
10 model.extTransition = MagicMock(name="extTransition", side_effect=criticalsData)
11 model.extTransition(1, 2)
12 => 1
13 model.extTransition(2, 3)
14 => 2

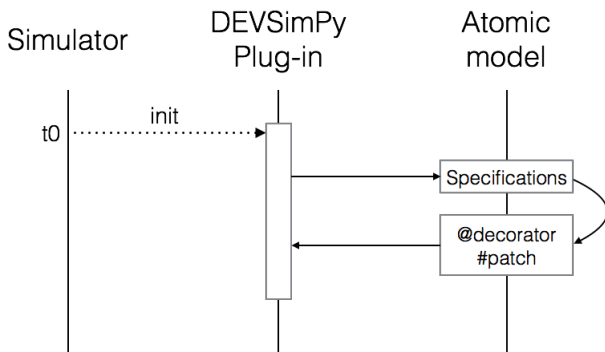
```

**Figure 9.** Simple method patching example.

In Figure 9 we suppose that the external transition of the DEVS model to be tested has not been already implemented. From the specification we have been able to deduce the critical data which are expressed as follows: for inputs 1 and 2, the output should be 1 while for inputs 2 and 3, the outputs should be 3. The use of patches and mocking objects is highlighted by the introduction of a MagicMock object (from line 10 to 14 of Figure 9). The external transition (“extTransition” function) is patched with the previously mentioned MagicMock object). This example points out how the behavior of DEVS functions can be defined using mocking objects when a DEVS model has not been totally implemented.

### 4.3. Test Execution

Figure 10 describes how the DEVSImPy plug-in: (i) is integrated into the simulation kernel, (ii) is able to select between the injection of decorators or patches (mocking objects).



**Figure 10.** Sequence diagram of the DEVSImPy plug-in.

In order to select which kind of injection has to be performed during the simulation, we use an important property of the Python programming language called dynamic introspection: it is possible to know if a python object is completely implemented or not at any time of the execution of a given python script. Figure 10 details how this property is used in order to dynamically (during the simulation phase) detect if an given DEVS function has been (or has not been) already implemented. At time  $t_0$  (initialization of the simulation phase), the DEVSImPy plug-in is executed:

1. It refers to the specification associated with the DEVS

model being implemented and tested.

2. It executes the BTG engine which allows to transform the previous specification into decorators or patches according to model introspection.
3. It ends by returning into the DEVSImPy simulation kernel in order to go on with the simulation.

The definition of this DEVSImPy plug-in as described above permits a user to perform design and test of DEVS models as presented in Figure 4.

## 5. CONCLUSION AND PERSPECTIVES

The paper introduced an approach allowing to perform the test of DEVS models at the very early phases of the Design. The main idea is to apply the concepts which have been defined in the Software Engineering domain in the framework of the Agile community. We described how the BDD (Behavior Driven Development) design and test methodology stemming from the software engineering domain can be applied to the design of DEVS models. We presented how we have adapted the three main steps of the BDD methodology: (i) definition of a semi-formal language allowing test specification using the DEVSSpecL language, (ii) generation of the test scenarios using the definition of a BTG using the notion of decorators and patches, (iii) execution of the test scenario using the concepts of dynamic introspection and mocking objects. The resulting Design and Test DEVS approach has been validated in the framework of the DEVSImPy environment. The three previous steps have been integrated into a DEVSImPy plug-in. We have validated the proposed solution on pedagogical examples which point out the feasibility of the approach. The future work will concentrate in the validation using the design of a more complex DEVS models.

## REFERENCES

- BDD python style. <https://github.com/behave/behave>.
- Decorators for functions and methods. <http://legacy.python.org/dev/peps/pep-0318/>.
- Mock - mocking and testing library. <http://www.voidspace.org.uk/python/mock/index.html>.
- Simpleparse a parser generator for mxtxttools v2.1.0. <http://simpleparse.sourceforge.net/>.
- J.-S. Bolduc and H. Vangheluwe. pythonDEVs : A modeling and simulation package for classical hierarchal DEVs. In *Rapport technique, MSDL, Université de McGill*, juin 2001.
- J. H. Byun, C. B. Choi, and T. G. Kim. Verification of the DEVs model implementation using aspect embedded DEVs. In *Proceedings of the 2009 Spring Simulation Multiconference*, pages 151:1–151:7, San Diego, CA, USA, 2009. URL <http://dl.acm.org/citation.cfm?id=1639809.1655380>.
- L. Capocchi, J. F. Santucci, B. Poggi, and C. Nicolai. DEVSimPy: A collaborative python software for modeling and simulation of DEVs systems. In *20th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pages 170–175, June. doi: 10.1109/WETICE.2011.31.
- A. Cockburn. *Agile Software Development*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002. ISBN 0-201-69969-9.
- J. de Lara and H. Vangheluwe. AToM3: A tool for multi-formalism and meta-modelling. In *Proceedings of FASE'02*, pages 174–188, London, UK, 2002. Springer-Verlag. ISBN 3-540-43353-8.
- S. Fraser, K. Beck, B. Caputo, T. Mackinnon, J. Newkirk, and C. Poole. Test driven development (tdd). In *Proceedings of the 4th International Conference on Extreme Programming and Agile Processes in Software Engineering, XP'03*, pages 459–462, Berlin, Heidelberg, 2003. Springer-Verlag. ISBN 3-540-40215-2. URL <http://dl.acm.org/citation.cfm?id=1763875.1763973>.
- K. J. Hong and T. G. Kim. Devspecl: {DEVs} specification language for modeling, simulation and analysis of discrete event systems. *Information and Software Technology*, 48(4):221 – 234, 2006. ISSN 0950-5849. doi: <http://dx.doi.org/10.1016/j.infsof.2005.04.008>. URL <http://www.sciencedirect.com/science/article/pii/S0950584905000650>.
- X. Hu, B. P. Zeigler, M. H. Hwang, and E. Mak. DEVs systems-theory framework for reusable testing of I/O behaviors in service oriented architectures. In *IRI*, pages 394–399. IEEE Systems, Man, and Cybernetics Society, 2007. URL <http://dblp.uni-trier.de/db/conf/iri/iri2007.html#HuZHM07>.
- H. P. Langtangen. *Python Scripting for Computational Science (Texts in Computational Science and Engineering)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. ISBN 3540294155.
- X. Li, H. Vangheluwe, Y. Lei, H. Song, and W. Wang. A testing framework for dev formalism implementations. In *Proceedings of the 2011 Symposium on Theory of Modeling & Simulation: DEVs Integrative M&S Symposium*, pages 183–188, San Diego, CA, USA, 2011. Society for Computer Simulation International. URL <http://dl.acm.org/citation.cfm?id=2048476.2048500>.
- C. Solis and X. Wang. A study of the characteristics of behaviour driven development. In *Software Engineering and Advanced Applications (SEAA), 2011 37th EUROMICRO Conference on*, pages 383–387, Aug 2011. doi: 10.1109/SEAA.2011.76.
- T. Ville. Méthodes de tests comportementaux de modèles devs et mise en oeuvre dans DEVSimPy. Technical report, SPE UMR CNRS 6134, University of Corsica, June 2013.
- B. Zeigler and H. S. Sarjoughian. *Guide to Modeling and Simulation of Systems of Systems*. Springer, London, 2012. ISBN 978-0-85729-864-5. doi: 10.1007/978-0-85729-865-2.
- B. P. Zeigler. *Theory of Modeling and Simulation*. Academic Press, 1976.

# CPN-SIMULATION METHODOLOGY FOR THE BOARDING PROCESS OF AIRCRAFT

Miguel Mujica Mota<sup>(a)</sup>, Idalia Flores<sup>(b)</sup>, Daniel Guimarans<sup>(c)</sup>

<sup>(a)</sup> Aviation Academy, Amsterdam University of Applied Sciences, The Netherlands

<sup>(b)</sup> Facultad de Ingeniería, Universidad Nacional Autónoma de México, México

<sup>(c)</sup> Optimisation Research Group, National ICT Australia (NICTA), Australia

<sup>(a)</sup> [m.mujica.mota@hva.nl](mailto:m.mujica.mota@hva.nl), <sup>(b)</sup> [idalia@unam.mx](mailto:idalia@unam.mx), <sup>(c)</sup> [daniel.guimarans@nicta.com.au](mailto:daniel.guimarans@nicta.com.au)

## ABSTRACT

The boarding process of an aircraft is one of the identified bottlenecks in the turnaround when the aircraft arrives to an airport. The present article presents a methodology that models the key aspects of a boarding process using coloured Petri nets in order to understand in detail what are the main micro-dynamics and cause-effect relationships that hinder the smooth boarding process. Then the different models are integrated at a different level using a discrete-event program to simulate in detail the boarding process. The combination of both techniques reinforce mutually in such a way that the development process is very efficient for developing micro-models when the dynamics are difficult to understand. This kind of approach can be used for testing different boarding configurations and improve the efficiency of the studied process while reducing the cost associated to the process.

Keywords: simulation, discrete event simulation, coloured petri nets, aircraft boarding process

## 1. INTRODUCTION

The turnaround of an aircraft is the temporary space between consecutive flights when the aircraft is in the airport. Depending on the type of company, the time and space available, the turnaround will be more or less long (Bazargan 2004). Furthermore this length will depend upon the business model of the company as all flight schedules are different and the operations vary from business model to business model. In addition, the times of the ground handling operations result very difficult to homogenise due to the different factors that influence the operations such as handling agents, congestion at the airside, etc.

The steps followed by a typical stopover of an aircraft are:

- Prior preparation for boarding: the lines of passengers are organised and all their hand luggage and documentation is checked.
- The plane arrives at the parking stand.
- Block-In is performed.
- The passengers and bags disembark.

- The plane is fuelled.
- Whether there is a scheduled cleaning, the cleaning team will proceed to clean the plane.
- When the last passenger leaves and the cleaning services have finished, the passengers for the next flight shall be boarded. Simultaneously the bags shall start to be loaded for the new flight.
- During the boarding of passengers, the coordinator shall deliver the necessary documentation to the captain.
- As soon as the plane is loaded with fuel, bags and passengers, the doors are closed.
- The chocks are removed.
- The plane performs the taxiing towards the corresponding runway for the take-off.

The business model of the low cost carriers (LCC) forces them to take out some of the operations in order to reduce turnaround times and thus making them more competitive. For this reason the boarding/deboarding operations acquire more and more importance since those operations are the ones that consume most of the time when the aircraft is performing the stopover. Figure 1 illustrates the different processes that compose the turnaround of an aircraft (Airbus 2014).

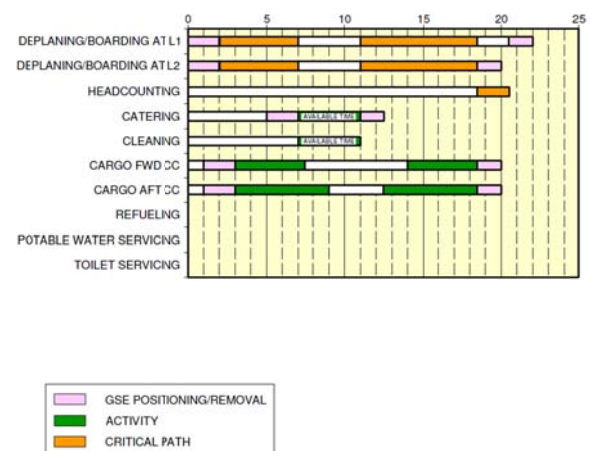


Figure 1: The outstation turnaround time for the A-320

From Figure 1 it can be noticed that the boarding and deboarding operations are currently in the critical path. Therefore in order to make the operations more profitable it is necessary to reduce the turnaround operations.

The current paper presents a methodology that will be useful for improving the boarding process using a combination of coloured Petri nets with simulation.

### 1.1. Review on Boarding Operations

Different authors have put focus on ways to improve the boarding process. Most of them put their efforts in managing the seats and the schedule of the boarding process in the best possible way. Some of the initial efforts for testing the seat scheduling are the ones presented by Marelli, Mattocks, and Merry (1998). These authors simulated the allocation of seats by two different states (On-Off). Using this approach they analysed the results when the boarding is performed using a window, middle aisle (WILMA) scheme. Their conclusion is that the boarding time can be decreased compared to a random assignment boarding. That experiment was followed by defining a different one for the WILMA BLOCK, which uses the same approach but enhanced with separating the boarding in blocks, first the ones at the back, then the ones in the middle, and then those in the front.

Another relevant study is the one presented by Van Landeghem and Beuselinck (2002), through the use of EXCEL and ARENA they simulated some scenarios and conclusions were drawn from the discrete event approach focusing mainly on the boarding time.

Other authors have put focus on policies that could help reduce the boarding time and they use different techniques to test their policies. One interesting example is the work done by Steffen (2008), in which he finds the optimal allocation of passengers using an approach that uses Markov Chains and Monte-Carlo Simulation. Furthermore Soolaki et al. (2012) present an approach that models the problem using linear programming and genetic algorithms. Finally, Steffen and Hotchkiss (2012) tested different configurations using a real-size fuselage of a Boeing 757 for improving the use of the aisle to reduce the boarding time.

The review performed reveals that most of the studies have used different modelling techniques ranging from abstract models developed using linear programming to real-size sets that try to copy the real boarding situations. In the previous years the studies focused mainly on the boarding strategies while in recent years scientific community is paying more attention to the main drivers that are behind the processes that could reduce the boarding times. These drivers are inherent to human nature such as age, companions, family relationships, passengers with bags, disabilities, and even psychological conditions. For these reasons the researchers are investigating with scenarios that are as close as possible to a real situation. However real-size experiments cannot be performed

anytime the researcher wants because they are expensive and it would result difficult to coordinate all the needed elements. This situation puts in evidence the need for the use of digital models that allow the researchers to integrate more characteristics that play a role in the process and result close to real situations. The paper will focus on a methodology that avoids using the traditional abstract approaches such as Monte Carlo simulation or linear programming. Instead, the approach goes one step further using coloured Petri nets (CPN) together with simulation for integrating in a straight forward way the characteristics that play an important role in the boarding process and at the same time allowing the modeller to integrate as much elements as needed.

## 2. CPN-SIMULATION APPROACH

In this paper CPN is used together with simulation for developing a methodological approach that allows modellers to cope with the inherent complexity present in systems whose performance depend on multiple factors. The advantage of using a modelling formalism with the simulation approach is that the different micro-dynamics can be easily understood and the simulation model can be developed in a structured fashion. Different situations, such as the ones present when the passengers with bags must allocate them in the upper compartments of the cabin, or those that appear when the passengers already sit block the ones that need to sit in the same section, can be modelled in a structured way resulting in a more reliable and robust simulator.

### 2.1. Coloured Petri Nets

Coloured Petri Nets is a simple yet powerful modelling formalism that allows to properly modelling discrete-event dynamic systems that present a concurrent, asynchronous and parallel behaviour (Moore and Gupta 1996; Jensen 1997; Christensen, Jensen, Mailund, and Kristensen 2001). CPN can be graphically represented as a bipartite graph which is composed of two types of nodes: the place nodes and the transition nodes. The entities that flow in the model are known as tokens and they have attributes known as colours.

The formal definition is as follows (Jensen 1997):

$$CPN = (\Sigma, P, T, A, N, C, G, E, I) \quad (1)$$

where:

- $\Sigma = \{C_1, C_2, \dots, C_{nc}\}$  represents the finite and not-empty set of colours. They allow the attribute specification of each modelled entity.
- $P = \{P_1, P_2, \dots, P_{np}\}$  represents the finite set of place nodes.
- $T = \{T_1, T_2, \dots, T_{nt}\}$  represents the set of transition nodes, such that  $P \cap T = \emptyset$ , which normally are associated to activities in the real system.



- $A = \{A_1, A_2, \dots, A_{na}\}$  represents the directed arc set, which relate transition and place nodes such as  $A \subseteq P \times T \cup T \times P$ .
- $N$  is the node function  $N(A_i)$ , which is associated to the input and output arcs. If one is a place node then the other must be a transition node and vice versa.
- $C$  is the colour set function,  $C(P_i)$ , which specifies the combination of colours for each place node such as  $C: P \rightarrow \Sigma$ :

$$C(P_i) = C_j : P_i \in P, C_j \in \Sigma \quad (2)$$

- $G$  is the guard function, associated to transition nodes,  $G(T_i)$ ,  $G: T \rightarrow EXPR$ . It is normally used to inhibit the event associated with the transition upon the attribute values of the processed entities.
- $E$  is the set of arc expressions,  $E(A_i)$ , such as  $E: A \rightarrow EXPR$ . For the input arcs they specify the quantity and type of entities that can be selected among the ones present in the place node in order to enable the transition. When it is dealing with an output place, they specify the values of the output tokens for the state generated when transition fires.
- $I$  is the initialization function,  $I(P_i)$ . It allows the value specification for the initial entities in the place nodes at the beginning of the simulation. It is the initial state of a particular scenario.
- $EXPR$  denotes logic expressions provided by any inscription language (logic, functional, etc.)

The state of every CPN model is also called the *marking*, which is composed by the expressions associated to each place  $P$  and they must be closed expressions i.e. they cannot have any free variables.

### 3. THE CPN-SIMULATION APPROACH

A discrete-event system (DES) model has been developed in which the micro-operations such as identifying a person in the block seat, moving forward-backward, putting the luggage on the overhead compartment, or defining different attributes for the type of passenger that board the aircraft are based on the CPN modelling formalism. The advantage of developing a DES simulation model based on the CPN model is that the different micro-operations that vary depending on the type of passenger are clearly identified and therefore a robust simulation model can be developed, which would be too complex for a DES simulator alone. The CPN elements have been mapped and implemented using the technique presented by Mujica and Piera (2011). Following the same logic, the simulation model has been developed using the software SIMIO.

The CPN model models the different interactions that a passenger experience when they are boarding the

aircraft. The initial interactions modelled are the ones that appear when a passenger needs to sit in his respective seat and he is blocked by passenger that previously have arrived. These relations are modelled using the CPN model and they illustrate that if more interactions are identified then is just necessary to modify the respective transition in the CPN model and then implement it in the simulator. After the interactions have been identified and modelled with the CPN model then the implementation in the simulation model is straightforward.

As it has been mentioned, the interactions modelled are the ones at the seat-level in the aircraft. The relations will be implemented in a model of the seat of the aircraft and the total simulation model will be constructed using the simulation program and replicating the seat models so that a final model of the cabin is obtained. Figure 2 illustrates the methodological approach.

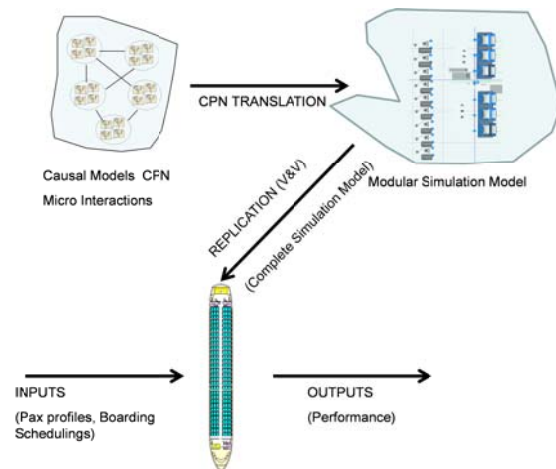


Figure 2: The methodological Approach

The CPN model of the seat-interaction is composed by 16 transitions and 3 place nodes. The definition of colours is presented in Table 1.

Table 1: Colour Definition and Description

Colour	Definition	Description
X	Integer	It describes the row number of the seat block
Y	{000, 001, 010, 100, 011, 101, 111}	It describes the seats occupied by the passengers
Z	Integer	It represents the amount of people waiting in the aisle for the passenger to sit.
R	Integer	It represents the row where the passenger is supposed to be sat.
W	{001, 010, 100}	It represents the seat location of the passenger.
D	{0,1}	It represents if the waiting person is seated in the middle.

The information related to the place nodes is presented in Table 2.

Table 2: Definition of Place Nodes

Place	Colour Set	Description
S	Product X*Y	This place represents the information of the left block of seats of one row of the cabin.
W	Z, D	This place holds information about the passengers standing up and waiting for the passenger to sit.
P	Product R*W	This place holds the information of the passenger.

The following figures present examples of the different transitions that compose the model. These transitions are grouped in the transitions that model the different sitting situations that a passenger can face when boarding a plane. The remaining interactions are modelled taking advantage of the characteristics of the simulation program once the replication of the modules has been performed.

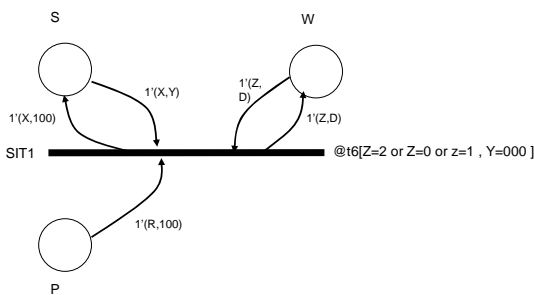


Figure 3: Transition for Sitting

Figure 3 illustrates the situation when a passenger has to sit at the window and the row of seats is empty. This situation is modelled by the value of the colour W for the passenger ( $W=100$ ). In this example 100 means that the passenger seat is the one in the window and the other two are not his (using a binary coding). The value of the variable Y is 000 which represents that the row is empty and the value of the colour Z can be either 2, 1 or 0. For this situation the result is the same, either no one is waiting for the passenger to sit or there are passengers that stood up for letting the passenger reach the window ( $Z=1$  or  $Z=2$ ). In this example the corresponding time consumption can be associated to the value of the variable t6, but it would depend on the correspondent study and maybe on the characteristics of the passenger sitting. Once the passenger is sat, the new colour value for Y is assigned with the output arc to S ( $Y=100$ ).

Another example is presented in Figure 4, which represents the situation in which a passenger must sit at a window seat ( $W=100$ ) and the middle seat is occupied by another passenger ( $Y=010$ ) who was previously sat and must walk out to let the former reach his window seat.

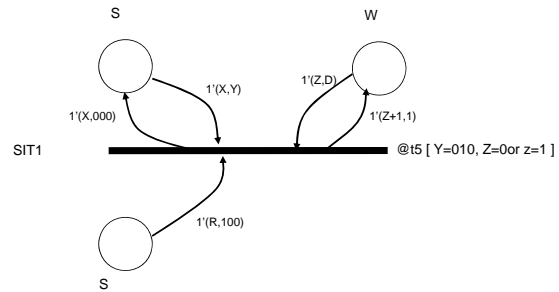


Figure 4: The Walking Out of a Passenger Seated in the Middle

In this model a unit is added to the colour Z of the token in the waiting place node and the variable D turns to 1; in this way track is kept for the number of passengers waiting until the correspondent passenger is seated. Once this event takes place the correspondent person has left the middle seat ( $Y=000$ ) and the corresponding time consumption is modelled by the value of t5. For the remaining transitions the authors encourage the reader to contact them but using the same logic the model can be developed.

Using the CPN approach it is possible to identify clearly the cause-effect relationships that sometimes hinder the smooth flow of passengers inside the cabin during the boarding or deboarding process. Moreover the colours can also be used to model characteristics such as age, size, number of bags, disabilities etc. and those characteristics can be used to simulate in a more accurate way the micro-processes that generate emergent dynamics once the people is interacting with each other.

#### 4. IMPLEMENTATION OF THE CPN MODEL

The CPN models can be integrated with the software tool making a mapping or implementing some logic such as the one presented by Mujica and Piera (2011). In this example the micro-simulation of the interaction present in the system is implemented following the methodology previously presented. Since the software uses an object-oriented modelling paradigm the simulation of the complete cabin is performed efficiently. As it has been mentioned, firstly, the different transitions of the CPN model are used to develop the modular object that represents a row inside the cabin. Secondly, advantage is taken from the use of a modular approach when the different rows of the cabin are put together in order to make a complete model for the cabin that takes into account not only the micro-interaction between passengers but also the interaction that occurs at higher levels e.g. aisle blocking, speed reduction duo to the passengers, etc.

Figure 5 illustrates the elements used in the module that simulates the interaction of people at seat-level. The methodology proposed by Mujica and Piera (2011) is used to implement the different transitions that occur during the seating process. The transitions implemented in the model are evaluated using the *Separator* objects

from SIMIO. The elements are configured in cascade using *Connectors* so that the logic associated to each object (CPN transitions) is evaluated at once and only those transitions that satisfy the different restrictions are fired thus performing the simulation with high accuracy. In Figure 5 the *TransX* objects correspond to transitions 1 to 10 and the logic behind each transition is implemented using the processes window of SIMIO.

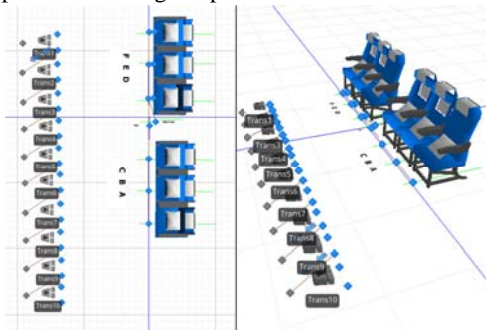


Figure 5: The Elements of the DES Software Model

Some SIMIO elements are used to model the place nodes and transition nodes. As mentioned earlier, the logic behind the transition nodes is coded using different *steps* within the *Processes* windows from SIMIO and the place nodes are modelled using the elements called *Stations*. Figure 6 presents the different stations used in the object; some of these stations are just used to store the entities that simulate that passengers sat in the cabin seats. Other stations, namely *HOLD* and *PAX*, are used to represent the place nodes *W* and *P* of the CPN model, respectively.

Station	
▲ SF	Station
▲ SE	Station
▲ SD	Station
▲ SA	Station
▲ SB	Station
▲ SC	Station
▲ HOLD	Station
▲ PAX	Station

Figure 6: Station Elements

Figure 7 illustrates the implemented logic for transition called SIT2 (General Sit). First the step *DECIDE* evaluates if the passenger wants to sit (**ModelEntity.Seat=100**), and if the seat row is empty (**Binary\_OccupiedL=000**) and if the people waiting is zero, one or two (**Hold. Contents**):

```
ModelEntity.Seat=="100"&&Binary_OccupiedL==0&&(Hold.contents==0||Hold.contents==1||Hold.contents==2)
```

If the condition is fulfilled then the new values are assigned (*ASSIGN*) to the variable that stores the availability of the row. The remaining steps are used to send the entity to the corresponding seat and the time is consumed depending on the characteristics of the entity and the length of the path (this is managed by the software itself).

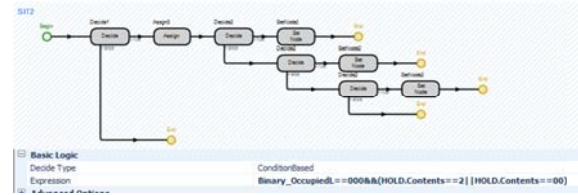


Figure 7: The CPN Logic Coded in SIMIO Steps

Once the module of the row is constructed, the development of the whole cabin is performed in a straightforward way. The top figure of Figure 8 illustrates the whole model of a cabin once the different row-modules are put.

Every time an entity (passenger) enters to a module the CPN logic behind the model will govern the simulation while the rest of the time the dynamics will be governed by the interaction of the entities themselves and with the different objects. Furthermore and for aesthetic purposes it is possible to hide those elements of the module such as the *Separators* that evaluate the different conditions of the transitions. In the bottom part of Figure 8 the sequence when the passenger needs to sit in the middle is illustrated. In the first snapshot 2 passengers are sit in the first row while another one approaches. The second snapshot illustrates the situation when the passenger in the aisle has to stand up so that the approaching passenger can get to his seat and in the following snapshots the passengers sit and let the remaining ones to continue to their corresponding seats. All of those events were modelled at the micro-level using the CPN model.

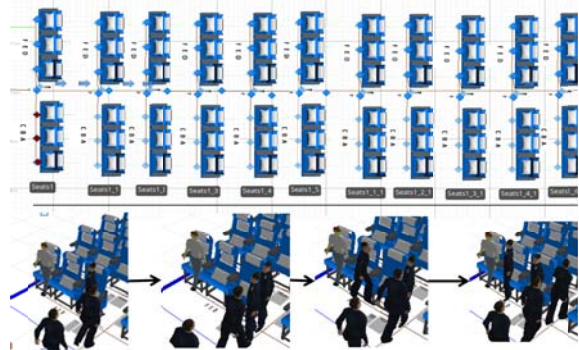


Figure 8: The different modules put together

## 5. CONCLUSIONS

A methodology for developing high-detailed simulators for the different operations at a cabin has been presented. Using this methodology is possible to construct detailed simulation models considering interactions between elements that otherwise could not be taken into account, such as age, disabilities, number of bags etc. The cause-effect relationships present between the different elements are modelled using the coloured Petri net formalism and afterwards the CPN models are integrated with a discrete-event software

called SIMIO to evaluate the emergent dynamics once the elements that participate in the boarding process are put together. The advantage of using this approach is that using the CPN formalism, the cause-effect relationships are revealed and modelled in a straightforward way, so that is very easy to integrate them in the digital model when developing the commercial simulator. Using this approach it would be possible to take these kinds of simulations one step further for analysing and optimizing processes where the interaction of small elements becomes important for the optimisation of the whole procedure.

The model presented here only uses the most basic micro-interactions such as passenger blocking, passenger standing up, speed of passengers however future implementations will take into account particular characteristics of the passengers such as carried luggage, aptitude for boarding, familiar relationships among others and those will be easily implemented using the CPN-Simulation approach.

#### ACKNOWLEDGEMENTS

NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

The work was also supported by the grant UNAM-PAPIIT No. IN116012

#### REFERENCES

- Airbus, 2014. *A320/A320NEO Aircraft Characteristics*. Technical Report.
- Bazargan, M., 2004. *Airline operations and Scheduling*. Burlington, USA: Ashgate Publishing Company.
- Christensen, S., Jensen, K., Mailund, T., Kristensen, L.M., 2001. State Space Methods for Timed Coloured Petri Nets. *Proceedings of 2nd International Colloquium on Petri Net Technologies for Modelling Communication Based Systems*, 33-42. September, Berlin, Germany.
- Jensen, K., 1997. *Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use*. Berlin, Germany: Springer-Verlag.
- Marelli, S., Mattocks, G., Merry, R., 1998. The Role of Computer Simulation in Reducing Airplane Turn Time. *AERO Magazine* 1.
- Moore, K.E., Gupta, S.M., 1996. Petri Net Models of Flexible and Automated Manufacturing Systems: A Survey. *International Journal of Production Research* 34(11): 3001-3035.
- Mujica, M., Piera, M.A., 2011. Integrating Timed Coloured Petri Net Models in the SIMIO Simulation Environment. *Proceedings of 2011 Summer Computer Simulation Conference*, 91-98. June, The Hague, The Netherlands.
- Soolaki, M., Mahdavi, I., Mahdavi-Amiri, N., Hassanzadeh, R., Aghajani, A., 2012. A New Linear Programming Approach and Genetic Algorithm for Solving Airline Boarding Problem.

*Applied Mathematical Modelling* 36(9): 4060-4072.

- Steffen, J., 2008. Optimal Boarding Method for Airline Passengers. *Journal of Air Transport Management* 14: 146-150.
- Steffen, J.H., Hotchkiss, J., 2012. Experimental test of airplane boarding methods. *Journal of Air Transport Management* 18: 64-67.
- Van Landeghem, H., Beuselinck, A., 2002. Improving Passenger Boarding Times in Airplanes: a Simulation Based Approach. *European Journal of Operations Research* 142: 294-308.

#### AUTHORS BIOGRAPHY

**Miguel Mujica Mota** studied Chemical Engineering in the Metropolitan Autonomous University of Mexico. He also studied a MSc. in Operations Research at the National Autonomous University of Mexico. After spending years in industry, in the year 2011 he obtained a PhD with highest honours in Industrial Informatics from the Autonomous University of Barcelona, and a PhD in Operations Research at the National Autonomous University of Mexico. His research interests lie in the use of simulation, modelling formalisms and heuristics for the performance analysis and optimization of aeronautical operations, manufacture and logistics.

**Idalia Flores** She received her Ph.D. in Operations Research at the Faculty of Engineering of the UNAM. She has been a referee and a member of various Academic Committees at CONACYT, and referee for journals such as Journal of Applied Research and Technology, the Center of Applied Sciences and Technological Development, UNAM and the Transactions of the Society for Modeling and Simulation International. Her research interests are in simulation and optimization of production and service systems. She is a full time professor at the Postgraduate Program at UNAM.

**Daniel Guimarans** is a researcher in the Optimisation Research Group and the Infrastructure, Transport and Logistics team at NICTA, a research centre of excellence in IT in Australia. He holds a PhD in Computer Science from the Autonomous University of Barcelona. His main research is devoted to solving logistics and transportation problems, specially focused on air and road transportation. His research has been aimed at hybridising different optimisation techniques, namely Constraint Programming and several heuristics and metaheuristics. Other interests include the development of Simheuristics, combining simulation and optimisation methods to obtain more robust solutions for stochastic systems.

# TOOL PROFILES EVALUATION BASED ON VIBROACOUSTICAL SIGNALS GENERATED BY FRICTION-STIR WELDING

Blanco-Fernández J. <sup>(a)</sup>, Jiménez-Macías E. <sup>(b)</sup>, Sanchez-Orozco, M. <sup>(c)</sup>, Sanchez-Roca A <sup>(d)</sup>, Carvajal-Fals, H. <sup>(e)</sup>

<sup>(a)</sup> Department of Mechanical Engineering, University of La Rioja, Logroño, Spain.

<sup>(b)</sup> Department of Electrical Engineering, University of La Rioja, Logroño, Spain.

<sup>(c,d,e)</sup> Faculty of Mechanical Engineering, University of Oriente, Santiago de Cuba, Cuba

<sup>(a)</sup> [julio.blanco@unirioja.es](mailto:julio.blanco@unirioja.es), <sup>(b)</sup> [emilio.jimenez@unirioja.es](mailto:emilio.jimenez@unirioja.es), <sup>(c)</sup> [mario@cepca.uo.edu.cu](mailto:mario@cepca.uo.edu.cu), <sup>(d)</sup> [sanchez@rect.uo.edu.cu](mailto:sanchez@rect.uo.edu.cu),  
<sup>(e)</sup> [carvajal@corpus.uo.edu.cu](mailto:carvajal@corpus.uo.edu.cu)

## ABSTRACT

This paper presents a study to assess the relationship between the tool design and vibro-acoustic signals of the friction stir welding process in the aluminium AA1050. The characterization in time and frequency domains (Martinez-de-Pison et al., 2012) of vibro-acoustical signal and the statistical analysis have been carried out in order to correlate them with the design of two different tools (Jimenez-Macias et al., 2010). Vibro-acoustical signals have been filtered in order to eliminate the noise of the friction stir welding machine. Additionally, the Wavelet Transform has been used as an efficient tool for filtering vibration signal waveforms. The statistical analysis has confirmed that the vibro-acoustical signals were significantly affected by the tool design. It was found that there is valid information to assess changes in the tool profile in the frequency range of 0 to 100Hz. Analysis techniques based on the Wavelet Transform constitute a tool that can be efficiently used in the assessment of changes in the tool profile.

Keywords: Friction stir welding, signal processing, tool design, Wavelet Transform

## 1. INTRODUCTION

Friction stir welding (FSW) is a solid-state welding process with several advantages such as: energy efficiency, environment friendly, and versatile. Furthermore, it is a versatile joining technique, energy-efficient and with low environmental impact. The process is relatively simple: the pressure, friction and stir generated by tool, composed of a probe and a shoulder moving forward throughout the line of joint, plasticizes and forges the surrounding metal, achieving the welded joint. The material plastically deformed is transferred from the advancing side of the tool to the retreating side which allows the joint of two metal sheets in solid state (Threadgill 2007).

This process is considered to be the most relevant in terms of metal welding over the last two decades. It is widely used in automotive and aeronautics applications with remarkable results (Uday 2010; Aldanondo 2011;

Martinez-de-Pison et al., 2010). Many researches have been devoted to FSW of aluminium alloys (Hassan 2010; Szkodo2010; Xunhong 2008; Cavaliere 2008).

Hassan (2010), reported the effect of the tool rotation speed and travel speed on the microstructure and the mechanical characteristics of the aluminium weld made of different metal sheets from the alloys A139 and the laminated A356 by friction stir welding. The authors proved that an increase in the rotational speed reduces the resistance to traction but increases the softness of the joint.

Szkodo (2010), assessed the parameters of the friction stir welding process of the aluminium alloy AW7075-T651 by means of a non-destructive technique so they could analyse and prove the relationship between the deformation in the mixture and the thermo-mechanically affected zones by the parameters of the process.

Several studies have covered the influence of the welding parameters and the tool profile in the microstructure and the properties of welded joints of aluminium (Cederqvist 2009; Mahmoud 2009; Hattingh 2008; Padmanabanand Balasubramanian 2009; Vijay and Murugan 2010).

Mahmoud (2009) analysed the effects of three different diameters and four different tool probes in the production of aluminium surfaces reinforced with SiC particles.

Hattingh (2008), developed a characterization of the influence of the FSW tool profile in the forces that take place in the process and the tensile strength of the welded joint. They took into consideration the influence of the tool profile as well as the visual information of the interaction between tool profile and plastic stir zone.

Valdameri and Esmerio (2009), performed an experiment in which they analysed the influence of the tool profile and the parameters of the FSW process of the aluminium alloy AA5052 and compared their results with other welding with the same metal sheets but using the MIG process. To achieve that, they used three different tools designs and different welding conditions. Finally, through a macro-graphic analysis and a set of tests of the mechanical properties, they reached to the



conclusion that the AA5052 metal sheets welded by FSW with the proper tool and parameters have better mechanical properties than those welded with the MIG process (Martinez, Eguia and Godinez, 2012).

Tool profile is considered one of the main parameters in the control of the flow of material, heat generation and joint quality. Rai (2011), asserted that, within the tool profile, factors such as, shoulder diameter, shoulder surface angle, probe design and the nature of the tool surface play a key role in the process.

The feasibility of the application of the analysis of the vibro-acoustic emission generated by the FSW process is a topic of great interest for the field of Engineering due to the need of non-invasive techniques for the characterization and control of this kind of process. Nowadays, there are not enough studies available regarding the subject (Suresha, Rajaprakash and 2009; Burford 2010; Soundararajan, Atharifar and Kovacevic 2006; Fernández et al, 2012; Orozco et al, 2013,2013b).

Soundararajan, Atharifar and Kovacevic(2006) and Macias et al (2013), evaluated the possibility of the application of techniques of analysis of acoustic emission (AE) to monitor the FSW process. Based on the AE signal generated, the author determined the correlation between the loss of contact of the tool and the parts and the state of the welding process. The characteristics of the signal corresponding to the interaction with the parts and the state of the welding were studied using the Fast Fourier Transform (FFT) and the Discrete Wavelet Transform (DWT). The author explained how through the identification of frequencies during the process and the analysis of the signal by means of the Wavelet Transform, it is possible to monitor effectively the contact or lack of contact of the tool, or the transition state of the welding, and thereby accurately identify the changes in the process.

Other studies from Chen, Kovacevic and Jandgric (2003), show the application of the AE signals in the FSW processes. The authors assert that the FFT methods are not appropriate for these processes. Their research focuses on processing AE signals to determine how valid their use is to determine the abrupt changes on the tool profile.

All the bibliography consulted focuses on FSW for aluminum alloys, their parameters and their influence in the micro-structural changes and mechanical properties of the joint. However, there is no information available of the influence of the tool profile analysed through AE techniques.

The main objective of this work is the assessment of the influence of tool profile in the vibro-acoustic signals in order to prove the possibility of the application of this technique to characterize FSW processes (Roca et al., 2007, 2008, 2009).

## 2. MATERIALS AND METHODS

The metal sheets used for the Research were made of aluminium AA1050 H24 with the following dimensions: 200 mm (L), 100 mm (A) and 3 mm (E). This alloy is well known for its conductivity and it is

commonly used in the automotive industry, for the equipment of the chemical and food industry, light reflectors, and heat transfer units. This alloy is also used in other studies about the FSW process.

The chemical composition (wt%) of the AA1050 H24 used was as follows: Si = 0,25; Fe = 0,40; Cu = 0,05; Mn = 0,05; Mg = 0,05; Zn = 0,07; Ti = 0,05, Al = 99,5 and 0,03 corresponding to other elements.

The FSW process was made by means of a conventional CNC machine with position control in the three axes. The plates were butt-joined and locked in position using mechanical clamps and the direction of welding was normal to the rolling direction.

An on-line monitoring system was implemented to detect the changes of the tool profile reflected in the vibroacoustic signal (Figure 1).

Two vibroacoustic sensors (piezoelectric accelerometers) were placed at right angles in the upper (axis Z) and lateral (axis Y) surfaces of support plate. The sensors were placed in the base in order to make the experiments to be as close to an industrial environment as possible. A threaded coupling was used in order to guarantee a proper vibroacoustic transduction between the support plate and the sensor.

The values of the acceleration of the vibroacoustic signal in the directions Z and Y were taken using an NI USB-9234 instrument connected to a PC. All signals were sampled at 51.2 KS/s and processed using a software designed specifically for this purpose. Initially the generated vibroacoustic signal was acquired while the machine worked in a vacuum, with the objective of characterizing the noise contained in the signal.

The diameters of the probe and shoulder were 3 and 10 mm respectively. Two different probe designs were considered: a conventional grooved cylindrical probe (Tool 1) and a cylindrical probe with horizontal flutes (Tool 2). The shoulder profile for these tools was concave and similar for both tool designs.

During the tests, three levels of rotation speed ( $\omega$ ) and travel speed ( $v$ ) were used (two boundaries and a central level). A factorial experimental design was used. To reduce the experimental error, the welding was carried out three times under the same conditions.

### 2.1. Methodology to evaluate the tool profile using the vibro-acoustic signals

Initially, a comparison of the vibro-acoustic signal in the time domain was made for each one of the analysed tool profiles. Also, a characterization of the frequency range of the background noise was performed in order to remove it from the vibro-acoustic signals.

Taking into account that the analysed signals are stationary, the assessment was made by means of the Wavelet Transform as an efficient tool for this kind of signal processing (Chen, Kovacevic and Jandgric (2003). It allows obtaining a higher resolution at the low frequencies, which turns to be of interest in the analysed process, taking into account the range of rotation speed evaluated in the research.



In this paper, for the decomposition of vibro-acoustic signals in the frequency ranges of interest, a level-8 Wavelet and a Daubechies mother Wavelet (Db5) were used.

Finally, the statistical analysis was carried out for the validation of experimental results. The results of the statistical tests were used to demonstrate the relationship between the design of the tool and the vibro-acoustic signals generated by the FSW process.

### 3. RESULTS AND DISCUSSION

In the FSW process, tool geometry is one of the main causes of mixing and flow of material. Probe and shoulder design, and the relative dimensions of their geometry are decisive (Widener, Burford and Jurak 2010).

Taking into account the above statement, it was verified that the tool profile may influence the flow of the material during the FSW process as well as the vibro-acoustic signal generated as a result of the physical phenomena taking place. The results for selected conditions are shown below.

Figure 1 shows the vibro-acoustic signals corresponding to two conditions with the same process parameters ( $\omega = 450$  rpm,  $v = 100$  mm•min<sup>-1</sup>) and for the two different tool profiles evaluated.

Figure 3a corresponds to the results of the use of the conventional grooved cylindrical probe (Tool 1) and Figure 3b corresponds to the cylindrical probe with horizontal flutes (Tool 2).

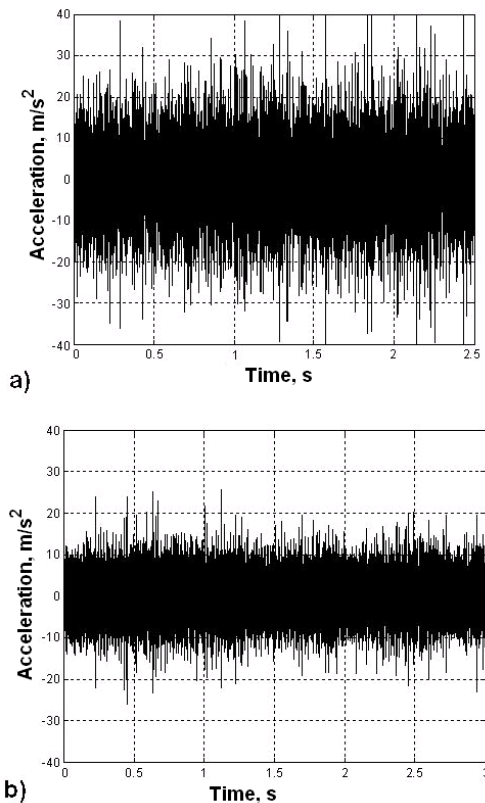


Figure 1: Vibro-acoustic signals for the two tool profiles ( $\omega = 450$  rpm,  $v = 100$  mm•min<sup>-1</sup>). a) Tool 1; b) Tool 2

As shown in Figure 1, the vibro-acoustic signal level is higher for  $\omega = 450$  rpm,  $v = 100$  mm•min<sup>-1</sup> and Tool 1. Vibro-acoustic signals are generated mainly by the friction between the tool and the material and the deformation undergone by the material.

This increase in the vibration levels may be associated with an increased contact surface between the tool and the material. Table 1 shows the values of the statistical parameters of the analyzed signals, and allows observing the changes of the tool profile in the parameters of the vibro-acoustic signals obtained during the process.

Table 2: Statistical characterization of vibro-acoustic signals.

Condition	RMS	Variance
Tool 1	6.35	40.34
Tool 2	4.01	16.11

In the case of cylindrical probe with horizontal grooves, much of the material flow occurs by simple extrusion, and the material does not seem to have vertical movement, which is apparently necessary to stabilize the rotational zone and provide enough deformation of the material to obtain a high quality welding.

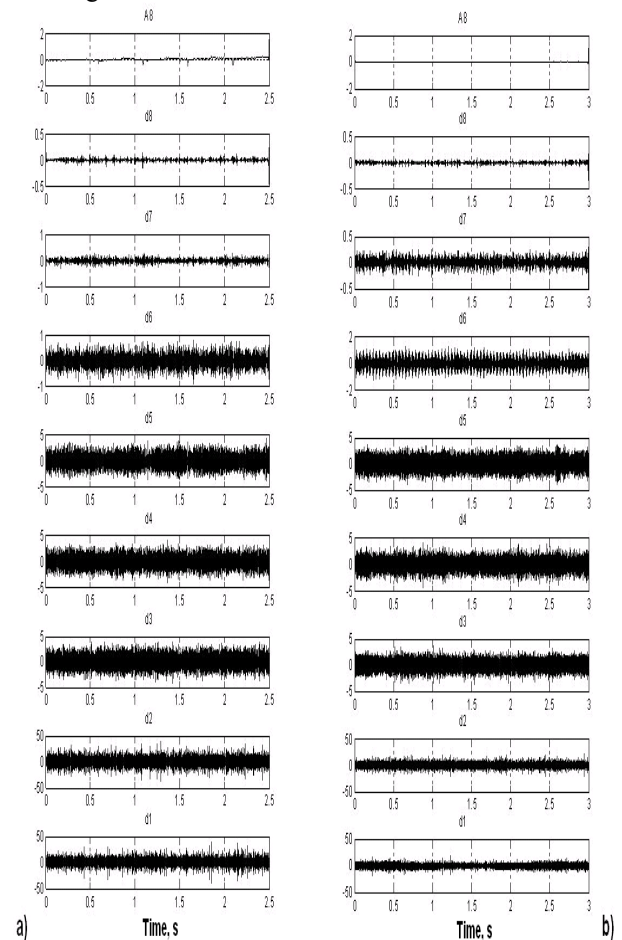


Figure 2: Wavelet transform for the Tool 1.

The presence of the grooves in Tool 1 improves the vertical flow and mixing, causing an increment in the vibro-acoustic signal levels, as shown in Figure 1

### 3.1. Wavelet Transform

Considering that the FFT cannot properly describe the characteristics of the AE signal at low frequencies, it was decided to use other signal processing techniques that are feasible for this purpose.

The result of the DWT is a series of decomposed signals belonging to different frequency bands.

The graphs of Figures 2 and 3 shows the Wavelet transforms of the vibro-acoustic signals for the analyzed conditions. In this case, the DWT decomposition is applied using 8 levels and a Mother Wavelet Daubechies (Db5).

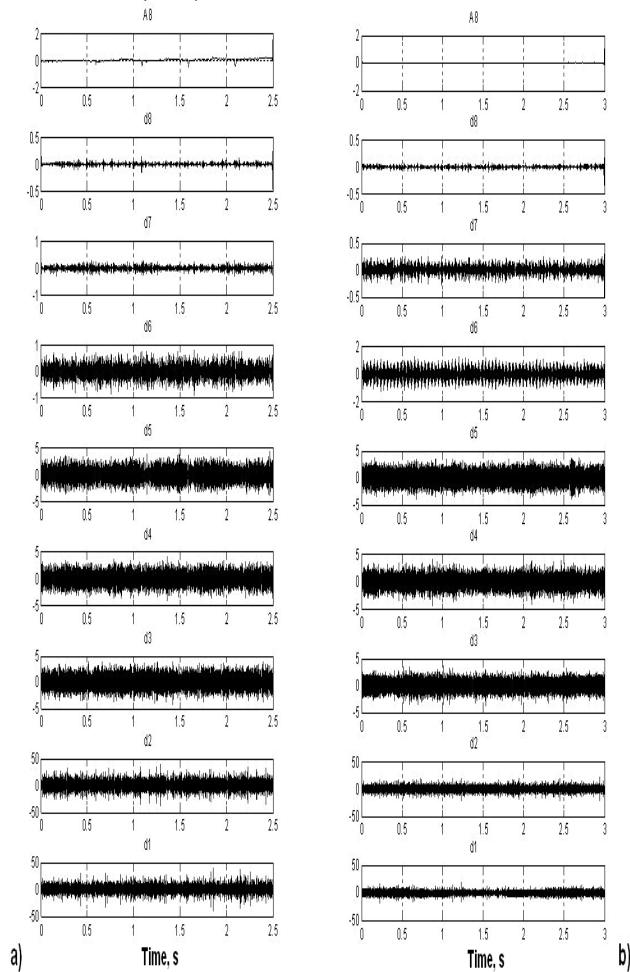


Figure 3: Wavelet transform for the Tool 2

A8 signal (Figures 2 and 3) can be considered as the result of making a lowpass filter to the vibro-acoustic signal with a resulting bandwidth of [0, 100 Hz]. D1-D8 correspond to the frequency bands [100, 200 Hz], [200, 400 Hz], [400, 800 Hz], [800 Hz, 1.6 kHz], [1.6, 3.2 kHz], [3.2, 6.4 kHz], [6.4, 12.8 kHz] and [12.8, 25.6 kHz], respectively.

A similar behavior for the two types of tool analyzed showed the decompositions D2, D3, D4, and D5. A comparison of D8 decomposition in the two

conditions studied shows a difference in the signal frequency band [12.8, 25.6 kHz].

Therefore, in this work the Wavelet Transform has been used to obtain a series of vibro-acoustic signals decomposed in the frequency ranges of interest.

Figure 4 shows the approximation A8 for both conditions corresponding to the two tool profiles analyzed.

The analysis of the A8 approximations presented in Figures 4a and 4b (with a bandwidth [0, 100 Hz]) shows a visible difference at low frequencies between signals of the different tool profiles.

With the objective of comparing the two types of tool profile and correlate them with vibro-acoustic signals at low frequencies some signal characteristics have been extracted, as shown in Table 2. The root mean square (RMS) and the variance are the statistical parameters that have represented the changes in the signals extracted from the DWT.

Table 2: Characterization of the A8 DWT of the vibro-acoustic signals.

Condition	RMS	Variance
Tool 1	0.09	0.009
Tool 2	0.03	0.0008

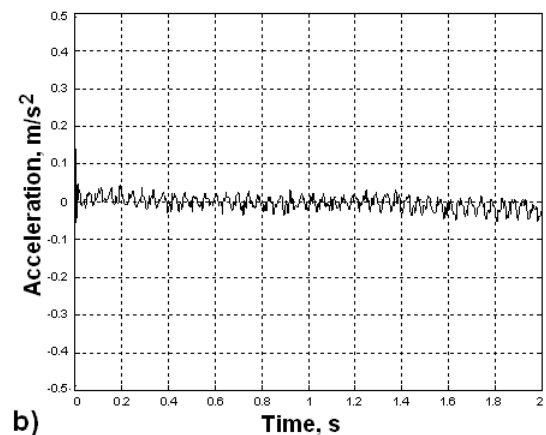
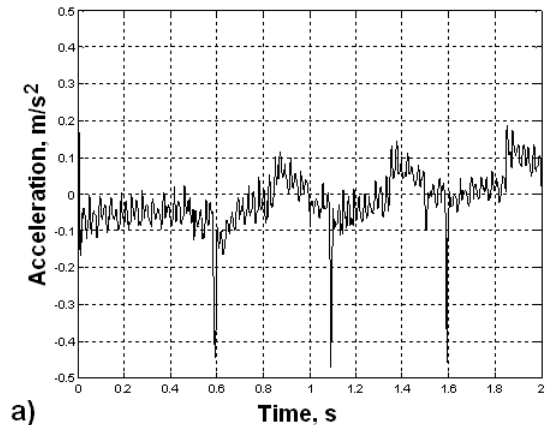


Figure 4. Approximation A8 for the tool profiles ( $\omega=450$  rpm y  $v = 100$  mm·min<sup>-1</sup>). a) Tool 1; b) Tool 2.

Based on the statistical analysis, the fact that the tool profile significantly affects signals studied at low frequency (A8) is proved. This result may be related to the presence of the three grooves in Tool 1. This feature of the tool increases the resistance offered by the motion of the probe through the work piece.

An assessment of the effect of changes in parameters of the FSW process in the vibro-acoustic signals, weld quality, and mechanical properties of the joint, will be made in successive works as a continuation of this subject.

#### 4. CONCLUSIONS

This research examines the impact of the tool profile on vibro-acoustic signals generated during FSW of the aluminum alloy AA1050 H24. From the analysis made it is concluded that:

1. The features of the vibro-acoustic signals in the frequency range of 0-100 Hz hold valid information to evaluate the changes in tool profile during friction-stir welding.
2. Vibro-acoustic signals may be applied as an effective on-line monitoring method of the changes in tool profile.
3. Junctions free of wormhole-type defects were obtained with the tool that incorporates a conventional grooved cylindrical probe; it was corroborated in the macro structural analyses made to all the test-tubes.
4. Analysis techniques based on the Wavelet Transform constitute a valid instrument that can be efficiently used to evaluate the changes in tool profile.

#### ACKNOWLEDGMENTS

This paper has been partially supported by the project of the Ministry and Innovation “DPI2011-25007. Soldadura por fricción-agitación de materiales disimilares. Caracterización mediante técnicas de emisión acústica e inteligencia artificial”.

#### REFERENCES

Aldanondo, E., 2011. Avances en la tecnología de soldadura por fricción “STIR”. *DYNA Ingeniería e Industria*, 86(6): 614-618.

Burford, D., 2010. Evaluation of friction stir weld process and properties for aerospace application: e-NDE for friction stir processes. *FAA Joint Advanced Materials & Structures (JAMS)*. 6th Annual Technical Review Meeting, 1-27. May 19–20.

Cavaliere, P., Panella, F., 2008. Effect of tool position on the fatigue properties of dissimilar 2024-7075 sheets joined by friction stir welding. *J. of Mater. Proc. Technol.*, 206: 249-255.

Cederqvist, L., 2009. Improved process stability during friction stir welding of 5 cm thick copper canisters through shoulder geometry and parameter studies. *Sci. Technol. Weld. Join.*, 14(2): 178-184.

Cesar Sanchez-Orozco, M., Blanco-Fernandez, J., Jimenez-Macias, E., & Perez-de-laParte, M. (2013). VIBROACOUSTICAL SIGNALS TO EVALUATE TOOL PROFILES ON FRICTION STIR WELDING PROCESS. *Dyna*, 88, 318-324. doi: 10.6036/5326

Chen, Ch., Kovacevic, R., Jandgric, D., 2003. Wavelet transform analysis of acoustic emission in monitoring friction stir welding of 6061 aluminum. *Int. J. of Mach Tools & Manuf.*, 43: 1383–1390.

Fernández, J.B., Roca, A.S., Fals, H.C., Macías, E.J., de la Parte, M.P. 2012. Application of vibroacoustic signals to evaluate tools profile changes in friction stir welding on AA 1050 H24 alloy. *Science and Technology of Welding and Joining* 17(6):501-510.

Hassan, A.S. 2010. Friction stir welding of dissimilar A319 and A356 aluminium cast alloys. *Sci. Technol. Weld. Join.* 15(5): 414-422.

Hattingh, D.G., 2008. Characterization of the influences of FSW tool geometry on welding forces and weld tensile strength using an instrumented tool. *J. of Mater. Proc. Technol.*, 203: 46-57.

Javier Martinez-de-Pison, F., Sanz, A., Martinez-de-Pison, E., Jimenez, E., & Conti, D. (2012). Mining association rules from time series to explain failures in a hot-dip galvanizing steel line. *Computers & Industrial Engineering*, 63, 22-36. doi: 10.1016/j.cie.2012.01.013

Jimenez Macias, E., Sanchez Roca, A., Carvajal Fals, H., Blanco Fernandez, J., & Perez de la Parte, M. (2010). Time-frequency diagram applied to stability analysis in gas metal arc welding based on acoustic emission. *Science and Technology of Welding and Joining*, 15, 226-232. doi: 10.1179/136217109x12665778348588

Jiménez-Macías, E., Sánchez-Roca, A., Carvajal-Fals, H., Blanco-Fernández, J., Martínez-Cámara, E. 2014. Wavelets application in prediction of friction stir welding parameters of alloy joints from vibroacoustic ANN-based model. *Abstract and Applied Analysis*. Volume 2014, Article number 728564.

Macías, E.J., Roca, A.S., Fals, H.C., Fernández, J.B., Muro, J.C.S.-D. 2013. Neural networks and acoustic emission for modelling and characterization of the friction stir welding process. *RIAI-Revista Iberoamericana de Automatica e Informatica Industrial* 10(4): 434-440.

Mahmoud, E., 2009. Effect of friction stir processing tool probe on fabrication of SiC particle reinforced composite on aluminium surface. *Sci. Technol. Weld. Join.*, 14(5): 413-425.

Martinez, E.C., Miguel, V.E., Godinez, C.S., 2012. Análisis de la penetración en la soldadura de una estructura metálica. *DYNA Ingeniería e Industria.*, 87(1): 90-97

- Martinez-de-Pison, F. J., Pernia, A., Jimenez-Macias, E., & Fernandez, R. (2010). Overall model of the dynamic behaviour of the steel strip in an annealing heating furnace on a hot-dip galvanizing line. *Revista De Metalurgia*, 46, 405-420. doi: 10.3989/revmetalm.0948
- Orozco, M.S., Macías, E.J., Roca, A.S., Fals, H.C., Fernández, J.B. 2012. Optimisation of friction-stir welding process using vibro-acoustic signal analysis. *Science and Technology of Welding and Joining* 18 (6): 532-540
- Padmanaban, G., Balasubramanian, V., 2009. Selection of FSW tool pin profile, shoulder diameter and material for joining AZ31B magnesium alloy – An experimental approach. *Mater and Design*,30: 2647–2656.
- Rai, R., 2011. Review: Friction Stir Welding Tools. *Sci. Technol. Weld. Join.*, 16(4): 325 – 342.
- Roca, A. S., Fals, H. C., Fernandez, J. B., Macias, E. J., & Adan, F. S. (2007). New stability index for short circuit transfer mode in GMAW process using acoustic emission signals. *Science and Technology of Welding and Joining*, 12, 460-466. doi: 10.1179/174329307x213882
- Sanchez Roca, A., Carvajal Fals, H., Blanco Fernandez, J., Sanz Adan, F., & Jimenez Macias, E. (2008). Stability analysis of the gas metal arc welding process based on acoustic emission technique. *Soldagem & Inspecao*, 13, 61-69.
- Sanchez Roca, A., Fals, H. C., Fernandez, J. B., Macias, E. J., & de la Parte, M. P. (2009). Artificial neural networks and acoustic emission applied to stability analysis in gas metal arc welding. *Science and Technology of Welding and Joining*, 14, 117-124. doi: 10.1179/136217108x382981
- Soundararajan, V., Atharifar, H., Kovacevic, R., 2006. Monitoring and processing the acoustic emission signals from the friction-stir-welding process. *Proc. of the Inst. of Mech. Eng.* 220-10: 1673-1686. March.
- Suresha, C.N., Rajaprakash, B.M., Sarala, U., 2009. Applicability of acoustic emission in the analysis of friction stir welded joints. *Int. J. of Recent Trends in Eng.*, 1(5): 86-89.
- Szkodo, M., Adamowski, J., Sitko, A., 2010. Assessment of FSW Welds Made of Aluminum Alloy AW7075-T651. *Solid State Phenom.*165: 201-206.
- Threadgill, P.L., 2007. Terminology in friction stir welding. *Sci. Technol. Weld. Join.*12(4): 357-360.
- Uday, M.B., 2010. Advances in friction welding process: a review. *Sci. Technol. Weld. Join.* 15(7): 534-558.
- Valdameri, T., Emerio, J.A., 2009. Avaliação da Geometria de Ferramenta e Parâmetros do Processo FSW na Soldagem da Liga de Alumínio AA 5052. *Soldagem & Inspeção*, 14(3): 215 – 227.
- Vijay, S.J., Murugan, N., 2010. Influence of tool pin profile on the metallurgical and mechanical properties of friction stir welded Al–10 wt.% TiB2 metal matrix composite. *Mater. and Design.*, 31: 3585–3589.
- Widener, C., Burford, D., Jurak, S., 2010. Effects of Tool Design and Friction Stir Welding Parameters on Weld Morphology in Aluminum Alloys. *Mat. Sci. Forum.* 638–642: 1261-1266.
- Xunhong, W., 2008. Comparison of fatigue property between friction stir and TIG welds. *Journal Univ. Sci. Technol. Beijing.*15(3): 280-284.

# EVALUATION OF THE EROSIWE WEAR OF THE AA 6082 T6 ALLOY TREATED WITH FRICTION STIR PROCESSING

Martínez-Cámara E. <sup>(a)</sup>, Blanco-Fernández J. <sup>(b)</sup>, Chacón, Y. P. <sup>(c)</sup>, Sanchez-Roca A. <sup>(d)</sup>, Carvajal-Fals, H. <sup>(e)</sup>

<sup>(a,b)</sup> Department of Mechanical Engineering, University of La Rioja, Logroño, Spain.

<sup>(c,d,e)</sup> Faculty of Mechanical Engineering, University of Oriente, Santiago de Cuba, Cuba

<sup>(a)</sup> [ecamara@eolicas.com](mailto:ecamara@eolicas.com), <sup>(b)</sup> [julio.blanco@unirioja.es](mailto:julio.blanco@unirioja.es), <sup>(c)</sup> [yuni@fim.uo.edu.cu](mailto:yuni@fim.uo.edu.cu), <sup>(d)</sup> [sanchez@rect.uo.edu.cu](mailto:sanchez@rect.uo.edu.cu),  
<sup>(e)</sup> [carvajal@corpus.uo.edu.cu](mailto:carvajal@corpus.uo.edu.cu)

## ABSTRACT

This paper makes an evaluation of the influence of the parameters of friction stir process: number of passes and tool rotation speed, in erosive wear resistance of the alloy AA6082 T6. Samples were processed with one, two and three passes with an overlap of 100%. For the wear tests, a mixing vessel tribometer was used with rotary motion for tilt angles of particles of 30° and 90°. The number of passes, the tool rotation speed and the tilt angle of the particles has an influence on the resistance to erosive wear of the alloy AA6082T6. The best performance in terms of erosive wear resistance for the tilt angle of 30° is achieved when the material is processed at 1800 rpm with three passes. When the tilt angle of the particles was 90°, samples treated with two passes showed better erosive wear behaviour, regardless of the tool rotation speed.

Keywords: Surface, friction stir processing, wear.

## 1. INTRODUCTION

The superficial modification by Friction-Stir Processing (FSP) causes an intense plastic deformation and generates a mild heating in the processed area; these thermodynamic conditions favour the development of recrystallized fine grains, with excellent properties, resistance to fatigue, corrosion and superficial wear. This process is considered to be the most relevant in terms of metal welding over the last two decades. It is widely used in automotive and aeronautics applications with remarkable results (Uday 2010; Aldanondo 2011). Many researches have been devoted to FSW of aluminium alloys (Hassan 2010; Szkodo 2010; Xunhong 2008; Cavaliere 2008, Fernandez et al, 2012; Orozco et al, 2012, 2013; Jimenez et al, 2010, 2014).

Santella (2005) studied the effect of the application of FSP over the mechanical properties of molten aluminium alloys (A356 and A319) and observed a fine equiaxial microstructure in the stir zone; the tensile strength, ductility and resistance to fatigue of both alloys improved.

The application of FSP with multiple passes for the creation of superplastic 7075 aluminium was studied by Johannes and Mishra (2007).

Magdy (2012) reported an increase of the hardness with the increase of processing speed, which decreases with the increase of the number of passes in the

aluminium alloy AA 6082-T9. They also concluded that with the increase of the number of passes the tensile strength increases.

Most researches study the influence of the application of the FSP in the microstructure, micro-hardness and tensile strength (Martinez-de-Pison et al, 2010, 2012); and there is a lack of studies about the influence of the FSP parameters in the resistance to wear. This paper assesses the influence of the parameters: number of passes and tool rotation speed in the erosive wear of the aluminium alloy AA 6082-T6 modified by the FSP technique.

## 2. MATERIALS AND METHODOS

The material used for this research was the aluminium alloy AA 6082-T6. The chemical composition (wt%) of the material used for this research was as follows Si = 1,2; Fe = 0,5; Mn = 0,95; Mg = 1; Cu = 0,10; Cr = 0,25; Zn = 0,20; Al = 95,65 and 0,15 corresponding to other elements. The dimensions of the test samples were 150x150x3 mm.

The FSP process was made by means of a CNC milling machine with position control and three tool rotation speeds were used (710, 1120 and 1800 rpm) and a constant tool advancing speed of (100mm/min), with a variation from 1 to 3 passes and a 100% overlapping for each pass.

The processing of the material was made perpendicular to the laminating direction. The inter-pass temperature was constant and room temperature was always used.

The tool used was made of AISI D2 steel, and was designed with a fluted cylindrical probe, 3mm diameter and 1.6 mm long. The tool shoulder has a diameter of 10mm with a concave profile.

Measuring of the micro-hardness was made in the different regions of the transversal section of the material treated. The measuring was made using a micro-hardness tester "Shimadzu Micro Hardness Tester Type M". The weight used to measure micro-hardness was of 25 grams with duration of five seconds.

The tests of erosive wear were made with a tribometer of mixture recipient with rotatory motion and two devices were used with a tilt angle of 30° and 90°. The area exposed to erosive wear was removed from the processed zone, without any influence of the base metal.



Figure 1 shows a representative image of the FSP process and erosion test.

The erosive mixture was composed of distilled water (1.8 litres) and abrasive particles of silica sand (300 g/litre). The size of the abrasive particles ranged from 30 and 50  $\mu\text{m}$  diameter (Figure 2) (Roca et al., 2007, 2008, 2009). An on-line monitoring system was implemented to detect the changes of the tool profile reflected in the vibroacoustic signal (Figure 1).

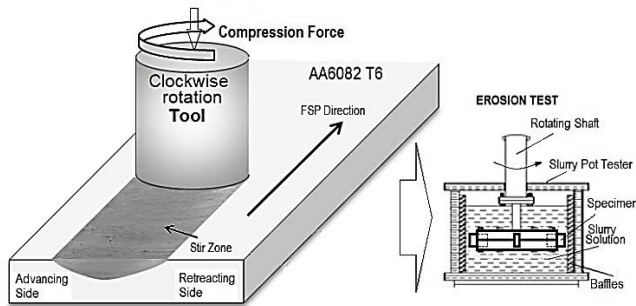


Figure 1. Representation of Friction Stir Process and Slurry Pot Tester

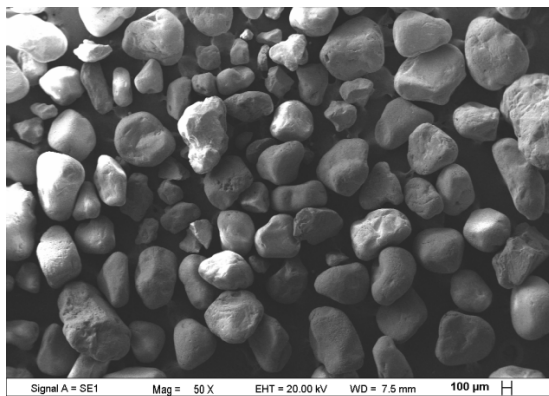


Figure 2: Silica abrasive particles

The impact speed of the particles was of 4.15 m/s and the tests were made for 3 hours, with 30 minute intervals. The wear was determined by the difference in weight of the samples before and after each pass; the erosive mixture was substituted by each interval and all the tests were repeated 3 times.

### 3. RESULTS AND DISCUSSION

Figure 3 shows the different areas in the process zone, such as, nugget zone (NZ), thermo mechanically affected zone (TMAZ), and heat affected zone (HAZ). In Figure 3, one can observe the good quality of the regions. The macro-images of the process zone presented in Fig. 3 show no defects and macro-structural continuity.

The tool rotation speed and the number of passes influenced the microstructural transformations that took place in the surface of the material. In the nugget zone,

a total recrystallization of the grains took place and in the thermo-mechanically affected zone, a partial recovery.

An image corresponding to the experimental condition of 710 rpm with one pass, representative macrostructure is shown in Figure 3. Figure 1 shows the vibro-acoustic signals corresponding to two conditions with the same process parameters ( $\omega = 450 \text{ rpm}$ ,  $v = 100 \text{ mm}\cdot\text{min}^{-1}$ ) and for the two different tool profiles evaluated.

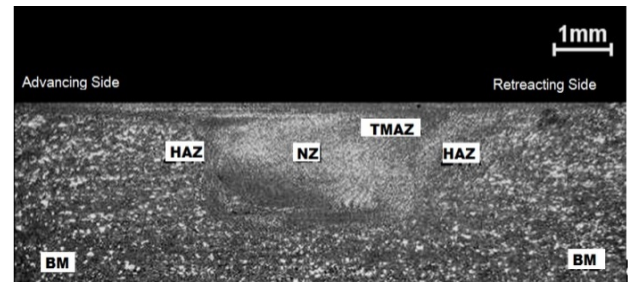


Figure 3. Representative macrograph image of the cross section of specimens processed with 710 rpm and one pass.

A central zone or core can be seen, which corresponds to the material that is subjected to stirring in plastic state where a high refining of the size of the grain occurs. Later on, a thermo-mechanically affected zone where the material suffers a high deformation and the elongated grains of the base metal are deformed in an advancing flow pattern around the core. Following that, a thermo-mechanically affected zone can be seen at both sides of the footprint left by the tool, characterized by the occurrence of a thermal cycle where the grains keep the same structure of the base metal and finally the area of the base metal where one can see the microstructure of the original material.

In both zones, the fine precipitates ( $\beta\text{-Mg}_5\text{Si}_6$ ), present in the base metal and responsible for the hardening of the original material, were dissolved in the matrix and did not re-precipitate when using 710 rpm and 1200 rpm.

A precipitate of a bigger size was formed in the thermo-mechanically affected zone ( $\beta\text{-Mg}_1.7\text{Si}$ ), corroborating the results of Moreira (2009). The processes of recrystallization, microstructural recovery and solubilisation (710 rpm y 1200 rpm), and the reformation of new precipitates (1800 rpm) caused changes in the strength of the material and thus influenced the behaviour of the erosive wear of the aluminium alloy AA 6082-T6.

Figure 4 shows a representation of the micro-hardness profile of the cross section of one of the samples for the condition with 710 rpm and three passes.



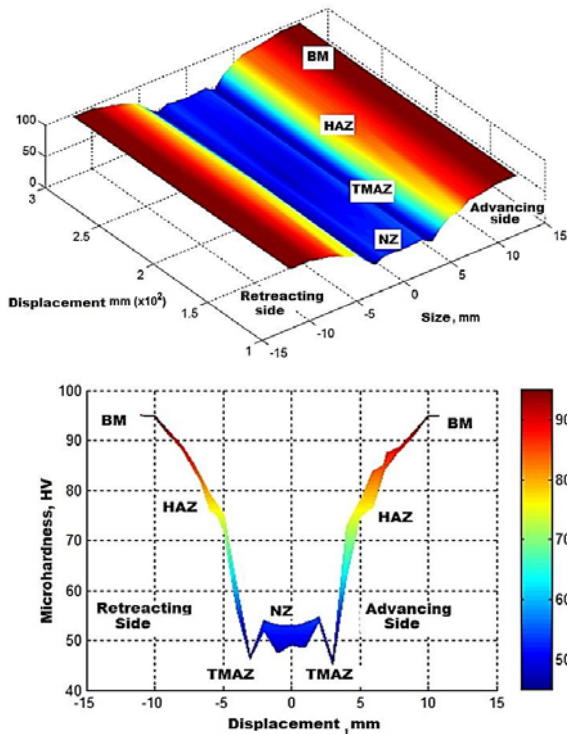


Figure 4: Micro-hardness behaviour for 710 rpm and three passes. UP) Aerial view; DOWN) Cross section.

The core zone (ZN) experiences a decrease in micro-hardness due to the deformation of the material and the changes resulting from the dynamic recrystallization that takes place in the process. It can also be seen that in the thermo-mechanically affected zone (ZTMA), there is a minimum of micro-hardness compared to the ZN more inclined to the advancing side of the tool.

According to Sterling (2005), the distribution of the micro-hardness profiles is due to the different morphologies, grain sizes and the bands of deformed material. These authors assert that, the microstructural result is due to the processes of dynamic restoration-recrystallization reached, in accordance with the thermal cycle and the variation of the deformation speed. The latter decreases rapidly as the tool pin separates due to the viscous dissipation of the plastically deformed material in the core and the ZTMA, both in a horizontal and vertical way (Nandan, 2008).

Figure 5 shows the micro-hardness maps in accordance with the number of passes for the tool rotation speed of 710rpm.

Figure 5 shows that for the condition 710 rpm the higher values of micro-hardness are reached when the material is processed with one pass, which is attributed to the refining of the size of the grain of the superficial layer of the process of dynamic recrystallization that takes place in the zone. This corroborates the results of Surekha (2009), Mishra (2005) and Mroczka and Pietras (2009) who assert that, with a lower heat input, that is, a low rotation speed, a structure with a quite refined grain is obtained.

On the other hand, with the increase of the number of passes at the same tool rotation speed, the micro-hardness decreases due to the growth of the grain of the superficial layer resulting from the thermal cycles accumulated in each pass, parallel to the continuous dynamic recrystallization that takes place. This result ratifies what Johannes and Mishra (2007) have stated in previous works.

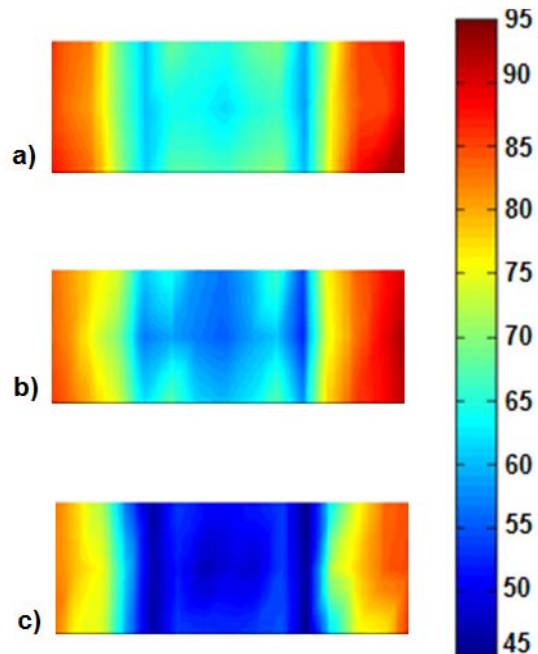


Figure 5: Micro-hardness maps for 710 rpm of tool rotation speed. a) 1 pass, b) 2 passes, c) 3 passes.

Figure 6 shows the results of the accumulated erosive wear for the different combinations of parameters and tilt angle of the particles of 30°.

The curves obtained (Figure 6) show that the increase in tool rotation speed and number of passes are influential factors in the increase of resistance to wear. In general, it can be asserted that the best behaviour for the 30° tilt angle takes place when the material is processed with three passes at 1800 rpm.

Likewise, Figure 7 shows a graph with the results of the accumulated erosive wear, in the form of accumulative mass loss, for the different combinations of parameters and tilt angle of the particles of 90°.

In the curves shown in Figure 7, it can be seen that the increase in tool rotation speed diminishes the resistance to erosive wear. When the tilt angle of the particles was 90°, the samples with the best behaviour to erosive wear were those processed with two passes, regardless of the tool rotation speed; and when three passes are made, the resistance to erosive wear diminishes. The best behaviour for the tilt angle of 90° takes place when the material is processed with two passes at 710 rpm of tool rotation speed.

It was thus proved that the tilt angle of the particles influences significantly the loss of weight, which is related to the change in the wear mechanism (Wang, 2008). It could be seen that the loss of weight by

erosive wear with tilt angle of the particles of 30° was higher in the samples processed at 710 rpm with two passes as well as the samples processed at 1200 rpm with two and three passes.

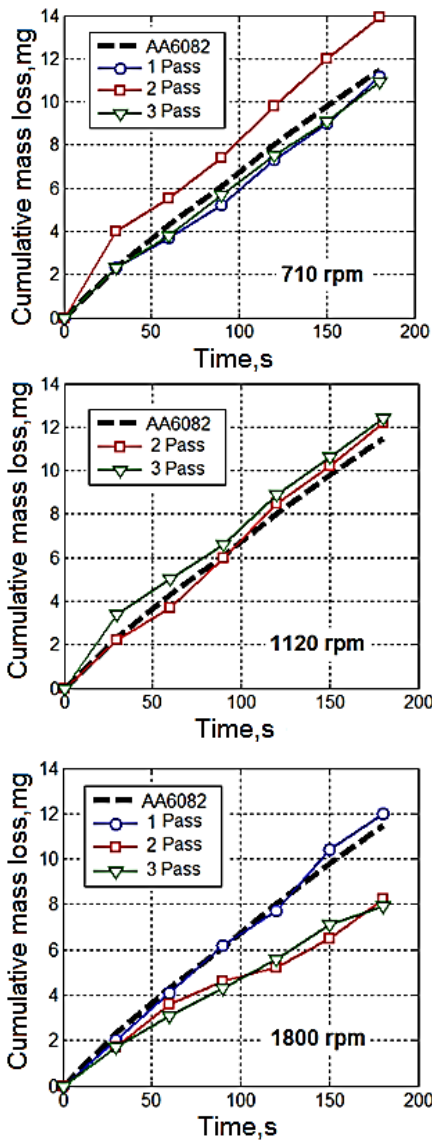


Figure 6: Graph of the loss of weight accumulated in the erosive wear for a tilt angle of the particles of 30°.

Under such processing conditions, a loss of superficial hardness occurred because of the solubilisation of the precipitates responsible for the hardening of the material.

This behaviour corroborates the results obtained by Magne (1991). This author explains that the materials with lower hardness have lower resistance to erosive wear when the tilt angle of the particles is small, with a predominance of the wear by micro-cuts (figure 8). Figure 8 and 9 shows a micrograph of the worn surface where this kind erosive wear mechanism can be seen.

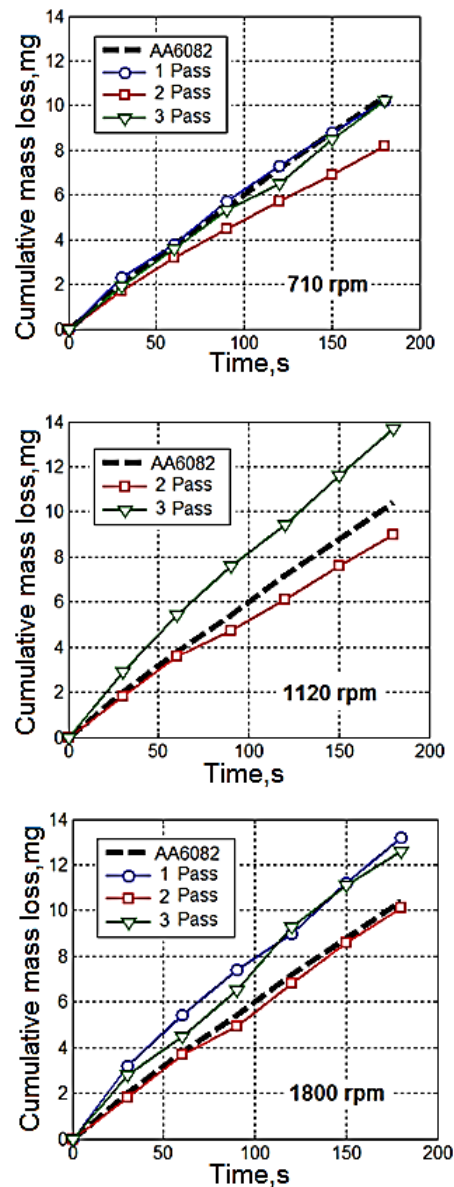


Figure 7: Graph of the loss of weight accumulated in the erosive wear for a tilt angle of the particles of 90°.

For the test samples processed at 1800 rpm, it was proved that the higher value of erosive wear takes place when the tilt angle of the particles is of 90°.

When the process uses 1800 rpm tool rotation speed, there is a change in the properties of the material, increasing the superficial hardness as a result of the formation of new precipitates in the processed zone. The process temperature increase and a lower cooling temperature, when 1800 rpm were used, causes the formation of new fine precipitates, probably type  $\square$ Mg<sub>5</sub>Si<sub>6</sub>, which constitute obstacles to the process of deformation of the material, and lead to the hardening of the stir zone by precipitation.

The superficial processing of the material at 1800 rpm, caused a change in the mechanism of erosive wear, taking place mainly a wear by deformation with extruded edges (Figure 9).

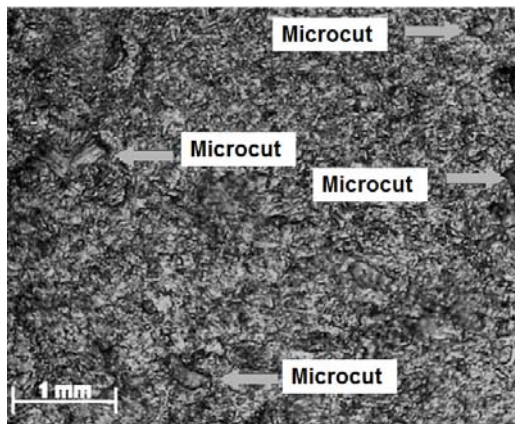


Figure 8. Micrograph of the erosive wear surface for a tilt angle of the particles of  $30^{\circ}$  and the condition of 710 rpm and 3 passes.

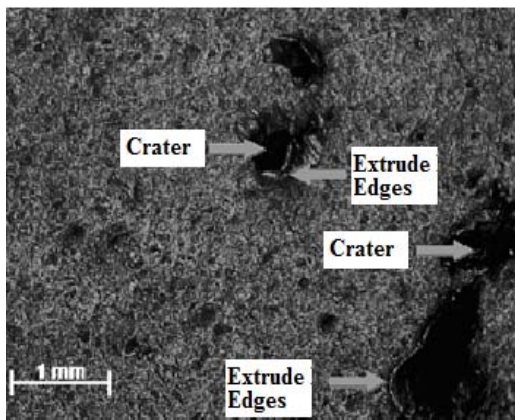


Figure 9: Micrograph of the erosive wear surface for a tilt angle of the particles of  $30^{\circ}$  and the condition of 710 rpm and 3 passes.

It is thus evident that the wear mechanism corresponds to the findings published in the literature consulted (Magne 1991), and is related to the way of eliminating the wear surface; in these cases, the wear comes as a result of the successive impact of particles that are normally projected and cause primarily the impregnation in the surface of the material. It is possible that in these cases the plastic and elastic deformations of the worn surfaces coincide. This means that extruded edges are formed from plastically deformed zones that come in the form of wear particles.

Finally, it can be said that, the number of passes with a 100% overlapping, the tool rotation speed and the tilt angle of the particles, influence the resistance to erosive wear of the aluminium alloy AA 6082 T-6, treated with friction stir processing.

It can also be concluded that when the tilt angle of the particles is of  $30^{\circ}$ , the increase of the tool rotation speed and the number of passes cause an increase in the resistance to erosive wear of the aluminium alloy AA 6082 T-6. Furthermore, with a  $90^{\circ}$  tilt angle of the particles, the increase of the number of passes at a tool

rotation speed of 710 rpm caused an increase in the resistance to erosive wear of the samples analysed.

#### 4. CONCLUSIONS

This research paper shows the influence of the parameters, number of passes and tool rotation speed, in the resistance to erosive wear of the aluminium alloy 6082-T6 modified by the FSP process. From the analyses conducted, it can be concluded that:

1. The number of passes with a 100% overlapping, the tool rotation speed and the tilt angle of the particles influence the resistance to erosive wear of the aluminium alloy 6082-T6 treated with the Friction Stir Process.
2. When the tilt angle of the particles is  $30^{\circ}$ , the increase in tool rotation speed and number of passes cause an increase in the resistance to erosive wear of the aluminium alloy 6082-T6.
3. The best behaviour for the tilt angle of  $30^{\circ}$  occurs when the material is processed with two and three passes at a tool rotation speed of 1800 rpm.
4. For a tilt angle of the particles of  $90^{\circ}$ , the increase in the number of passes at a tool rotation speed of 710 rpm caused an increase in the resistance to erosive wear in the analysed samples.
5. The best behaviour for the tilt angle  $90^{\circ}$  occurs when the material is processed with three passes at a tool rotation speed of 710 rpm.

#### ACKNOWLEDGMENTS

This paper has been partially supported by the project of the Ministry and Innovation "DPI2011-25007. Soldadura por fricción-agitación de materiales disimilares. Caracterización mediante técnicas de emisión acústica e inteligencia artificial".

#### REFERENCES

- Aldanondo, E., 2011. Avances en la tecnología de soldadura por fricción "STIR". DYNA Ingeniería e Industria. 86(6): 614-618.
- Cavaliere, P., Panella, F., 2008. Effect of tool position on the fatigue properties of dissimilar 2024-7075 sheets joined by friction stir welding. J. of Mater. Proc. Technol., 206: 249-255.
- Fernández, J.B., Roca, A.S., Fals, H.C., Macías, E.J., de la Parte, M.P. 2012. Application of vibroacoustic signals to evaluate tools profile changes in friction stir welding on AA 1050 H24 alloy. Science and Technology of Welding and Joining 17(6):501-510.
- Hassan, A.S. 2010. Friction stir welding of dissimilar A319 and A356 aluminium cast alloys. Sci. Technol. Weld. Join. 15(5): 414-422.
- Jiménez-Macías, E., Sánchez-Roca, A., Carvajal-Fals, H., Blanco-Fernández, J., Martínez-Cámara, E. 2014. Wavelets application in prediction of

- friction stir welding parameters of alloy joints from vibroacoustic ANN-based model. *Abstract and Applied Analysis*. Volume 2014, Article number 728564.
- Johannes, L.B., Mishra, R.S., 2007. Multiple passes of friction stir processing for the creation of superplastic 7075 aluminium. *Materials Science Engineering A*, 464: 255 – 260.
- Magdy, M., Ehab, A., 2012. The influence of multi-pass friction stir processing on the microstructural and mechanical properties of aluminium alloy 6082. *Journal of Materials Processing Technology*, 212: 1157 – 1168.
- Magne A., *Etude des processus d'usure de matériaux métalliques soumis à l'abrasion et à l'érosion*. Thesis, University of Liege.
- Mishra R.S., Ma, Z.Y., 2005. Friction stir welding and processing. *Material Science and Engineering R*, 50: 1- 78.
- Moreira, P.M., et al. 2009. Mechanical and metallurgical characterization of friction stir welding joints of AA6061-T6 with AA6082-T6. *Materials and Design*, 30: 180–187.
- Mroczka K., Pietras A., 2009. FSW characterization of 6082 aluminum alloys sheets. *Archives of Materials Science and Engineering*, 40(2): 104 - 109.
- Nandan R. et al., 2008. Recent advances in friction-stir welding – Process, weldment structure and properties. *Progress in Materials Science*, 53 (6) : 980–1023,
- Orozco, M.S., Macías, E.J., Roca, A.S., Fals, H.C., Fernández, J.B. 2012. Optimisation of friction-stir welding process using vibro-acoustic signal analysis. *Science and Technology of Welding and Joining* 18 (6): 532-540
- Santella M.L., Storjohann E., Pan, T.Y., 2005. *Scripta Material* 53:201.
- Su J., Nelson T., Sterling C., 2005. Microstructure evolution during FSW/FSP of high strength aluminum alloys. *Materials science and engineering, A*, 405: 277–286,
- Szkodo, M., Adamowski, J., Sitko, A., 2010. Assessment of FSW Welds Made of Aluminum Alloy AW7075-T651. *Solid State Phenom.*165: 201-206.
- Surekha, B. Murty, K., 2009. Effect of processing parameters on the corrosion behaviour of friction stir processed AA 2219 aluminium alloy. *Solid State Sciences* 1: 907–917.
- Uday, M.B., 2010. Advances in friction welding process: a review. *Sci. Technol. Weld. Join.* 15(7): 534-558.
- Wang Y, Yang Z (2008) Finite element model of erosive wear on ductile and brittle materials *Wear* 265: 871 – 878.
- Xunhong, W., 2008. Comparison of fatigue property between friction stir and TIG welds. *Journal Univ. Sci. Technol. Beijing.*15(3): 280-284.
- Cesar Sanchez-Orozco, M., Blanco-Fernandez, J., Jimenez-Macias, E., & Perez-delaParte, M. (2013). VIBROACOUSTICAL SIGNALS TO EVALUATE TOOL PROFILES ON FRICTION STIR WELDING PROCESS. *Dyna*, 88, 318-324. doi: 10.6036/5326
- Javier Martinez-de-Pison, F., Sanz, A., Martinez-de-Pison, E., Jimenez, E., & Conti, D. (2012). Mining association rules from time series to explain failures in a hot-dip galvanizing steel line. *Computers & Industrial Engineering*, 63, 22-36. doi: 10.1016/j.cie.2012.01.013
- Jimenez Macias, E., Sanchez Roca, A., Carvajal Fals, H., Blanco Fernandez, J., & Perez de la Parte, M. (2010). Time-frequency diagram applied to stability analysis in gas metal arc welding based on acoustic emission. *Science and Technology of Welding and Joining*, 15, 226-232. doi: 10.1179/136217109x12665778348588
- Martinez-de-Pison, F. J., Pernia, A., Jimenez-Macias, E., & Fernandez, R. (2010). Overall model of the dynamic behaviour of the steel strip in an annealing heating furnace on a hot-dip galvanizing line. *Revista De Metalurgia*, 46, 405-420. doi: 10.3989/revmetalm.0948
- Roca, A. S., Fals, H. C., Fernandez, J. B., Macias, E. J., & Adan, F. S. (2007). New stability index for short circuit transfer mode in GMAW process using acoustic emission signals. *Science and Technology of Welding and Joining*, 12, 460-466. doi: 10.1179/174329307x213882
- Sanchez Roca, A., Carvajal Fals, H., Blanco Fernandez, J., Sanz Adan, F., & Jimenez Macias, E. (2008). Stability analysis of the gas metal arc welding process based on acoustic emission technique. *Soldagem & Inspecao*, 13, 61-69.
- Sanchez Roca, A., Fals, H. C., Fernandez, J. B., Macias, E. J., & de la Parte, M. P. (2009). Artificial neural networks and acoustic emission applied to stability analysis in gas metal arc welding. *Science and Technology of Welding and Joining*, 14, 117-124. doi: 10.1179/136217108x382981



# DESIGN AND OPERATION OF A DAIRY PLANT BY MEANS OF A DECISION SUPPORT TOOL BASED ON THE PETRI NETS PARADIGM

Latorre-Biel JI<sup>(a)</sup>, Jiménez-Macías E<sup>(b)</sup>, Blanco-Fernández J<sup>(c)</sup>, Martínez-Cámara E<sup>(d)</sup>,  
Sánchez-Díez JC<sup>(e)</sup>, Pérez-Parte M<sup>(f)</sup>

<sup>(a,c,d,f)</sup> Department of Mechanical Engineering. University of La Rioja. Logroño. Spain.

<sup>(b,e)</sup> Department of Electrical Engineering. University of La Rioja. Logroño. Spain.

<sup>(a)</sup> [juan-ignacio.latorre@unirioja.es](mailto:juan-ignacio.latorre@unirioja.es), <sup>(b)</sup> [emilio.jimenez@unirioja.es](mailto:emilio.jimenez@unirioja.es), <sup>(c)</sup> [julio.blanco@unirioja.es](mailto:julio.blanco@unirioja.es),  
<sup>(d)</sup> [eduardo.martinezc@unirioja.es](mailto:eduardo.martinezc@unirioja.es), <sup>(e)</sup> [juan-carlos.saenz-diez@unirioja.es](mailto:juan-carlos.saenz-diez@unirioja.es), <sup>(f)</sup> [mercedes.perez@unirioja.es](mailto:mercedes.perez@unirioja.es)

## ABSTRACT

Despite a worldwide presence, the characteristics of dairy industry may vary largely from one country to another and also between the different regions of a country. Farmers can form co-operatives or operate individually, while processors may transform very different amounts of milk into a diversity of dairy products.

Moreover, the variety of dairy products that can be found in the market, the short-term expiry date of some of them, as well as the strong competition for a saturated market, lead to a complex environment, where the design and operation of the dairy factories require decision support tools for helping the decision makers. In this paper, it is considered the development of a decision support system for improving the survival chances of dairy plants in a competitive worldwide market. This tool is based in the implementation of an optimisation process, where the performance evaluation of the feasible decisions is performed by means of simulation. The objective of the decision making covers the complete life cycle of the dairy plant, from its design to its operation (Jimenez et al. 2014).

Keywords: decision support system, dairy plant, Petri nets, discrete event systems, modeling and simulation, exclusive entities, alternatives aggregation Petri net.

## 1. INTRODUCTION

Despite a worldwide presence, dairy industry may vary largely from one country to another and also between the different regions of a country. Farmers can form co-operatives or operate individually, while processors may transform very different amounts of milk into dairy products.

Moreover, the variety of dairy products, the short-term expiry date of some of them, as well as the strong competition for a saturated market, lead to a complex environment, where the design and operation of the dairy factories require decision support tools for helping in the associated decision making.

In this paper, it is considered the development of a decision support system for improving the survival chances of dairy plants in a competitive worldwide

market. A variety of decision support systems can be found in the scientific literature (Bruzzone and Longo, 2010), (Longo et al. 2013). This tool is based in the implementation of an optimization process, where the performance evaluation of the feasible decisions is performed by means of simulation, (Mújica et al. 2010). Several tools for decision support in the food industry have already been applied successfully (Latorre et al. 2014a), (Latorre et al. 2013b) (Jimenez et al., 2014).

Furthermore, the simulation model of the system is performed using the paradigm of the Petri nets (Piera et al. 2004), (Latorre et al. 2013a). So far, Petri nets have been applied to very different systems with success (Xiao and Ming, 2011), (Zaitsev and Shmeleva, 2011). In particular, considering a usual situation in the design of dairy plants, where the system presents a variety of alternative Petri nets or feasible structural configurations, the formalism used for developing the model of the system needs to include a set of exclusive entities (Latorre et al. 2010)(Jimenez et al, 2005).

The methodology proposed in this paper presents a model of a general middle-sized dairy factory in process of being designed (Macias and Parte, 2004). The freedom degrees in the model of the factory are associated to alternative structural configurations.

Moreover, the optimization process is based on the simulation of a set of selected feasible decisions, chosen from a solution pool. The choice of the most promising decisions is performed by means of a search methodology guided by means of a genetic algorithm (Latorre et al. 2014a).

As a result of the application of this methodology, a practical decision support tool for the design and operation of a medium-sized dairy factory is available for improving the chances of survival and success of such a facility in a competitive worldwide market.

The paper is organized as follows. Section 2 presents the main production activities developed in a dairy plant, the objective of the modeling and decision making considered in this paper.

Furthermore section 3 is devoted to the description of the modeling formalism, based on the formalism of the Petri nets. In particular, due to the fact that the discrete event system to be modeled includes several

alternative structural configurations, it is convenient that the chosen formalism include a set of exclusive entities, such a compound Petri net, a set of alternative Petri net, an alternatives aggregation Petri net, or a disjunctive colored Petri net.

Section 4 will introduce the model of the dairy plant, discussing the freedom degrees, represented by undefined parameters in the Petri net model of the system.

The next section, will deal with the decision support tool based on the simulation of the model described in the previous section. In this section, it is discussed the statement of a simulation-based optimization problem, as well as its further solution, A heuristic will be used to search for the optimum in the pool of feasible solutions.

Last section refers the conclusions of the research described in this paper, where future research lines will be commented.

## **2. PRODUCTION IN A DAIRY PLANT**

### **2.1. Importance of milk**

Milk is one of the most complete foods that exist. Despite the fact that 90% of its composition is water, it contains many minerals and vitamins, in addition to fat, lactose and three proteins: casein, lactalbumin, and lactoglobulin. The latter compounds contain all the essential amino acids that are necessary for a balanced diet for human beings. These substances need to be supplied by the ingested food, since human body by itself cannot synthesize them.

Ideally, from the nutritional point of view, the healthiest approach consists in serving consumers milk so little modified as possible, i.e. raw milk. This is possible in the places close to those where production is performed. However, appropriate processes are necessary for guarantee the conservation when the period between production and consumption may be large.

### **2.2. Types of milk to be produced**

There are several kinds of milk that might be found in the market, according to the industrial process developed for obtaining them:

#### 1) Whole milk:

This product contains all the properties of cow's milk. The main treatment it receives is sterilization.

#### 2) Semi-skimmed and skimmed milk:

From whole milk, by removing a certain percentage of fat, the different types of semi-skimmed and skimmed milk are produced. These products are suitable for those needed of a diet low in lipids.

#### 3) Concentrated milk:

This type of milk is prepared by evaporating nearly 65% of the water contained in its composition. The concentration of the milk is performed in vacuum. Adding the same amount of extracted liquid, the original product can be reconstituted.

#### 4) Condensed milk:

Adding a significant proportion of sugar (17 to 18 %) to a quantity of milk already concentrated it is possible to obtain sweetened condensed milk. The sweetened condensed milk is not subjected to a heat treatment after the filling and sealing of packaging. Their conservation properties are attributed to its content of lactose and other soluble sugars capable of preventing the growth of the populations of the main types of bacteria.

#### 5) Powdered milk:

This product is obtained after a similar phase of concentration in vacuum to the concentrated milk. However, to obtain powdered milk, the first stage is followed by a process of drying in cameras. The overall process consists in the evaporation of the water at low pressure and temperature of less than 50 °C. At the end of the process, a powder is obtained that retains all the nutritious elements of the original fluid and most of its vitamins.

#### 6) Fermented milk:

In fermented milk, germs are eliminated by the action of heat. After this process, the product is inoculated with ferments, mainly lactic, in order to obtain the resulting dairy product.

### **2.3. Industrial processes applied to milk.**

There are different types of industrial processes for the treatment of the milk in order to extend its shelf life once packed for its consumption:

#### A) Processes based on the application of cold:

The implementation of a temperature of approximately 4 °C prevents microbial blooms. However, the cold does not improve the bacteriological quality of the milk, because the germs are not usually destroyed.

#### B) Processes based on the application of heat:

To date, different conservation treatments for milk based on heat transfer have been applied:

##### B.1) Pasteurization at low temperature:

This technique consists in the application of a temperature of approximately 63 °C in the milk for a period of half an hour. The main drawback is the alteration of the properties of the product that represents a time interval so long.

##### B.2) Pasteurization at high-temperature, short-time (HTST):

In this case the temperature applied to the flow of milk is about 72°C. This process lasts from 15 to 20 seconds. The pasteurization provides protection for about 3 days to the product.

##### B.3) Pasteurization for extended shelf life (ESL):

The sterilization of the milk is achieved in this case, by applying a temperature of 90 to 95 °C for a few seconds.

##### B.4) Sterilization:

While pasteurization aims at reducing the number of pathogens in the milk so they are unlikely to cause disease, sterilization is intended for the elimination of all the microorganisms present in the milk. This result is achieved by the application of a temperature around



115°C during a time that varies between 15 and 20 minutes.

#### B.5) UHT:

In this case, the milk is processed by a ultra high temperature. This process raises the temperature of the milk to over 130 °C or more, during one or more second.

### 2.4. Essential stages in the production of milk.

The production of the milk passes through several essential stages. The first of these stages is milking. At this step the harvested milk should be kept in refrigerated tanks at 3°C of temperature, preventing the proliferation of microorganisms in the product.

The next stage is described below:

#### A) Transport:

To keep the milk in perfect conditions, this process must be performed at a constant temperature, equal to the storage temperature of the product in the previous stage. The shipping of the milk from the milking facilities to the processing and packing plant is performed on a daily basis.

#### B) Pretreatment of milk:

This second stage prepares the product for sterilization. It consists of the following phases:

##### B.1) Checking the acidity:

This parameter should be known since if the acidity of the milk is very high, the subsequent processing will be defective. This is due to the fact that the following stage of heat treatment of acid milk would cause precipitation of certain substances on the inner walls of the ducts of the heat exchangers. For this reason it is not convenient to use milk with a high acidity.

##### B.2) Filtering:

The milk is subjected to a filtration or to centrifuge purification in order to remove macroscopic impurities that could eventually be contained. This clarification responds to a double purpose: the first is a commercial one, while the second is more technical. The first is obvious, because this process improves the quality of the product under the eyes of the consumer. The second objective of this phase consists of minimizing the macroscopic impurities on the internal surfaces of the ducts of the heaters.

##### B.3) Standardize:

The percentage of fat content of the milk received in dairy plants may vary in significant proportions. For this reason, the milk is subjected to a process of standardization by special separators after heating the product to 30 or 35 °C. The quantities of cream (fat) recovered in this way are delivered for consumption or used for the manufacture of butter.

##### B.4) Creaming:

This process is an alternative to the previous one because that consists of a implementation of the a similar but more intense process. It is performed on centrifugal separators producing milk with a low fat content (skimmed or semi skimmed) and fat. This second product is used for the production of butter.

The last of the stages of processing milk consists of sterilization and packaging:

#### C) Treatment:

The packaging of milk requires a prior stage of sterilization to minimize the risk of subsequent microbial proliferation. Milk is one of the foods that best responds to the treatment known by the name of high temperature – short time (HTST). In the particular case of the milk, this process is called UHT treatment. It consists of a sterilization by continuous flow in which the product reaches a temperature over 135 °C for a few seconds. This method consists of the following phases:

##### C.1) Warming up:

During this stage the temperature of the milk is increased as quickly as possible until the value required by the sterilization of the product is reached.

##### C.2) Sterilization:

The high temperature of the food is kept for a minimum period of time that ensures the total destruction of the microorganisms present in the milk. Their duration should be the minimum possible to preserve as much as possible the properties of the milk.

##### C.3) Cooling:

To proceed to the packaging, the flow of warm milk should be cooled to a temperature under 35 °C. In particular an appropriate value is about 4 °C. In the same way as the previous stage of warming up, this step should be done as quick as possible. The goal of this process is to minimize the time at which the milk is exposed to high temperatures, preserving, in this way, the product from nutritional and organoleptic alterations by effect of the temperature.

#### D) Sterilization of the packaging:

This step differs from the sterilization of the milk on the technology used in its development. In this case, a chemical sterilization is applied. The bactericidal agent should be easy to apply and to remove as well as fast and efficient in its operation. The substance most commonly used for sterilizing the packages is the oxygenated water (hydrogen peroxide). The bactericidal action of this substance is effective at high temperatures.

#### E) Packaging:

The flow of sterilized milk can be packed in accordance with the rate of sterilization. Another possibility consists of the storage in aseptic conditions of the sterilized milk, before packing. This latter technique allows greater freedom in the choice of the filling speed. The process of filling and sealing should be performed in completely aseptic conditions, in order to not neglect the sterilization process of the milk and its container previously made.

#### F) Distribution:

This phase consists of the storage of packaged milk, its transport, and subsequent distribution to the points of sale.

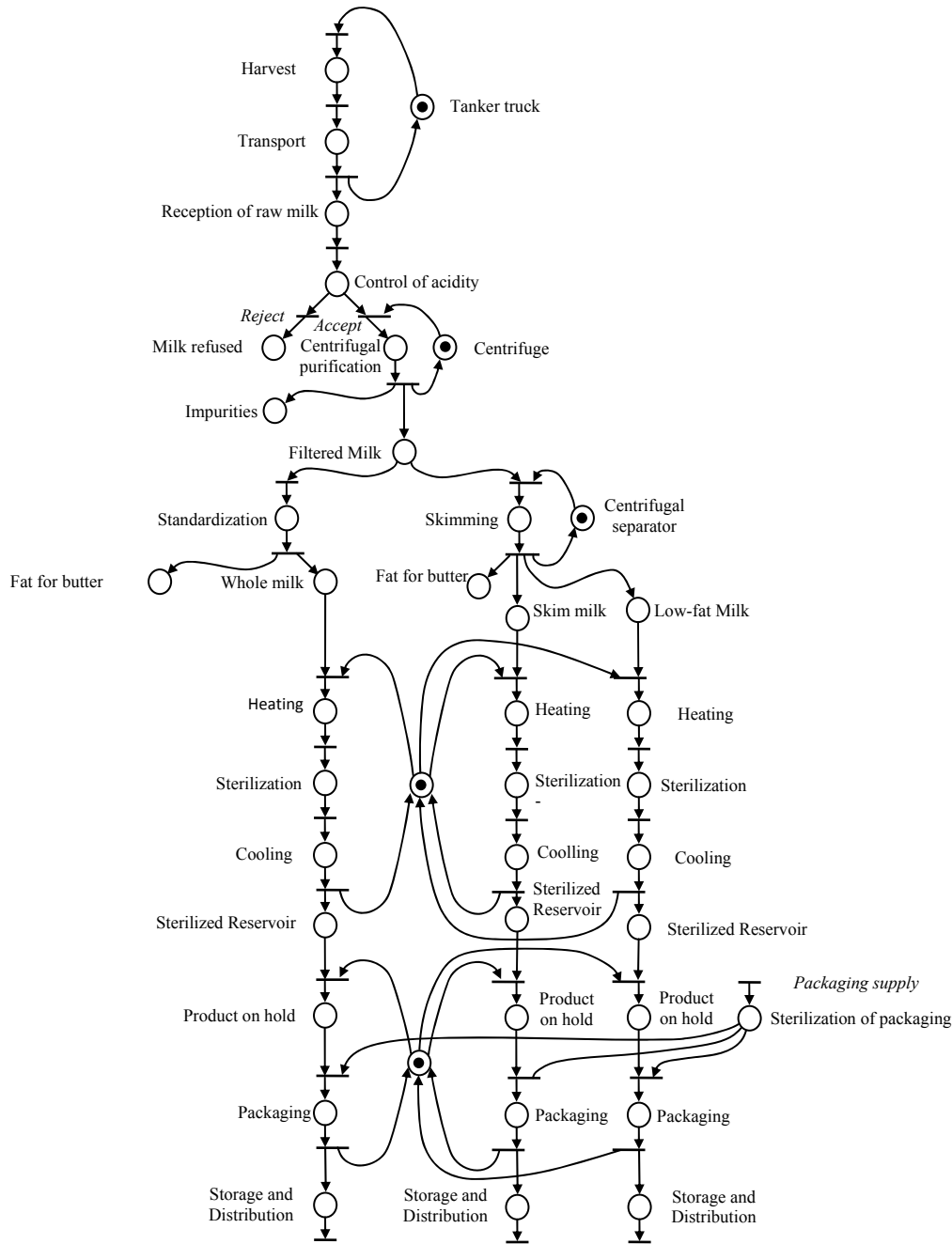


Figure 1. Petri net model of a dairy plant.

### 3. MODELING FORMALISM

In order to develop a model suitable for a decision support system based on simulation, an appropriate formal language has to be chosen.

The formalism chosen in this paper is based in the paradigm of the Petri nets, due to its proven suitability for the analysis, simulation, and optimization of discrete event systems (Silva 1993), (David and Alla 2005), (Jensen and Kristensen 2009).

**Definition 1.** A Petri net system is a 5-tuple  $R = (P, T, \text{pre}, \text{post}, \mathbf{m}_0)$  such that:

i)  $P$  is a non-empty set of places.

ii)  $T$  is a non-empty set of transitions and  $P \cap T = \emptyset$ .

iii) pre and post are functions that associate a weight to the directed arcs between the elements of the sets  $P$  and  $T$ , in the following way:

iv)  $\text{pre}: P \times T \rightarrow \mathbb{N}^*$  and  $\text{post}: T \times P \rightarrow \mathbb{N}^*$ , where  $\mathbb{N}^*$

is the set of natural numbers, excluding zero.

v)  $\mathbf{m}_0$  is the initial marking, such that  $\mathbf{m}_0: P \rightarrow \mathbb{N}^*$ .

□

A discrete event system in process of being designed may contain a set of freedom degrees, some of them corresponding to the structure of the system. The model of such a system should be able to describe with the alternative structural configurations of the discrete event system.

An intuitive approach for modelling a discrete event system with alternative structural configurations may be to develop so many models as structural configurations (Latorre *et al* 2012).

Nevertheless, this approach uses to be ineffective, since the whole set of alternative Petri nets usually shows large amounts of redundant information among the models (Latorre *et al* 2014b).

A more appropriate formalism for developing efficient simulations is the alternatives aggregation Petri nets. The mechanism for the decision making associated to the choice of one of these alternative structural configurations is represented explicitly by means of the so called choice variables (Latorre *et al.* 2013c).

**Definition 2.** Given a discrete event system with  $n$  alternative structural configurations, a set of choice

variables can be defined as  $S_A = \{ a_i \text{ Boolean} \mid \exists! k \in$

$\mathbb{N}^*, k \leq n, \text{ such that } a_k = 1 \wedge \forall j \in \mathbb{N}^*, j \leq n, j \neq k \text{ it is}$

verified  $a_j = 0 \}$ , where  $|S_A| = n$ , and the assignment  $a_k = 1$  is the result of a decision.

□

Once the concept of choice variable has been introduced as a way to represent in a Petri net model of a discrete event system a set of alternative structural configurations, it will be defined the formalism called alternatives aggregation Petri nets.

**Definition 3.** An alternatives aggregation Petri net can be defined as a 10 tuple  $R^A = (P, T, \text{pre}, \text{post}, \mathbf{m}_0, S_\alpha, S_{\text{valnstr}\alpha}, S_A, f_A, R_{\text{val}\gamma})$ , where

- i)  $S_\alpha$  is a set of undefined parameters.
- ii)  $S_{\text{valnstr}\alpha}$  is the set of feasible combination of values for the undefined parameters in  $S_\alpha$ .
- iii)  $S_A$  is a set of choice variables,  $S_A \neq \emptyset$  and  $|S_A| = n$ .
- iv)  $f_A: T \rightarrow f(a_1, \dots, a_n)$  is a function that assigns a function of the choice variables to each transition  $t$  such that  $\text{type}[f_A(t)] = \text{boolean}$ .
- v)  $R_{\text{val}\gamma}$  is a binary relation between  $S_{\text{val}\gamma}$  and  $R_A$ .

On the other hand,  $f_A: S_A \rightarrow T$ , assigns a choice variable to a single or several transitions of the Petri net, and if  $S_A' = \{ a_1, a_2, \dots, a_k \}$  is the set of every choice

variables associated to the transition  $t$ , then the guard function of the mentioned transition is  $g_A(t) = a_1 + a_2 + \dots + a_k$ .

□

#### 4. MODEL OF A DAIRY PLANT

In order to develop an appropriate model for a general purpose dairy plant, the different subsystems that can be implemented in such a facility will be overviewed in the following paragraphs.

1) Milking facilities:

The facilities in which the animals are kept and/or milked present a degree of automation that increases with time. It is necessary to count on refrigerated tanks with controlled temperature to store the harvested milk up to the time of their transport.

B) Trucks:

Isotherm tankers carry the milk to the dairy facilities.

C) Cooled tank:

Until the time of its processing, the milk is kept in conditioned deposits that optimize the preservation of the product. A temperature of 3 °C prevents the growth of the microbial content of the product.

D) Centrifugal separator:

This equipment allows removing from the product a desired amount of fat, as well as developing the stages of filtration and standardization. For decreasing the viscosity of the milk and to increase the efficiency of the process of separation, this system is usually operated at a temperature around 35 °C.

E) Heat exchanger:

A treatment received by the milk before its packaging consists of a thermal sterilization in a heat exchanger. Another heat exchanger cools the output milk under 35 °C.

F) Sterile Reservoir:

It consists of a reservoir that adapts the rates of sterilization of the milk and its packaging. This stages provides flexibility to the production process.

G) Sterilization equipment for the packaging:

The implementation of the chemical agent bactericidal, usually hydrogen peroxide, is carried out in different forms, depending on the circumstances of the manufacturing process:

H) Aseptic filler:

This equipment must operate under aseptic conditions. Performs the filling and subsequent closure of the sterile packing.

A graphical representation of the Petri net model of a general dairy plant can be seen in figure 1.

#### 5. DECISION MAKING

This paper has the purpose of describing the construction of a decision support system for a dairy plant, from its design process to its operation.

In the previous section, a model of the system has been described. In the following section, the rest of the elements that constitute a decision support tool will be overviewed.

The decision support system will be based on the statement of an optimization problem, where the objectives of the optimization will be quantified in the so called cost or objective function.

A heuristic search, such as the genetic algorithms, or the ant colony, will be used for finding the most promising solutions in the solution space. Simulating the behavior of the model under the scenario imposed by the selected solutions it is possible to compare the relative quality of them and, so, to choose the best one.

## 6. CONCLUSIONS

As it has been seen in this paper, a model of a dairy plant can be developed, using the paradigm of the Petri nets, with the purpose of developing a decision support system for the design and operation of such a facility.

As open research lines, it can be considered the application of this methodology to a variety of dairy plants, as well as to improve the described model by including other uses of the milk harvested, such as yoghurt, cheese, cream, or butter manufacturing.

## REFERENCES

- Bruzzone 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.
- Jimenez, E., Martinez, E., Blanco, J., Perez, 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-Transactions of the Society for Modeling and Simulation International*, 90, 143-161.
- Jiménez, E., Pérez, M., & Latorre, I. (2005). On deterministic modelling and simulation of manufacturing systems with petri nets. Paper presented at the I3M 2005, European Modeling Simulation Symposium, EMSS 2005.
- Jimenez, E., Tejada, A., Perez, M., Blanco, J., & Martinez, E. (2012). Applicability of lean production with VSM to the Rioja wine sector. *International Journal of Production Research*, 50, 1890-1904. doi: 10.1080/00207543.2011.561370
- 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*. 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*. 9(3). 267-278.
- 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*. In press. Published online, May 2014.
- 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)Vienna, 2012.
- 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., 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.
- 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., 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).
- 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
- Macias, E.J., & de la Parte, M.P. (2004). Simulation and optimization of logistic and production systems using discrete and continuous Petri nets. *Simulation-Transactions of the Society for Modeling and Simulation International*, 80:143-152.
- 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.À., 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.
- Xiao, Z. and Ming, Z. 2011. A method of workflow scheduling based on colored Petri nets. *Data & Knowledge Engineering* 70, pp. 230–247. Elsevier. 2011.
- Zaitsev, D.A., Shmeleva, T.R. 2011. A Parametric Colored Petri Net Model of a Switched Network. *Int. Journal of Communications, Network and System Sciences*, 4, 65-76, Scientific Research Publishing Inc., 2011.

# MIXING FORMALISMS BASED ON PETRI NETS FOR IMPROVING THE EFFICIENCY OF MODELLING AND SIMULATION OF DES

Latorre-Biel JI<sup>(a)</sup>, Pérez-Parte M<sup>(b)</sup>, Sáenz-Díez JC<sup>(d)</sup>, García-Alcaraz JL<sup>(c)</sup>, Jiménez-Macías E<sup>(d)</sup>

<sup>(a,b)</sup> Department of Mechanical Engineering, University of La Rioja, Logroño, Spain.

<sup>(c)</sup> Department of Industrial and Manufacturing Engineering, Autonomous University of Ciudad Juarez

<sup>(d)</sup> Department of Electrical Engineering, University of La Rioja, Logroño, Spain.

<sup>(a)</sup> [juan-ignacio.latorre@unirioja.es](mailto:juan-ignacio.latorre@unirioja.es), <sup>(b)</sup> [mercedes.perez@unirioja.es](mailto:mercedes.perez@unirioja.es), <sup>(c)</sup> [jorge.garcia@uacj.mx](mailto:jorge.garcia@uacj.mx),  
<sup>(d)</sup> [juan-carlos.saenz-diez@unirioja.es](mailto:juan-carlos.saenz-diez@unirioja.es), <sup>(e)</sup> [emilio.jimenez@unirioja.es](mailto:emilio.jimenez@unirioja.es)

## ABSTRACT

The use of formal languages for describing discrete event systems provides with high quality tools for validation, verification, simulation, or optimization. Choosing and using an appropriate formalism for modeling a system is crucial for these purposes, since this choice might influence the easiness of modeling or the speed of simulation. For example, the performance of simulation-based optimization is usually very sensitive to the size of the model of the system. Large models require more computational effort and they are very common in realistic modeling and simulation.

In large models, it might be recommendable the use of different formalisms for every subsystem. This paper deals with the theoretical analysis of this situation, focussed on a paradigm, such as the Petri nets, which offer a wide range of formalisms, and an application field, such as the discrete event systems with alternative structural configurations, which usually provide with large models for developing decision support systems.

Keywords: workstation design, work measurement, ergonomics, decision support system

## 1. INTRODUCTION

The use of formal languages for describing discrete event systems provides the researchers and practitioners with high quality tools for relevant tasks, such as validation, verification, simulation, or optimization (Bruzzone and Longo, 2010), (Xiao and Ming, 2011), (Longo et al. 2013).

Choosing and using an appropriate formalism for developing the model of the system is a crucial stage in the application of strategies of structural analysis or performance evaluation (Balbo and Silva 1998). According to the characteristics of the system itself, it might be adequate to select a certain formal language, for improving the certain operations, such as the modeling stage or the simulation of the evolution of the system (Piera et al. 2004), (Mújica et al. 2010).

Simulation-based optimization is a usual example of operation that may require high computational effort, when the modeled system is made realistic by means of a large size of the model (Zaitsev and Shmeleva, 2011),

(Latorre et al. 2013a, 2014b, 2013b) (Macias and Parte, 2004). The development of large models is usually coped by means of a modeling strategy, such as top-down or bottom-up. The resulting subsystems in which the final model can usually be divided may present very different characteristics (Silva 1993).

An efficient choice of the appropriate formalism to describe the model of a large discrete event system might require a multiple choice in the sense that diverse subsystems should be modeled using different formal languages (Latorre et al 2014a)(Jimenez et al. 2014).

This paper deals with the theoretical analysis of this situation, focused on a paradigm, such as the Petri nets, which offer a wide range of formalisms, and an application field, such as the discrete event systems with alternative structural configurations, which usually provide with large models (Silva 1993), (Recalde et al 2004), (David and Alla 2005), (Jensen and Kristensen 2009) (Jimenez et al, 2005, 2006, 2009).

The present paper is organized as follows. First of all, the concept of exclusive entity and monotypic set of exclusive entities are introduced. In section 3, a series of definitions of diverse formalisms, based on the Petri net paradigm, able to represent a set of exclusive entities is presented. In section 4 several propositions prove the ability of the mentioned formalisms for representing a set of exclusive entities. The following section deals with a key concept in this paper, the polytypic set of exclusive entities, as a formal way to introduce a variety of formalisms for describing different parts of a Petri net model. Moreover, sections 6, 7, and 8, show three case studies, where models representing monotypic, as well as polytypic sets of exclusive entities are presented. Finally, a conclusion section summarizes the achievements of the research line described in this paper, and outlines open research lines.

## 2. EXCLUSIVE ENTITIES

Several formalisms, based on the Petri net paradigm, have been developed for modelling discrete event systems with alternative structural configurations, such as the alternatives aggregation Petri nets, disjunctive Petri nets, compound Petri nets, or a perhaps more

intuitive set of alternative Petri nets (Latorre et al. 2013c), (Latorre et al. 2012). All of these models share in common to have an associated set of exclusive entities, representing the exclusive nature of the different structural alternative configurations (Latorre et al. 2013d).

Different transformation algorithms have been developed for converting the model of a discrete event system from one of these formalisms to another one, being both of the models equivalent, since the graphs of reachable states are the same or at least isomorphous.

This property and knowledge open the door to deal with the so called polytypic set of exclusive entities, associated to a model of the system containing subsystems expressed using different formal languages (Latorre et al. 2014c). In fact, it is possible to combine in the same model all the four formalisms mentioned in the previous paragraphs.

For example, it is possible to profit from the specific abilities of the diverse formalisms to ease the modelling of certain type of systems to deal with a bottom-up modelling process by using diverse formalisms for different subsystems. If it is required, once the modelling process has been finished, it is possible to convert, in an automatic fashion, the formalisms describing the subsystems present in the model into a single one (Latorre et al. 2014c).

More in detail, the general concept of exclusiveness, based in the idea of mutually exclusive evolution, is defined in a formal and general way by means of the concept of set of exclusive entities. Different forms of these exclusive entities allow defining different representations of disjunctive constraints equivalent to a set of alternative Petri nets; always preserving the exclusiveness between the different elements or entities.

**Definition 1. Monotypic set of exclusive entities.**

Given a discrete event system  $D$ , a monotypic set of exclusive entities associated to  $D$  is a set  $S_x = \{ X_1, \dots, X_n \}$ , which verifies that

i) The elements of  $S_x$  are exclusive, that is to say, only one of them can be chosen as a consequence of a decision.

ii)  $\forall i, j \in \mathbb{N}^*, i \neq j$  and  $1 \leq i, j \leq n$  it is verified that  $X_i \neq$

$X_j$ .

iii) The elements of  $S_x$  are of the same type.

iv)  $\exists f: S_x \rightarrow S_R$  such that

iv.a)  $S_R = \{ R_1, \dots, R_n \}$  is a set of alternative Petri nets, feasible models of  $D$ .

iv.b)  $f$  is a bijection  $\Rightarrow \forall X_i \in S_x \exists! f(X_i) = R_i \in S_R$  such that  $R_i$  is a feasible model for  $D$  and  $\forall R_i \in S_R \exists! f^{-1}(R_i) = X_i \in S_x$ .

□

The purpose of a set of exclusive entities is to represent in different ways the exclusive Petri net models of a discrete event system. A supposition is made on these alternative models: they have a different static structure. For this reason, as it happens in the real practice, if the difference is, for example, the initial state, the models are supposed to be the same.

The exclusive entities may be understood as an abstraction of the feasible formal representations for the alternative Petri nets that can be chosen for a DES. This doctoral thesis is devoted to find different forms of the exclusive entities, their properties, transformation algorithms between them, their feasible combinations and their performance in the solution of optimization problems.

**3. PETRI NET FORMALISMS ABLE TO REPRESENT EXCLUSIVE ENTITIES**

Several formalisms, based on the paradigm of the Petri nets, have been defined and used for the description of discrete event systems with alternative exclusive entities.

The most natural and intuitive way to represent a discrete event system with freedom degrees related to alternative structural configurations is a set of alternative Petri nets. According to this approach, it is possible to obtain a different model for the system associated to every one of the structural configurations.

An important property verified by the elements of a set of alternative Petri nets is the following.

**Definition 2. Mutually exclusive evolution.**

Given two different Petri nets  $R$  and  $R'$ . They are said to have mutually exclusive evolutions if it is verified:

i) If  $\mathbf{m}(R) \neq \mathbf{m}_0(R) \Rightarrow \mathbf{m}(R') = \mathbf{m}_0(R')$

ii) If  $\mathbf{m}(R') \neq \mathbf{m}_0(R') \Rightarrow \mathbf{m}(R) = \mathbf{m}_0(R)$

□

Based on the previous definition, it is possible to define a set of alternative Petri nets in the following way:

**Definition 3. Set of alternative Petri nets.**

Given a set of Petri nets  $S_R = \{ R_1, \dots, R_n \}$ ,  $S_R$  is said to

be a set of alternative Petri nets if  $n > 1$  and  $\forall i, j \in \mathbb{N}^*$

such that  $i \neq j, 1 \leq i, j \leq n$ , it is verified that

i)  $R_i$  and  $R_j$  have mutually exclusive evolution.

ii)  $\mathbf{W}(R) \neq \mathbf{W}(R')$ . That is to say, the structure of the Petri nets, represented by their incidence matrices, is different.

$R_i$  is called the  $i$ -th alternative Petri net of  $S_R$ .

□

Another formalism, able to represent a set of exclusive entities, is the compound Petri net. It can be seen as a parametric Petri net, where some of the parameters belong to the structure of the net and not to its behaviour.



**Definition 4.** Compound Petri net.

A compound Petri net is a 7-tuple

$$R^c = \langle P, T, W^+, W, \mathbf{m}_0, S_\alpha, S_{val\alpha} \rangle, \text{ where}$$

i)  $P$  and  $T$  are disjoint, finite, non-empty sets of places and transitions respectively.

ii)  $W: P \times T \rightarrow \mathbb{N}$  is the pre-incidence or input function.

iii)  $W^+: T \times P \rightarrow \mathbb{N}$  is the post-incidence or output

function.

iv)  $\mathbf{m}_0$  is the initial marking of the net that represents the initial vector of state and is usually a function of the choice variables.

v)  $S_\alpha$  is the set of undefined parameters of  $R^c$ .

vi)  $S_{str\alpha} \neq \emptyset$  is the set of undefined structural parameters of  $R^c$ , such that  $S_{str\alpha} \subseteq S_\alpha$ .

vii)  $S_{val\alpha}$  is the feasible combination of values for the undefined parameters .

□

Furthermore, a formalism, able to represent sets of exclusive entities, and appropriate to model discrete event systems, which verify that the alternative Petri nets share complete subnets is the alternatives aggregation Petri net.

**Definition 5. Alternatives aggregation Petri net system.**

An alternatives aggregation Petri net system,  $R^A$ , is defined as the 10-tuple:

$$R^A = \langle P, T, W^+, W, \mathbf{m}_0, S_A, f_A, R_{val\gamma} \rangle$$

Where,

The first four elements in the definition are the same of the ones seen in the previous definition of compound Petri net.

i)  $\mathbf{m}_0$  is the initial marking of the net and is usually a function of the choice variables.

ii)  $S_A$  is a set of choice variables such that  $S_A \neq \emptyset$  and  $|S_A| = n$ .

iii)  $f_A: T \rightarrow f(a_1, \dots, a_n)$  is a function that assigns a function of the choice variables to each transition  $t$  such that  $\text{type}[f_A(t)] = \text{boolean}$ .

iv)  $R_{val\gamma}$  is a binary relation between  $S_{val\gamma}$  and  $R_A$ .

□

#### 4. FORMS OF A SET OF EXCLUSIVE ENTITIES

The set of exclusive entities  $S_x$  can take different forms. On the other hand, all the elements of  $S_x$  must be of the same nature or structure: Petri nets, binary or Boolean variables, colours of tokens or integer numbers.

In this section, several propositions will be given, for justifying the fact that the formalisms based on the

Petri net paradigm, enumerated in the previous section can, in fact, represent a set of exclusive entities.

**Proposition 1.** A set of alternative Petri nets as a monotypic set of exclusive entities.

Let  $D$  be a discrete event system.

Let  $S_R$  be a set of alternative Petri nets such that  $\forall R_i \in S_R, R_i$  is a feasible model of  $D$ .

It is verified that  $S_R$  is a set of exclusive entities associated to  $D$ .

**Proof**

Being  $S_R$  a set of alternative Petri nets, models of  $D$  with exclusive evolution, they are alternative models for  $D$ . This statement means that in order to define univocally the model of  $D$  it is necessary to make a decision on a  $R_i \in S_R$ . As a consequence i) and iii.a) are verified.

The elements of  $S_R$  are Petri nets hence they belong to the same type. Thus, ii) is verified.

□

**Proposition 2.** A set of choice variables as a monotypic set of exclusive entities.

Let  $S_R$  be a set of alternative Petri nets, feasible models of a discrete event system  $D$ .

Let  $S_A$  be a set of choice variables.

If  $|S_A| = |S_R| \Rightarrow S_A$  is a monotypic set of exclusive entities.

**Proof**

By definition, the elements of the set  $S_A$  are exclusive, since only one of them can be set to 1 as a consequence of a decision. Then i) is verified.

The elements of  $S_A$  are binary or Boolean variables hence they belong to the same type.

As  $|S_A| = |S_R|$  is possible to define a bijection  $f: S_A \rightarrow S_R$  such that

$\forall a_i \in S_A \exists! f(a_i) = R_i \in S_R$  such that  $R_i$  is a feasible model for  $D$

$\forall R_i \in S_R \exists! f^{-1}(R_i) = a_i \in S_A$ .

□

**Proposition 3.** A set of feasible combinations of values for a set of undefined structural parameters as a monotypic set of exclusive entities.

Let  $R^c$  be a compound Petri net, developed as model of a DES.

Let  $S_{valstr\alpha}$  be a set of undefined structural parameters of a compound Petri net.

Let  $S_R$  be a set of alternative Petri nets, associated as models of a discrete event system  $D$ , that arise when the undefined structural parameters of  $R^c$  take values from the different feasible combinations of values belonging to  $S_{valstr\alpha}$ .

$S_{valstr\alpha}$  is a monotypic set of exclusive entities.

**Proof**

By definition, the elements of the set  $S_{valstr\alpha}$  are exclusive, since only one of them can be chosen at a time as a consequence of a decision.

The elements of  $S_{valstr\alpha}$  are integers. As they belong to the same type, ii) is verified.

By construction of  $S_R |S_{valstrad}| = |S_R|$ , then it is possible to define a bijection  $f: S_{valstrad} \rightarrow S_R$  such that

$\forall cv_i \in S_A \exists! f(cv_i) = R_i \in S_R$  such that  $R_i$  is a feasible model for  $D$

$\forall R_i \in S_R \exists! f^{-1}(R_i) = cv_i \in S_{valstrad}$ .

□

Notice that in this last proposition a set of alternative Petri nets has been built up from a compound Petri net. The proof may be performed as well by the construction of the compound Petri net from the merging of a set of matching alternative Petri nets from an original set of alternative Petri nets.

## 5. POLYTYPIC SET OF EXCLUSIVE ENTITIES

### Definition 4. Polytypic set of exclusive entities.

Given a discrete event system, a polytypic set of exclusive entities associated to it is a set  $S_x^p = \{X_1, \dots, X_n\}$ , which verifies that

i) The elements of  $S_x$  are exclusive, that is to say, only one of them can be chosen as a consequence of a decision.

ii)  $\exists S_x, S'_x \subseteq S_x^p$  such that the elements in  $S_x$  and in  $S'_x$  are of different types.

iii)  $\exists f: S_x^p \rightarrow S_R$  such that

iii.a)  $S_R = \{R_1, \dots, R_n\}$  is a set of alternative Petri nets, feasible models of  $D$ .

iii.b)  $f$  is a bijection  $\Rightarrow \forall X_i \in S_x^p \exists! f(X_i) = R_i \in S_R$  such that  $R_i$  is a feasible model for  $D$  and  $\forall R_i \in S_R \exists! f^{-1}(R_i) = X_i \in S_x^p$ .

□

Note that the only difference between “Definition 1. Monotypic set of exclusive entities” and “Definition 4. Polytypic set of exclusive entities” is the condition ii). In other words, the monotypic set includes exclusive entities of a single type, for example Petri nets, choice variables or feasible combinations of values for the undefined structural parameters. On the other hand, a polytypic set of exclusive parameters contains elements of different types, for example Petri nets and choice variables.

In order to illustrate the concept of polytypic set of exclusive entities and its potential for easing the modeling of large and complex discrete event systems, in the following sections, a set of case studies will be presented.

## 6. CASE STUDY 1

As first example in the illustration of the concepts presented in this paper, it will be developed a set of alternative Petri nets.

Let us consider the design of a certain discrete event system, where, after a first stage in the design process, a number of twelve alternative structural configurations have been identified.

A first, natural, and intuitive way to proceed for obtaining a model of this system consists in the development of a set of alternative Petri nets. Every

structural configuration will have a counterpart as an alternative Petri net.

However, this approach for the modeling process of a discrete event system with alternative structural configurations may be too intensive in time consumption, since a model of the system should be developed for every feasible solution in the pool of alternative structural configurations.

The set of exclusive entities in this model is a monotypic set of exclusive entities, since all the exclusive entities are alternative Petri nets; hence, they are of the same type.

In figures 1 to 3 some of the twelve alternative Petri nets are shown, only in their graphical representation. However, obtaining from this figures the matrix-based representation is trivial.

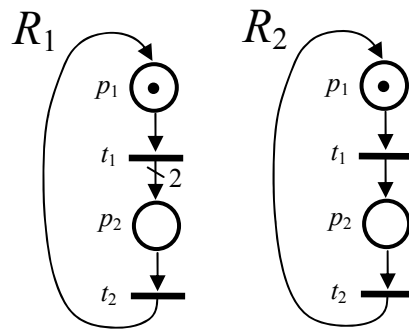


Figure 1: Alternative Petri nets  $R_1$  and  $R_2$

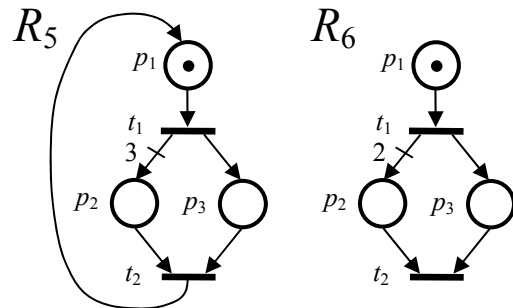


Figure 2: Alternative Petri nets  $R_5$  and  $R_6$

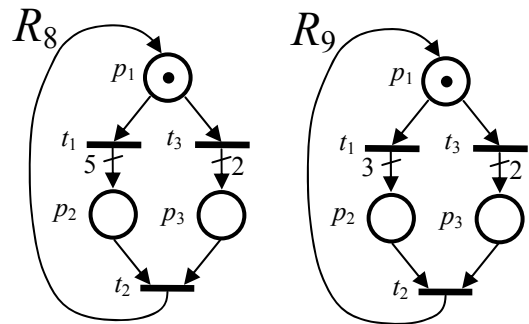


Figure 3: Alternative Petri net  $R_8$  and  $R_9$

## 7. CASE STUDY 2

In this section, the formalism of the compound Petri nets will be applied for profiting from its advantages for transforming the model of the discrete event system developed in the previous section into another one more compact and better for performing simulation-based operations, such as performance analysis or optimization.

Furthermore, for the modeling process illustrated in the previous section, the compound Petri nets might have been used in conjunction with the alternative Petri nets. Combining both formalisms, the modeling process is also simplified, since it is not necessary to develop different models for every alternative structural configuration. In fact, compound Petri nets allow developing a single Petri net model for a set of alternative structural configurations.

In figures 4 to 6, the three compound alternative Petri nets of the model can be found:  $S_R^c = \{ R_1^c, R_2^c, R_3^c \}$ .

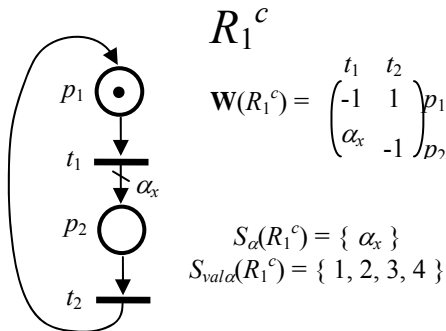


Figure 4: Alternative Petri net  $R_1^c$

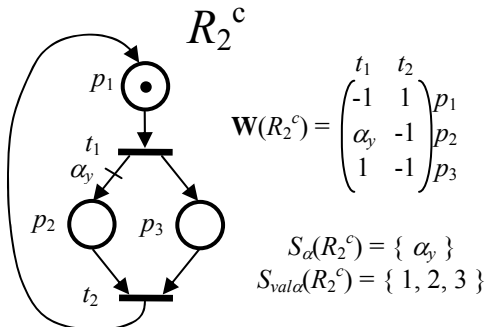


Figure 5: Alternative Petri net  $R_2^c$

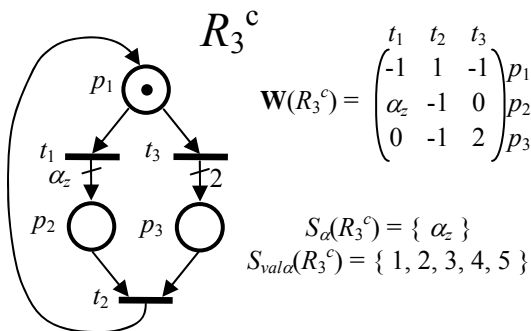


Figure 6: Alternative Petri net  $R_3^c$

## 8. CASE STUDY 3

The model presented in the previous section is a set of alternative compound Petri nets. In that model, some of the exclusive entities are represented by means of three alternative Petri nets, while the rest of the exclusive entities, totaling a number of 12, are represented by sets of feasible values for the undefined structural parameters of the compound Petri nets: 4, 5, and 3, respectively. As it has been proven, the model has an associated polytypic set of exclusive entities.

In this section, the three alternative compound Petri nets will be reduced into a single compound Petri net, including undefined structural parameters. In this case, the exclusive entities are not represented by alternative Petri nets anymore but the twelve of them by feasible combinations of values for the undefined structural parameters. Again, as it happened in section 6, the model of the original discrete event system, has included a monotypic set of exclusive entities.

The purpose of this transformation of the model, from a set of alternative compound Petri nets to a single compound Petri net may be, for example, to assess the improvement (or not) in the computer time required to perform a simulation for developing a performance evaluation.

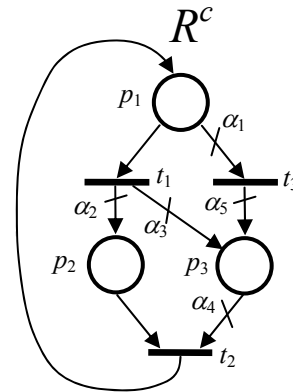


Figure 7: Graphical representation of the compound Petri net  $R^c$

$$W(R^c) = \begin{pmatrix} t_1 & t_2 & t_3 \\ -1 & 1 & -\alpha_1 \\ \alpha_2 & -1 & 0 \\ \alpha_3 & -\alpha_4 & \alpha_5 \end{pmatrix} \begin{matrix} p_1 \\ p_2 \\ p_3 \end{matrix}$$

Figure 8: Matrix-based representation of the compound Petri net  $R^c$

The set of undefined parameters of the compound Petri net contains five structural parameters  $S_\alpha(R^c) = S_{str\alpha}(R^c) = \{ \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5 \}$ . On the other hand, the set of feasible combinations of values for the undefined structural parameters has twelve elements:  $S_{valstr\alpha}(R^c) = \{ (0,1,0,0,0), (0,2,0,0,0), (0,3,0,0,0), (0,4,0,0,0), (0,1,1,1,0), (0,2,1,1,0), (0,3,1,1,0) \}$ .

(1,1,0,1,2) , (1,2,0,1,2) , (1,3,0,1,2) , (1,4,0,1,2) , (1,5,0,1,2) }.

In figure 7, it is possible to find the graphical representation of the resulting compound alternative Petri net, while in figure 8, its matrix-based representation is shown.

## 9. CONCLUSIONS

This paper has presented the concept of monotypic and polytypic sets of exclusive entities, as an important consideration for improving the efficiency of simulation-based operations, such as performance evaluation or optimization, using Petri net models for describing discrete event systems with alternative structural configurations.

The previous sections have overviewed a number of formalisms, able to cope with sets of exclusive entities, and some of them have been used in three case studies for showing both monotypic and polytypic sets of exclusive entities, as well as some of the advantages of using one approach or the other one.

In particular a set of 12 alternative Petri nets have been considered in the first case study, while in a second example, these Petri nets have been merged into three compound Petri nets profiting from the structural similarities of the original nets. As a result, a set of three alternative compound Petri nets have been obtained.

In the last case study, the three alternative compound Petri nets have been merged into a single compound Petri net, transforming the previous polytypic set of exclusive entities into a monotypic one.

From a theoretical point of view, the present paper has proven that any of the formalisms presented in this paper can represent a set of exclusive entities. These results are the basis that allow the combination of some of these formalisms into a single model of a discrete event system.

As open research lines, it can be considered the analysis of other combinations of formalisms not considered in this paper, as well as their application to real problems, where the size of the model can be increased in orders of magnitude.

## REFERENCES

Balbo G and Silva M (eds). 1998. Performance Models for Discrete Event Systems with Synchronizations: Formalisms and Analysis Techniques. Saragossa, Spain: Kronos, 1998.

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.

Jimenez, E., Martinez, E., Blanco, J., Perez, 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-Transactions of the Society for Modeling and Simulation International*, 90, 143-161

Jiménez, E., Pérez, M., & Latorre, I. (2005). On deterministic modelling and simulation of manufacturing systems with petri nets. Paper presented at the International Mediterranean Modeling Multiconference, I3M 2005, European Modeling Simulation Symposium, EMSS 2005.

Jiménez, E., Pérez, M., & Latorre, I. (2006). Industrial applications of petri Nets: System modelling and simulation. Paper presented at the International Mediterranean Modelling Multiconference, I3M.

Jiménez, E., Pérez, M., & Latorre, J. I. (2009). Modelling and simulation with discrete and continuous PN: Semantics and delays. Paper presented at the 21st European Modeling and Simulation Symposium, EMSS 2009.

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*. 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. 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. 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., 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., 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., 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., Jiménez, E., Pérez, M., “Petri nets with exclusive entities for decision making”.

International Journal of Simulation and Process Modeling, Special Issue on Multidisciplinary Approach to Complex Systems Design and Management: Modelling and Simulation based on Methodologies and Tools. 2013 Vol. 8 No. 1

- 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
- Macias, E. J., & de la Parte, M. P. (2004). Simulation and optimization of logistic and production systems using discrete and continuous Petri nets. Simulation-Transactions of the Society for Modeling and Simulation International, 80, 143-152.
- 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.À., 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
- Recalde, L.; Silva, M.; Ezpeleta, J.; and Teruel, E. 2004. Petri Nets and Manufacturing Systems: An Examples-Driven Tour. In: Lectures on Concurrency and Petri Nets: Advances in Petri Nets, pages 742-788. Volume 3098 of Lecture Notes in Computer Science / Desel, J.; Reisig, W.; and Rozenberg, G. (editors). Springer-Verlag, June 2004.
- Silva, M. "Introducing Petri nets", In Practice of Petri Nets in Manufacturing", Di Cesare, F., (editor), pp. 1-62. Ed. Chapman&Hall. 1993.
- Xiao, Z. and Ming, Z. 2011. A method of workflow scheduling based on colored Petri nets. Data & Knowledge Engineering 70, pp. 230–247. Elsevier. 2011.
- Zaitsev, D.A., Shmeleva, T.R. 2011. A Parametric Colored Petri Net Model of a Switched Network. Int. Journal of Communications, Network and System Sciences, 4, 65-76, Scientific Research Publishing Inc., 2011.

# DISYUNTIVE COLOURED PETRI NETS: A FORMALISM FOR IMPROVING THE APPLICABILITY OF CPN TO THE MODELING OF DES WITH ALTERNATIVE STRUCTURAL CONFIGURATIONS

Latorre-Biel JI<sup>(a)</sup>, Pérez-Parte M<sup>(b)</sup>, Jiménez-Macías E<sup>(c)</sup>

<sup>(a,b)</sup> Department of Mechanical Engineering, University of La Rioja, Logroño, Spain.

<sup>(c)</sup> Department of Electrical Engineering, University of La Rioja, Logroño, Spain.

<sup>(a)</sup> [juan-ignacio.latorre@unirioja.es](mailto:juan-ignacio.latorre@unirioja.es), <sup>(b)</sup> [mercedes.perez@unirioja.es](mailto:mercedes.perez@unirioja.es), <sup>(c)</sup> [emilio.jimenez@unirioja.es](mailto:emilio.jimenez@unirioja.es)

## ABSTRACT

Colored Petri nets is a well-known formalism for constructing models of discrete event systems with subsystems presenting structural similarities. The folding of these common structures, described by means of ordinary or generalized Petri nets leads to compact and easy-to-understand models. The disjunctive colored Petri nets, can be considered as an extension of the colored Petri nets, making this formalism able to cope with the modeling of a discrete event system with alternative structural configurations. This modeling may be very useful for the task of designing a discrete event system, where some freedom degrees in the structure of the system in process of being designed lead to a set of alternative configurations for the system. This paper presents the disjunctive colored Petri nets, provides with some of their characteristics, as well as an algorithm for constructing models, and explains case study for illustrating its applicability.

Keywords: disjunctive colored petri nets, manufacturing facility, design, decision support system, discrete event system.

## 1. INTRODUCTION

Colored Petri nets is a well-known formalism, based on the paradigm of the Petri nets, (Silva 1993), for the development of models of discrete event systems with subsystems presenting structural similarities (Jensen and Kristensen 2009). The folding of these common structures, described by means of ordinary or generalized Petri nets leads to compact and easy-to-understand models (Macias and Perez, 2004). This is a main advantage of colored Petri nets, which make them very popular among practitioners (Xiao and Ming 2011), (Zaitsev and Shmeleva 2011), (Piera et al. 2004).

The folding process mentioned in the previous paragraph transfers the redundant information of repeated subsystems into attributes of the tokens. The values of the attributes of a given colored token allow knowing through which one of the original subsystems it belongs, despite in the colored model there is a single subsystem, equivalent to a set of them in the original model (David and Alla 2005).

Dealing with models of discrete event systems, such as those attainable by means of the formalism of the colored Petri nets, allow developing processes of structural analysis, and specially performance evaluation, simulation, and optimization for decision making (Mújica et al. 2010). Petri nets are not the only formalism able to cope with these tasks, as it can be seen in (Bruzzone and Longo 2010) and (Longo et al. 2013), however, they have been applied with success in many applications: generalized Petri nets (Latorre and Jiménez 2013a), (Latorre et al. 2013c), as well as colored Petri nets (Piera and Mušič 2011), (Zaitsev and Shmeleva 2011), (Piera et al. 2004).

The disjunctive colored Petri nets can be considered as an extension of the colored Petri nets, which make them able to cope with the modeling of a discrete event system with alternative structural configurations (Latorre et al. 2014a), (Latorre et al. 2010). This modeling may be very useful for the task of designing a discrete event system, where some freedom degrees in the structure of the system in process of being designed lead to a set of alternative configurations for the structural freedom degrees (Latorre et al. 2012). This formalism has been proven very useful for obtaining compact models for decision making (Latorre et al. 2014b)(Jimenez et al, 2014).

This paper presents the disjunctive colored Petri nets, provides with some of their characteristics, as well as an algorithm to construct a model of a discrete event system based on the mentioned formalism, and explains an example for illustrating its applicability.

## 2. DISJUNCTIVE COLORED PETRI NETS

There are several ways to define a disjunctive colored Petri nets. In all the possible definitions of disjunctive colored Petri nets there should be a mechanism to model a set of exclusive entities, modeling the associated set of alternative structural configurations for the modeled discrete event system.



**Definition 1. Set of boolean choice colours.**

$S_C = \{c_1, c_2, \dots, c_n \mid c_i \text{ is boolean and } \exists! c_i = \text{true}, i \in \mathbb{N}^*\}$

,  $1 \leq i \leq n \wedge c_j = \text{false} \forall j \neq i, j \in \mathbb{N}^*, 1 \leq j \leq n$  }, and the

assignment  $c_i = \text{true}$  is the result of a decision.  $\square$

In CPM ML language, the set of boolean choice variables would lead to the following sentences.

First, the color set for a single boolean choice variable is defined:

```
colset CHOICE = bool;
```

The choice color set of an token is a n-tuple of so many boolean values as the cardinality of the set of boolean choice colors  $S_C$ .

```
colset DECISION = product CHOICE * CHOICE * ... * CHOICE;
```

Where the color set `choice` appears  $|S_C|$  times.

Of course, there are other ways to represent a set of exclusive entities by means of the color of a token. However, the one presented here is enough to illustrate the concept.

The following definition will deal with the definition of a set of exclusive entities, as a way to represent in a Petri net model, a set of alternative structural configurations for a discrete event system.

**Definition 2. Monotypic set of exclusive entities.**

Given a discrete event system  $D$ , a monotypic set of exclusive entities associated to  $D$  is a set  $S_x = \{X_1, \dots, X_n\}$ , which verifies that

i) The elements of  $S_x$  are exclusive, that is to say, only one of them can be chosen as a consequence of a decision.

ii)  $\forall i, j \in \mathbb{N}^*, i \neq j$  and  $1 \leq i, j \leq n$  it is verified that  $X_i \neq$

$X_j$ .

iii) The elements of  $S_x$  are of the same type.

iv)  $\exists f: S_x \rightarrow S_R$  such that

iv.a)  $S_R = \{R_1, \dots, R_n\}$  is a set of alternative Petri nets, feasible models of  $D$ .

iv.b)  $f$  is a bijection  $\Rightarrow \forall X_i \in S_x \exists! f(X_i) = R_i \in S_R$  such that  $R_i$  is a feasible model for  $D$  and  $\forall R_i \in S_R \exists! f^{-1}(R_i) = X_i \in S_x$ .

$\square$

In (Latorre *et al.* 2014b) it has been proven that the conditions of the definition of monotypic set of

exclusive entities are verified for any set of Boolean choice variables.

**Definition 3. Monochrome choice marking.**

Let  $R = \langle N, \mathbf{m}_0 \rangle$  be a colored Petri net system.

Let us consider a feasible marking  $\mathbf{m}$  of  $R$ , reached from the initial marking  $\mathbf{m}_0$  when the sequence of transitions  $\sigma(R)$  is fired.

Let  $S_C$  be a set of boolean choice colors such that  $|S_C| = n$ .

If every token of  $\mathbf{m}$  verifies that  $\forall c_i \in S_C, c_i$  is constant, then the marking  $\mathbf{m}$  of the Petri net system  $R$  is said to be a monochrome choice marking.  $\square$

It might happen that in a certain application, a colored Petri net contains other colors, different from the choice colors. In that case, only the choice color has to be monochrome.

Furthermore, **definition 3** refers to a monochrome marking presenting constant color at a certain stage in the evolution of the Petri net. Nevertheless, there is not any restriction to the possible change of color when a transition is fired. This constraint is included in **definition 4**, which deals with the concept of a disjunctive colored Petri net. The choice marking should be constant for every marking since it will be associated to a certain decision and a decision constraints the complete evolution of a certain system.

**Definition 4. Disjunctive colored Petri net**

A disjunctive colored Petri net  $R = \langle N, \mathbf{m}_0 \rangle$  is a twelve-tuple

$CPN = \langle P, T, F, \mathbf{m}_0, \Sigma, V, c, g, e, i, S_\alpha, S_{val\alpha} \rangle$ , where:

1.  $P$  is a finite set of places.
2.  $T$  is a finite set of transitions  $T$  such that  $P \cap T = \emptyset$ .
3.  $F \subseteq P \times T \cup T \times P$  is a set of directed arcs.
4.  $\mathbf{m}_0$  is the initial marking that is a monochrome choice marking.
5.  $\Sigma$  is a finite set of non-empty color sets, such that verifies one of the following two conditions:
  - 5.a.  $\exists S_C$  set of boolean choice variables such that  $S_C \in \Sigma$ .
  - 5.b.  $\exists (c, C)$  a natural choice color such that  $C \in \Sigma$ .
6.  $V$  is a finite set of typed variables such that  $\text{type}[v] \in \Sigma$  for all variables  $v \in V$ .
7.  $c : P \rightarrow \Sigma$  is a color set function that assigns a color set to each place.
8.  $g : T \rightarrow \text{EXPR}_V$  is a guard function that assigns a guard to each transition  $t$  such that  $\text{type}[g(t)] = \text{Boolean}$ .
9.  $e : F \rightarrow \text{EXPR}_V$  is an arc expression function that assigns an arc expression to each arc  $a$  such that  $\text{type}[e(a)] = c(p)_{MS}$ , where  $p$  is the place connected to the arc  $a$ .
10.  $S_\alpha$  is a set of undefined parameters.
11.  $S_{val\alpha}$  is a set of feasible combination of values for the undefined parameters.

And it is verified that  $\forall \mathbf{m} \in rs(N, \mathbf{m}_0)$ ,  $\mathbf{m}$  is a monochrome choice marking and every token of  $\mathbf{m}$  verifies that  $c$  is constant and  $\forall c_i \in S_C$ ,  $c_i$  is constant.  $\square$

In the following section, it will be described an algorithm to construct a disjunctive colored Petri net model of a discrete event system with alternative structural configurations.

### 3. MODELING ALGORITHM

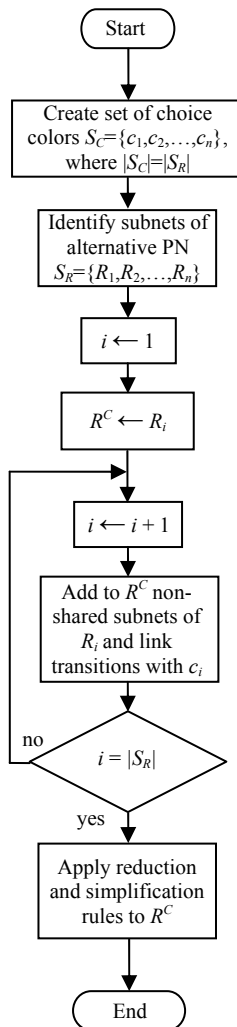


Figure 1: Algorithm for constructing a disjunctive colored Petri net from a set of alternative Petri nets.

In order to produce a disjunctive colored Petri net model for a discrete event system with alternative structural configurations, it is very convenient the use of a systematic methodology. The following algorithm is proposed for the case of a DES with alternative structural configurations, where every configuration is modeled by means of a low-level Petri net (non-colored) and the resulting set of alternative Petri nets presents a certain number of shared subnets.

In the aforementioned algorithm, an iterative procedure is described for constructing a single disjunctive colored Petri net.

Let us consider,  $S_R = \{R_1, R_2, \dots, R_n\}$ , a set of  $n$  alternative Petri nets, where  $n \in \mathbb{N}$ , and  $\mathbb{N}$  is the set of natural numbers, and let us create a set of choice colors  $S_C = \{c_1, c_2, \dots, c_n\}$ , such as  $|S_C| = |S_R|$ . As a consequence of having the same cardinality both sets, it is possible to define a bijection between them and, hence, to associate one and only one choice color from  $S_C$  to every alternative Petri net from  $S_R$ . As a general rule it will be associated  $c_i$  to  $R_i$ , where  $1 \leq i \leq n = |S_R| = |S_C|$ .

The first of the stages consists of decomposing the different alternative Petri nets into subnets and link transitions. A fast and usually effective way of coping with this task consists in considering as subnets the Petri net models of physical subsystems present in the manufacturing facility.

As a second stage, it is envisaged assigning the first alternative Petri net  $R_1$  to the disjunctive colored Petri net  $R^C$ . In this process, the link transitions are associated to a guard function that consists of the choice variable corresponding to the first alternative Petri net. Furthermore, the initial marking, conditioned by the mentioned choice color, will be composed exclusively by tokens of this choice color.

The following step in building up the disjunctive colored Petri net will be adding the subnets of the alternative Petri net  $R_2$  not contained by  $R_1$ , also called subnets not shared by  $R_2$ . Afterwards, the link transitions, with guard functions consisting of the choice color associated to  $R_2$  should be added to  $R^C$ . This last step should be repeated so many times as alternative Petri nets have not been considered yet, in fact  $|S_R| - 2$ .

The resulting Petri net will be a disjunctive colored Petri net. The appellation “colored” is due to the fact that the tokens may have attributes or colors and the adjective “disjunctive” is explained because the set of colors includes a subset of choice colors, which is a set of exclusive entities.

In the following section an example of application of this algorithm will be applied to the process of design of a manufacturing facility.

### 4. EXAMPLE OF APPLICATION

In the case study described in this section, it will be illustrated how the modeling process of a discrete event system with alternative structural configurations can be developed by the use of a disjunctive colored Petri net.

The system that will be considered is a manufacturing line in process of being designed, which will be composed of three stages: the raw materials supply, the machining of the semifinished parts, and the assembly and packing of the resulting products.

The raw materials reception system has already been chosen by the decision makers involved in the design process and will be called subsystem “A”. Moreover, the machining process can be implemented by means of two alternative subsystems, offered by two different suppliers, which will be called “B” and “C”

respectively. In the same way, the assembly and packing cell can be built up by means of other two alternative subsystems, called “D” and “E”.

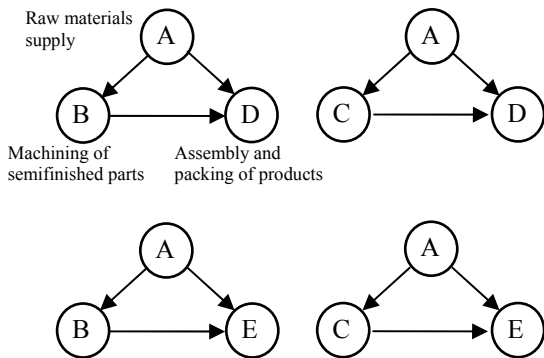


Figure 2: Alternative Configurations for the DES in Process of Being Designed.

The design of the resulting manufacturing facility will require choosing a single solution from the pool of four alternative systems obtained by the different combinations of the alternative subsystems.

The mentioned four solutions have been represented in a simplified way in figure 2, where the different subsystems are depicted by labeled circles, while the arrows inform about material flow in the manufacturing process.

A natural modeling process of the resulting discrete event system consists of obtaining a Petri net for every one of the alternative structural configurations of the system. This resulting system, modeled in the form of a set of alternative Petri nets, may be inefficient for tasks such as performance evaluation and optimization, since there is usually redundant information that can be removed.

As a result of this modeling process, it is possible to obtain four different alternative Petri nets, represented in a simplified way by means of subnets, depicted by clouds, and link transitions between some of them.

The four alternative Petri net models,  $S_R = \{R_1, R_2, R_3, R_4\}$ , have been represented in the figure 3.

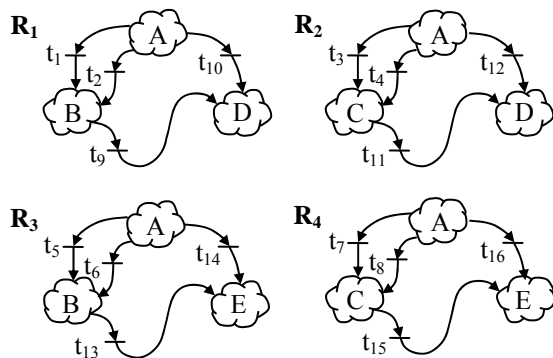


Figure 3: Simplified Representation of the Four Alternative Petri Nets decomposed into subnets and link transitions.

As it can be seen in figure 3, the subnets represented in the alternative Petri nets, which correspond to real subsystems in the DES in process of being designed, are shared by different nets. For example, subnet “A” is shared by the four alternative Petri nets, while subnet “B” is shared by  $R_1$  and  $R_3$ . For this reason, the set of four models represented in figure 3 include redundant information. This redundant information arises due to the subnets shared by several alternative Petri nets. This redundant information may reduce the computational performance of a decision making algorithm implemented to cope with the decision making in the design process of the manufacturing facility.

One way to remove the redundant information of the nets is by using CPN, where the attributes or colors of the tokens will avoid losing information when this removal is applied.

However, two considerations should be made before dealing with this modeling process. First of all, a conventional CPN is not appropriate for modeling a discrete event system with alternative structural configurations. For this use, a disjunctive colored Petri net is much more adequate, since in its definition it is included a subset of choice colors, which is a set of exclusive entities, as the set of alternative Petri nets is. Secondly, a folding of shared subnets, while a certain number of them are present and also a given number of Petri nets are taken into account may be complicated without a clear and simple methodology.

This example will illustrate the application of a technique, described in the previous section, able to cope with this problem.

The starting point of the application of the algorithm for constructing a disjunctive colored Petri net from the original discrete event system is a decomposition of the alternative Petri nets into subnets and link transitions as it has been shown, in a simplified way, in figure 3. The criterion followed for achieving the mentioned decomposition is to consider as subnets, the models of the subsystems present in the manufacturing system: the raw materials supply, the machining centers, and the assembly and packing system.

The following step in the application of the algorithm is to consider, in a first iteration, the first alternative Petri net as the disjunctive colored Petri net that will be constructed. A guard function, which corresponds to the choice color related to the first alternative Petri net,  $c_1$ , is associated to the link transitions  $\{t_1, t_2, t_9, t_{10}\}$  of  $R_1$ . In figure 4 it is possible to see the result.

As a second step in the construction of the disjunctive colored Petri net, as model of the discrete event system consists of including in  $R^c$  the subnets of  $R_2$  that are not present in  $R^c$ .

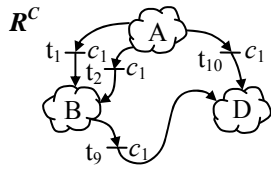


Figure 4: Simplified Representation of the disjunctive colored Petri net after the application of the first step of the algorithm.

At this early stage in the application of the construction algorithm is the same as saying that the subnets to be included in  $R^c$  should be present in  $R_2$  but not in  $R_1$ . In particular, the subnets in which, by decision of the modeler, the alternative Petri net  $R_2$  has been decomposed are  $\{A, C, D\}$ . It is a fact that  $\{A, D\}$  are shared by  $R_1$  and  $R_2$ , however,  $\{C\}$  belong to  $R_2$  but not to  $R_1$ ; hence, it should be included in  $R^c$  as well as all the link transitions of  $R_2$ , which are  $\{t_3, t_4, t_{11}, t_{12}\}$ . In figure 5 it can be seen the result of the application of this step of the construction algorithm.

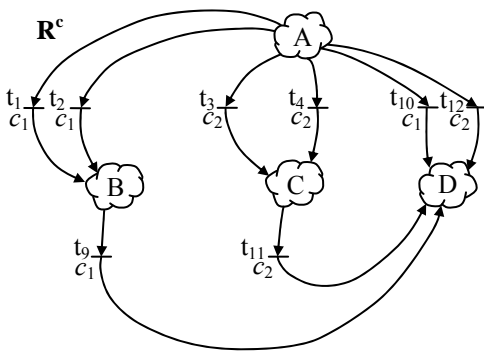


Figure 5: Simplified Representation of the disjunctive colored Petri net after the application of the second step of the algorithm.

The third step in the application of the algorithm consists in including in the disjunctive colored Petri net, the subnets of  $R_3$  that do not belong to  $R^c$  so far, that is to say  $\{E\}$ . Moreover, all the link transitions of  $R_3$  should also be included:  $\{t_5, t_6, t_{13}, t_{14}\}$ . The result of these operations can be found in figure 6.

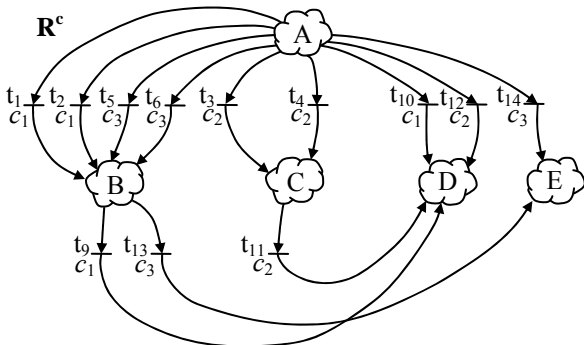


Figure 6: Simplified Representation of the disjunctive colored Petri net after the application of the third step of the algorithm.

As fourth step in the application of the algorithm, it has to be considered the alternative Petri net  $R_4$ . All the subnets in which  $R_4$  has been decomposed already belong to  $R^c$  so far. For this reason, the application of this step only implies adding to  $R^c$  the link transitions of  $R_4$ ; in other words, the transitions of  $R_4$  that does not belong to any of the subnets in which this alternative Petri net has been decomposed.

The result of this fourth step of the algorithm has been represented in figure 7.

The last step in the application of the algorithm consists of simplifying the last model obtained from the development of the previous steps in order to try to limit the number of link transitions with the purpose of reduce the size of the model.

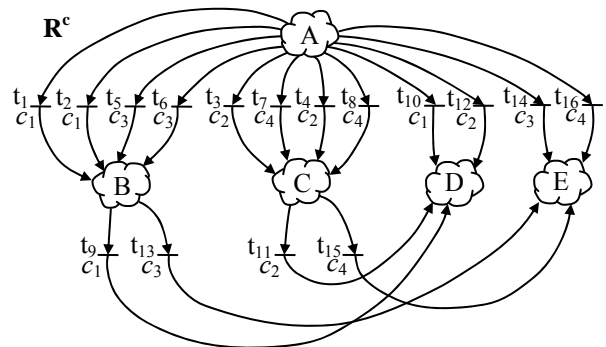


Figure 7: Simplified Representation of the disjunctive colored Petri net after the application of the fourth step of the algorithm.

Even though the details of every subnet of  $R^c$  are not shown in this paper, since they are not essential for illustrating the construction of a disjunctive colored Petri net, it is possible to state that in the example that has been considered the transitions  $\{t_1, t_5\}$  are quasi-identical. This fact means that they present input and output arcs of the same weight from and to the same places. The only difference between quasi-identical transitions is the guard functions associated to them. In the case of  $t_1$ , the guard function is the choice color  $c_1$ , while in the case of  $t_5$ , the guard function is  $c_3$ .

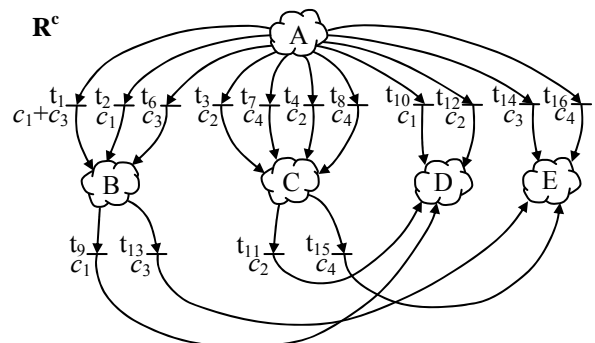


Figure 8: Simplified Representation of the disjunctive colored Petri net after the application of a reduction rule to the quasi-identical transitions  $\{t_1, t_5\}$ .

Two or more quasi-identical transitions can be reduced into a single one by creating a guard function

that combines by means of the logic operator “or” the guard functions of the quasi-identical transitions. In the case of  $\{t_1, t_5\}$ , the new transition will be called  $t_1$  and the associated guard function will be  $c_1 + c_3$ .

The result of the application of the reduction rule to the couple of quasi-identical transitions  $\{t_1, t_5\}$  can be found in figure 8.

It is possible to continue applying the reduction rule mentioned in the previous paragraphs. In doing so, it is possible to identify the following couples of quasi-identical transitions:  $\{t_2, t_6\}$ ,  $\{t_3, t_7\}$ ,  $\{t_4, t_8\}$ ,  $\{t_1, t_5\}$ ,  $\{t_{10}, t_{12}\}$ , and  $\{t_{14}, t_{16}\}$ .

The result in the application of the reduction rule to the couples of quasi-identical transitions has been represented in figure 9. This resulting disjunctive colored Petri net can be compared with the one presented in figure 7, where it had not been applied any reduction rule yet.

Furthermore, it is also possible to compare the disjunctive colored Petri net of figure 9 with the set of four alternative Petri nets depicted in figure 3.

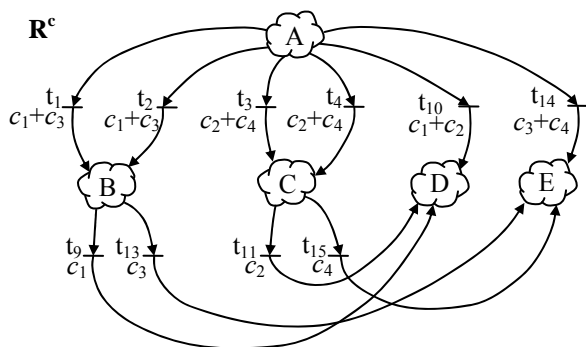


Figure 9: Simplified Representation of the disjunctive colored Petri net after the application of a reduction rule to all the quasi-identical transitions of  $R^c$ .

## 5. CONCLUSIONS

In this paper, an extension of the colored Petri nets, the disjunctive colored Petri nets have been introduced as a way to construct models of discrete event systems with alternative structural configurations.

Furthermore, an algorithm, describing the steps to be followed for constructing such model, is also detailed, as well as an application example, where a manufacturing facility in process of being designed has been modeled.

As a result, it can be said that the formalism of the disjunctive colored Petri nets is a very promising one for the description of discrete event systems with alternative structural configurations for diverse purposes, such as structural analysis, performance evaluation, or optimization, for example for the development of decision support systems.

As future lines of research, it can be envisaged the application of this formalism to a wider range of sectors and discrete event systems.

## 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.
- Jimenez, E., Martinez, E., Blanco, J., Perez, 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-Transactions of the Society for Modeling and Simulation International*, 90, 143-161.
- 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. 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., 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., 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.
- 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., 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).
- 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
- Macias, E. J., & de la Parte, M. P. (2004). Simulation and optimization of logistic and production

systems using discrete and continuous Petri nets. *Simulation-Transactions of the Society for Modeling and Simulation International*, 80, 143-152.

- 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.À., 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.
- Xiao, Z. and Ming, Z. 2011. A method of workflow scheduling based on colored Petri nets. *Data & Knowledge Engineering* 70, pp. 230–247. Elsevier. 2011.
- Zaitsev, D.A., Shmeleva, T.R. 2011. A Parametric Colored Petri Net Model of a Switched Network. *Int. Journal of Communications, Network and System Sciences*, 4, 65-76, Scientific Research Publishing Inc., 2011.



# DEVELOPMENT OF A PETRI NET MODEL FOR A RECONFIGURABLE INTELLIGENT SYSTEM BASED ON EXPERIMENTAL DATA

Juan Ignacio Latorre <sup>a</sup>, Karim El-Laithy <sup>b</sup>, Martin Bogdan <sup>c</sup>, Emilio Jiménez <sup>d</sup>

<sup>(a,d)</sup> University of La Rioja. High Technical School of Industrial Engineering. C/ Luis de Ulloa 20, 26004 Logroño, Spain  
<sup>(b,c)</sup> Leipzig University. Faculty of Mathematic and Computer Science. Postfach 10 09 20, 04009 Leipzig, Germany

<sup>(a)</sup> [juan-ignacio.latorre@unirioja.es](mailto:juan-ignacio.latorre@unirioja.es), <sup>(b)</sup> [kellaithy@informatik.uni-leipzig.de](mailto:kellaithy@informatik.uni-leipzig.de), <sup>(c)</sup> [bogdan@informatik.uni-leipzig.de](mailto:bogdan@informatik.uni-leipzig.de), <sup>(d)</sup> [emilio.jimenez@unirioja.es](mailto:emilio.jimenez@unirioja.es)

## ABSTRACT

This paper proposes a methodology for solving a double problem. On the one side, it is considered the construction of a Petri net model of a system using data recovered from its behaviour. On the other side, this model should have the capacity of reconfiguration, in order to be able to modify substantially its behaviour depending on its learning experiences. As an open research line related with this methodology, an application on the biological behaviour of neural systems is envisaged

## 1. Introduction

Several methodologies have been proposed, by different authors, for constructing a Petri net model of a system (Recalde et al., 2004; Piera et al., 2005; Gradišar and Mušič, 2007; Jiménez et al., 2009; Latorre et al., 2011b). However, rarer are the references that deal with constructing a model from scratch by the use of experimental data, which may be incomplete (Lorenz et al., 2007).

On the other hand, the construction of a reconfigurable Petri net system has been the objective of many researchers as well.

The early work of Badouel and Darondeau (1997) introduces the stratified Petri nets as a subclass of the self-modifying Petri nets defined by Valk (1978). The stratified nets change by self-modification and can be used for modelling dynamical processes whose structure evolves along computation. Moreover, Ehrig et al. (2008) describe a theory of transformations of Petri nets derived from graph transformations or graph grammars, applied as stepwise modifications in the static structure. These ideas contribute to the formal basis of the top-down modelling, PN verification and variations in the development process. The concept generalizes the idea

of place fusion of nets by the union or gluing of subnets using interfaces.

The contribution Perez-Palacin and Merseguer (2010) develops an approach to model self-reconfigurable open-world software systems with stochastic Petri nets. The mentioned systems may automatically make decisions regarding their own behaviour, evaluating the effectiveness of the reconfiguration strategies.

A relevant topic in the analysis of the transformations of a Petri net consists of the modifications that do not vary the behaviour of the system. For example, Berthelot (1987) introduced some transformations in Petri nets that preserve certain properties and allow simplifying or refine a system. Recalde et al. (1997) deals with the topic of transformation and decomposition techniques for improving the applicability of structural techniques for liveness analysis of Petri nets. The transformations are aimed at obtaining a Petri net for which well-known results exist for the characterization of certain structural properties.

Several matrix-based operations, applied to the incidence matrices of a set of alternative Petri nets, are presented in Latorre et al. (2011a). These operations correspond to transformations of the alternative Petri nets that preserve their behaviour.

In the following section the goals of the paper will be described. Section 3 will be focused on the chosen formalism, the experimental Petri nets, developed for modelling a real system. The following section describes the procedure to construct a model of a real system, based on the formalism of the experimental Petri nets, from streams of experimental data. The last section presents a research line, focused on the analysis of neural systems, where the modelling methodology presented in this paper will be applied.

## 2. Objectives to be fulfilled

The methodology proposed in this paper is aimed for the construction of a Petri net model of a system, using data recovered from its behaviour. This data can be generated by means of experiments or by observation of an existing system, whose behaviour should be mimicked by the Petri net model.

Due to limitations in the observation, for example due to a constrained time of field work, or due to limitations in the available experiments, the experimental information may be incomplete for reflecting the real behaviour of the system in the model.

In the Petri net model that will be presented in this paper, the lack of knowledge in the behaviour of the system will be handled in the same way as the capacity of the model of the system to evolve and self-reconfigure as a consequence of its experience. This experience and its implications in behavioural changes of the system may be considered as a learning process of the system. Section four deals with the advantages of the use of the methodology based on the experimental Petri nets,

## 3. Chosen formalism: experimental Petri nets

The formalism considered to build the model of a real system is the experimental Petri nets, defined in the following way:

$$R = \langle P, T, W, W^+, m_0, avtf, freqf \rangle$$

Where

$P$  is the set of places.

$T$  is the set of transitions. Notice that  $P \cap T = \emptyset$  and  $P \cup T \neq \emptyset$ .

$W$  is the input incidence matrix.

$W^+$  is the output incidence matrix. Both incidence matrices represent the arcs between the set of arcs and transitions.

$m_0$  is the initial marking.

$avtf$  is a function that associates an average firing time to every transition in  $T$ .

$freqf$  is a function that associates a frequency of appearance of a transition in a training process to every transition in  $T$ .

Given a set of transitions associated to a given conflict  $T_{cj} = \{t_a, t_b, \dots, t_q\} \in T$  a priority of firing is associated to them according to the associated value from  $freqf$ , being the highest priority the one which corresponds to the greater value of frequency of appearance in the training process and so on.

A model constructed using this formalism can be trained from a set of experimental data for representing

the discrete event system that produces the mentioned data.

## 4. Advantages of this approach

There are other formalisms that allow representing a real system by means of a quantitative model, such as the neural networks. As an improvement from that kind of formalisms that model a real system by means of "black boxes", experimental PN explicitly may show the internal structure of the real system.

This explicit representation of the internal structure of the real system may allow a researcher to know the present state of the system and the probabilities to change to a new state.

Moreover, when the internal structure of the model is known, it is possible to perform a structural analysis of the model of the system, as well as performance analysis.

Finally, this insight in the behaviour of the model might lead to a better understanding on the behaviour of the real system and maybe new results in the research on the behaviour of the system.

## 5. Procedure for the construction of a model of the real system

The stage of building up a timed Petri net model is performed from experimental data in the form of  $n$  sequences of states, including time information of the change of states, associated to  $n$  training events:

$$\begin{aligned} seq_1 &= \{ \sigma_{a1} \xrightarrow{time(a_1, b_1)} \sigma_{b1} \xrightarrow{time(b_1, c_1)} \sigma_{c1} \\ &\dots \} \\ seq_2 &= \{ \sigma_{a2} \xrightarrow{time(a_2, b_2)} \sigma_{b2} \xrightarrow{time(b_2, c_2)} \sigma_{c2} \\ &\dots \} \\ &\dots \\ seq_n &= \{ \sigma_{an} \xrightarrow{time(a_n, b_n)} \sigma_{bn} \xrightarrow{time(b_n, c_n)} \sigma_{cn} \\ &\dots \} \end{aligned}$$

All the sequences configure the set  $S_{seq} = \{ seq_1, seq_2, \dots, seq_n \}$ .

This data should verify some characteristics to allow the development of an accurate model of the behaviour and structure of the real system:

- a) Every possible state and change of state should appear in the experimental data for obtaining an accurate model of the behaviour of the system.
- b) Experimental data that may seem redundant might be, on the contrary, useful, for example if the same sequence appears several times in the experimental data; hence, it should not be removed, since some information to construct

the model of the system depends on the frequency of occurrence of every change of state. That is the case, for example, of the priority of firing transitions belonging to a conflict in the Petri net model.

- c) As a consequence of the previous considerations, the larger is  $n$ , the most accurate the model may be expected to be.

Some assumptions can be made before the model is built up:

- a) Every state is characterised by a single value of energy and a given stable value of energy is associated to a single state.
- b) There are two kinds of transitions between stable states:
  - b.1) The ones that are produced by known causes, are driven by deterministic relationships of cause and effect.
  - b.2) The ones, whose occurrence cannot be modelled deterministically because of lack of knowledge and hence introduce the need of a probabilistic approach for parts of the model of the system.

Once the experimental data has been gathered, it is time for constructing the experimental Petri net:

**Step 1.** Count the number  $m$  of different states  $\sigma_i$  in the experimental data.

**Step 2.** Create two matrices of  $m$  rows and 1 column, called  $\mathbf{W}^-$  and  $\mathbf{W}^+$  respectively, which will represent the input and output incidence matrices of the model.

Create an array **avt** with a single element that will contain the average firing time of the transition,  $avt(\sigma_{aj} \rightarrow \sigma_{bj})$ . Every element of this array will be associated to the columns in  $\mathbf{W}^-$  and  $\mathbf{W}^+$  placed in the same position (meaning to the same transition).

Create an array **freq** with a single element, which will represent the frequency of firing the transition  $freq(\sigma_{aj} \rightarrow \sigma_{bj})$ . Every element of this array will be associated to the columns in  $\mathbf{W}^-$  and  $\mathbf{W}^+$  placed in the same position (meaning to the same transition).

Let us call  $\mathbf{W}_{i,j}^-$  a generic element of  $\mathbf{W}^-$  placed in the  $i$ th row and  $j$ th column, while  $\mathbf{W}_{i,j}^+$  is the element of  $\mathbf{W}^+$  placed in the same position.

Let us call  $q$  the number of columns either in  $\mathbf{W}^-$  or  $\mathbf{W}^+$ .

**Step 3.** For every sequence of states  $seq_i$  belonging to the set  $S_{seq}$ , repeat the following steps:

**Step 3.1.** For every column of either in  $\mathbf{W}^-$  or  $\mathbf{W}^+$  (meaning for every transition already

included in the model constructed so far) repeat:

**Step 3.1.1.** If there is a change of state  $\sigma_{aj} \rightarrow \sigma_{bj}$  in  $seq_i$ , such that it is possible to find a column  $k$  in  $\mathbf{W}^-$  and  $\mathbf{W}^+$  verifying  $\mathbf{W}_{a,j,k}^- \neq 0$  and  $\mathbf{W}_{b,j,k}^+ \neq 0$  (meaning that the transition has been already included in the model of the system) then modify the values associated to the transition in the following way:

$$avt(\sigma_{aj} \rightarrow \sigma_{bj}) \leftarrow [avt(\sigma_{aj} \rightarrow \sigma_{bj}) \cdot freq(\sigma_{aj} \rightarrow \sigma_{bj}) \cdot i + time(a_j, b_j)] / [freq(\sigma_{aj} \rightarrow \sigma_{bj}) \cdot (i+1)]$$

$$freq(\sigma_{aj} \rightarrow \sigma_{bj}) \leftarrow [freq(\sigma_{aj} \rightarrow \sigma_{bj}) \cdot (i-1) + 1] / i$$

else modify the values associated to the transition in the following way:

$$freq(\sigma_{aj} \rightarrow \sigma_{bj}) \leftarrow [freq(\sigma_{aj} \rightarrow \sigma_{bj}) \cdot (i-1)] / i$$

**Step 3.2.** For every change of state  $\sigma_{aj} \rightarrow \sigma_{bj}$  belonging to the sequence  $seq_i$  not considered in the previous **step 3.1.1** repeat:

**Step 3.2.1.** Add a new column in the matrices  $\mathbf{W}^-$  and  $\mathbf{W}^+$ , completed with zeroes, except the element in the  $a_j$ th row of  $\mathbf{W}^-$  and the element in the  $b_j$ th row of  $\mathbf{W}^+$ , which should be 1.

**Step 3.2.2.** Add a new element in the array **avt** with the following value:

$$avt(\sigma_{aj} \rightarrow \sigma_{bj}) \leftarrow time(a_j, b_j)$$

**Step 3.2.2.** Add a new element in the array **freq** with the following value:

$$freq(\sigma_{aj} \rightarrow \sigma_{bj}) \leftarrow 1 / i$$

The Petri net model, obtained from this procedure can evolve if more streams of data are used for update the parameters of the model.

Once the model has been updated, it can be used with constant values for the parameters for different purposes such as structural analysis, performance evaluation, and behavioural analysis, just to give a few examples.

## 6. Application of the methodology to develop a Petri net model of a neural system.

Considering a brain as a biological neural computing system, (El-Laithy and Bogdan, 2011) proposes an interpretation of the behaviour of different features and components of neural systems, such as synapsis, spiking activity and neural states under an energy-based framework. According to this approach, a hypothetical energy function is proposed for dynamic synaptic models.

El-Laithy (2011) shows that a synapse exposes stable operating points and proposes that synapses in a network operating at these stable points can drive this network to an internal state belonging to discrete sets of neural states. The neural states exist as long as the network sustains the internal synaptic energy.

La figura 1 ilustra los estados y las transiciones entre estados que han sido extraídos de la actividad de la red y representa cada estado por medio de la letra griega  $\sigma$ . El eje de ordenadas representa el nivel de estado definido como la distancia de la actividad a una actividad de referencia. Conforme pasa el tiempo, el comportamiento de la red muestra una diversidad de estados transitorios que se han representado por la letra griega  $\sigma$ . Las transiciones entre estados se han ilustrado utilizando flechas. La secuencia de transiciones de estados ilustrada en la figura es el rendimiento del modelo computacional novedoso propuesto para la máquina de estado finita temporal. Estos estados y las transiciones de estados reflejan, de una manera abstracta, los algoritmos generales disponibles para las computaciones en la red como en una máquina de estados. Los estados que duran más señalizan más situaciones de procesamiento final que aquéllos que se mantienen menos tiempo.

Figure 1 illustrates the states and transitions between states that have been extracted from the network activity, and represents each state by the Greek letter  $\sigma$ . The vertical axis represents the level of state defined as the distance from the activity to a reference. As time goes on, the behavior of the network shows a variety of transient states. Transitions between states are illustrated using arrows. The sequence of state transitions shown in the figure is the performance of the novel computational model proposed for the temporary finite state machine. These states and state transitions presents, in an abstract way, the general algorithms available for the computations in the network as a state machine. The states that last longer represent more situations of final processing than those remaining less time.

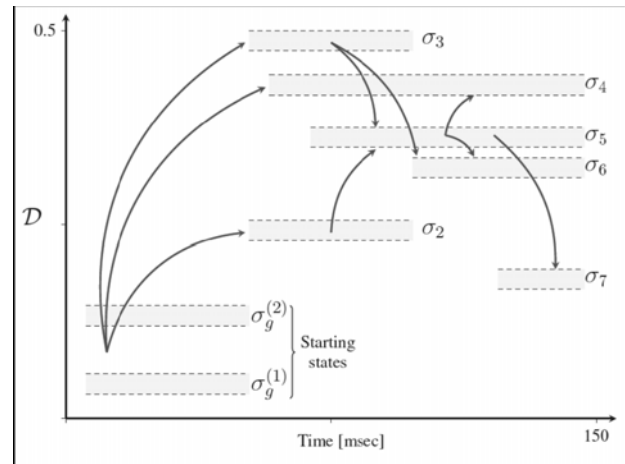


Figura 1: Result of the previous basic analysis developed (El-Laithy and Bogdan, 2011b).

This system, whose behaviour might be approximated by a discrete event system, constitutes a perfect objective for the application of the methodology based on experimental Petri nets. As a consequence, an open research line that the authors of this paper have undertaken consists in the development of a model of the mentioned biological system, which may allow analysing and understanding better the behaviour of the human brain.

This extended abstract presents the main line of the paper that is proposed, based on the results obtained up to now in the application of the methodology to develop a Petri net model of a neural system based on experimental and simulated data.

## References

- Badouel, E., Darondeau, P. "Stratified Petri nets". Research Report No. 3128, INRIA (Institut national de recherche en informatique et en automatique), 1997.
- Berthelot, G., "Transformations and decompositions of nets" in "PetriNets: Central Models and Their Properties, Advances in Petri Nets". Lecture Notes in Computer Science, Brauer, W., Reisig, W., and Rozenberg, G. (Eds.), vol. 254-I, pp. 359–376. Springer, 1987.
- Ehrig, H., Hoffmann, K., Padberg, J., Ermel, C., Prange, U., Biermann, E. and Modica, T., "Petri Net Transformations", in Vedran Kordic (eds) "Petri Net. Theory and Applications", I-TECH Education and Publishing, Vienna, 2008.
- El-laithy, K. "Towards a Brain-inspired information processing system: Modeling and analysis of

- synaptic dynamics". PhD thesis, Leipzig University, 2011.
- El-Laithy, K. and Bogdan, M., "A hypothetical free synaptic energy function and related states of synchrony". In *Artificial Neural Networks and Machine Learning - ICANN 2011*, volume 6792 of *Lecture Notes in Computer Science*, pages 40-47. Springer, 2011.
- Gradišar, D. and G. Mušič. "Production-process modelling based on production-management data: a Petri-net approach". *International Journal of Computer Integrated Manufacturing*, Vol. 20, No. 8, pp. 794-810, 2007.
- Jiménez, E., Pérez, M., Latorre, J. I. "Modelling and Simulation with Discrete and Continuous PN: Semantics and Delays", *Proceedings of the 21st European Modelling and Simulation Symposium (EMSS 09)*. Puerto de la Cruz, Spain, vol. 2, pp. 14-19, September 2009.
- Latorre, J.I., Jiménez, E., Pérez, M., "Matrix-based operations and equivalence classes in alternative Petri nets" *Proceedings of the 23rd European Modelling and Simulation Symposium (EMSS 11)*. Rome, Italy, pp. 587-592, September 2011.
- Latorre, J.I., Jiménez, E., Pérez, M., "Simulation-Based Optimization for the Design of Discrete Event Systems Modeled by Parametric Petri Nets". *Proceedings of the UKSim 5th European Modelling Symposium on Mathematical Modelling and Computer Simulation (EMS2011)*, pp. 150-155. Madrid, November 2011.
- Lorenz, R.; Bergenthum, R.; Mauser, S.; and Desel, J. "Synthesis of petri nets from finite partial languages". *Proceedings of the 7th International Conference on Application of Concurrency to System Design (ACSD)*, pages 157–166, 2007.
- Perez-Palacin, D. and Merseguer, J. "Performance evaluation of self-reconfigurable service-oriented software with stochastic Petri nets". *Electronic Notes in Theoretical Computer Science (ENTCS)*. Volume 261, pp. 181-201. 2010.
- Piera, M. A.; Guasch, T.; Porcar, V.; and Bacardit, M. "Constraint Satisfaction in the Automobile Industry: A Coloured Petri Net Simulation Model Approach". In Bruzzone, A.G. and Williams, E. (Eds.): *Proceedings of the 2005 Summer Computer Simulation Conference*, pp. 267-273. 2005.
- Recalde, L., Teruel, E., Silva, M. "Transformation and Decomposition Techniques to Improve the Decision Power of the Rank Theorems". *Proceedings of the Workshop on Petri Nets in System Engineering (PNSE'97)*, Hamburg, September 25-26, 1997 / Farwer, B.; Moldt, D.; Stehr, M.-O.: Report FBI-HH-B-205, pages 111-122. Universität Hamburg, September 1997.
- Recalde, L.; Silva, M.; Ezpeleta, J.; and Teruel, E. "Petri Nets and Manufacturing Systems: An Examples-Driven Tour". In: *Lectures on Concurrency and Petri Nets: Advances in Petri Nets*, pages 742-788. Volume 3098 of *Lecture Notes in Computer Science / Desel, J.; Reisig, W.; and Rozenberg, G. (editors)*. Springer-Verlag, June 2004.
- Valk, R. "Self-Modifying Nets, a Natural Extension of Petri nets". *Proceedings of Icalp'78. Lecture Notes in Computer Science*. Vol 62. Pages 464-476. 1978

# PETRI NET REPRESENTATION WITH CIPHERED SUBNETS: DEFINITION OF PNML EXTENSIONS FOR SUBNETS REPRESENTATION AND USE OF XML ENCRYPTION FOR CIPHERING.

Iñigo León-Samaniego<sup>(a)</sup>, Juan-Carlos Sáenz-Díez<sup>(b)</sup>, Jorge Luis García<sup>(c)</sup>, Mercedes Pérez-Parte<sup>(d)</sup>

<sup>(a,b)</sup> University of La Rioja. Department of Electrical Engineering

<sup>(b)</sup> Department of Industrial and Manufacturing Engineering. Autonomous University of Ciudad Juarez

<sup>(d)</sup> University of La Rioja. Department of Mechanical Engineering

<sup>(a)</sup> [inigo.leon@gmail.com](mailto:inigo.leon@gmail.com), <sup>(b)</sup> [juan-carlos.sáenz-diez@unirioja.es](mailto:juan-carlos.sáenz-diez@unirioja.es), <sup>(c)</sup> [jorge.garcia@uacj.mx](mailto:jorge.garcia@uacj.mx), <sup>(d)</sup> [mercedes.perez@unirioja.es](mailto:mercedes.perez@unirioja.es)

## ABSTRACT

With Petri Nets we can model a great amount of systems. However, they are described in a comprehensive way, so we need to have the hole Petri Net to work with it. We are going to take advantage from previous works presented in EMSS 2011 and EMSS 2013 in which we defined subnets of a Petri Net and extract the interface in order to cut Petri Nets in disjoint subnets. We have two aims for this work. By one side, we want to choose a way of representation of Petri Nets that allows to include all this information. In this case we are going to select PNML for this objective, but it hasn't a way to represent subnets, so we will have to extend this language with an extension that allows it. By the other side, we will use XML Encryption [4] in order to hide subnets only exposing only its interface, hiding the rest of the subnet for every non receiver of the Petri Net.

## 1. INTRODUCTION

We start from a Petri Net  $R = \langle P, T, \alpha, \beta \rangle$  where  $P$  is the set of places (Samaniego et al., 2011, 2013),  $T$  is the set of transitions,  $\alpha$  is the pre-incidence function and  $\beta$  is the post-incidence function. We define  $R' = *P'$ ,  $T'$ ,  $\alpha'$ ,  $\beta'^*$  such that  $P' \in P$  and  $T' \in T$ ,  $\alpha'$  and  $\beta'$  are restrictions of  $\alpha$  y  $\beta$  over  $P' \times T'$  ( $P'$  and  $T'$  are not empty).

$$C = \begin{matrix} & \begin{matrix} t_1 & \cdots & t_s & t_{s+1} & \cdots & t_r \end{matrix} \\ \begin{matrix} p_1 \\ \vdots \\ p_m \\ p_{m+1} \\ \vdots \\ p_n \end{matrix} & \left( \begin{array}{ccc|ccc} a_{11} & \cdots & a_{1s} & a_{1(s+1)} & \cdots & a_{1r} \\ \vdots & R' & \vdots & \vdots & HN & \vdots \\ a_{m1} & \cdots & a_{ms} & a_{m(s+1)} & \cdots & a_{mr} \\ \hline a_{(m+1)1} & \cdots & a_{(m+1)s} & a_{(m+1)(s+1)} & \cdots & a_{(m+1)r} \\ \vdots & HT & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ns} & a_{n(s+1)} & \cdots & a_{nr} \end{array} \right) \end{matrix}$$

Is shown (Latorre et al., 2011b; Latorre and Jimenez, 2013a, 2013b) that we can reorder the rows and columns of the incidence matrix without loss of generality in order to put in the first rows and columns the places and transitions belonging to  $R'$ .

So we have separated the original Petri Net in to disjoint subnets  $R'$  and the rest. This process can be

repeated every time you want in order to separate the original Petri Net in several disjoint subnets (Jiménez, 2010; Jiménez et al., 2005; Jiménez et al., 2014). Once this is done, we can analyze the inputs and outputs of  $R'$  in order to extract the interface. For example, starting from this Petri Net, we want to extract  $R_1$  and its interface (Biel et al. 2011; Latorre et al., 2013b, 2013d):

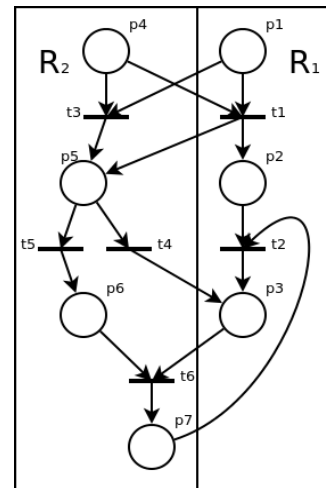


Figure 1: Petri Net divided into two subnets

This is the  $R_1$ 's interface extracted from the original Petri net [3]

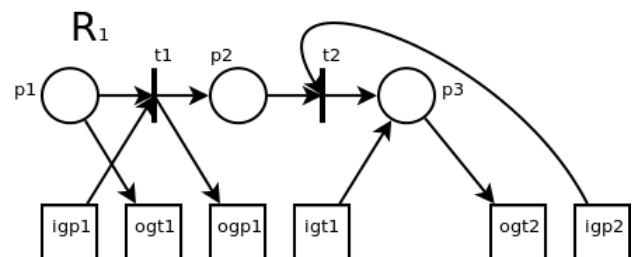


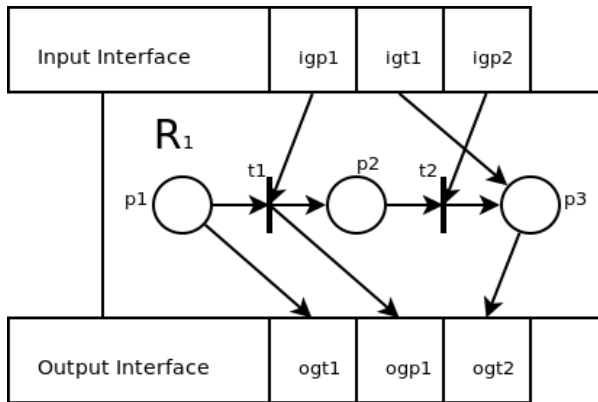
Figure 2:  $R_1$  with its interface

At this moment we have this Petri Subnet inputs and outputs. In the same way we can extract  $R_2$ 's



interface, but the process is the same. In order to have a more clear vision, we separate the input interface and the output interface in this way.

Figure 3:  $R_1$  with its input and output interface.



Once we have this interface extracted, the next step is to implement this information in a Petri Net representation in such way that we could hide the subnet, exposing only the interface. For this objective, we are going to use PNML.

## 2. PNML

Petri Net Marked Language (PNML) is an XML language created for represent Petri Nets (David and Alla, 2005; Jensen and Kristensen, 2009; Latorre et al., 2014). With this language we can take a Petri Net and store it into an XML file without loss of information.

But PNML hasn't got a way to represent subnets. So we are going to extend PNML language in order to get several goals:

1. Represent subnets of a Petri Net [3].
2. Include input and output interfaces for every subnet.

If we take the PNML grammar, there are lots of tags, but we are going to take some of them, without loss of generality. The official grammar is described and can be downloaded from the official page of PNML ([www.pnml.org](http://www.pnml.org)).

Inside a PNML document there are three main elements. This elements have required tags:

- place: define a place in the petri net with an id and a name. This corresponds with a column in the incidence matrix

```
<place id="p1">
 <name>
 <text>Place one</text>
 </name>
</place>
```

- transition: define a transition in the petri net, with an id and a name too. This corresponds with a row in the incidence matrix.

```
<transition id="t2">
 <name>
 <text>Transition 2</text>
 </name>
</transition>
```

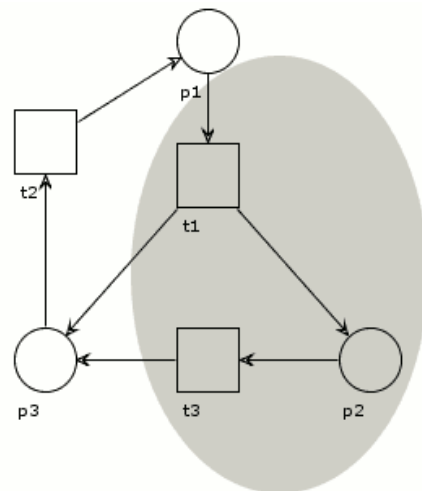
- arc: define an arc with an id, form a place to a transition or from a transition to a place, defined by their own. It corresponds with a non zero element in the incidence matrix. Here we have the arc source and target.

```
<arc id="a2" source="t1" target="p2">
 <inscription>
 <text>1</text>
 </inscription>
</arc>
```

There are other tags associated to each one of this elements, but, for simplicity, we obviate them because the process is exactly the same, but with more information.

In order not to occupy too much space in the article, let's take a simple Petri Net like this in which the subnet to be hidden is the gray ellipse.

Figure 4: Petri Net example



The PNML file representing this net is the one shown in Table1.

We want to specify that the elements inside the grey ellipse form a subnet, so we define a new tag for this objective `<subnet>`. The arcs joining elements inside the subnet are included inside the subnet. The arcs outside the subnet and the arcs entering or exiting the subnet stay outside the tag `<subnet>`. And we have this other PNML extended file (Table2):

Table 1: PNML file representing PN in Figure 4

<pre> &lt;?xml version="1.0" encoding="UTF-8"?&gt; &lt;pnml&gt;   &lt;net id="myNet"&gt;     &lt;place id="p3"&gt;       &lt;name&gt;         &lt;text&gt;p3&lt;/text&gt;       &lt;/name&gt;     &lt;/place&gt;     &lt;place id="p2"&gt;       &lt;name&gt;         &lt;text&gt;p2&lt;/text&gt;       &lt;/name&gt;     &lt;/place&gt;     &lt;place id="p1"&gt;       &lt;name&gt;         &lt;text&gt;p1&lt;/text&gt;       &lt;/name&gt;     &lt;/place&gt;     &lt;transition id="t3"&gt;       &lt;name&gt;         &lt;text&gt;t3&lt;/text&gt;       &lt;/name&gt;     &lt;/transition&gt;     &lt;transition id="t2"&gt;       &lt;name&gt;         &lt;text&gt;t2&lt;/text&gt;       &lt;/name&gt;     &lt;/transition&gt;     &lt;transition id="t1"&gt;       &lt;name&gt;         &lt;text&gt;t1&lt;/text&gt;       &lt;/name&gt;     &lt;/transition&gt;     &lt;arc id="a9" source="t3" target="p3"&gt;       &lt;inscription&gt; </pre>	<pre>         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a11" source="p2" target="t3"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a1" source="p1" target="t1"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a2" source="t1" target="p2"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a3" source="t1" target="p3"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a4" source="p3" target="t2"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a7" source="t2" target="p1"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;   &lt;/net&gt; &lt;/pnml&gt; </pre>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Table 2: PNML extended file

<pre> &lt;?xml version="1.0" encoding="UTF-8"?&gt; &lt;pnml&gt;   &lt;net id="myNet"&gt;     &lt;subnet id="sn1"&gt;       &lt;place id="p2"&gt;         &lt;name&gt;           &lt;text&gt;p2&lt;/text&gt;         &lt;/name&gt;       &lt;/place&gt;       &lt;transition id="t1"&gt;         &lt;name&gt;           &lt;text&gt;t1&lt;/text&gt;         &lt;/name&gt;       &lt;/transition&gt;       &lt;transition id="t3"&gt;         &lt;name&gt;           &lt;text&gt;t3&lt;/text&gt;         &lt;/name&gt;       &lt;/transition&gt;       &lt;arc id="a6" source="p2" target="t3"&gt;         &lt;inscription&gt;           &lt;text&gt;1&lt;/text&gt;         &lt;/inscription&gt;       &lt;/arc&gt;       &lt;arc id="a5" source="t1" target="p2"&gt;         &lt;inscription&gt;           &lt;text&gt;1&lt;/text&gt;         &lt;/inscription&gt;       &lt;/arc&gt;     &lt;/subnet&gt;     &lt;place id="p1"&gt;       &lt;name&gt;         &lt;text&gt;p1&lt;/text&gt;       &lt;/name&gt;     &lt;/place&gt;     &lt;place id="p3"&gt; </pre>	<pre>         &lt;name&gt;           &lt;text&gt;p3&lt;/text&gt;         &lt;/name&gt;       &lt;/place&gt;       &lt;transition id="t2"&gt;         &lt;name&gt;           &lt;text&gt;t2&lt;/text&gt;         &lt;/name&gt;       &lt;/transition&gt;       &lt;arc id="a1" source="p1" target="t1"&gt;         &lt;inscription&gt;           &lt;text&gt;1&lt;/text&gt;         &lt;/inscription&gt;       &lt;/arc&gt;       &lt;arc id="a3" source="t1" target="p3"&gt;         &lt;inscription&gt;           &lt;text&gt;1&lt;/text&gt;         &lt;/inscription&gt;       &lt;/arc&gt;       &lt;arc id="a4" source="p3" target="t2"&gt;         &lt;inscription&gt;           &lt;text&gt;1&lt;/text&gt;         &lt;/inscription&gt;       &lt;/arc&gt;       &lt;arc id="a7" source="t2" target="p1"&gt;         &lt;inscription&gt;           &lt;text&gt;1&lt;/text&gt;         &lt;/inscription&gt;       &lt;/arc&gt;       &lt;arc id="a8" source="t3" target="p3"&gt;         &lt;inscription&gt;           &lt;text&gt;1&lt;/text&gt;         &lt;/inscription&gt;       &lt;/arc&gt;     &lt;/net&gt;   &lt;/pnml&gt; </pre>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The elements inside the tag `subnet` are the elements of my subnet. For more clarity we will put, when possible, with dots the content of the subnet and the rest of the net.

Now we have to determinate how to represent the input and output interfaces. We define for this purpose five new tags called `interface`, outside the subnet, and `igp`, `igt`, `ogp` and `ogt` for input gates and output gates inside the subnet. In this case we have:

- One input gate from a place (`igp1`) from `p1` to `t1`
- Two output gates to a place from `t1` to `p3` (`ogp1`) and from `t3` to `p3` (`ogp2`)

In the same way we would define Input gates from transitions an Output gates to transitions (Latorre et al., 2011a, 2011c, 2014). It is clear that the only things that should be exposed outside the subnet are the gates, not the source of the output gates neither the target of the input gates. So we define in the PNML content the public interface elements inside the subnet. The rest of the subnet is encapsulated in a `<content>` tag. So we have the code in Table 3.

Table 3: PNML extended file

<pre> &lt;?xml version="1.0" encoding="UTF-8"?&gt; &lt;pnml&gt;   &lt;net id="myNet"&gt;     &lt;subnet id="sn1"&gt;       &lt;interface&gt;         &lt;igp id="igp1"/&gt;         &lt;ogp id="ogp1"/&gt;         &lt;ogp id="ogp2"/&gt;       &lt;/interface&gt;       &lt;content&gt;         &lt;place id="p2"&gt;           &lt;name&gt;             &lt;text&gt;p2&lt;/text&gt;           &lt;/name&gt;         &lt;/place&gt;         &lt;transition id="t1"&gt;           &lt;name&gt;             &lt;text&gt;t1&lt;/text&gt;           &lt;/name&gt;         &lt;/transition&gt;         &lt;transition id="t3"&gt;           &lt;name&gt;             &lt;text&gt;t3&lt;/text&gt;           &lt;/name&gt;         &lt;/transition&gt;         &lt;arc id="a6" source="p2" target="t3"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a5" source="t1" target="p2"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;       &lt;/content&gt;     &lt;/subnet&gt;     &lt;place id="p1"&gt;       &lt;name&gt;         &lt;text&gt;p1&lt;/text&gt; </pre>	<pre>           &lt;/name&gt;         &lt;/place&gt;         &lt;place id="p3"&gt;           &lt;name&gt;             &lt;text&gt;p3&lt;/text&gt;           &lt;/name&gt;         &lt;/place&gt;         &lt;transition id="t2"&gt;           &lt;name&gt;             &lt;text&gt;t2&lt;/text&gt;           &lt;/name&gt;         &lt;/transition&gt;         &lt;arc id="a1" source="p1" target="t1"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a3" source="t1" target="p3"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a4" source="p3" target="t2"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a7" source="t2" target="p1"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a8" source="t3" target="p3"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;       &lt;/net&gt;     &lt;/pnml&gt; </pre>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

The last step is to cut the arcs that enter or exit the subnet in the same way explained in the creation of the interface [3]. So we are going to have:

3. In case of an arc that enters into the subnet, we change the arc from the original source to the input gate associated to it (defined in the interface) and a new arc appears inside the subnet from the gate above to the original target.
4. In case of an arc that exits the subnet, we change this arc's original target to the output gate associated to it (defined in the interface) and a new arc appears inside the subnet from the original source to the gate above.

To represent the gates' ids, we take the subnet id and concatenate the gate's id, separated by '/'

So this is the final aspect of the PNML content that represents the example Petri net containing the described subnet in Table 4.

Table 4: PNML extended file

<pre> &lt;?xml version="1.0" encoding="UTF-8"?&gt; &lt;pnml&gt;   &lt;net id="myNet"&gt;     &lt;subnet id="sn1"&gt;       &lt;interface&gt;         &lt;igp id="igp1"/&gt;         &lt;ogp id="ogp1"/&gt;         &lt;ogp id="ogp2"/&gt;       &lt;/interface&gt;       &lt;content&gt;         &lt;place id="p2"&gt;           &lt;name&gt;             &lt;text&gt;p2&lt;/text&gt;           &lt;/name&gt;         &lt;/place&gt;         &lt;transition id="t1"&gt;           &lt;name&gt;             &lt;text&gt;t1&lt;/text&gt;           &lt;/name&gt;         &lt;/transition&gt;         &lt;transition id="t3"&gt;           &lt;name&gt;             &lt;text&gt;t3&lt;/text&gt;           &lt;/name&gt;         &lt;/transition&gt;         &lt;arc id="a6" source="p2" target="t3"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a5" source="t1" target="p2"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a9" source="sn1/igp1" target="t1"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a10" source="t1" target="sn1/ogp1"&gt;           &lt;inscription&gt;             &lt;text&gt;1&lt;/text&gt;           &lt;/inscription&gt;         &lt;/arc&gt;         &lt;arc id="a10" source="t3" target="sn1/ogp2"&gt;           &lt;inscription&gt; </pre>	<pre>           &lt;text&gt;1&lt;/text&gt;         &lt;/inscription&gt;       &lt;/content&gt;     &lt;/subnet&gt;     &lt;place id="p1"&gt;       &lt;name&gt;         &lt;text&gt;p1&lt;/text&gt;       &lt;/name&gt;     &lt;/place&gt;     &lt;place id="p3"&gt;       &lt;name&gt;         &lt;text&gt;p3&lt;/text&gt;       &lt;/name&gt;     &lt;/place&gt;     &lt;transition id="t2"&gt;       &lt;name&gt;         &lt;text&gt;t2&lt;/text&gt;       &lt;/name&gt;     &lt;/transition&gt;     &lt;arc id="a1" source="p1" target="sn1/igp1"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a3" source="sn1/ogp1" target="p3"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a4" source="p3" target="t2"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a7" source="t2" target="p1"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a8" source="sn1/ogp2" target="p3"&gt;       &lt;inscription&gt;         &lt;text&gt;1&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;   &lt;/net&gt; &lt;/pnml&gt; </pre>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

We have to notice that the subnet elements are only included inside the content of the subnet, so, outside the subnet, there is no mention to them: there are only mentions to the subnet interface. Now we can cipher the content and nothing is going to give clues about what there is inside the subnet.

This is the PNML extended format that we propose for separate a subnet from the original net. But this is only a process over an example Petri net. In further works, we will modify the PNML grammar rules ([www.pnml.org](http://www.pnml.org)) for a more formal study. PNML grammar consists in several rng files. These files are RELAX NG files. RELAX NG is a schema language for XML. Its specifications have been developed within OASIS by the RELAX NG Technical Committee.

### 3. XMLENCRYPTION

The next step in this study is to hide the private part of the subnet. For this goal, we are going to use standard technologies like XMLEncryption [4].

XMLEncryption (<http://www.w3.org/TR/xmlenc-core>) is a XML files standard cipher. We can use symmetric or asymmetric encryption, but in this case it's preferable to use symmetric ciphering because it is computationally less expensive.

The idea of this encryption is to replace the XML elements that we want to be encrypted by another piece of XML that contains the encrypted data and information about the algorithms used for encryption.

When a non XML file is encrypted, we only have not the option of encrypting it completely. When the encryption is applied to an XML file, this technology enables to define specific parts of the document to be encrypted or even transform the document before applying encryption.

Regardless of data source, the result is always an XML element. Typically, this document has all the information needed to be deciphered. Among this information can be found:

- Encryption Algorithm: The name of the method chosen to encrypt the information. It may not be included. Still need to be known by cipher and decipher.
- Ciphered information: It has been to be present always.
- Name of the password used: It is optional. Used when a set of keys known by both the part ciphering as the deciphering.
- Encrypted password: It is optional. The party that encrypts the document must have a public key and a private key. With this key you can encrypt the password used to encrypt the content. The party decrypts the document must have the other key.

Once we've got the PNML file in this extended format, we can apply XMLEncryption to the subnet <content> tag . The final file is shown in Table 5.

Table 5: PNML extended file

<pre> &lt;?xml version="1.0" encoding="UTF-8"?&gt; &lt;pnml&gt;   &lt;net id="myNet"&gt;     &lt;subnet id="sn1"&gt;       &lt;interface&gt;         &lt;igp id="igp1"/&gt;         &lt;ogp id="ogp1"/&gt;         &lt;ogp id="ogp2"/&gt;       &lt;/interface&gt;       &lt;content&gt; &lt;xenc:EncryptedData   xmlns:xenc="http://www.w3.org/2001/04/xmlenc#"   Type="http://www.w3.org/2001/04/xmlenc#Element"&gt;   &lt;xenc:EncryptionMethod     Algorithm="http://www.w3.org/2001/04/xmlenc#aes128-     cbc"     xmlns:xenc="http://www.w3.org/2001/04/xmlenc#" /&gt;   &lt;xenc:CipherData     xmlns:xenc="http://www.w3.org/2001/04/xmlenc#"&gt;     &lt;xenc:CipherValue       xmlns:xenc="http://www.w3.org/2001/04/xmlenc#"&gt;       WrlnjyJlYYOM91AYqcgCWkw2L4pUjQD2GGVoU       91VZ0wKqHY8y3l3GY8FY4i5K3G8grIe1HRFqe       7RtkFiXZgGMeYnQp6oB6ckKp3KFKHVqtucc9       rAVzOgC7XAw/oe61HRFqe6RRVzXjNm GY8FY4i5K       dId8NVPQmUSDx7NRtnRXZgGMeYnQp6oB6ckKp3       SWrlnjyJlYYOM91AYqcgCWkw2L4pU6ckKp3KFK       9h1VZ0wKqHY8y3l3GY8FY4i5K3G8grIe2xN4u7x       7fRtkFiXZgGMeYnQp6oB6ckKp3KFRRVzXjN       AtVzOgC7XAw/oe61HRFqe6RRVzXjNMLU5ZgGMeYn       Iy8NVPQmUSDx7NRtnR6YnQp6oB6GY8F       dWrlnjyJlYYOM91AYqcgCWkw2L4pUjQ3l3GY8FY       v91VZ0wKqHY8y3l3GY8FY4i5K3G8grIe2xN4u7x       b7RtkFiXZgGMeYnQp6oB6ckKp3KFKG8grIe2       nAVzOgC7XAw/oe61HRFqe6RRVzXjNMLU5T1HRFqe       LI8NVPQmUSDx7NRtnR68=     &lt;/xenc:CipherValue&gt;   &lt;/xenc:CipherData&gt; &lt;/xenc:EncryptedData&gt;       &lt;/content&gt;     &lt;/subnet&gt;     &lt;place id="p1"&gt;       &lt;name&gt; </pre>	<pre>         &lt;text&gt;p1&lt;/text&gt;       &lt;/name&gt;     &lt;/place&gt;     &lt;place id="p3"&gt;       &lt;name&gt;         &lt;text&gt;p3&lt;/text&gt;       &lt;/name&gt;     &lt;/place&gt;     &lt;transition id="t2"&gt;       &lt;name&gt;         &lt;text&gt;t2&lt;/text&gt;       &lt;/name&gt;     &lt;/transition&gt;     &lt;arc id="a1" source="p1" target="sn1/igp1"&gt;       &lt;inscription&gt;         &lt;text&gt;l&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a3" source="sn1/ogp1" target="p3"&gt;       &lt;inscription&gt;         &lt;text&gt;l&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a4" source="p3" target="t2"&gt;       &lt;inscription&gt;         &lt;text&gt;l&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a7" source="t2" target="p1"&gt;       &lt;inscription&gt;         &lt;text&gt;l&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;     &lt;arc id="a8" source="sn1/ogp2" target="p3"&gt;       &lt;inscription&gt;         &lt;text&gt;l&lt;/text&gt;       &lt;/inscription&gt;     &lt;/arc&gt;   &lt;/net&gt; &lt;/pnml&gt; </pre>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

## CONCLUSIONS

The conclusion of this work is to show that its possible to hide part of a Petri net in an easy way so that nobody can read the full Petri net unless he has the correct key to access it. This is a very important issue in some industrial applications such as automation systems, design of manufacturing plants (Biel et al. 2006; Jiménez et al., 2006; Macias and Parte, 2004), decision making based on simulation of the models (Bruzzone and Longo, 2010), etc.

Other applications may be the digital signing of a subnet in a similar way. Once signed the information we want, we will detect any non authorized modification.

## FURTHER WORKS

The next steps to continue this work are:

- Complete the PNML grammar for a more formal specification of the extended PNML described.
- Extend this process to more complex Petri nets.
- Apply digital signature ensure integrity, authentication and non repudiation of the content

## REFERENCES

- Biel, J. I. L., Macías, E. J., & De La Parte, M. P. (2006). Comparison of optimization techniques applied to a flexible manufacturing system. Paper presented at the International Mediterranean Modelling Multiconference, I3M.
- Biel, J. I. L., Macías, E. J., & Pérez De La Parte, M. (2011). Simulation-based optimization for the design of discrete event systems modeled by parametric Petri nets. Paper presented at the Proceedings - UKSim 5th European Modelling Symposium on Computer Modelling and Simulation, EMS 2011.
- Bruzzone 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.
- Jiménez, E. (2010). Modeling and simulation of periodic systems by ISS continuous PN. Paper presented at the 22th European Modeling and Simulation Symposium, EMSS 2010.
- Jiménez, E., Pérez, M., & Latorre, I. (2005). On deterministic modelling and simulation of manufacturing systems with petri nets. Paper presented at the International Mediterranean Modeling Multiconference, I3M 2005, European Modeling Simulation Symposium, EMSS 2005.
- Jiménez, E., Pérez, M., & Latorre, I. (2006). Industrial applications of petri Nets: System modelling and simulation. Paper presented at the International Mediterranean Modelling Multiconference, I3M.
- Jimenez, E., Martinez, E., Blanco, J., Perez, 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-Transactions of the Society for Modeling and Simulation International*, 90, 143-161.
- Latorre-Biel, J. I., & Jiménez-Macías, E. (2011a). The exclusive entities in the formalization of a decision problem based on a discrete event system by means of Petri nets. Paper presented at the 23rd European Modeling and Simulation Symposium, EMSS 2011.
- Latorre-Biel, J. I., & Jiménez-Macías, E. (2011b). Matrix-based operations and equivalent classes in alternative Petri nets. Paper presented at the 23rd European Modeling and Simulation Symposium, EMSS 2011.
- Latorre-Biel, J. I., & Jiménez-Macías, E. (2011c). Petri net transformation for decision making: Compound Petri nets to alternatives aggregation petri nets. Paper presented at the 23rd European Modeling and Simulation Symposium, EMSS 2011.
- Latorre-Biel, J. I., & Jiménez-Macías, E. (2013a). Comparison of incidence matrices to detect common patterns in petri nets. Paper presented at the 25th European Modeling and Simulation Symposium, EMSS 2013.
- Latorre-Biel, J. I., & Jiménez-Macías, E. (2013b). Modelling, analysis, and simulation of manufacturing processes with alternative structural configurations by parametric Petri nets. *Advanced Science Letters*, 19(2), 665-669.
- Latorre-Biel, J. I., Jiménez-Macías, E., & Pérez-Parte, 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.
- Latorre-Biel, J. I., Jiménez-Macías, E., Pérez De La Parte, M., Blanco-Fernández, J., & Martínez-Cámara, E. (2014). Control of discrete event systems by means of discrete optimization and disjunctive colored PNs: Application to manufacturing facilities. *Abstract and Applied Analysis*, 2014.
- Latorre, J. I., & Jiménez, E. (2013a). Petri nets with exclusive entities for decision making. *International Journal of Simulation and Process Modelling*, 8(1), 66-73.
- Latorre, J. I., & Jiménez, E. (2013b). Simulation-based optimization of discrete event systems with alternative structural configurations using distributed computation and the Petri net paradigm. *Simulation*, 89(11), 1310-1334.
- 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*, 89(3), 346-361.
- Macías, E. J., & de la Parte, M. P. (2004). Simulation and optimization of logistic and production systems using discrete and continuous Petri nets. *Simulation*, 80(3), 142-152.
- Samaniego, I. L., Jiménez-Macías, E., & Latorre-Biel, J. I. (2013). Analysis of information partial encryption options for exchanging petri nets systems. Paper presented at the 25th European Modeling and Simulation Symposium, EMSS 2013.
- Samaniego, I., Pérez, M., Martínez-Cámara, E., (2011). Security in sending and storage of petri nets by signing and encryption. Paper presented at the 23rd European Modeling and Simulation Symposium, EMSS 2011.



# PRODUCTION OF COMPOST FOR MUSHROOM CULTIVATION: A LIFE CYCLE ASSESSMENT STUDY.

Leiva-Lázaro, F.J. <sup>(a)</sup>, Blanco-Fernández, J. <sup>(b)</sup>, Martínez-Cámara, E. <sup>(c)</sup>, Jiménez-Macías, E. <sup>(d)</sup>

<sup>(a,b,c)</sup> Department of Mechanical Engineering. University of La Rioja. Logroño. Spain.

<sup>(d)</sup> Department of Electrical Engineering. University of La Rioja. Logroño. Spain.

<sup>(a)</sup> [Francisco.leiva@unirioja.es](mailto:Francisco.leiva@unirioja.es), <sup>(b)</sup> [julio.blanco@unirioja.es](mailto:julio.blanco@unirioja.es), <sup>(c)</sup> [ecamara@eolicas.com](mailto:ecamara@eolicas.com), <sup>(d)</sup> [emilio.jimenez@unirioja.es](mailto:emilio.jimenez@unirioja.es)

## ABSTRACT.

The composting process is an important part in the reuse of waste materials, especially those that are not recycled. Besides helping to reuse these products, the composting process significantly reduces air emissions and pollutants, compared to other processes. There are a wide variety of composting techniques. Windrow composting is the most used worldwide.

In this study the tunnel composting process with forced aeration for growing *Agaricus Bisporus* is studied. The composting process mainly consists of two phases, a tunnel composting phase and a pasteurisation phase in closed chambers. During the first phase a pile of compost with three turns, forced aeration, controlled temperature, humidity and CO<sub>2</sub> emission, is analyzed. In the second phase, the temperature, forced aeration, controlled temperature, humidity, CO<sub>2</sub> emission and the content of ammonia in the compost pile are controlled. The parameters for the entire process are constantly monitored and modified according to the needs of the process.

The results of this study indicate that the time required inside the composting tunnels and the weather conditions are directly related. A pile of compost needs more time inside tunnels in cold and wet seasons, as temperature and humidity are important parameters to control during the composting process.

**Keywords:** Compost, emissions, temperature, composting time, life cycle assessment.

## 1. INTRODUCTION

Mushroom cultivation in La Rioja has its origin in the early decades of the twentieth century. La Rioja region is well known by the red wine (Jimenez et al 2012) but it is actually the national leader in terms of production of mushrooms, presenting a strong production and marketing structure.

La Rioja has an output of 70,000 tons per year, with an income of 64.7 million Euros each year, a tenth of the total agricultural production. The biggest concentration of cultivation is in the municipalities of Pradejon, Ausejo and Calahorra. The growth capacity has increased with the Technology Mushroom Research Center created by the Government of La Rioja in the

heart of the production area and built as a great reference point and support for the local sector.

Composting is based on the action of various aerobic microorganisms (Haug, 1993), which transform the original organic matter. This process produces elevated temperatures, reduces residues' volume and weight and causes humidification and darkening. During this process various factors which provide proper microbial growth and therefore adequate mineralization of organic matter (Cronje et al., 2003), must be controlled. Since these reactions are governed by microorganisms' biological cycles involved in the process, composting requires a long process to transform the organic matter (Nakasaki et al., 2005). Therefore, a composting process that seeks to obtain a useful final product, cannot be done spontaneously, since all required variables and parameters must be controlled to ensure the total completion of the process in as a short time period as possible and a minimum costs (Hedegaard and Krüger, 1996).

The type of waste that is suitable for the composting process is large. According to Haug (1993) these include urban and industrial waste, manure, logging and agriculture residues and organic waste (Garcia et al 2014).

The life cycle assessment (LCA) provides a framework and methods to analyze and assess the environmental aspects and the potential impact of a material, a product or a service during the entire period of its life cycle (Lorenz, 2014; Roy et al., 2009; Elduque et al, 2014). The scope of the evaluation covers the extraction and processing of raw materials, manufacturing and assembly processes, energy consumption (Martinez et al 2009a, 2009b, 2010; Azofra et al 2014) product distribution, use, reuse, maintenance, recycling and final disposal (Nash and Stoughton, 1994). It is an approach which reviews the environmental effect of the above processes in a holistic way (Barton, 1996). The LCA is now considered as a powerful tool for sustainability (Tsalidis et al., 2014).

This study proposes the integration of LCA in the production system as a tool for decision-making to identify any existing or potential impact during the composting process. Likewise, the life cycle inventory (LCI) is conducted to analyze the input and output of

the process as well as any emission to study the environmental impact during composting.

## 2. MATERIALS AND METHODS.

### 2.1. Life cycle assessment.

An international standard for LCA developed by the International Organization for Standardization (ISO), has recently emerged and is under evaluation and review (Burgess and Brennan, 2001). The methodology framework used by the ISO is similar to the SETAC methodology. There are, however, some differences as it concerns the interpretation phase. The ISO include posterior as well as sensitivity analysis.

The LCA is used to determine the total impact of each product based on the resource consumption during the process. This philosophy of life cycle approach, is also known as “cradle to grave”, (Jiménez et al., 2014).

The LCA methodology has been developed to consider a wide range of resources as well as their potential environmental impact (Tsalidis et al., 2014):

- Climate change.
- Acidification.
- Eutrophication.
- Ecosystems toxicity.
- Ozone depletion.
- Smoke formation.
- Habitat destruction and alteration.
- Loss of biodiversity.
- Resources depletion.
- Water consumption.
- Land use.

LCA methodology is the most widely used tool to consider the systematic environmental impact along the life cycle of a product, process or activity (Sarasa et al, 2009). It is also a methodology according to the ISO 14040 (Jiménez et al., 2014).

### 2.2. Modeling and analysis of the production process.

Modeling and analysis of a production process involves data collection and calculation procedures to quantify the input and output of a system. It is an iterative process that must be repeated or modified, if necessary, to obtain useful information during its execution.

The system under study must be modeled as a complete sequence of unit operations that communicates with each other via input and output. It is necessary to build a real model able to represent, as accurately as possible, all the changes between the operations of the studied system. Data collection will create an inventory for each individual process, quantifying the input and output associated with the flow of the final unit output.

For this modeling phase the following steps are defined:

- Construction of the flow diagram in accordance with the limits of the systems in the definition of the objectives and scope of the study.
- Data collection of all activities in the production system through field measurements or literature.
- Calculation of the environmental loads associated with the functional unit of the final product.
- Standardization of data to consistently work with the same units in reference to the functional unit.
- Mass balance to interrelate the input and output between the different tasks of the overall process.
- Create flowcharts for each task, quantifying and indicating the input and output for each task and sub process.
- Global inventory, both input and output, process and machinery used.

## 3. RESULTS AND DISCUSSION. CASE STUDY: A COMPOSTING FACILITY IN LA RIOJA.

In this article a case study is conducted in a composting facility in La Rioja region. The tunnel composting process investigated in this study, is based on a windrow composting system in a closed container. The process takes place in a rectangular concrete tunnel 30 meters long by 6 wide and by 5 high, with a single door in the front for input loading and unloading. On the ground, the tunnel has an aeration system, with a large quantity of blowing holes through which the air is injected into the compost pile.

Each tunnel is treated as a separate unit with its own control and air system.

Leachates produced during the composting process are also used and recirculated in places where they are needed to moisten the compost mixture, as it is done in the straw pre-wet process.

### 3.1. Modeling and analysis of the composting process.

#### 3.1.1. Input mixing.

The composting process begins with the input and its storage in the company. Entries occur gradually throughout a campaign, from July to May to meet the requests of the farmers.

#### 3.1.2. Straw pre-wet process.

The next phase begins with a pre-wet process. The wheat straw is moistened for 3-9 hours to reach the required moisture for a proper composting process. This step is important because the straw has a low humidity and the composting process requires high humidity to rot the raw material correctly (Haug, 1993). At the same time, the mixing of poultry litter, gypsum, ammonium

sulphate, urea and calcium carbonate takes place. It is important to correctly adjust the properties of the substrate in this phase to make the mixture more homogenous and, thus, achieve the desired conditions of the substrate (Körner et al., 2003).

### 3.1.3. Filling the tunnels and building the compost pile.

Once the mixing phase is finished, the mixture is placed inside the tunnels. During the filling process water and leachates are introduced into the mixture to raise the moisture content and to allow the microorganisms to transform the organic matter (Jess et al., 2007).

### 3.1.4. Composting process inside the tunnels.

Once the filling phase is over, the composting process begins. Tunnel composting has 3 subphases (phase 1.1, 1.2 and 1.3), each lasting from 2 to 3 days, while the compost remains inside the tunnel. Once each subphase is completed the pile is turned to make the compost temperature homogeneous (Jess et al., 2007).

During composting, aeration times must be set, depending on the characteristics of the compost, since this will influence CO<sub>2</sub> concentration inside the tunnel. CO<sub>2</sub> concentrations will increase when the turbines are switched on. There are many odors that are emitted during composting. Therefore, the temperature of the compost should be kept elevated during this stage to avoid undesirable odors (Noble and Gaze, 1998) and to eliminate pathogens (Jess et al., 2007).

During composting, besides controlling the critical parameters such as temperature and humidity, other parameters such as pH, nitrogen, ash and C/N ratio as indicators of compost quality and productivity must be controlled (Noble and Gaze, 1996; Lyons et al., 2006).

### 3.1.5. Pasteurization phase.

Once composting is terminated, the compost pile is introduced into the pasteurization chamber using a filling machine. In agreement with the report of Noble and Gaze (1996) during pasteurization ammonia was found to be significantly removed from the compost pile. Ventilation during this process was continuous, in order for CO<sub>2</sub> concentrations to be homogenous.

### 3.1.6. Packaging and transportation of the final product.

After pasteurization, the pasteurized compost is taken to the packaging area. The final product is introduced in a packing machine with mycelium seeds to create the final product. All packets of compost are then palletized and transported to the growing areas.

### 3.1.7. Leachate collection.

During composting several wastes are generated in a liquid form, commonly known as leachates (Inbar et al., 1990; Krogmann and Woyczehowski, 2000). The leachate is a hazardous liquid because of its high concentration of pollutants. In some cases leachates contain heavy metals that must be first treated so they can be reused (Tabatabaei et al., 2012).

## 3.2. Results.

### 3.2.1. Composition of a standard compost pile.

The composition of a standard compost pile is shown in Table 1:

**Table 1**

Composition of a standard compost pile.

Material	Units	Quantity (kg)
Wheat straw	80 bales	32,180
Poultry litter	-	24,050
Urea	6 bags	240
Gypsum	-	375
Ammonium sulphate	100 bags	2,500
Calcium carbonate	-	600
Seeds	36 packets	632.16

### 3.2.2. Temperature.

Temperature is one of the critical parameters to control during composting. Temperature control is very important to destroy pathogens, to optimize breathing rates, to remove moisture and to stabilize the compost. In order to achieve this, a proper aeration control is essential (Ekinici et al., 2004).

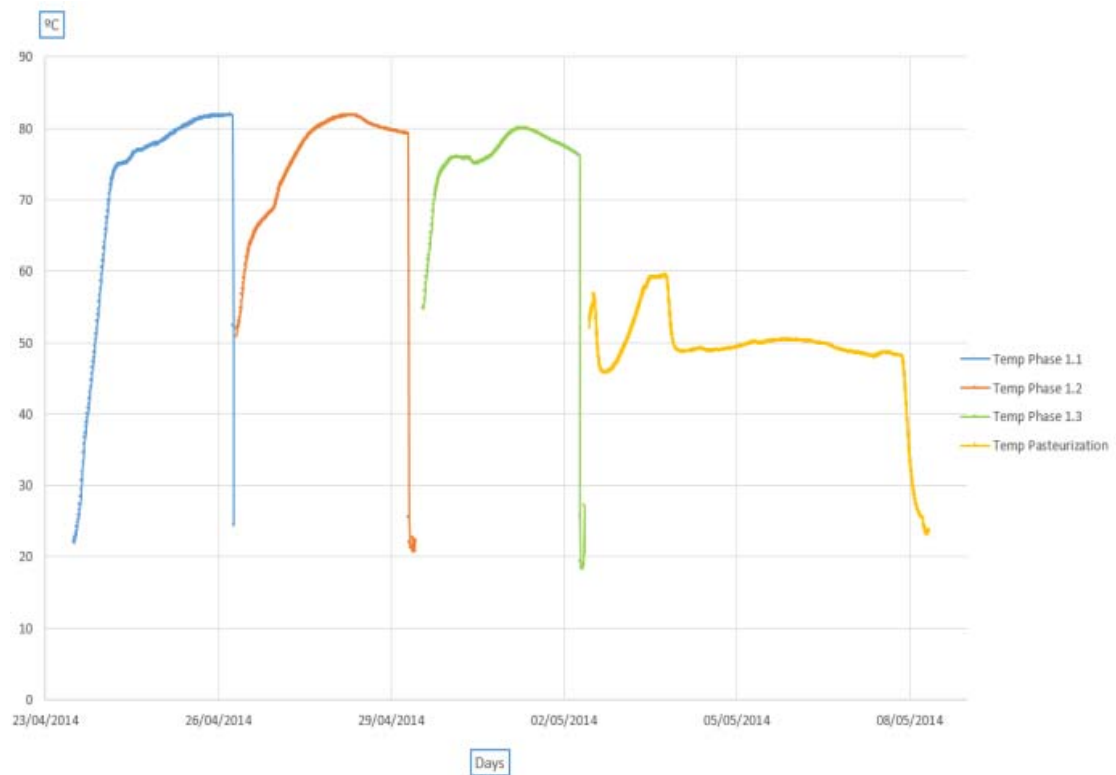
Tunnel composting enables a high control during the process, particularly of the temperature. This is due to the fact that aeration can be easily controlled. During the process in the tunnel, temperature stabilizes at 80 °C, and this occur in the middle of each subphase shown in Fig. 1. The oscillations depend largely on the incidence of air in the pile of compost, the outdoor humidity and the characteristics of the material that compose the compost pile.

Table 2: Average compost temperature.

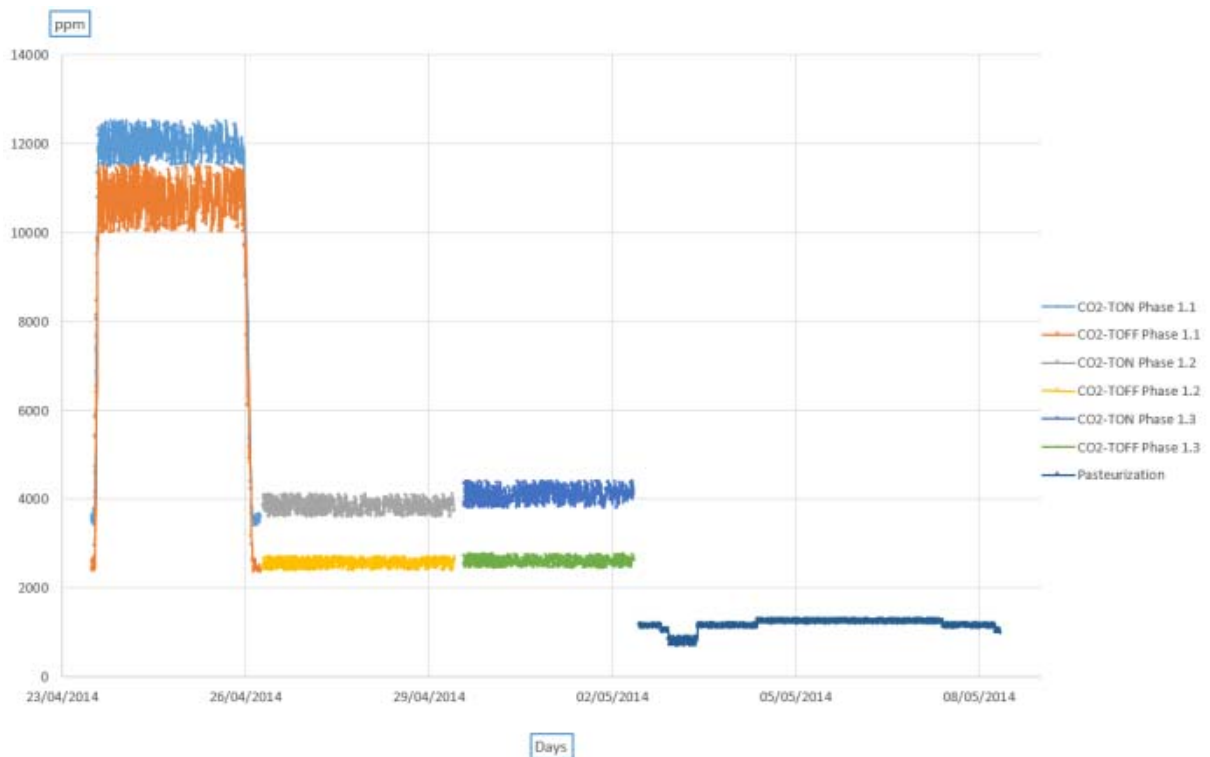
Composting phase		Pasteurization phase
Phase 1.1	71.21 °C	48.69 °C
Phase 1.2	73.5 °C	
Phase 1.3	74.33 °C	

### 3.2.3. CO<sub>2</sub> emissions.

CO<sub>2</sub> emissions are also measured throughout the process. During the first subphase, phase 1.1, the tunnel is closed so that the concentration, in ppm, of CO<sub>2</sub> is significantly higher than in other subphases, phase 1.2 and 1.3. Two different tendencies for each phase of the composting process can be distinguished in Fig.2. During the composting process the compost pile was aerated for a period of time previously established to maintain the temperature steady and to avoid the compost pile cool down. During periods of aeration the concentration of CO<sub>2</sub> is higher because it is released faster from the interior of the pile and, therefore, the concentration increases rapidly. During pasteurization emissions are more stable because the turbine is running continuously at a determined frequency.



**Fig 1.** Compost temperature



**Fig. 2.** CO<sub>2</sub> emissions.

**Table 2.**  
Average CO<sub>2</sub> emissions.

Composting phase		Pasteurization phase
Phase 1.1	10562 ppm	1169 ppm
Phase 1.2	3199 ppm	
Phase 1.3	3348 ppm	

#### 3.2.4. Time required inside the composting tunnels.

The time required inside tunnels of compost piles for a full year is analyzed. The average time required in tunnels is approximately 9 days through the year, substantial differences across seasons are noted as illustrated in Table 3.

**Table 3.**  
Time required inside the composting tunnels according to season.

Season	Time (days)
Spring	8,92
Summer	8,59
Autumn	9,22
Winter	9,92

## 4. CONCLUSIONS

Tunnel composting offers great advantages over other systems. Among them:

- High odor control.
- High temperature control.
- High pathogens rate elimination.
- Reliable process.
- High reliability and consistency of the final product.
- Dramatically reduces the processing time (to 15-20 days).

The time required inside the composting tunnel, depends, among other parameters, on the external climatic conditions. The amount of time is higher in wet and low temperature seasons than in high temperature and dry seasons. Therefore, in winter days the required time inside the tunnel is higher, and conversely in summer the amount of time is lower.

To develop the composting process optimally, with the greatest pathogens rate elimination and optimal quality of the compost, temperature, humidity, C/N ratio and pH, among other parameters, must be monitored.

In this study, the temperature control through the aeration process is studied. This is relatively easy to accomplish in the first two composting subphases. However, in the third subphase the temperature is more difficult to control because of the reduction of organic matter.

The concentration of CO<sub>2</sub> is higher during periods of aeration in the tunnel composting phase than during the pasteurization phase because of the reduction of organic matter.

In future works, apart from studying the process, the facilities and machines will be analyzed in the life cycle assessment, taking into account the materials used (Aisa et al, 2006; Javierre et al 2006, 2013, 2014; Fernandez et al 2013), the way of fabrication, etc.

## ACKNOWLEDGMENTS

This paper has been partially supported by the ADER Project: "Producción sostenible del champiñón de La Rioja y mejora de la protección ambiental, a través de la investigación de Ecoindicadores del Análisis de Ciclo de Vida (ACV)".

## 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.
- Azofra, D., Martinez, E., Jimenez, E., Blanco, J., & Saenz-Diez, J. C. (2014). Comparison of the influence of biomass, solar-thermal and small hydraulic power on the Spanish electricity prices by means of artificial intelligence techniques. *Applied Energy*, 121, 28-37. doi: 10.1016/j.apenergy.2014.01.064
- Barton J.R. 1996. Life cycle assessment for waste management. *Waste Management*; 16:35-50.
- Burgess A.A., Brennan D.J. 2001. Application of life cycle assessment to chemical processes. *Chemical engineering Science* 56: 2589-2604.
- Cronje A., Turner C., Williams A. 2003. Composting under controlled conditions. *Environ. Technol.* 24 (10): 1221-1234.
- Ekinici, K., Keener, H.M., Akbolat, D., 2004. Effect of thermocouple location on the optimum composting rate. *Biosyst. Eng.* 89, 345e353.
- Elduque, D. et al., 2014. Life cycle assessment of a domestic induction hob: electronic boards. *Journal of Cleaner Production*.
- 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.
- Garcia, J. L., Alvarado, A., Blanco, J., Jimenez, E., Maldonado, A. A., & Cortes, G. (2014). Multi-attribute evaluation and selection of sites for agricultural product warehouses based on an Analytic Hierarchy Process. *Computers and Electronics in Agriculture*, 100, 60-69. doi: 10.1016/j.compag.2013.10.009
- Haug R.T. 1993. *The Practical Handbook of Compost Engineering*. Lewis Publishers. Boca Raton. Florida.
- Hedegaard M., Krüger I. 1996. Composting of agricultural wastes in Denmark in respect of potential, industrial process technology and environmental considerations. En: De Bertoldi M., Sequi P., Lemmes B., Papi, T. (Eds.). *The Science*

- of Composting, Vol. I, pp. 691-697. Blackie Academic & Professional, London.
- Inbar, Y., Y. Chen, and Y. Hadar, Humic substances formed during the composting of organic matter, *Soil Sci. Soc. Am. J.*, 54, 1316-1323, 1990.
- Javierre, C., Abad-Blasco, J., Camañes & V., F. D., 2013. Redesign of metallic parts with thermoplastic materials: Application example | Rediseño de componentes metálicos En materiales termoplásticos: Ejemplo de aplicación. *DYNA*, 88(2), pp. 197-205.
- Javierre, C., Fernández, A., Aísa, J. & Clavería, I., 2006. Criteria on feeding system design: Conventional and sequential injection moulding. *Journal of Materials Processing Technology*, 171(3), pp. 373-384.
- Javierre, C., Fernández, Á., Clavería, I. & Elduque, D., 2014. Textile injection process characterisation by means of a spiral mould. Measurement: Journal of the International Measurement Confederation, 47(1), pp. 100-109.
- Jess, S., Murchie, A.K., Bingham, J.F.W. Potential sources of sciarid and phorid infestations and implications for centralised phases I and II mushroom compost production. *Crop Protection*. Volume 26, Issue 4, April 2007, Pages 455-464.
- Jiménez, E., Martínez, E., Blanco, J., Pérez, M., Graciano, C. Methodological approach towards sustainability by integration of environmental impact in production system models through life cycle analysis: Application to the Rioja wine sector. Simulation: Transactions of the Society for Modeling and Simulation International. 143-161. 2014.
- Jimenez, E., Tejada, A., Perez, M., Blanco, J., & Martinez, E. (2012). Applicability of lean production with VSM to the Rioja wine sector. *International Journal of Production Research*, 50, 1890-1904. doi: 10.1080/00207543.2011.561370
- Körner, I., Braukmeier, J., Herrenklage, J., Leikam, K., Ritzkowski, M., Schlegelmilch, M., Stegmann, R. 2003. Investigation and optimization of composting processes-test systems and practical examples. *Waste Manag.*, 23: 17-26
- Krogmann, U., Woyczehowski, H. Selected characteristics of leachate, condensate and runoff released during composting of biogenic waste. *Waste Management and Research*. Volume 18, Issue 3, June 2000, Pages 235-248
- Lorenz, E. Life-cycle assessment in U.S. codes and standards. Volume 59, Issue 1, December 2014, Pages 49-54
- Lyons, G.A., Sharma, H.S.S., Kilpatrick, M., Cheung, L., Moore, S., 2006. Monitoring of changes in substrate characteristics during mushroom compost production. *J. Agric. Food Chem.* 54, 4658e4667.
- Martinez, E., Jimenez, E., Blanco, J., & Sanz, F. (2010). LCA sensitivity analysis of a multi-megawatt wind turbine. *Applied Energy*, 87, 2293-2303. doi: 10.1016/j.apenergy.2009.11.025
- Martinez, E., Sanz, F., Pellegrini, S., Jimenez, E., & Blanco, J. (2009). Life-cycle assessment of a 2-MW rated power wind turbine: CML method. *International Journal of Life Cycle Assessment*, 14, 52-63. doi: 10.1007/s11367-008-0033-9
- Martinez, E., Sanz, F., Pellegrini, S., Jimenez, E., & Blanco, J. (2009). Life cycle assessment of a multi-megawatt wind turbine. *Renewable Energy*, 34, 667-673. doi: 10.1016/j.renene.2008.05.020
- Nakasaka K., Nag K. Karita S. 2005. Microbial succession associated with organic matter decomposition during thermophilic composting of organic waste. *Waste Manag. Res.* 23 (1): 48-56.
- Nash J., Stoughton M.D. 1994. Learning to live with life cycle assessment. *Environmental Science and Technology* 28: 236A-237A.
- Noble, R., Gaze, R.H. Composting in aerated tunnels for mushroom cultivation: Influences of process temperature and substrate formulation on compost bulk density and productivity. *Acta Horticulturae*. Volume 469, 1998, Pages 417-426
- Noble, R., Gaze, R.H. Preparation of mushroom (*Agaricus bisporus*) composts in controlled environments: factors influencing compost bulk density and productivity. *International Biodeterioration and Biodegradation*. Volume 37, Issue 1-2, 1996, Pages 93-100
- Roy, P., Nei, D., Orikasa, T., Xu, Q., Okadome, H., Nakamura, N., Shiina, T. A review of life cycle assessment (LCA) on some food products. *Journal of Food Engineering*. 2009.
- Sarasa, J., Gracia, J. & Javierre, C., 2009. Study of the biodegradation of a bioplastic material waste. *Bioresource Technology*, 100(15), pp. 3764-3768.
- Tabatabaei, S.H., Najafi, P., Mirzaei, S.M.J., Nazem, Z., Heidarpour, M., Hajrasoliha, S., Afyuni, M., Harchegani, H.B., Landi, E., Akasheh, L., Zamanian, M., Barani, M., Amini, H. Compost leachate recycling through land treatment and application of natural Zeolite. *International Journal of Recycling of Organic Waste in Agriculture*. 2012. 1:2.
- Tsalidis, G.A., Joshi, Y., Korevaar, G., de Jong, W. Life cycle assessment of direct co-firing of torrefied and/or pelletised woody biomass with coal in The Netherlands. *Journal of Cleaner Production*. 2014



# MODELING OF ALTERNATIVES FOR IN SITU UNIDIRECTIONAL SLAB

Esteban Fraile-García<sup>(a)</sup>, Javier Ferreiro-Cabello<sup>(b)</sup>, Eduardo Martínez-Cámara<sup>(c)</sup>

<sup>(a)</sup> University of La Rioja. Mechanical Engineering Department. EDMANS Group

<sup>(b)</sup> Qualiberica SL and University of La Rioja. Mechanical Engineering Department

<sup>(c)</sup> University of La Rioja. Mechanical Engineering Department

<sup>(a)</sup> [estean.fraile@unirioja.es](mailto:estean.fraile@unirioja.es), <sup>(b)</sup> [ferrecabello@yahoo.es](mailto:ferrecabello@yahoo.es), <sup>(c)</sup> [eduardo.martinezc@unirioja.es](mailto:eduardo.martinezc@unirioja.es)

## ABSTRACT

Slab is an element in which we consider basic criteria that structures, like any other human ingenuity, must satisfy: functionality, safety, economy and aesthetics. Technological developments have provided us with materials with high resistant performance level, appearing simultaneously concepts such as recycling or additional benefits of comfort (acoustic, thermal, fire resistance ...). This piece of research focuses on the case of unidirectional floor fully implemented on site, usually called "In Situ", and analyses the combination of such important criteria..

Keywords: slab, environmental impact, multicriteria selection.

## 1. INTRODUCTION

Slab is an element in which we consider basic criteria that structures, like any other human ingenuity, must satisfy: functionality, safety, economy and aesthetics. The basic criterion of structural functionality (stiffness) is combined with additional ones, such as the environmental impact (Elduque et al, 2014), ease of maintenance and management, recyclability (Jimenez et al, 2012), etc., which become more important added in more developed economies. Clearly, the relative importance of each of the criteria depends on the type of concerned structure.

This piece of research focuses on the case of unidirectional floor fully implemented on site, usually called "In Situ". This type of unidirectional floor has as main characteristic the absence of prefabricated structural elements. In this case the nerve of the child element is set to work with deformed bars, if necessary reinforcement is placed to withstand extreme shear. The other components are the usual lightening element (vaults in different materials), mesh in the compression layer and negative reinforced.

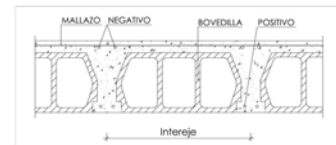


Figure 1 Unidirectional slab with in situ nerve

This configuration is quite wrought more flexible than those using prefabricated elements. In this case starting from a lightening element of constant width (56 centimeters) can be set different widths nerves and compression layers.

Moreover environmental issues become relevant in recent years and there are new concepts such as sustainable construction (Martinez et al 2009a, 2009b, 2010). In the shadow of the different ways that concept are established to quantify the environmental impacts of human activities on consumption (Azofra et al 2014a, 2014B) and its production. Carbon Footprint and Life Cycle Analysis (LCA) as parameters to measure these impacts; both values are used routinely. The carbon footprint is to quantify the total greenhouse gas (GHG) emitted by direct or indirect effect of a product / service or organization as a whole (eg CO<sub>2</sub>, methane, HFCs, sulfur hexafluoride, etc.). Determining the carbon footprint continues to be a study of Life Cycle Assessment (LCA) which calculates and considers only a single category of environmental impact, the Global Warming, whose measurement units are kg of CO<sub>2</sub> equivalent (Evans 1994) (Armengou et al 2012) (Baldo et al 2002).

Global warming is an environmental impact category that has gained great importance in recent years. In this study the impact of CO<sub>2</sub> equivalent emissions for various alternative unidirectional slabs (Carrascón et al 2007) CTE, EHE-08, EN 2006a, 2006b, 2012) (Danatzko et al 2013) is identified. This piece of information is added to the structural performance of different alternatives providing useful additional information for making correct decisions (Fernandez-Ceniceros et al 2013) (Ferreiro-Cabello et al 2013a, 2013B, 2013c, 2013d).

## 2. METHODOLOGY

The research methodology focuses on the analysis of structural performance and CO<sub>2</sub> emissions for the case of one-way (unidirectional) slabs executed in the work (in situ). The materials incorporated into the study alternatives are reduced to three, lightening elements, steel and concrete (Jönsson et al 1998) (Monahan and Powell 2011).

Starting from a base geometry for the lightening element and changing the other geometrical parameters to incorporate, the case study (Reza et al, 2011) (Ximenes and Grant 2013) are defined.

24 geometric arrangements in which we vary the structural depth between 15 and 40 centimetres are modeled. These geometric arrangements record three different rib widths (12/14/16 cm), four depth of regular slabs (15/20/25/35 centimetres) and two compression layers (5/10 centimetres), to determine a variation of the resistant section (% SR) which runs from the maximum of 53.33% to 29.41% minimum. This percentage represents the amount corresponding to the reinforced concrete section compared to the total.

The implementation of the cases using lightening materials of vibratory concrete and polystyrene foam, incorporates solutions with distinct structural impacts and response.

This structural element is modeled on a primary beam structure. For the analysis of the responses of the floor, it was considered a means and constants for the consumption of concrete and steel structure of this primary values. With the usual loads in buildings, wide beam (primary structural element) of 60 centimeters and a consumption of 11 kg Fe/m<sup>2</sup>. The evaluation is made for discrete cases 4, 5, 6 and 7 meters span slab to visualize the results according to level of structural stress.

From these changes in geometry, materials and lighting calculation a total of 192 models were obtained. The charges referred to are constant and correspond to 2.5 kN/m<sup>2</sup> for dead loads and 2 kN/m<sup>2</sup> for overhead use.

Structural analysis of different solutions is performed discarding those that do not comply with regulations. For structural analysis alternatives are modeled by CYPECAD software. Assumptions for calculating are made both for "Ultimate Limit States" (ELU) and for "Serviceability Limit States" (ELS), complying with regulations of the Technical Building Code (CTE).

Defined dimensions, materials, loads, and assumptions, to determine the next step we must calculate the 192 possible solutions through

professional program for calculating concrete structures, CYPECAD. The analysis of the stresses is performed by a calculation in 3D space by stiffness matrix methods, with a model consisting of all the elements that define the structure: columns, beams and slabs.

Compatibility of deformations at all nodes is established, considering 6 degrees of freedom, and the indeformability of the plane hypothesis of each plant is created to simulate the rigid behavior of the floor, preventing relative displacements between its nodes (rigid diaphragm). Therefore, each plant can only rotate and move as a whole (3 degrees of freedom).

For all charge states a static calculation is performed (except when considering earthquake dynamics actions, in which case the spectral modal analysis is used) and a linear material behavior is assumed and therefore, a calculation of the first order, in order to obtain displacements and efforts.

Through this structural analysis, the different armed elements are defined according to the stresses in each of the sections, applying the rules of Spanish concrete currently in force (EHE-08). Our objective is not to stop at this point, but trying to assess and compare the structural performance of each alternative studied. For this, stiffness is selected as the most significant parameter.

In engineering, the stiffness is the capacity of the structural element to withstand stresses without acquiring large deformations and / or displacements. Therefore, if we control the deflections and deformation we will know how rigid the structure is. In reinforced concrete constructions in which the deformation of the considered as structural elements can affect those considered as non-structural, it is necessary to control the values of active and total deflection at infinite time in these structural elements.

Active deflection of a structural element referred to one damageable as nonstructural means hath deflection occurs in the first from the existence of the second one.

The total deflection at infinite time is the deflection obtained as the sum of the active deflection more deflection developing the structural element (slabs, beams ...) until the moment of the construction of damageable element (partitions, walls, party walls ...).

Since the active deflection is what causes the damage to nonstructural elements, it has been selected as a parameter for the evaluation of different alternatives.

Given current regulations EHE-08 and CTE-DB-SE as to the limitations of time, we have considered to take the following in our research: limitation of deflection active for one-way slabs:  $F_{Act} = L/500$  or  $L/1000$   $F_{Act} = +0.5$ .

Additionally, in all cases it has been considered a limit deflection at infinite time:  $\text{Inf FPlazo} +1 = L/500$  or  $L/250$ .

The numerical value of the active deflection in centimetres allows comparison between different alternatives by determining the degree of goodness of the solution in order to stiffness.

The construction, like any other activity of human character, has effects on the environment with which it interacts, as it requires a significant amount of resources and energy.

The environmental impact of different proposals was evaluated according to the CML methodology and software package SIMAPRO. The research details the construction sector are incorporated by Spanish databases (BEDEC-CYPE) and a more extended and specific stroke worldwide (ECOINVENT), complementing all this information with EPD.

From a scientific point of view, academic analysis suggests the use of CML and to compare the results of impact categories, these being those shown in Table 1 with their respective units.

Impact category	Unit
Abiotic depletion	kg Sb eq
Global warming (GWP100)	kg CO2 eq
Ozone layer depletion (ODP)	kg CFC-11 eq
Human toxicity	kg 1,4-DB eq
Fresh water aquatic ecotox.	kg 1,4-DB eq
Marine aquatic ecotoxicity	kg 1,4-DB eq
Terrestrial ecotoxicity	kg 1,4-DB eq
Photochemical oxidation	kg C2H4
Acidification	kg SO2 eq
Eutrophication	kg PO4--- eq

Table 1 Impact categories depending on CML

As the ultimate goal is to compare different solutions, the task of comparing these 10 CML categories separately for the 192 solutions and obtaining optimal solutions is neither trivial nor recommended. The Global Warming (GW equivalent CO2 emissions) category is selected as the representing one.

For the development of the LCA a division in main structural components of the proposed solution is established, which allows to incorporate the contribution of each of the components at different stages of the life cycle, from cradle to grave. To achieve this goal the manufacture of the components is taken into account, and the transport to site, installation and assembly, and processing of the waste at the end of its useful life.

The functional unit is considered one m2 structure executed in reinforced concrete, being outside the limits of field of study the following activities: foundation, auxiliary transport elements within the work, different residential uses, values of operations that are constant in the construction process, and final dismantling of the structure of its useful life.

During the development of life cycle inventory, and to facilitate the final calculation of the environmental impact associated with each of the different structural solutions raised, a division in three main components has been established: rebar, concrete and lightening elements.

With this parameterized information on the basic components of the studied structural solutions, assigning a numerical value to the environmental impact produced by each solution is achieved, in this case the Global Warming (GW Emissions of CO2 equivalent). With this value we can compare the different solutions from the point of view of environmental impact.

At this point we have a database in which the amounts of materials used for each alternative and its benefits in stiffness are reflected, as a result of the structural analysis. Moreover we know the CO2 emissions produced by the realization of the solution proposed, as a result of environmental analysis. From this database, we can graph the results as reflected in the following section.

### 3. RESULTS

The presentation of the results using graphs includes information corresponding to each material emissions, and the structural response of the solution. This is repeated for each of the studied span, which represent different levels of structural stress.

The representation presents on the vertical axis the value of the equivalent CO2 emissions in kg. This information is presented fractionated, making possible to appreciate the emissions from each of the elements of the slab. For each case the top reports the stiffness value in centimeters.

Meanwhile on the horizontal axis the percentage of resistant section and the summary of the case definition for forging are presented. The option of using lightening concrete vibratory element (H) or expanded polystyrene (P) are shown in each case. A specific case is illustrated in Table 2.

%SR	29,41%	H
Bov+CC Intereje	30+5 68	P

Table 2 Identification of slab and resistant section percentage

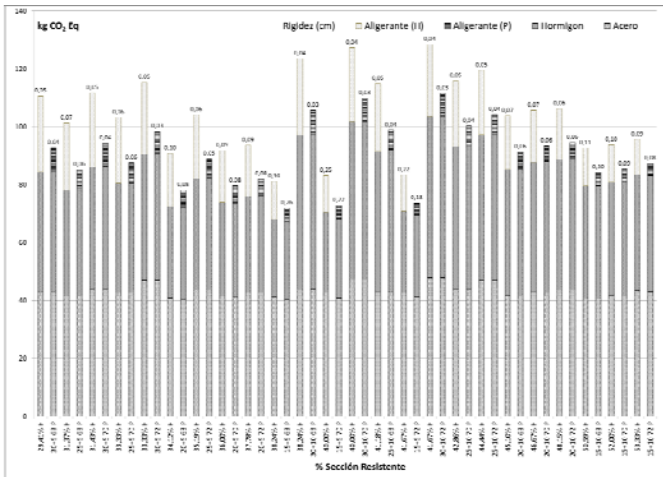


Figure 2 Results of Stiffness/Emissions (4 meters span slab)

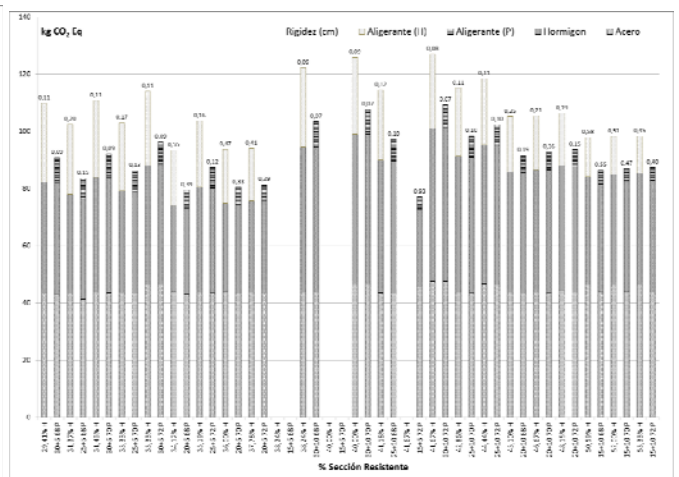


Figure 3 Results of Stiffness/Emissions (5 meters span slab)

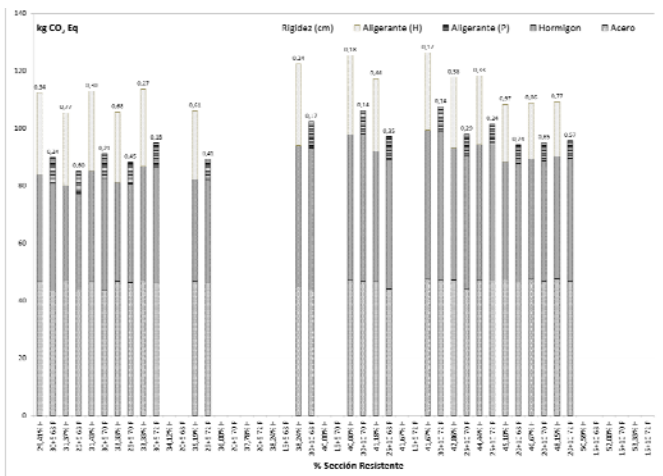


Figure 4 Results of Stiffness/Emissions (6 meters span slab)

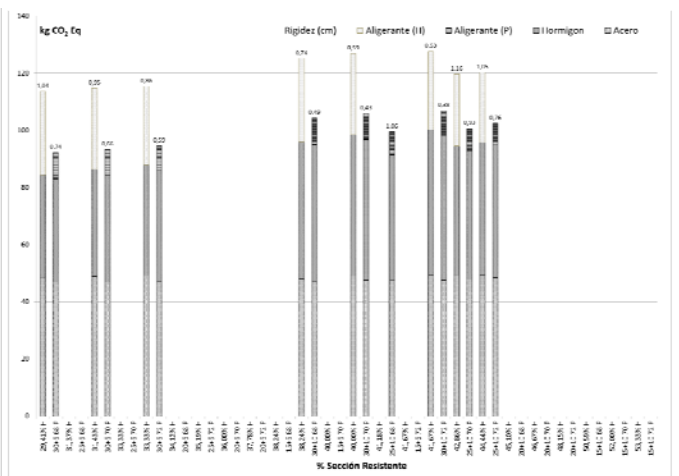


Figure 5 Results of Stiffness/Emissions (7 meters span slab)

For the case of low beams (four meters), given the level of stresses, 100% of cases are technically feasible. Solutions with lower CO2 emissions correspond to the configuration of 15+5 68 using both vibratory as expanded polystyrene concrete with total emissions of 81.26 and 72.02 kg CO2 Eq/m2. Emissions steel used for the same values ranging from the difference in emissions attributable to lightening element. In the section of the proposed stiffness, options with lower emissions cause major deformations, in this case 0.26, and 0.34 using expanded polystyrene if raised by lightening of vibratory concrete. Minimum deflexion provides similar results being 0.03 (P) and 0.04 (H) the estimated values (Figure3).

In the case of five meters span an increase in the level of the stresses is produced, and slabs of 20 centimeters of total depth are viable only using expanded polystyrene and 16 inches wide rib lightening element. Solutions with lower CO2 emissions are:

configuration 15+5 72 using expanded polystyrene with total CO2 emissions of 77.06 kg Eq/m2 and 20+5 68 configuration using vibratory concrete with total emissions of 93.42 Eq/m2 kg CO2. The emissions for the steel used in this case are lower in cases (P), equaling slightly on the high ridges. Regarding stiffness the options with lower emissions assume major deformations, in this case being 0.93, and 0.55 using expanded polystyrene, if raised by lightening of vibratory concrete. Minimum deflexion provides similar results, being 0.07 (P) and 0.08 (H) with 109.29 and 127.02 emissions respectively (Figure 4).

In the case of six meters span it occurs a further increase in the level of stresses, and technically viable slabs are reduced to 62.5% of the initial options. Solutions with lower CO2 emissions correspond to the configuration 25+5 68 using both vibratory as expanded polystyrene concrete with total emissions of 85.14 and 105.53 kg CO2 Eq/m2. The emissions for the steel used in this case have a lower swing, only 2.91 kg of CO2

emitted in cases (H) and 3.53 in cases (P). Regarding stiffness, the options with lower emissions assume major deformations, in this case 0.74, and 1.04 using expanded polystyrene if raised by lightening of vibratory concrete. Minimum deflection now presents similar results, being 0.14 (P) and 0.17 (H) with 107.50 and 126.28 emissions respectively (Figure 5).

Finally, in the case of seven meters span, an extreme level is presented for the stresses, and technically feasible slabs are reduced to 35.4% of the initial options. Solutions with lower CO<sub>2</sub> emissions correspond to the configuration of 68 to 30+5 lightening the two materials, with total emissions of 92.26 and 113.63 kg CO<sub>2</sub> Eq/m<sup>2</sup>. The emissions for the steel used in this case have a lower swing only 1.6 kg of CO<sub>2</sub> emitted in the cases H (47.9 to 49.5) and 1.75 where P (46.75 to 48, 5). Regarding stiffness, the options with lower emissions assume major deformations, in this case 0.74, and 1.04 using expanded polystyrene if raised by lightening of vibratory concrete. Minimum deflection now shows different results being 0.38 (P) and 0.53 (H) with 106.91 and 127.72 emissions respectively.

#### 4. CONCLUSIONS

This research reflects on alternatives for one-way slabs in situ concrete structures, justifying the need for further multi-methods as decision making support. The current means and techniques of modeling and simulation allow us to incorporate existing metadata in the planning stages.

The conclusion from this study is that, as the stresses increase, solutions with lower emissions are achieved by decreasing the resistant section (singing increases without increasing widths nerve).

Indeed the rigidity optimum and the emission optimum do not match. It is necessary to implement selection algorithms to obtain the best definitions for each level of stress.

Emission of 1 kg of expanded polystyrene are greater than 1 kg of vibratory concrete. But incorporating lightening elements, less densely, derives in lower emissions caused in cases (P).

With the increase in the stresses, the system moves to higher emissions and lower rigidities. This phenomenon is more pronounced in cases (H) due to the weight of the implemented solutions.

#### REFERENCES

Armengou J., Aguado A., Ormazábal G. (2012) "Sistema integrado para toma de decisiones en el diseño de estructuras de hormigón". Informes de la construcción, ISSN 0020-0883, Vol. 64, N°. 527, págs. 391-400.

Azofra, D., Jimenez, B., Martinez, E., Blanco, J., & Saenz-Diez, J. C. (2014). Wind power merit-order and feed-in-tariffs effect: A variability analysis of

the Spanish electricity market. *Energy Conversion and Management*, 83, 19-27. doi: 10.1016/j.enconman.2014.03.057

Azofra, D., Martinez, E., Jimenez, E., Blanco, J., & Saenz-Diez, J. C. (2014). Comparison of the influence of biomass, solar-thermal and small hydraulic power on the Spanish electricity prices by means of artificial intelligence techniques. *Applied Energy*, 121, 28-37. doi: 10.1016/j.apenergy.2014.01.064

Baldo, G.L., Rollino, S., Stimmeder, G., Fieschi, M. (2002) "The use of LCA to develop eco-label criteria for hard floor coverings on behalf of the European flower" *International Journal of Life Cycle Assessment* 7 (5), pp. 269-275.

Carrascón Ortiz S., Aguado de Cea A., Josa A. (2007) "Evaluación medioambiental de productos de hormigón mediante el análisis del ciclo de vida". *Cemento Hormigón*, ISSN 0008-8919, N°. 898, págs. 40-48.

CTE "Código Técnico de la Edificación". Real Decreto 314/2006, de 17 de marzo, del Ministerio de Vivienda (España).

CYPECAD (2012) "CYPE Ingenieros Software para Arquitectura, Ingeniería y Construcción". CYPE Ingenieros S.A.

Danatzko, J.M., Sezen, H., Chen, Q. (2013) "Sustainable design and energy consumption analysis for structural components" *Journal of Green Building* 8 (1), pp. 120-135.

EHE-08 "Instrucción de Hormigón Estructural", Real Decreto 1247/2008, de 18 de julio, del Ministerio de Fomento (España).

Elduque, D., Javierre, C., Pina, C., Martinez, E., & Jimenez, E. (2014). Life cycle assessment of a domestic induction hob: electronic boards. *Journal of Cleaner Production*, 76, 74-84. doi: 10.1016/j.jclepro.2014.04.009

Evans, W.A., 1994. Approaches to intelligent information retrieval. *Information Processing and Management*, 7 (2), 147-168.

Fernandez-Ceniceros, J., Fernandez-Martinez, R., Fraile-Garcia, E., Martinez-De-Pison, F.J. (2013). "Decision support model for one-way floor slab design: A sustainable approach" *Automation in Construction* 35 PP. 460 - 470.

Ferreiro-Cabello J., Fraile-García E., Martínez-Cámara E., Jiménez-Macías E. (2013). "Modelling and Simulation of Slabs Combining Performance, Costs, and Life Cycle Assessment". 8th EUROSIM Congress on Modelling and Simulation IEEE ISBN 978-0-7695-5073-2 Cardiff, Wales.

Fraile-García E., Ferreiro-Cabello J., Martínez-Cámara E., Jiménez-Macías E. (2013) "Reducción de impacto ambiental mediante optimización en la edificación: Indicadores representativos para optimización de forjados reticulares". VIII Conferencia Científica Internacional Medio Ambiente Siglo XXI, MASXXI. Cayo Santa María, Cuba.

- Fraile-García E., Ferreiro-Cabello J., Martínez-Cámara E., Jiménez-Macías E. (2013) "Incorporación del Análisis de Ciclo de Vida en el Modelado de forjados reticulares en edificación". VIII Conferencia Científica Internacional Medio Ambiente Siglo XXI, MASXXI. Cayo Santa María, Cuba.
- Fraile-García, E., Fernández-Ceniceros, J., Martínez-de Pisón, E., Martínez-de Pisón, F.J. (2013). "Selection of one-way floor slabs based on technical, economic and sustainability criteria". *Informes de la Construcción* 65 (530) PP. 183 – 194.
- Jimenez, E., Tejada, A., Perez, M., Blanco, J., & Martinez, E. (2012). Applicability of lean production with VSM to the Rioja wine sector. *International Journal of Production Research*, 50, 1890-1904. doi: 10.1080/00207543.2011.561370
- Jönsson, Å., Björklund, T., Tillman, A.-M. (1998) "LCA of concrete and steel building frames". *International Journal of Life Cycle Assessment* 3 (4), pp. 216-224.
- Martinez, E., Jimenez, E., Blanco, J., & Sanz, F. (2010). LCA sensitivity analysis of a multi-megawatt wind turbine. *Applied Energy*, 87, 2293-2303. doi: 10.1016/j.apenergy.2009.11.025
- Martinez, E., Sanz, F., Pellegrini, S., Jimenez, E., & Blanco, J. (2009). Life-cycle assessment of a 2-MW rated power wind turbine: CML method. *International Journal of Life Cycle Assessment*, 14, 52-63. doi: 10.1007/s11367-008-0033-9
- Martinez, E., Sanz, F., Pellegrini, S., Jimenez, E., & Blanco, J. (2009). Life cycle assessment of a multi-megawatt wind turbine. *Renewable Energy*, 34, 667-673. doi: 10.1016/j.renene.2008.05.020
- Monahan, J., Powell, J.C. (2011). "An embodied carbon and energy analysis of modern methods of construction in housing: A case study using a lifecycle assessment framework". *Energy and Buildings* 43 (1), pp. 179-188.
- Reza, B., Sadiq, R., Hewage, K. (2011). "Sustainability assessment of flooring systems in the city of Tehran: An AHP-based life cycle analysis". *Construction and Building Materials* 25 (4), pp. 2053-2066.
- UNE-EN 15804 (2012) "Sostenibilidad en la construcción. Declaraciones ambientales de producto. Reglas de categoría de productos básicas para productos de construcción"
- UNE-EN ISO 14040 (2006) "Gestión ambiental. Análisis de ciclo de vida. Principios y marco de referencia" (ISO 14040:2006).
- UNE-EN ISO 14044 (2006) "Gestión ambiental. Análisis de ciclo de vida. Requisitos y directrices" (ISO 14044:2006).
- Ximenes, F.A., Grant, T. (2013) "Quantifying the greenhouse benefits of the use of wood products in two popular house designs in Sydney, Australia". *International Journal of Life Cycle Assessment* 18 (4), pp. 891-908.



# MODEL ENGINEERING FOR CYBER COMPLEX ADAPTIVE SYSTEMS

Saurabh Mittal

Dunip Technologies LLC, Colorado, USA  
[smittal@duniptech.com](mailto:smittal@duniptech.com)

## ABSTRACT

Sociotechnical systems pervade every facet of our life today. These IT systems interact with live users that result in emergent behaviors leading to their classification as complex adaptive systems (CAS). Cyber-CAS (CyCAS) exists in contemporary society when such systems have Internet as their platform. Modeling and Simulation (M&S) for CyCAS is extremely difficult: partly due to inherent complex nature of enterprise IT architectures and partly due to lack of sufficient formal processes used in enterprise systems architecting. Fundamentally, the interaction between various elements is hard to pin down as enterprise architecture components are often treated as black-box with limited information about their internal behaviors. In this article, I provide an overview of the M&S approach for enterprise frameworks, introduce CyCAS, the tools and metrics for various CyCAS Views and the sandbox requirements for CyCAS M&S.

Keywords: Cyber Complex Adaptive Systems, CyCAS, DUNIP, second order cybernetics, DEVS

## 1. INTRODUCTION

Today the sociotechnical systems pervade every facet of life, such as economics, politics, medicine, agriculture, energy, and so on. The information is distributed across the entire landscape and there are a million eyes instantly reacting to a new technology or a blog post that turns viral any minute. Modeling such a system seems impossible, as the notion of system cannot be formally identified even though everybody talks of a ‘system’ in general terms. These Information Technology (IT) systems have different architectures for different purposes and as human interact with these systems, both the IT systems and the interacting users adapt to the dynamic landscape. To manage the design of an Enterprise IT system, each underlying architecture subscribes to an enterprise architecture framework that anchors various enterprise architectures towards a common goal. Modeling an enterprise framework for the formal modeling and simulation effort remains a vision and the problem of executable architecture has not yet been completely solved. Agent based modeling (ABM) is the strongest contender towards modeling and simulation of system that require many autonomous or semi-autonomous entities. However, ABM alone cannot describe these complex adaptive systems in the cyber-world.

In this article, I will define Cyber Complex Adaptive Systems (CyCAS) and how the foundation of formal modeling and simulation could be applied towards developing a framework to study CyCAS. I will elaborate on the various perspectives CyCAS can be described through and then present the tools and metrics required to support the development of these perspectives.

The article is organized as follows. Section 2 provides a background overview of foundational M&S theory, systems framework and interoperability considerations. Section 3 describes the M&S framework for tooling and various enterprise architecture frameworks that make the problem even more complex. A discussion of complexity and system-of-system (SoS) perspective is provided in Section 4. Section 5 introduces CyCAS with Section 6 proposing various CyCAS Views within the CyCAS enterprise framework. Section 7 elaborates on the sandbox requirements for CyCAS modeling. Finally, conclusions are presented.

## 2. M&S ARCHITECTURE, SYSTEMS FRAMEWORK AND INTEROPERABILITY

As we all know, model is an abstraction of reality with some valid assumptions and the simulator is an algorithm or a program that manifests the behavior of model. The discipline of designing a simulator is called Simulation engineering and considerable effort has been spent in the last few decades in developing robust simulation engines that leverage high performance computing (HPC). On the contrary, model engineering (ME) aims at setting up a systematic, standardized, and quantifiable engineering methodology to manage the data, knowledge, activities, processes and organizations/people involved in the full life cycle of the model, in order to obtain the general term of credible model theory, methods, technology, standards and tools with the minimum cost.

Model engineering has been in use since the mid-1980s through the Model-based Engineering (MBE) and Model-based Systems Engineering (MBSE) paradigms (Zeigler 1976, Zeigler 2000, Mittal and Martin 2013b). Progress in ME should profit from the progress made in model-based and model-driven methods developed in various systems and software engineering communities. At the same time, it should address the distinct challenges posed in the modeling and simulation (M&S) domain clarified by its theoretical and conceptual frameworks.

M&S engineering can be best understood by the layered M&S architecture initially proposed by Zeigler, Praehofer and Kim (2000), and later elaborated by Mittal et.al (2008) in the context of system-of-systems engineering, and Linguistic Levels of Interoperability (i.e., pragmatic, semantic and syntactic levels). Simulation is associated with the computational domain while Modeling is generally associated with the semantic domain (Figure 1a). They both come together at the pragmatic level where the purpose and limitations are defined. The development of an M&S solution is guided through the pragmatics that involve various considerations for M&S engineering, such as, scope, collaboration, decision, search-space, and computational requirements. Computer and Software Engineers largely handle the discipline of simulator engineering while the discipline of model engineering accounts for the subject matter expert (SMEs) versed in domain theory and the model developer. In some cases, either SMEs take the plunge to learn various programming languages to encode the model (and the associated model editing environment) that serve their purposes that go beyond the existing simulation tools or they prefer to stay within the confines of mathematics. Regardless of the approach, the scope of model is defined by the questions it is meant to answer or phenomenon it should describe. Certainly, the question of which model is accurate is a difficult question as it is dependent on the pragmatics (Figure 1b). When model accurately represents the pragmatics, it is called a valid model and the process is called Validation. When the model behavior is correctly implemented by the simulator, the model is verified. This process is called Verification. For any model engineering effort, the verification and validation (V&V) (Figure 1b) studies are as critical as the M&S engineering process and any M&S effort without V&V raises questions on the credibility of the model.

The simulators are computational entities built using various programming languages with a defined syntax. Communication and integration between different simulators has been pursued in the derived discipline of Distributed Simulation with an objective of achieving simulation interoperability. Various standards (DIS, HLA, TENA, CORBA, etc.) are in use that have facilitated syntactic interoperability. However, the model interoperability, as it happens in the semantic domain has remained elusive, primarily for the reason that model pragmatics are different for different models as assumptions are seldom aligned beforehand. When the pragmatics are aligned, model interoperability can be achieved as the composite model can now be validated.

While conceptually, this is easy to understand, the problem becomes compounded with complexity when the models and simulators are engineered with distinct architectures, some proprietary, some open-source, some considering model and the simulator as separate elements and some coupling them together, making them “silos” with no system-of-systems vision and extensibility. As they are used in the field, they are called “simulation systems” and become black-boxes thereby masking away the V&V process that need to be defined for the system-of-system (SoS) that they begin to participate in. As elaborated in Mittal et.al. (2008), the component-based nature of SoS needs a next level of specification provided by an architecture framework that allows specification and integration of architectures. These architectures within the enterprise framework can then be subjected to the formal Test and Evaluation (T&E) procedures thereby validating their designs (Figure 1c). Verification in SoS remains elusive and it is largely assumed that the component system accurately performs the claimed functionality.

Model engineering in an enterprise environment need to be investigated in more detail and the V&V process need to be defined as well.

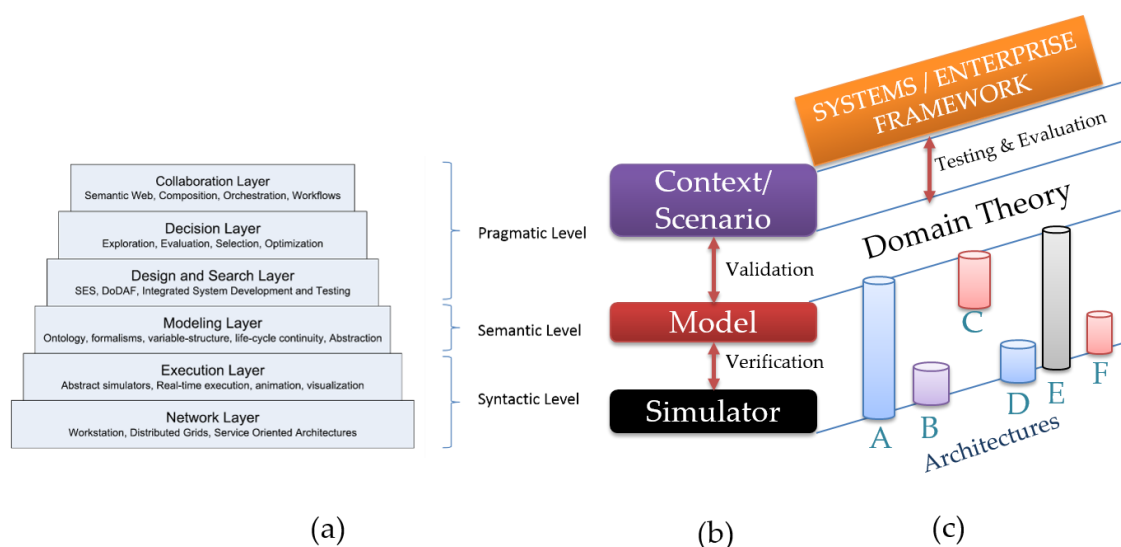


Figure 1: M&S Theory, Architectures, V&V, T&E and Interoperability

### 3. MSVC FRAMEWORK AND ENTERPRISE ARCHITECTURE FRAMEWORKS

Modeling editors and simulation engines are distinct pieces of an M&S solution and require different engineering approaches. However, when there are considered one and lack any formal engineering effort, they are programmed as one “simulation software” intending to serve only the purpose at hand or maybe just rapid prototyping.

When the model and simulator development is considered distinct, and when coupled with formal Software Engineering practices, they can be best be summarized by the Model-Simulator-View-Controller (MSVC) architecture framework (Mittal, Mak, Nutaro 2006) (Figure 2). As stated earlier, the pragmatics provide model’s objective, the semantics provide the model definition and the syntax provides the model’s computational implementation. In the Software Engineering realm, Model-View-Controller (MVC) architecture is the most widely used architecture to build software that has a decoupled Presentation layer. Many extensible Integrated Development Environment (IDE) frameworks are available that utilize MVC architecture

framework (for example Netbeans Rich Client Platform (RCP), Eclipse RCP, Enterprise J2EE, Microsoft .NET framework, etc.)

The Model in MVC is where the actual semantics are constructed, the View is where the presentation and user-input of model is created and the Control is where the communication between the Model and the View is engineered. In MSVC, there is an additional component of Simulator that represents the underlying platform over which the model is executed. Moreover, the Controller aspect is made rich and takes active role as it pertains to the pragmatics aspect of the M&S development exercise. The MSVC framework is discussed at length by Mittal, Mak and Nutaro (2006) where they associate it with DEVS M&S theory.

While usage of RCP platforms provide a solution for a desktop environment and the resulting solution can effectively be a considered a black-box, enterprise level solutions bring more complexity in the mix where cohesion and loose-coupling provide a different design paradigm.

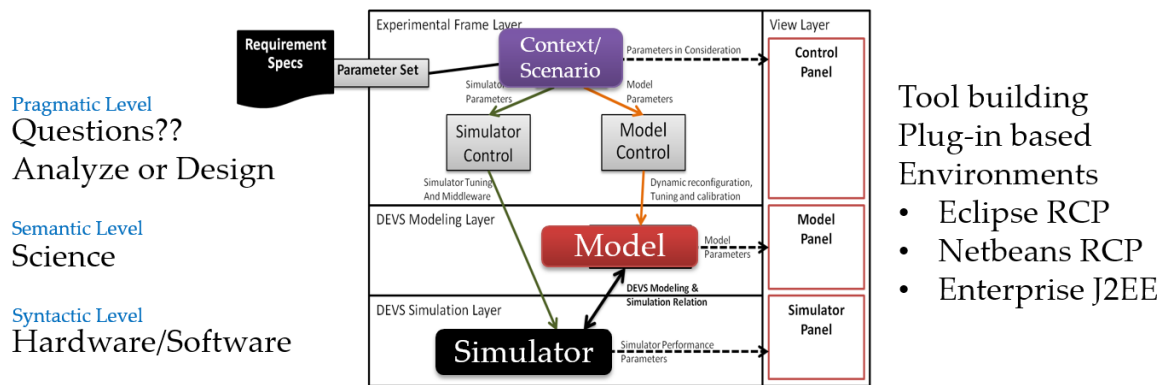


Figure 2: Model Engineering using MSVC Framework and Open Source tools

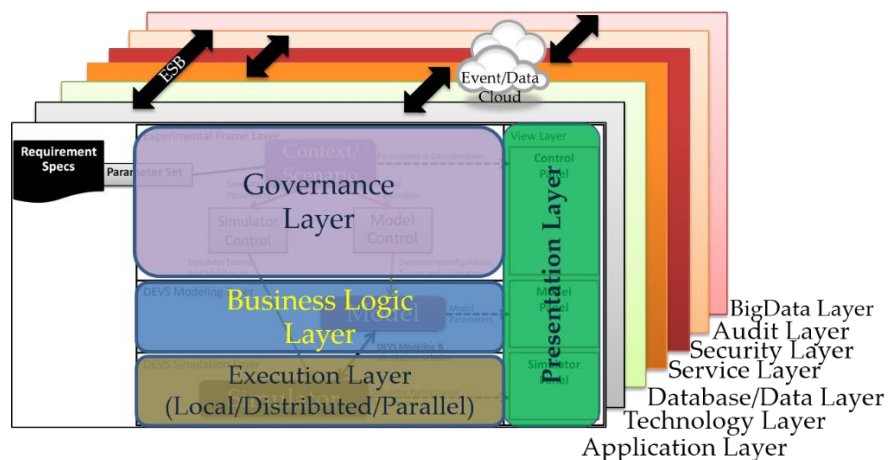


Figure 3: MSVC Framework and Enterprise Architectures

### 3.1. Architecture Frameworks

Enterprise frameworks are organized in various layers where each layer can have its own architecture. These layers are identified on the function they provide in serving the enterprise. Clearly, separation of concerns is at the heart of an enterprise framework, which in turn results in specific technologies being developed at each layer to enhance its functionality. Some of the identified layers are:

- Governance:
- Business Logic
- Execution
- Application
- Technology
- Database
- Service
- Security
- Audit
- Big Data

These layers, which have their own architectures, communicate using enterprise service buses (ESBs) that provide atomicity and various other message-based reliable communication services. This communication is largely in a netcentric domain where many Standards exist, facilitating solid integration. Event-driven Architectures (EDAs) are the latest in the list of enterprise architectures that provide a shared memory-store (an Event cloud) of user-specified queries and patterns that can access various other connected layers (Taylor, Yochem, and Phillips 2009).

Figure 3 shows the relation of enterprise framework with the MSVC pattern. The Controller is analogous to the Governance layer, the Model to the Business Logic layer, the Simulator to the Execution layer, and the View to the Presentation Layer. All the other layers are orthogonal to the MSVC layers. The MSVC is coupled together as an Application layer, as shown in Figure 3.

A framework enables the development of an architecture and it is imperative to have a framework to manage and utilize various architectures. The reader is encouraged to review frameworks like:

1. DoDAF (CIO 2010): Department of Defense Architecture Framework.
2. MoDAF (MoD 2012): Ministry of Defense Architecture Framework
3. TOGAF (TOG 2011): The Open Group Architecture Framework
4. Zachmann Framework (Zachmann 2008)
5. NAF (2013): NATO Architecture Framework
6. DNDAF (2010): Department of National Defense/ Canadian Armed Forces Architecture Framework
7. FEAF (2012): The Federal Enterprise Architecture Framework

While it is encouraging to know that these framework exists, the major problem that now arises is

that there are too many frameworks based on the domains they serve. The problem initially thought to be contained within the confines of a framework, seems compounded by the presence of multiple architectures and multiple frameworks. Again, no wonder why various agencies who like to share information and data are plagued with the inherent complexity. Various ontology and data-dictionary-based solutions are utilized that try to harmonize the knowledge between two or more different frameworks. Efforts are ongoing (Gorman, Haperen & Bailey 2013).

### 4. COMPLEXITY AND SOS

Section 2 provided an overview on the fundamental problem of V&V with multiple M&S architectures after becoming components within an enterprise framework. While there are many examples of enterprise frameworks as enumerated in Section 3.1, currently there exists none for enabling an M&S enterprise architecture. To that end, Joint Live, Virtual and Constructive (JLVC) Vision 2020 brings together Cloud-enabled Modeling and Simulation (CEMS) services (Weter 2012) and Joint Training Enterprise Architecture (JTEA), are under development (Edgren 2012) for enterprise level Joint Force training (Irwin 2012). These assume that the underlying LVC environment is indicative of enterprise M&S architecture. Technically, the integration problem is solved but semantic interoperability remains elusive with V&V from M&S perspective still questionable. Moreover, it is still not clear how causality, control and emergent behaviors need to be addressed in a distributed discrete event simulation environment (Rainey and Mittal 2014). Indeed, a cloud-based simulation framework continues to be a vision where fundamental issues of ownership of state, shared environment, truth and simulation remain unaddressed (Tolk and Mittal, 2014). There is a fundamental difference between software-based discrete event simulation and systems-based discrete event simulation. While the former is strictly based on object-oriented software engineering paradigm, the latter enforces mathematical System Theory on the object-oriented discrete event simulation engine. A successful solution must incorporate ideas from Systems Theory and M&S engineering practices. Even though all the above mentioned can be used for developing an M&S architecture, the Execution Layer (Figure 3) needs to be strengthened by distributed simulation architecture based on Systems Theory in a netcentric domain as described by Mittal and Martin (2013a, b). Until system-based approaches are encouraged, the vision of developing an executable architecture, where an architecture model can be executed over an enterprise simulation infrastructure, will continue to be a vision.

The problem of emergent behaviors when such architectures are brought together as a SoS has gained widespread attention in recent years. In addition to the five SoS characteristics defined by Maier (1998), where the constituent systems are operationally independent, managerially independent, have evolutionary nature

within their own systems development life cycle, are geographically distributed and together they portray emergent behavior, *an SoS also has a purpose*, to begin with. Assuming each constituent architecture is a 100% functional system, passing all the verification and validation tests in isolation, it cannot be determined a priori if the same system can perform with an SoS and can deliver: (a) the emergent behavior that is desired of an SoS and (b) fulfills the overall goal i.e. validation of SoS. Each constituent system or an architecture becomes an 'agent' that has a unique level of autonomy.

Finally, the biggest issue of all is the presence of humans as an actor in an enterprise solution thereby giving a dynamic nature to the entire landscape wherein everything is evolving and adapting (which includes, the technology, the users, the developers, the knowledge, the architectures and the systems themselves). This is how a complex adaptive system in an enterprise environment comes into being and there exist no mechanisms to perform T&E before such SoSs are made live with real users.

## **5. CYBER COMPLEX ADAPTIVE SYSTEM (CYCAS)**

Complex Adaptive Systems (CAS) are systems that display strong emergent behavior, have positive- and negative-feedback loops and have large number of adaptive agents. Examples of such systems include ant colonies, the biosphere, the brain, the immune system, the biological cell, businesses, communities, social systems, stock market, financial systems, High Frequency Trading systems, enterprise systems, etc. Some of these examples can also be classified as *open systems*. One of the characteristic features of these systems is the generation of new knowledge within the system that has causal properties resulting in adaptive behavior of the system components itself. CyCAS occurs in contemporary society where computational systems interact with both live human agents and virtual agents (software bots). The resulting complex system, for example, the World Wide Web, evolves in unknown ways not initially thought out by designers. Various types of communities emerge, that use the system in ways unthinkable by the system designers. Usage of Facebook to organize protests, Twitter to coordinate and You Tube to showcase demonstrations were critical in the Arab Spring revolution. On another note, usage of You Tube as a learning system makes everybody an educator. We frequently use do-it-yourself (DIY) videos on You Tube for our personal advancement and Massive Open Online Courses (MOOCs) give these videos a formal structure. Learning amidst such CAS happens in leaps and bounds and aptly so, the technology to create such systems advances in lock-step fashion. While we have technologies to engineer such netcentric complex systems today, we are short of methodologies to understand the emergence of adaptive stance in these engineered systems. They are difficult to study due to unavailability of a 'CAS' sandbox.

The engineering process is a formal process that is applied to solve a particular problem and deliver a system that addresses it. With CyCAS, even the initial objective takes a spiral upturn or downturn (based on one's perspective) and continuously evolves. In contemporary society, every Cyber-system must have a component of computational emergence that studies the life-cycle of such systems.

I identify eight characteristics of CyCAS as follows:

### **5.1. Human-in-the-System**

Every cyber system has a human as a critical component of the system. A CyCAS model accordingly must have infrastructure to manage user model profiles, the usage spread (distribution of users with respect to their roles), the capability to understand system responses for various usage spreads and the mechanisms to perform role switching for each user. Various human behavior modeling approaches and context quantification approaches (Mittal and Zeigler, 2014) need to be incorporated from fields like cybernetics, cognitive psychology, agent-based modeling etc.

### **5.2. Multi-agent-system**

A CyCAS is a multi-agent system (MAS) where the agents can be heterogeneous, homogeneous, deliberative or communicative in varying degrees and combinations. MAS (Weiss 2000) is a fairly new discipline of Distributed Artificial Intelligence (DAI) (Bond, Gasser 1988) that focuses on behavior management and design of agent-based systems. Canonically, such systems are multi-level algorithms subjected to multi-level control in a real-time environment.

### **5.3. Control and communications in a netcentric environment**

The communication aspect of CyCAS cannot be understated as it is the source of all emergent behavior. This Standards-based communication in a netcentric environment can be point-to-point, inter and intra-enterprise, and may include blackboard architectures with shared memory infrastructure. The presence of communication links at various levels establishes positive and negative feedback loops and unless such aspect is dealt per Systems Theory-based "closure under composition" principle, it is hard to establish who is in control of state and time for the resulting non-linear system.

### **5.4. Resource-constraints and economy of scale**

Each constituent system operate within the confines of real world and has finite resources that need to be used efficiently and in a priority-based manner as dynamic nature of pragmatics within CyCAS re-purpose existing resources. Scalability and resourcefulness need to be addressed in a formal manner as failure of a constituent system may result in cascaded failures. Constraint-graphs facilitate the evaluation of resource availability and limitations.



### 5.5. Emergent Attention and Second order cybernetics

As the CyCAS design is motivated by the pragmatics (overall SoS goal), various observers (models) and auditing systems need to exist within the CyCAS to ensure that the resultant emergent behavior is truly the intended one. Usage of BigData technologies to engineer advanced pre- and post-analytics system must conform to Second-order cybernetics philosophy (Heylighen and Joslyn, 2001) where both the observer and the observed define the actual system. The emergent attention (Mittal and Zeigler, 2014) becomes a critical aspect of such systems that indicates the most active constituent system (or a group of sub-systems) in a resource-limited and time-constrained environment.

### 5.6. Phase Transition

As SoS manifests emergent behaviors, the constituent systems may respond in unknown ways due to their black-box nature. This may lead to emergence of unknown behaviors not specified in the design. Sometime these behavior are benign and sometimes puts the entire SoS into an undesired trajectory. Consequently, a CyCAS model must be able to attend to these phase transitions through computational emergence studies that provides information about the adaptive stance of CyCAS and whether it continues to display the intended emergent behavior.

### 5.7. Structure of knowledge

Each of the constituent system being a black-box implements its own knowledge-base. While they can be syntactically integrated successfully, achieving semantic interoperability is the holy-grail of systems interoperability and CyCAS. Before semantic interoperability is achieved, knowledge and ontological structures have to be ensured and validated by the SMEs. Issues like knowledge sharing, transformation, evolution, computational representation (ontology), interoperability, context switching according to emergent attention, and new knowledge synthesis have to be addressed at multiple operational levels. Knowledge management and evolution roadmap is a critical component of CyCAS and demarcates the difference between a *weak* or a *strong* emergent system (Figure 4) (Mittal, 2013; Mittal and Zeigler, 2014).

### 5.8. Resilient or Anti-fragile

Finally, CyCAS undergo a lot of dynamic reconfiguration due to their component nature. It needs to be formally determined if the engineered CyCAS is fail-fast, resilient/robust or anti-fragile where it continues to function, albeit evolving, in unknown ways.

## 6. CYCAS FRAMEWORK

Based on the eight characteristics identified in the previous section, this section presents a framework to engineer a CyCAS. Following are the assumptions:

1. CyCAS is a digital complex adaptive system where the notion of Object is ubiquitous at every level of specification, and is rapidly being

replaced by the notion of a ‘smart’ object that may have sense-making capabilities.

2. System behavior is indicative of events that are manifested externally by a constituent system and are communicated using structured messages in a netcentric environment
3. Human in the most complex and the least predictive element. However, it is the most critical element that transforms a netcentric system into a CyCAS.

CyCAS Framework consists of eight views corresponding to each of the characteristics. Table 1 lists the eight views, the capabilities of tools that implement these views and the metrics that ensure that the tools are valid.

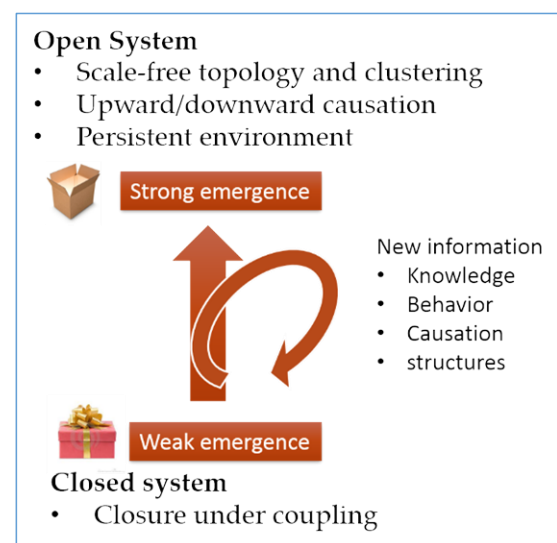


Figure 4: Strong and Weak Emergence

## 7. SANDBOX REQUIREMENTS AND THE ENGINEERING PROCESS

As we strive to adhere to Systems Theory for a CyCAS sandbox, we must be able to classify the eight CyCAS views into structure and behavior. This will allow us to separate the static and dynamic nature of the views. However, one has to remember the fact that the CyCAS structure continuously changes as the CyCAS behavior evolves. In a CyCAS, variable structure modeling is a critical capability (Mittal, 2013). Following is the classification:

### I. Structure

- a. Control and Communication View
- b. Resource and Constraints View
- c. Knowledge View
- d. Resilience View

### II. Behavior

- a. Human View
- b. Multi-agent View
- c. Emergence View
- d. Phase Transition View



Table 1: CyCAS Views, tool capabilities and various metrics for model engineering

ID	CyCAS Views	Tool Capabilities	Metrics
5.1	<b>Human View</b>	Cognitive architectures, Live, Virtual and Constructive (LVC) environments, user behavior modeling	Behavior quality, spread and quantification, cognitive plausibility, contextual realism, quantized context
5.2	<b>Multi-agent View</b>	Agent structure, behavior and interactions with other agents or environment, closed-under-composition	Ease of model-transformation and model integration, partial observability, group cohesion, shared goals
5.3	<b>Control and Communication View</b>	Hierarchical organization, logical structures and supervisory control (similar to DoDAF OV-7 and SV-7)	Degree of control (from centralized and totalitarian to completely decentralized and autonomous), feedback loops
5.4	<b>Resource and Constraints View</b>	Similar to DoDAF 2.0 System View 4, 5 resources. Constraints similar to policy considerations in DoDAF OV-6a	Utilization, availability, limitations, affordance
5.5	<b>Emergence View</b>	Multi-level instrumentation, Big Data, expected behaviors, causal behaviors, novel behaviors	Multi-level behavior validation and recognition, computational emergence, agent adaptation
5.6	<b>Phase Transition View</b>	System behavior transition matrix (similar to DoDAF SV-3)	Multi-level transition probabilities, credit assignment and new behavior detection and encoding
5.7	<b>Knowledge View</b>	Ontologies (data and its relationships)	Semantic network, semantic validity through SME and keyword-rank
5.8	<b>Resilience View</b>	Experimental frames	Degree of robustness at multiple-levels

While providing further details on each of the views is outside the scope of the article and is left for future work, I would like to describe the Multi-agent View in a bit more detail. As a CyCAS is a MAS at the fundamental level, there are many agent-based modeling (ABM) tools available. For example, JAMES II, Repast, NetLogo, Mason, Flame, Soar, Swarm, etc. To the best of my knowledge, none conforms to the Systems Theory's closure under composition principle and infact, many leverage the flat nature of agent based systems with limited encapsulation and hierarchy. The problem is rooted in usage of only the Software Engineering practices and ignoring the Systems Theory.

In the classical paradigm, and not going too far back for the purpose of this discussion, let us look at the 4<sup>th</sup> generation languages such as C. There is the *main* function and there are subroutines that invoke each other. Then came the Object-Oriented paradigm which brought forward concepts like inheritance, data encapsulation, and polymorphism, and grouped the set of subroutines within an Object as its 'behavior'. Capitalizing on this, the ABM community made the Object an Agent by adding constructs like purpose, behaviors and capabilities (i.e. a smart object). They also provided an agent execution environment where these agents could manifest their behavior with a virtual environment. All the existing ABM tools with the exception of JAMES II, have this notion of software agent with varying implementations and integration with a real or synthetic environment. Sometimes, there is a scientific theory

behind the ABM tool, for example SOAR has cognitive psychology as its foundation (Laird, 2012). On the other hand, a Systems-based agent, for example, a DEVS agent, has the notion of a system attached to it (Figure 5). It is built on formal semantics and adheres to the Systems Theory (Zeigler, Praehofer and Kim, 2000, Mittal, 2013). It also provides a simulator, a simulation protocol and a distributed high performance engine for agent/system model's execution and ensures that Systems Theory is not violated. The DEVS M&S theory (Zeigler, Praehofer and Kim, 2000) is applicable to both discrete and continuous hybrid systems and has the notion of abstract time. It models the dynamical behavior and has the notion of elapsed time. Work by Mittal and Martin (2013a,b) packages all these functionalities in a netcentric DEVS Virtual Machine (VM) that provides an agent-execution environment in a CyCAS setting.

To develop a CyCAS sandbox, Table 2 provides an overview of the DEVS M&S elements as applicable to implementation of the structural and behavioral components of CyCAS. The DEVS Unified Process and the related work by Mittal, Martin and Zeigler in the last decade (Zeigler 2013) bring technologies like DEVSMML Stack, DEVSVSM, metamodeling, domain specific languages (DSLs) to be integrated with the DEVS M&S environment based on dynamical Systems Theory towards the pursuit of multi-paradigm modeling, a prime requirement for integrating multiple modeling architectures. The CyCAS sandbox based on such ideas will be reported in future work.

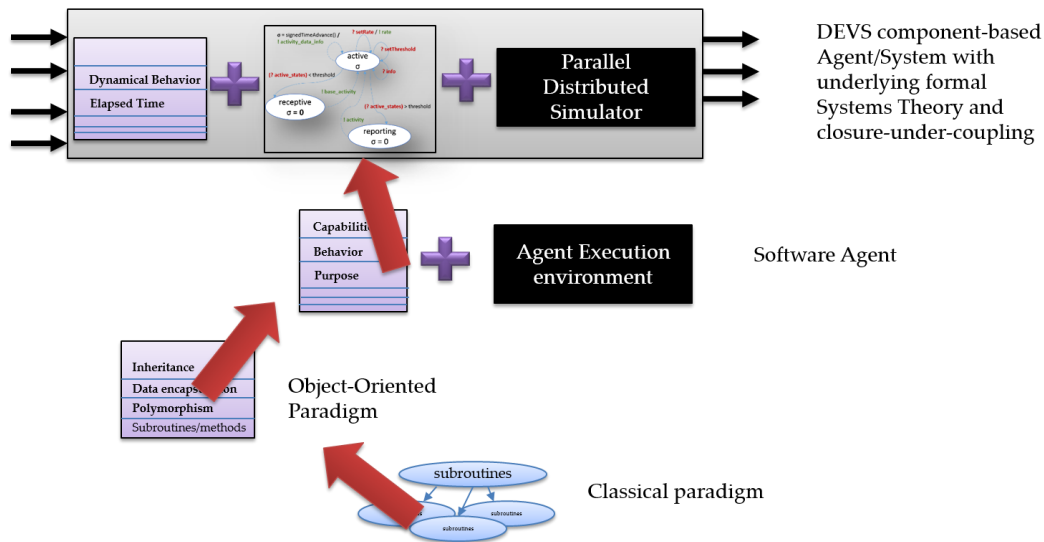


Figure 5: From a Software Agent to a DEVS Agent

Table 2: CyCAS and Model Engineering

CyCAS View	DEVS Element	Alternatives
Human	Dynamical agent behavior modeling in Atomic DEVS subject to SME validation	Cognitive agents, software agents, DEVS behavior models, SOAR agents
Multi-agent	Atomic, Coupled and Multi-Level DEVS	Object interaction diagrams
Control and Communication	Levels of Systems specifications	Subsumption architectures
Resource and Constraints	Atomic DEVS and network-delay models	External systems with defined I/O interfaces, constraint graphs
Emergence	Atomic observers at higher levels of hierarchy, Experimental Frames	Pattern matchers
Phase Transition	Atomic observers at multiple levels of system specifications, Experimental Frames	Pattern matchers
Knowledge	Representation through pragmatic frames within the System Entity Structure (SES) theory	Triple stores, RDF stores, etc.
Resilience	Experimental Frame, Atomic DEVS	-

## 8. CONCLUSIONS

Many enterprise architecture frameworks provide various views and perspectives to facilitate architecture development and capture the true purpose behind the enterprise architecture. Enterprise IT Systems underwent a lot of transformation in the last decade and robust technologies exist that deliver scalable enterprise solutions. Multiple enterprise architecture frameworks exist that serve different niche and efforts are underway to consolidate them towards a unified framework and none mandate the formal M&S engineering methodology to develop an executable architecture. Enterprise systems modeling is a different problem altogether and new approaches are needed to model such complex multi-faceted systems. Cyber complex adaptive systems are specialized cases of CAS in the cyber domain where both the systems and the system users adapt in unknown ways. Consequently, it is impossible to conduct studies beforehand as both sides of the equation are in constant flux. In this article, I introduced the CyCAS concept, essential elements of CyCAS and various perspectives that are

needed to understand it. The CyCAS perspectives can be used in conjunction with the existing architecture frameworks and as of this article, they are by no means the only perspectives that are needed to understand the holistic behavior of CyCAS. I also introduced the sandbox requirements for modeling and simulation of CyCAS and related it to the fundamental Systems Theory. I also stressed the importance of using frameworks like DEVS Unified Process that is based on formal discrete event systems (DEVS) theory to be applied to CyCAS M&S. Further details will be presented in the future work.

## REFERENCES

- Bond, A., L. Gasser, 1988. A Survey of Distributed Artificial Intelligence., Available at: <http://www.exso.com/nsurvo.pdf>, Last accessed: July 31, 2014
- CIO, DoD, The DoDAF Architecture Version 2.02, Available at: <http://dodcio.defense.gov/TodayinCIO/DoDArchit>

- [ectureFramework.aspx](#), Last accessed, July 31, 2014
- DNDAF, 2010. DND/CF Architecture Framework (DNDAF) Version 1.7 Home Page. Available at: <http://trak-community.org/index.php/wiki/DNDAF>, Last accessed: July 31, 2014
- Edgren, M. 2012. Joint M&S Strategy, WJTSC M&S Working Group, Defense Technical Information Center. Available at: [http://www.dtic.mil/doctrine/training/conferences/wjtsc12\\_2/wjtsc12\\_2\\_jmsc\\_msstrategy.pdf](http://www.dtic.mil/doctrine/training/conferences/wjtsc12_2/wjtsc12_2_jmsc_msstrategy.pdf), Last accessed July 22, 2014
- FEAF, 2012. The Common Approach to Federal Enterprise Architecture. Available at: [http://www.whitehouse.gov/sites/default/files/omb/assets/egov\\_docs/common\\_approach\\_to\\_federal\\_ea.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/common_approach_to_federal_ea.pdf), Last accessed: July 31, 2014
- Gorman, P., K. Haperen, I. Bailey, 2013. An Update on the Convergence of MOD and NATO Architecture Frameworks, Available at: <http://nafdocs.org/wp-content/uploads/2013/07/Gorman-Bailey-Van-Haperen-IEA14-v2.pdf>, Last accessed: July 31, 2014
- Heylighen, F., and C. Joslyn, 2001. Cybernetics and Second Order Cybernetics, in: R. A. Meyers (ed.), *Encyclopedia of Physical Science & Technology*, 4, 155-170, Academic Press, New York
- Irwin, T., 2012, Enabling Training Technologies for Joint Force 2020, Available at: [http://www.ndia.org/Resources/OnlineProceedings/Documents/21M0/Joint\\_Forces\\_2020/Intro\\_and\\_Overview\\_Thomas\\_Irwin.pdf](http://www.ndia.org/Resources/OnlineProceedings/Documents/21M0/Joint_Forces_2020/Intro_and_Overview_Thomas_Irwin.pdf), Last accessed, July 22, 2014
- Laird, J. 2012, *The SOAR Cognitive Architecture*, MIT Press, Cambridge, MA
- Maier, M.W. 1998. "Architecting Principles for Systems-of-Systems." *Systems Engineering* 1 (4): 267-284.
- Mittal, S., E. Mak, and J.J. Nutaro. 2006. "DEVS-Based Dynamic Model Reconfiguration and Simulation Control in the Enhanced doDAF Design Process." *Journal of Defense Modeling and Simulation* 3 (4).
- Mittal, S., B.P. Zeigler, J.L.R. Martin, F. Sahin, and M. Jamshidi. 2008. "Modeling and Simulation for System of Systems Engineering." In *System of Systems Engineering for 21st Century*, by M. (ed.) Jamshidi. Wiley.
- Mittal, S., 2013. Emergence in Stigmergic and Complex Adaptive Systems, *Journal of Cognitive Systems Research*, 21, 26-39.
- Mittal, S., and J. L. R. Martin. 2013a. *Netcentric System of Systems with DEVS Unified Process*. Boca Raton, FL: CRC Press.
- Mittal, S., and J. L. R. Martin. 2013b. Model-driven Systems Engineering in a Netcentric Environment with DEVS Unified Process, Winter Simulation Conference
- Mittal, S., and B. P. Zeigler, 2014. Context and Attention in Activity-based Intelligent Systems", ITM Web of Science, 2.
- MoD, UK. 2012. MOD Architecture Framework. Available at: <https://www.gov.uk/mod-architecture-framework>, Last accessed: July 31, 2014
- NAF, 2013. NATO Architecture Framework: MODEM Review. Available at: [http://nafdocs.org/wp-content/uploads/2013/07/20130816\\_MODEM\\_to\\_NAF\\_Review\\_V1\\_1-U.pdf](http://nafdocs.org/wp-content/uploads/2013/07/20130816_MODEM_to_NAF_Review_V1_1-U.pdf), Last accessed: July 31, 2014
- Rainey, L., S. Mittal, eds., 2014. Agent-based modeling for understanding complexity: Lessons learned from defense domain where lack of adherence to Systems theory lead to unpredictable behavior, Special Issue Journal of Defense Modeling and Simulation, Call for Papers. SCS.
- Taylor, H., A., Yochem, L. Phillips, F. Martinez, 2009. *Event-Driven Architecture: How SOA enables the real-time enterprise*, Boston, MA., Addison-Wesley
- TOG 2012. An Introduction to ToGAF 9.1, The Open Group. Available at: <https://www2.opengroup.org/ogsys/catalog/w118>, Last accessed: July 31, 2014
- Tolk, A., and S. Mittal, 2014 "A Necessary Paradigm Change to Enable Composable Cloud-based M&S Services", Winter Simulation Conference, Savannah, GA, USA
- Weiss, G., (ed.) 2000. *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, MIT Press, Cambridge, MA
- Weter, D., 2012, Joint Staff J7 Modeling & Simulation Program Update, available at: [http://www.rolands.com/jtls/j\\_iuc2012/JSJ7\\_Weter.pdf](http://www.rolands.com/jtls/j_iuc2012/JSJ7_Weter.pdf), last accessed: July 22, 2014
- Zachmann, J.A., 2008. The Zachmann Framework Evolution, Available at: <https://www.zachman.com/ea-articles-reference/54-the-zachman-framework-evolution>, Last accessed: July 31, 2014
- Zeigler, B.P. 1976. *Theory of modeling and simulation*. Wiley Interscience.
- Zeigler, B.P., H. Praehofer, and T.G. Kim. 2000. *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*. New York, NY: Academic Press.
- Zeigler, B. P. 2013. "Grand Challenges in Modeling and Simulation, What can DEVS Theory do to meet them? Parts 1 and 2" *Seminar to School of Automation Science and Electrical Engineering*, Beihang University, Beijing, China.

## BIOGRAPHY

SAURABH MITTAL is the founder and president of Dunip Technologies LLC, CO USA. He earned his MS (2003) and PhD (2007) in Electrical and Computer Engineering from the University of Arizona, Tucson. His current interests include modeling and simulation of complex natural systems, distributed artificial intelligence, multi-agent systems and complex adaptive systems. He can be reached at [smittal@duniptech.com](mailto:smittal@duniptech.com)

# SERIOUS GAMES AT INCREASED IMPACT ON CULTURE AND TOURISM

Francesco Longo<sup>(a)</sup>, Letizia Nicoletti<sup>(b)</sup>, Stefano Vena<sup>(c)</sup>, Antonio Padovano<sup>(d)</sup>

<sup>(a) (c) (d)</sup>DIMEG, University of Calabria, Italy

<sup>(b)</sup>CAL-TEK Srl, Italy

<sup>(a)</sup>f.longo@unical.it, <sup>(b)</sup>l.nicoletti@cal-tek.eu, <sup>(c)</sup>stefano.vena@unical.it, <sup>(d)</sup>a.padovano@msc-les.org

## ABSTRACT

This paper devoted to present some preliminary activities and results of an ongoing research project, SG-ICT, carried out by the MSC-LES laboratory of the University of Calabria. The SG-ICT main goal is to develop two groundbreaking applications for mobile devices: INSIDE VIRTUAL and OUSIDE VIRTUAL. The former is a mobile app, based on Serious Games and Simulation, designed to allow its users to enjoy digital contents through interactive and virtual experiences. The latter is another mobile app, based on augmented reality and on an intelligent personal assistant, aimed at creating new patterns of interaction with real contents and historical findings during a real tour in a museum or in an archaeological site. After an extensive analysis of the state of the art, the INSIDE VIRTUAL and OUSIDE REAL apps are discussed and the SG-ICT architecture is introduced. Lastly some preliminary results are shown.

Keywords: serious games, simulation, intelligent agents, cultural heritage

## 1. INTRODUCTION

Some preliminary activities and results of an ongoing research project called SG-ICT (Serious Games at Increased Impact on Culture and Tourism) are introduced. The SG-ICT project aims at providing groundbreaking technological solutions in the field of cultural heritage. In particular, the main project goal is to promote the cultural heritage in two different but complementary ways: allow people to enjoy digital contents through interactive and virtual experiences and set new patterns of interaction with real contents and historical findings during a real tour in a museum or in an archaeological site thanks to augmented reality and to an intelligent personal assistant and knowledge navigator.

In this perspective, the proposed solution integrates emerging technologies and innovative paradigms in order to enhance the cultural heritage enjoyment. As a result, on one hand SG-ICT creates new learning processes by exploiting the informative, educational and training potentials of Serious Games (SG) while on the other hand it creates new interaction patterns that can

make the experience in archaeological sites or museums more exciting and fascinating.

Therefore, the SG-ICT project includes two applications that are meant to work synergistically: the former is called INSIDE VIRTUAL and is based on SG while the latter is called OUTSIDE-VIRTUAL and is based on augmented reality and on an intelligent personal assistant. In other words, INSIDE VIRTUAL grounds on users' entertainment and engagement to increase their desire to visit firsthand the hystorical/cultural sites and the artifacts they encounter during the game; OUTSIDE-VIRTUAL supports the user in gathering the information he/she is interested in during an on-site visit.

As a result, the joint use of both applications seeks to enhance cultural content enjoyment using new emerging ICT technologies. Indeed the INSIDE-VIRTUAL and OUTSIDE-REAL apps will be developed for mobile devices in order to allow users to gain the maximum freedom in using (wherever and whenever) thereof.

## 2. STATE OF THE ART: OVERVIEW

The SG-ICT project represents a "trait d'union" between past, present and future since it seeks to use emerging technologies to preserve and enhance the historical and cultural heritage as well as to create new interaction patterns and learning processes.

From a technological and methodological point of view, the reference paradigms that SG-ICT relies on include Modeling & Simulation (M&S), Serious Games (SG), Intelligent Agents (IA), Computer Generated Actions (CGA), Virtual Environments (VE), and Augmented Reality (AR) in order to create two different Mobile Applications (MA).

M&S has proved to be a powerful methodology to support both decision making and operational processes in various application domains including manufacturing systems, logistics, health and defense. In these areas, several successful applications have been proposed highlighting the potentials of M&S for problem solving in a variety of complex systems (Longo, 2013; Longo et al., 2012). In addition, over the past few years, M&S has been used extensively as a very effective and useful training tool for professionals at different levels,

ranging from managers to operators (Bruzzone and Longo, 2013).

The combination of simulation and gaming technologies for purposes beyond the pure entertainment has brought forth the concept of serious games that are being increasingly popular and increasingly implemented in training and learning activities pertaining to many application domains, from defense to industry (Zyda, 2005).

To date, digital technologies applied to cultural heritage have been mainly used for reconstructing the original appearance of artifacts as well as for information storage. M&S can integrate digital technologies to create "artificial stories" (however faithful to the reality of a given historical period or a particular culture) to ensure greater involvement of users while they enjoy cultural objects.

The analysis of the state-of-the-art allows pointing out that, most of the time, researchers and research communities are mainly concerned about cultural heritage preservation through modeling techniques, graphics and virtual reality (Addison, 2000). As a matter of fact Anderson et al. (2010) show that the use of M&S and Serious Games for teaching/learning activities and for improving the visits in museums is still poorly considered.

Anyway, when dealing with the cultural and historical heritage, virtual museums and information kiosks (with the aim to make the cultural contents appealing) cannot be neglected.

In this regard, a state-of-the-art proposed by Sylaiou et al. (2009) shows that museums have different at their disposal many tools and methodologies, mostly based on virtual and augmented reality, to make the experience of their visitors more engaging.

As for virtual reality, solutions range from simple interactive screens to full-immersive systems such as projection systems and 3D stereoscopic glasses; whereas as for augmented reality applications, one of the first solutions is dated back to 1995 and was proposed by Bederson. This solution is based on the imposition of audio content based on the user's location. Anyway, since 1995, many solutions have been presented (see for instance Mase et al. (1996) and White et al. (2004)) and even solutions based on the coexistence of real and virtual objects (Mixed Augmented Reality), see Hall et al. (2001) and Hughes et al. (2004).

It is possible to note that Serious Games applied to cultural heritage are gaining increasing attention. According to the taxonomy adopted by Mortara et al. (2013), these SG can have various purposes including cultural awareness, historical reconstruction and awareness of the cultural heritage and can vary according to learning objectives, genre (e.g. adventure, simulation) and application context. To this end some meaningful examples can be found in Froschauer et al. (2010).

The combination of M&S and SG becomes even more innovative when thinking of using Intelligent

Agents that allow the simulated scenario to change according to a sequence of events (actions and reactions) generated by the reciprocal and continuous interaction between real users and Intelligent Agents. To this end, a meaningful example in the field of healthcare can be found in Bruzzone et al. (2012)

Indeed, agents are a powerful metaphor for the modeling and analysis through simulation of complex systems in various application domains (Bruzzone, 2013; Michel et al., 2009) and for the implementation of complex software systems (Jennings, 2000). According to this paradigm, a system is modeled as consisting 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).

These intelligent agents are the basis for implementing the SOPHOS system (a SIRI-like system) that acts as a personal assistant and knowledge navigator. In the literature, there are some examples of agent-based environments devoted to enhance the cultural experience in museums. I.e. Costantini et al. (2008) used the agent technology to create a significant experience during a visit. Specifically, the system offers suggestions in line with the profile and interests of visitors by monitoring (via satellite) the movements of visitors within an archaeological park and providing such information to agents.

As far as the use of Augmented Reality is concerned, several articles focused on AR applied to cultural heritage are reported in literature (some of them newer than others) to name a few: Sylaiou et al. (2009), Liarokapis et al. (2008), Brogni et al. (1999). It is evident that the use of AR requires an effort of software modeling, algorithms development and integration driven by the need to perceive and recognize the external environment and attach it to the appropriate source of "augmented" information.

Therefore, the literary background of the Digital Cultural Heritage shows a great variety and heterogeneity in the methodologies and technologies used so far.

At the international level, many projects have been proposed and developed in the field of cultural heritage. Anderson et al. (2010) refers to a set of projects on the visualization and virtual reconstruction of ancient historical sites, including Nova Roma (Frischer, 2008), Ancient Pompeii (Maim et al., 2007) and the Parthenon (Debevec, 2005). Nova Roma is a multi-year project on the 3D reconstruction of Ancient Rome in 320 AD; the model has been integrated with a game engine in order to be interactive and include non-player avatars. Ancient Pompeii is a project based on procedural modeling for the reconstruction of ancient Pompeii; within the reconstruction, entities able to show a realistic behavior of the ancient Romans that are simulated in real-time. In both cases margins of improvement may be highlighted: the system should ensure a greater interactivity, 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 sees the synergy between different technologies (multi-agent architecture, audio-visual technologies, sensors for 3D visualization, etc..) in order to enhance the experience of enjoyment of cultural heritage 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 at defining a knowledge base, a common language and an ontology dedicated to the creation of virtual museums as well as to identify innovative visualization and interaction tools for the virtual museums of the future. The 3D-COFORM project deals with all the aspects of the 3D digitalization, 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 aims at digitalizing 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 of cultural goods which are as much as possible in line with the interests of the visitors.

### 3. THE SG-ICT APPS

The SG-ICT project (*Serious Games at increased Impact on Culture and Tourism*) combines in an innovative way Modeling & Simulation (M&S), Intelligent Agents (IA), Computer Generated Actions (CGA), Virtual Environments (VE) and Augmented Reality (AR) for:

- creating an interactive and immersive way to produce, use and distribute cultural contents in museums and archaeological sites (to this end, M&S, IA, CGA, VE and MA are used to develop the INSIDE-VIRTUAL app);
- creating an easy-to-use and effective way to receive cultural contents while visiting real museums and archaeological sites (to this end, AR and MA will be used to develop the OUTSIDE-REAL app that is integrated with a "SOPHOS" intelligent personal assistant and knowledge navigator).

INSIDE-VIRTUAL and OUTSIDE-REAL are two sides of the same coin and, once they will be ready as a product or service, they can be easily sold on the market in two different ways:

- a low-cost mobile application for INSIDE-VIRTUAL (that may run on tablets, smart-phones,

AR glasses, head-mounted display, etc.) that can be bought by all those people interested in enriching their cultural knowledge about museums and archeological sites through an immersive, interactive and intelligent experience in 3D virtual environments;

- a low-cost mobile application for OUTSIDE-REAL (that runs on tablets, smart-phones, AR glasses, head-mounted display) that can be sold by museums and archeological sites to their visitors that are interested in trying a unique augmented reality experience and an intelligent personal assistant and knowledge navigator during their visit.

SG-ICT combines multiple methodologies and cutting-edge technologies to guarantee:

- an interactive and immersive cultural experience (the immersion is provided by the 3D VE while the interaction is provided by the IA and CGA);
- the production and intelligent automation of cultural contents (by using both IA and the SOPHOS assistant);
- the digital distribution, mobility and interoperability of cultural contents (by developing multi-players and web based Mobile Applications).

#### 3.1. The SG-ICT Inside-Virtual App

INSIDE-VIRTUAL is a multi-player, avatar-based, web-based mobile app in which the player follows a learning process as a virtual visitor in a museum or in an archaeological site. Within this app, the joint use of Intelligent Agents and M&S increases the levels of interactivity and immersion offering a unique learning and entertainment experience.

By using INSIDE-VIRTUAL, the user can easily understand cultural contents; specific avatars (driven by Intelligent Agents) provide the user with the most important information related to cultural heritage through questions and answers, videos and so on. For example, consider the case of an encounter with a virtual avatar that explains the history and significance of an archaeological find such as the Riace Bronzes. The avatar will illustrate the history of the discovery, the assumptions on dating, the provenance, who are the sculptors and the mythology. The avatar could also present the user with more options in terms of information to display or may suggest getting in touch with other avatars (e.g. the Head of the Magna Grecia area).

#### 3.2. The SG-ICT Outside-Real App

OUTSIDE-REAL is a mobile app that could be provided by museums or archaeological sites to visitors throughout their tour. The visitor can use the application installed on his/her mobile device and receive, through the use of AR, "augmented information". The application is also be equipped with a software



(SOPHOS) that recreates the features of a personal assistant and knowledge navigator. The SOPHOS software provides the user with appropriate information and, at the same time, collects information about the users' preferences that may be used by the museum or the archaeological site to improve the arrangement of its cultural contents.

#### 4. The SG-ICT Apps Architecture

Basic components of the SG-ICT Inside-Virtual App are hereunder described:

- *SG-ICT INSIDE-VIRTUAL Game GUI* represents the graphics interface that allows the user to interact with the application. This interface includes suitable menus, a control panel and is also able to operate both in stand-alone mode or remotely (connected on the web);
- *SG-ICT M&S Game Engine* is based on simulation and drives the evolution of the game according to the player's actions and decisions;
- *Virtual Agents* are avatars driven by Intelligent Agents and interacting with players through the *Game Engine*. Rules and logic behind the IA are implemented in a dedicated module called *IA-Inside*;
- *Virtual Environment* recreates a virtual museum or an archaeological park and gives the player the feeling of being inside them. The *Virtual Environment* is populated by *3D Geometric Models* representing both the environment and the cultural heritage (e.g. both the building housing the museum and all the findings contained within the museum itself);
- *SG-ICT Common Database* is a database where the information and data required for running the apps as well as the data that are generated while using the app are stored;
- *Museum Informative System* is a possible connection between the INSIDE-VIRTUAL application and the information system of the museum. This connection is carried out to ensure the app development and its maintenance;
- *SG-ICT OUTSIDE-REAL Connection* represents a connection gateway with the OUTSIDE-REAL app. In this sense, some of the functionalities implemented within INSIDE-VIRTUAL could be used to provide additional functionalities to the users of the OUTSIDE REAL app. While using OUTSIDE REAL during a visit to a museum the user can obtain, by means of the augmented reality and of the personal navigator, more information about a specific finding. The user may want to display a 3D model of the specimen (in order to be able to rotate or zoom in and then have a better view). The 3D models are part of the app INSIDE-VIRTUAL; so the user will have access to the sought information (3D models) through the connection gateway.

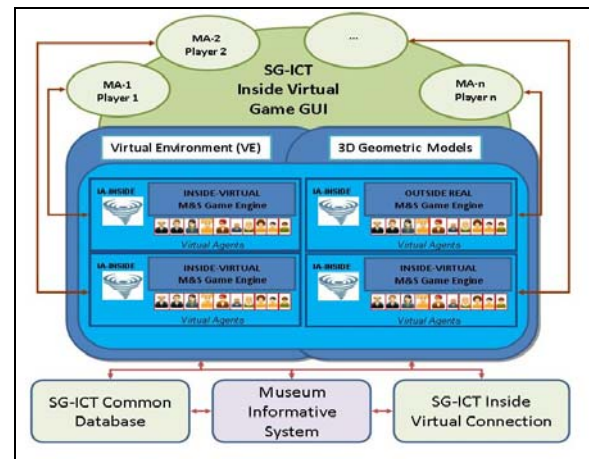


Figure 1 - INSIDE-VIRTUAL General Architecture

In order to provide a more detailed description of the INSIDE-VIRTUAL app architecture, consider the point of view of a player using the app. The user will see a three-dimensional virtual environment of a museum or an archaeological site on the screen of his/her tablet or smartphone.

The user can interact with the app through the *Game GUI*, request information and perform specific actions (e.g. establish a relationship with one of the Intelligent Agents and inquire about a specific exhibit inside the museum). Depending on the actions taken by the user, various "Course of Actions" are generated and, consequently, the user is led towards the knowledge and the enjoyment of the cultural heritage of the museum.

Intelligent Agents (which are the domain experts the user interacts with) may provide multiple options/answers to enrich the cultural experience of the user. By this approach, the game evolves dynamically according to the decisions/actions taken by the user (each user will then live an individual experience of enjoyment of the cultural heritage).

In order to make the cultural experience even more interactive and engaging, different levels of difficulty and scenarios are implemented to allow a vehicled access to a richer information content.

Basic components of the SG-ICT Outside-Real app are described in the sequel:

- *SG-ICT OUTSIDE-REAL GUI* represents the graphics interface that allows the user to interact with the application during a visit to a museum or an archaeological site. This interface include suitable menus, a control panel and is also able to operate both in stand-alone mode or remotely (connected on the web);
- *OUTSIDE-REAL Engine* is the application engine that enables the integration of the Augmented Reality (AR) module with the Intelligent Agents (IA) which in turn should manage the SOPHOS personal assistant and knowledge navigator;

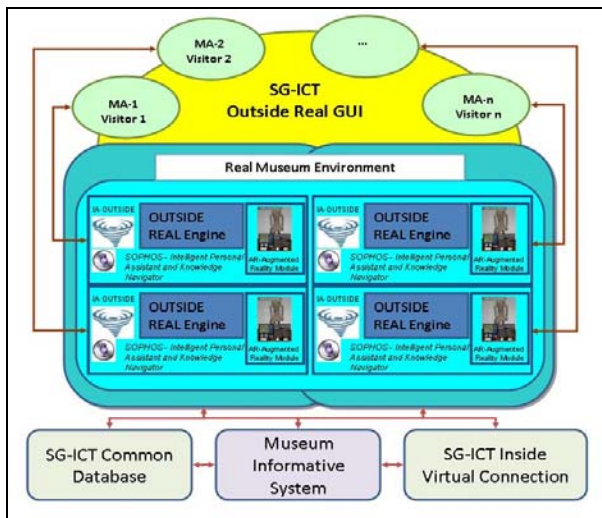


Figure 2 - OUTSIDE-REAL General Architecture

- *SOPHOS* is an *Intelligent Personal Assistant and Knowledge Navigator* that the visitor can use during the visit to the museum in order to get quickly, easily and efficiently further information about the cultural object of interest through questions orally made. Intelligent Rules and logic that characterize the navigation system of knowledge are implemented in a dedicated module called *IA-Outside*;
- *Augmented Reality Module* allows the user to get “augmented” information about the cultural object of interest through the mobile device (smartphone or tablet), pointing the camera to the find of interest;
- *SG-ICT Common Database* is a database in which are stored all the information and data required for the functioning of the app combined with all the data that are generated during the functioning of the app itself;
- the *Museum Informative System* is the possible connection between the *OUTSIDE-VIRTUAL* application and the information system of the museum. The rationale behind such a connection is the same of those reported for the *INSIDE-VIRTUAL* application. In fact, this connection is devoted to ensure the development of the app and its maintenance.
- the *SG-ICT INSIDE-VIRTUAL Connection* represents a connection gateway with the *INSIDE-VIRTUAL* app. As previously highlighted, some of the information, used in the *INSIDE-VIRTUAL* app, could be used in the *OUTSIDE-REAL* application. Besides 3D models (the information flow goes from *INSIDE-VIRTUAL* to *OUTSIDE-REAL*), reviews, statistics, users’ preferences are collected through the use of the *OUTSIDE-REAL* app (in particular through the *SOPHOS* system). In this sense, such information could be viewed comfortably by other users who use the *INSIDE-VIRTUAL* app. (the information flow goes from *OUTSIDE-REAL* to *INSIDE-VIRTUAL*).

In order to provide a more detailed description of the *OUTSIDE-REAL* app architecture, consider the point of view of a user who is using the app during a visit to a museum or archaeological park. The visitor can download the *OUTSIDE-REAL* app for free, or paying a fee, and install it on his/her mobile device before starting the visit to the museum. During the visit, just pointing the smartphone or the tablet towards the object of interest, the user receives additional information in augmented reality about the exhibit. There are different kinds of “augmented” information and they are proposed to the user through different modes (e.g. text, images, videos, audio files).

In addition, the user can request specific information through the *SOPHOS* software, driving consequently his learning process according to their needs.

Another important aspect is the scalability and replicability of the *SG-ICT* apps. In fact, museums and archaeological sites are constantly trying to define a knowledge base and a common language (especially for the *Virtual Museum* concept and for the identification of new tools for visualization and interaction).

The scalability of the *SG-ICT* apps is guaranteed by the capacity to operate correctly both in the case of larger system and in case of smaller ones. While *Intelligent Agents*, *Augmented Reality* algorithms and the *SOPHOS* software continue to work regardless of the size of the system, virtual environments and databases can easily be expanded/reduced. So the transition from a little to a larger museum will only need the upgrade of virtual environments and databases.

The repeatability is ensured by the definition of specific ontologies for the *Intelligent Agents*, for *Augmented Reality* algorithms and for the *SOPHOS* system. Such ontologies ensure the capacity of the various components to be replicated and operate in different contexts than those for which it was originally designed.

The *SG-ICT* project opens the way for the application of similar *OUTSIDE-REAL* approaches in *Marine Protected Areas*, with the adoption of new waterproof technologies to be used and installed on devices with the purpose of a sustainable diving tourism. Even the *INSIDE-VIRTUAL* app could be used to create interactive virtual underwater environments. In addition, both *OUTSIDE-REAL* and *INSIDE-VIRTUAL* are very interesting tools to produce, distribute and enjoy scientific, ecological, cultural and educational contents (which are the most expensive activities in *Marine Protected Areas*). To this end, the *Protected Marine Area of Portofino* has expressed its interest in the *SG-ICT* project, indicating clearly the potential of the project.

In particular, the *SG-ICT* project involves several museums as end users:

- the Museum of “Brettii” and Sea of Cetraro;
- the Operation Avalanche Museum of Eboli;
- the Silk Museum of Mendicino.

#### 4.1. Intelligent Agents and SOPHOS System

As already pointed out, the SOPHOS software is a personal assistant performing the role of an intelligent navigator and, therefore, able to answer to a specific question.

Initially, the SOPHOS ability to comprehend and provide coherent answers to the questions posed by the user is limited not only by the available technologies but also by the willingness to develop a system capable of responding coherently to a list of keywords or simple questions made up of keywords that can be used by the user in an intuitive, fast and effective way. The SOPHOS system is able to collect information and then to transfer such information from the Common Database to the Museum Information System (via remote connection). Such information can be useful for the museum management to improve the service and the usability of the cultural heritage (e.g. modify the information available for each exhibit, change the layout of the museum or of a particular exhibition area).

#### 5. THE SG-ICT PRELIMINARY DEMO

Currently an SG-ICT preliminary demo has been developed as shown in Fig. 3,4,5,6. In particular Fig 3 and Fig.4 refer to the INSIDE VIRTUAL mobile app while Fig 5 and Fig. 6 pertain to the OUTSIDE REAL app. As previously discussed INSIDE VIRTUAL is a Serious game based on simulation and is devoted to provide the user with an engaging and interactive experience while he/she is gaining information about the cultural heritage of a particular site or museum. In addition, in the INSIDE VIRTUAL app the scenario evolution evolves thanks to Intelligent Agents that are implemented in the form of avatars. Therefore, Fig 3 shows the user and its interaction with a virtual avatar that is providing information on a particular finding and is trying to get insights about the user's interests in order to provide him/her with the information he/she is interested in and drive suitably its visit within the museum based on his/her feedbacks. In Fig.4, the point of view is changed and the user can give a look into a museum room to have just a quick overview of the historical artefacts it contains before deciding whether to get into the room or not so that he/she can assess if there is something that could be of interest for him/her.

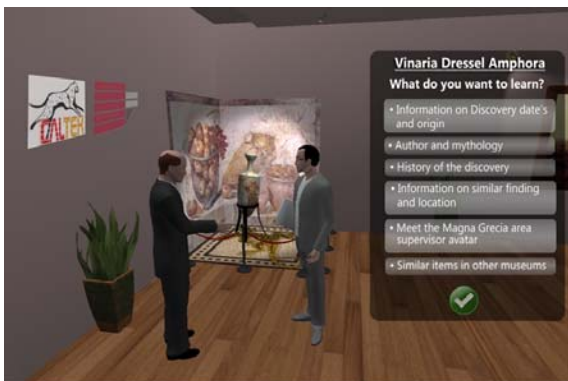


Figure 3, Example of digital Mock-up for INSIDE-VIRTUAL: the users interacts with an avatar virtually



Figure 4, Example of Digital Mock-up for INSIDE-VIRTUAL: view from a museum room

Fig. 5 and 6 show some digital Mock-up of the OUTSIDE-REAL app. As discussed in section 3.2 this app is based on Augmented Reality and includes an intelligent software navigator SOPHOS. This way, the app is able to provide the user with the information he/she wants about a particular finding during a visit in a real museum or archeological site. As a matter of fact, Fig.5 and Fig. 6. show a user while he/she is getting information on a particular amphora simply pointing his/her mobile device toward it. At this stage the knowledge navigator SOPHOS that is a siri-like system can be directly asked for further information (through a dedicated button placed in the lower left corner) on the particular object being considered, or even the user can ask for the location and the correct position of other artefacts.



Figure 5, Digital Mock-up for OUTSIDE-REAL: a visitor is using the app installed on a tablet in a real museum





Figure 6, Digital Mock-up for OUTSIDE-REAL: view of a visitor that uses Augmented Reality glasses.

## 6. CONCLUSIONS

The SG-ICT project (*Serious Games at increased Impact on Culture and Tourism*) is devoted to create an interactive and immersive way to produce, use and distribute cultural contents in museums and archaeological sites and to create an easy-to-use and effective way to receive cultural contents while visiting real museums and archaeological sites. To this end SG-ICT combines Modeling & Simulation, Intelligent Agents, Computer Generated Actions, Virtual Environments and Augmented Reality. The project goals will be achieved thanks to two different but complementary applications for mobile devices INSIDE VIRTUAL and OUTSIDE REAL.

INSIDE-VIRTUAL is a multi-player, avatar-based, web-based mobile app in which the player follows a learning process as a virtual visitor in a museum or an archaeological site.

OUTSIDE-REAL is a mobile app that will be provided to visitors throughout their tour in a real museum or site. The visitor can use the application installed on his/her mobile device and receive, through the use of AR, "augmented information". The application will also be equipped with a software (SOPHOS) that recreates the features of a personal assistant and knowledge navigator.

A preliminary demo of both the apps has been developed and some digital mock-ups are given to demonstrate the potentials of the SG-ICT framework.

## REFERENCES

- Addison A., 2000. Emerging trends in virtual heritage. *IEEE Multimedia, Special Issue on Virtual Heritage*, 7 (2): pp. 22–25.
- Amigoni, F., Schiaffonati, V., 2009. The Minerva System: A Step Toward Automatically Created Virtual Museums. *Applied Artificial Intelligence* 23(3): pp. 204-232.
- Anderson, E.F., McLoughlin, L., Liarokapis, F., Peters, C., Petridis, P., and de Freitas, S., 2010. *Developing Serious Games for Cultural Heritage: A State-of-the-Art Review*. *Virtual Reality* 14 (4): pp. 255-275.
- 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. (2013) Intelligent agent-based simulation for supporting operational planning in country reconstruction. *International Journal of Simulation and Process Modelling*, 8 (2-3), pp. 145-159.
- Bruzzone A.G., Longo F., 2013. 3D simulation as training tool in container terminals: The TRAINPORTS simulator. *Journal of Manufacturing Systems*, 32: pp. 85-98.
- 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
- 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.
- Ikeuchi K., Nakazawa A., Hasegawa K., Ohishi T. 2003, The Great Buddha Project: Modeling Cultural Heritage for VR Systems through Observation. In: *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, IEEE Computer Society Washington, DC, USA.
- 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., Massei M., Nicoletti L., 2012. An application of modeling and simulation to support industrial plants design. *International Journal of Modeling, Simulation, and Scientific Computing*, 3, pp. 1240001-1-1240001-26.
- Longo, F. (2013) On the short period production planning in industrial plants: A real case study. *International Journal of Simulation and Process Modelling*, 8 (1), pp. 17-28.
- Luck M., McBurney P., Preist C., 2004. A Manifesto for Agent Technology: Towards Next Generation Computing. *AAMAS*, 9(3): pp. 203–252, Springer.
- Maim J., Haegler S., Yersin B., Mueller P., Thalmann D., Van Gool L., 2007. Populating ancient pompeii with crowds of virtual romans. In: *Proceedings of VAST07: the 8th international symposium on virtual reality, archaeology and intelligent cultural heritage*, pp 109–116
- 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
- Vilbrandt C., Pasko G., Pasko A., Fayolle P.-A., Vilbrandt T., Goodwin J. R., Goodwin J. M., Kunii T. L., 2004. Cultural Heritage Preservation Using Constructive Shape Modeling. *Computer Graphics Forum*, 23(1), pp. 25–41.
- 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.
- Zyda M. (2005) From visual simulation to virtual reality to games. *IEEE Computer*, 38(9): pp. 25–32.

#### ACKNOWLEDGMENTS

The research presented in this paper is financed by the Italian Ministry of University and Research (MIUR) under the program “Bando STARTUP”, PAC02L2\_00192 SG-ICT Project.

# DISASTER AND EMERGENCY MANAGEMENT SIMULATION IN INDUSTRIAL PLANTS

Agostino G. Bruzzone  
Simulation Team, DIME, University of Genoa, Italy  
*Email: agostino@itim.unige.it - URL: www.itim.unige.it*

Marco Frascio  
DISC, University of Genoa, Italy  
*Email: mfrascio@unige.it - URL: www.disc.unige.it*

Francesco Longo, Alessandro Chiurco  
MSC-LES, University of Calabria, Italy  
*Email: {f.longo, a.chiurco}@unical.it - URL: www.msc-les.org*

Simone Zanoni, Lucio Zavanella  
DIMI, University of Brescia, Italy  
*Email: zanoni@ing.unibs.it, lucio.zavanella@unibs.it, - URL: www.unibs.it*

Paolo Fadda, Gianfranco Fancello  
DICAAR, University of Cagliari, Italy  
*Email: {fadda, fancello}@simulationteam.com - URL: dipartimenti.unica.it/ingegneriacivileambientaleearchitettura*

Domenico Falcone, Fabio De Felice  
DICEM University of Cassino, Italy  
*Email: {falcone, defelice}@unicas.it - URL: www.dii.unicas.it*

Antonella Petrillo  
Department of Engineering - University of Naples "Parthenope", Italy  
*Email: antonella.petrillo@uniparthenope.it*

Pasquale Carotenuto  
IAC CNR, Italy  
*Email: p.carotenuto@iac.cnr.it - URL: www.iac.rm.cnr.it*

## ABSTRACT

This paper presents an approach to model and simulate industrial plant accidents as well as the related emergency management; interoperable simulation is proposed as approach for applying High Level Architecture in this context. The authors are focusing their attention on the disaster simulation and its interaction with the emergency management. Modern simulation, by using technology enables such as mobile solution, enhance its support to first responders, the dynamic reaction to crisis evolution as well as the improvement in training and management of safe routing and handling of injured people. This paper introduces these elements as part of new research track devoted to get benefits from interoperable simulation in federating multidisciplinary models for industrial plant emergency management.

**Keywords:** *Interoperable Simulation, High level Architecture, Industrial Plants, Emergency Management, Safety and Security*

## 1. INTRODUCTION

Emergency Management is a problem where simulation represents a crucial element since many years; hence specific organizations and events are consolidating in this area to address the phenomena complexity (Wybo et al. 1996; Bruzzone and Kerckhoff 1996); indeed the complexity of accidents due to physical aspects and

interactions among many elements are usually promoting simulation as the main investigation approach.

Therefore it becomes evident today the opportunity to get additional benefits from current advances in this area for enhancing the potential of modeling and simulation (M&S). Indeed M&S has been extensively applied in Industry both for supporting decision making on for training purposes (example of M&S applications in industrial plants can be found in Latorre-Biel and Jiménez, 2013; Longo, 2013, Bruzzone and Longo, 2013, Pereira et al. 2013). From this point of view there are different elements to be considered, from one side the technology enablers, such as Internet of Things (IoT), allows today to distribute easily simulation; this concept is even more stressed by introducing the Simulation as a Service (SaaS) paradigm that could allow to transform any mobile device (Amram 2009; Bruzzone et al. 1999).

From other point of view, disasters in Industrial Plants are critical issues not only for the internal complexity of phenomena by their selves, but also for the necessity to combine them together; for instance it is growing the interest in further develop existing models in order to be able to address combined cases such as dust explosion and fire or interaction among seismic activities and firefighting (Dastidar et. al. 2005; Hinrichs et al. 2011; CSB Reports 2005 and 2008). From this point of view the use of interoperable simulation allows to federate different models and to let them to interact dynamically (Kuhl et al.



1999); this approach enables also to combine models reproducing the people reaction as well as the actions carried out by the different responders (Mujica and Piera 2012); in this sense it could be possible to obtain interesting results by applying intelligent agents (Shi 2008; Taboada et al. 2011); in the proposed approach it is expected to federate in the interoperable simulator the IACGF (Intelligent Agent Computer Generated Forces) developed by Simulation Team and already tested over earthquake scenarios and in health care sector (Bruzzone Massei 2010; Bruzzone, Novak, Madeo 2012).

Due to these reasons the authors are currently developing an innovative research line for applying interoperable simulation to disaster and emergency management within industrial plants and in this paper it is proposed the general approach as well as the federation architecture. Indeed this research is currently supported within the DIEM-SSP Project (Disasters and Emergencies Management for Safety and Security in industrial Plants) co-sponsored by the Italian Department for Research (MIUR) and carried out by several Universities and Institutions (i.e. DICAAR University of Calabria, DII University of Cassino, Simulation Team, DIME University of Genoa, DIMI University of Brescia, DISC University of Genoa, IAC CNR, MSC LES University of Calabria); indeed the main goal of DIEM-SSP is to carry out research and developments activities to deal with the complexity resulting from emergency management in industrial plants and critical infrastructures.

Currently the authors are working on this research track in strong international collaborations (both in medicine and engineering area); for instance, the following institutions have already expressed their interest in taking part in the DIEM-SSP project: the Harvard Medical School (MA, USA), the Dartmouth University (NH, USA), the Old Dominion University, Virginian Modeling, Analysis and Simulation Center, VMASC (VA, USA), in addition to several Firefighters Departments.

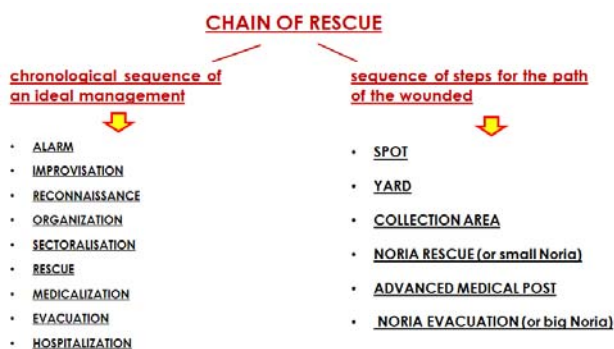


Figure 1- Chain of Rescue during a Disaster

## 2. INDUSTRIAL PLANTS, SAFETY & SECURITY

Many industrial plants represent Critical Infrastructures (CI), and it is evident the complexity of these CIs exponentially increases when emergency occurs, both for safety or security issues. In this sense there are, among

others, different crucial aspects to be addressed in emergency management within this context:

- Disaster dynamics related to the specific plant and disaster on going as well as to interactions among the different elements: e.g. domino effect in fires and explosions within a chemical plant (Babrauskas 2003; Crowl 2003)
- Medical issues intended as the need to have emergency procedures allowing medical staff to operate efficiently in stressing and complex environments.
- Emergency management issues intended as capabilities to reduce human errors as much as possible (during emergency management) and properly route any patients with severe trauma to the proper First Aid Facility.

The proper and effective management of an emergency scenario is usually based on work scheme in which specific roles held by doctors nurses and rescue workers (with adequate training) are identified.

This management approach requires to combine technical and medical aids provided by the different actors (e.g. firefighters, police, health care systems, etc.) so it becomes evident the importance to have multiple models addressing these different elements. In addition and based on this assumption, it is important to define the chain of rescue that support the process as proposed in figure 1.

Hence the disaster could emerge as accident or as a manmade event, due to these reason it could be pretty interesting to model not only the safety systems and to use the proposed simulation as support for safety engineering and procedure design, but also the security systems; it is evident that malicious behaviour represents a threat for industrial plants that could be pretty challenging to be mitigated. Simulation supports the vulnerability reduction in this context by allowing evaluation of the effectiveness of introducing innovative technologies, new solutions as well as new security procedures.

In addition to these problems the accidents represent another very important element to be introduced, even in this case sometime the human factor affect the relevance of the event as well as its evolution; so it is evident the necessity to model these elements.

Due to these reason it is evident the importance to develop a new generation of interoperable simulators addressing these problems; indeed it become necessary to create simulators able to combine different models addressing these different elements.

In addition it is evident the necessity to develop decision support systems able to face simultaneous all these crucial aspects: efficient procedures are needed both on the emergency site and in hospitals as well along the entire supply chain (Merkuryeva et al. 2011); these procedures should be combined with a proper critical patients routing toward the most suitable First-Aid Facilities should be designed and prepared in order to improve effectiveness and to reduce the human errors during the crisis. Furthermore, new procedures or methodologies for emergency management must be carefully tested before used: the need to assess their validity before their application in a real emergency situation strongly depends

on the capability to have a clear picture of the procedures effectiveness based on simulation (Sell, Braun 2009).

Due to these reasons it emerged the opportunity to develop interoperable simulators, based on the IEEE 1516 High Level Architecture (HLA) and to test them on specific case studies for validating the approach.

In order to complete successfully these aspects it is necessary to address several specific elements:

- Modelling of the disaster evolution in dynamic interaction with humans
- Modelling Human Behaviour Modifiers in order to evaluate their impact on the crisis evolution; in particular to evaluate the possibility to reduce the number of patients with severe trauma as well as the damages to critical infrastructures by adopting using emergency procedures which take into account the human factor/error
- Modelling new emergency procedures to be used in hospitals for critical patients management (the patients with severe trauma coming from the emergency site)
- Modelling new technological aid able to support emergency management (i.e. RFID and/or face recognition to track and manage patients)
- Development of algorithms to guarantee the proper routing of critical patients from the emergency site to the most suitable First Aid Facilities through the optimal location of proper infrastructures (including hospitals) and logistic network redesign (Lin et al. 2009)
- Development of Training Solution based on Simulation to support education and training of first responders, emergency management planner, civil protection, firefighters and other actors (Raybourn 2011).
- Development of Decision Support Systems based on Simulation to support emergency management on the field adopting edge maturity level (Bruzzone et al. 2009).

This research obviously strongly focuses on the necessity to model the complex issues related to emergencies management in industrial plants and CIs; to achieve this goal it is necessary to be able to create models that rely on the capability to acquire knowledge and collect data related to this context; for instance it is crucial to be able to properly model emergency procedures to be carried out in healthcare facilities or directly in the emergency site, as well as to assess their impact on the performance of such structures; another critical element is provided by the necessity to model the handling of injured people, the optimal healthcare facilities identification selection, as well as the human behaviour modifiers that could affect the crisis and potential solutions to mitigate their effects.

For instance a critical aspect in emergency management is related to the efficiency of first aid facilities. It is critical to have accurate symptoms assessment and quick decision-making to match critical patient needs. Real-time points-of-care (based on new approaches/methodologies, i.e. ultrasound) are the real added value for the assessment of critical patients both in pre-hospital and in-hospital situations. Although recommendations in terms of education and decision making have been proposed by a variety of specialties, today there are still scattered and

limited examples of standards for critical and intensive care professionals (Raybourn 2012). The potential of simulation is very high in this sector providing the capability to visualize the situation and the emergency evolution in clear way (Jezek et al. 2012).

From an engineering point of view, in case of emergency in a critical infrastructure, evacuation and routing of injured people to first aid facilities (that should have as a consequence optimal capabilities as well as convenient location) must be properly addressed, since many experiences have shown that inefficiencies at this stage lead to serious problems both on people and infrastructures (CSB 2005). In addition, the “human factor” significantly contributes in creating the emergency as well as in the emergency management; many studies have been carried out in this area and they usually consider separately human factor and systems reliability, without considering any interactions among themselves as in the conventional reliability allocation methods (Bruzzone et al. 2008). Considering high risk industrial areas, it is clear the need of combining risk analysis with human factor evaluation methodologies, known under the name of Human Reliability Analysis (Vuorio et al. 1987; Bieder et al. 1998; Reer et al. 2004).

It is worth saying that in such a context, any new emergency procedure or methodology need to be extensively tested and validated before use. To this end, simulation has been widely recognized as the most suitable approach for investigating, analyzing and solving complex problems in real systems such as industrial plants and critical infrastructures. The authors acquired experience confirming that simulation is very effective in emergency management even if much more can be done by using interoperable and distributed simulation and the authors are currently working on using architecture for distributed simulation such as HLA (High Level Architecture) to further empower this approach (Bruzzone and Massei 2010). In this case different interoperable simulation models recreating the behaviour of different entities acting in the same scenario could be used to recreate the crisis framework and to test the interactions among all the actors involved; by this approach it becomes possible to evaluate the effectiveness of the simulated procedures and methodologies as well as their efficiency; for instance it is possible to use a simulation model to reproduce the evacuation processes and the critical patients routing to first aid facilities during an accident in a chemical plant, in this case the simulation model could reproduce the disaster evolution as well as the application of emergency procedures on site and in the first aid facilities (Vonolfen et al. 2011).

### **3. HEALTH SUPPORT AS CRUCIAL ELEMENT IN INDUSTRIAL PLANT DISASTERS**

The health support is one of the most complex aspects in emergency management in industrial plants, because the structures must be able to ensure, in the early hours, a rapid response in order to give aid to as many people as possible.

In response to these complex requirements in Italy was created a macro system able to address of the integrated management of Emergency Care; this organization is supported by National Health System (DPR 27/3/1992), which can be accessed for free through the 118 telephone number unique on the whole national Italian territory (Cossiga 1992). The core of this system is the Operational Territorial Centre able to immediately activate locally the first public health response to an emergency situation. In this way it is possible to achieve seamless integration between community care and hospital care in its various forms. In addition to the Operational Territorial Centres the macro system for medical emergency management is composed by mobile rescue units (road transportation vehicles, special emergency vehicles for fast response), fixed and mobile points for first responders located on decentralized areas, emergency assets as we (Chen 2010); as multilevel Departments (i.e. I and II level). In the event of a disaster, the integration of all the actors (Police, Fire Brigade, Health Personnel) is crucial for guarantee successful management of the emergency. The specific activities of each actor include among the others area demarcation, police services, radio link activation, rescue operations, rescue, health care. The goal of all these phases is the collection of the victims and their subsequent treatment. The first triage (triage sweeping), corresponding to the first quick immediate action, is performed within the area of the disaster, if accessible, with the aim of selecting the victims at the end of their evacuation and of their initial treatment on site if possible (basic actions necessary to ensure survival). The method adopted for the first triage system is called START (simple rapid triage and treatment); it should be noted that next to the START method it is applied also CESIRA (which don't include black code) intended for first responder teams that don't include medical component and that is unable to address deaths on the field (Higgins 1987; Benson et al.1996; Bateman 2014). the Advanced Medical Post (AMP), the seat of second triage, health treatment, including procedures for stabilization and subsequent evacuation to the most suitable hospital centres is located at the outer edge of the safety area, or in a central area respect to the front of the event, protected from any risk of evolution and near roads, is compiled and managed by the staff of the 118.

Hence the main objective of the proposed simulators is to be able to carry out experimentation to address the complexity that comes from emergency management in industrial plants and in critical infrastructures; indeed this elements becomes even more critical when the emergencies are resulting from voluntary manmade disasters; it is evident that one of the first objectives to be pursued is creating a simulator supporting planning from a multi-disciplinary point of view (Rauner 2012). In this case it becomes fundamental to promote new "points of care" in order to maximize the basic and advanced skills in the evaluation and management of critical patients; from this point of view the simulator could support evaluation of new methods proposed for covering diagnosis over crisis site as well as in (e.g. ECO FAST); in addition the

simulator could allow to estimate effectiveness and efficiency provided by the introduction of new telemedicine technologies and procedures focused on improving early diagnosis (Lagi et al. 2009). The diagnostic and therapeutic paths followed by patients with multiple trauma within the hospital will be evaluated by considering the different existing protocols for multidisciplinary management.

Therefore the definition of new emergency procedures is not the only aspect to be considered to improve the management of emergencies in case of accidents or disasters in critical infrastructures; in fact, it is necessary to improve the activities carry out outside of emergency rooms, in order to properly route the critical patients to the appropriate facility (including both optimal location of first aid facilities that critical infrastructure and efficient logistics network design) and to reduce human errors during the emergency phase; to this end, all patients will be evaluated with specific diagnostic score, a score correlated to treatment outcome.

The proposed simulator could serve as a tool to investigate the problem by identifying bottlenecks and most sensitive parameters in addressing the emergency management respect a specific scenario becoming an useful decision support system as well as an effective planning aid and training equipment.

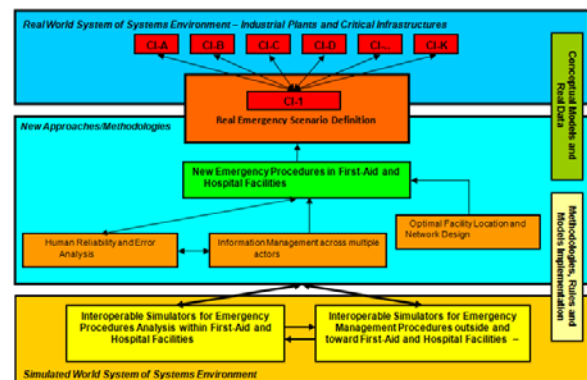


Figure 2- Disaster Modelling Multi Layer Approach

#### 4. GENERAL APPROACH

In order to succeed in modeling disasters in this complex environment as well the related emergency management it is proposed an approach based on a 3-layer approach (see the following figure); indeed the approach is based on approaching the real world as a system of systems; usually it includes multiple interconnected Critical Infrastructures (CIs) that represent the first layer; the approach move toward to the definition of new emergency procedures to be used in case of disasters/accidents in CIs both within First-Aid and Hospital Facilities (with the aim of improving the critical patients pathway) and outside to support the optimal location of most effective First-Aid and Hospital Facilities; the routing of critical patients and the reduction of human errors in the emergency management represent the second layer in the proposed approach. Finally, the third layer is a simulated system of

systems environment that includes different simulators to be used for testing the new emerging procedures.

The three layers are interconnected at two different levels that also represent the main research effort of the research:

- innovative emergency procedures will be defined (in the second layer) by strictly considering real CIs (first layer) and they will be tested and validated by using the interoperable simulators provided in the third layer
- test and validation of the emergency procedures will require the implementation of interoperable simulators (third layer); such implementation will require the definition of conceptual models and data collection from the real system (first layer) and the correct translation of the approaches and methodologies (defined in layer 2) in a form able to be executed through interoperable simulators.

As clearly shown in the modeling approach, these activities require joint engineering-medicine research efforts and knowledge. Accurate assessment, rapid decision-making and ad-hoc procedures are essential to save lives and improve performances in emergency situations and it is necessary to properly identify the critical issues to model working side by side with subject matter experts. It is evident that creating a virtual simulation reproducing these dynamics represents a very interesting support for education and training people operating in First-Aid and Hospital Facilities: these people receive critical patients coming from an accident area, so a high level of quality is necessary and it should be guaranteed by appropriate education, experience, credentialing, quality control, continuing education, and professional development. Although educational recommendations have been proposed over the time, to date they are still scattered and limited examples of standards and emergency procedures for critical and intensive care professionals; improvements in these areas could be guaranteed by the proposed simulators. In particular, the challenge of providing adequate decision-making, as encouraged by major medical societies, is made even more difficult by the diversity of critical care methods utilization and by the lack of well-defined guidelines in the advanced trauma life support. Therefore the first goal to be pursued within First-Aid and Hospital Facilities is to plan a multidisciplinary “Critical Care Pathway” for the management of a critical ill patient (arriving from a disaster scenario) and promote specific points-of-care program with the aim of maximize basic and advanced competences in assessing and managing the critical patients.

However, while new points-of care programs are the correct way to define new emergency procedures to be used within first-aid and hospital facilities and tremendously contribute to save human lives, it is also important to understand the effect in terms First-Aid and Hospital performances when these new emergency procedures are applied. The definition of new emergency procedures within First-Aid and Hospitals facilities (and their impact on facilities performances when applied) is not the only aspect to be considered for improving emergency management in case of emergency in CIs. Indeed, much more should be done outside the First-Aid and Hospital

facilities for supporting the routing of the critical patients toward the First-Aid and Hospital Facilities (optimal facilities localization and logistic network design) and also for reducing the human errors during the emergency management. In fact, there are many examples of accidents/disasters in CIs where many lives have not been saved for human errors (above all at leadership level). Also in this case it is critical not only to define, but also to understand “a-priori” the effect that new methodologies for routing optimization and optimal facilities location as well as for human errors reduction may have on the emergency scenario management. Therefore while a research effort will be devoted to define new emergency procedures within First-Aid and Hospital Facilities, to investigate the effects of human errors in case of emergencies management in CIs and to study routing problems and optimal facilities localization, an additional research effort will be carried out to implement two interoperable simulators that will be used as test-bed and released as demonstrators based on specific case studies. This will give the possibility to simulate different scenarios including medical and para-medical units or other operators both outside and within the First-Aid and Hospital facilities; the correct definition of the human behaviours will require extensive information in terms of human reliability assessment and error analysis and a great contribution from medical experts is also expected to implement correctly the new emergency procedures.

From this point of view the simulator should be able to model the different phases of the disaster including:

- Simulation of the Cause of the Disaster
- Simulation of the Disaster
- Simulation of the Effects related to the Disaster
- Simulation of the behaviour of systems of intervention

The methodological approach that will be used for implementing the new interoperable simulators will be based on the standards for interoperable simulation projects IEEE HLA 1516 (High Level Architecture). Initially the authors plan to develop two families of interoperable simulators will be released : the first one devoted to investigate the effects of the new procedures within the First-Aid and Hospital facilities, while second one focusing on simulating what happens in a CI (therefore before the arrival of patients to the hospital); in this second case the simulators will investigate the effects of human errors and for testing new methodologies for optimal hospital facilities location and for optimal patients routing toward hospital facilities.

Furthermore in order to increase the success of this research the authors are creating a Simulation Quality Assurance (SQA) group devoted to act as responsible – during the whole research– for the Verification, Validation and Accreditation (VV&A) of the new simulators (Frydman et al. 2009). To prevent developers' bias, the SQA group will involve people and subject matter experts (both engineers and doctors) that are not directly involved in the definition of the new emergency procedures as well as in the implementation. The proposed simulators are expected to improve reaction capabilities of industrial plants and critical infrastructure respect man-made and/or

natural disasters also from a social and economic perspective; these aspects aim at increasing security and response capabilities of industrial plants (and generally speaking, of critical infrastructures) in emergency situations through advanced technologies based on M&S.

## **5. MODEL DEVELOPMENT PLAN**

The proposed research is currently articulated in different main activities devoted to achieve, step by step, the main objectives including new simulator experimentation

### ***5.1 Data Collection and Knowledge Acquisition on major disasters in critical infrastructures and Industrial Plants***

This activity will survey the major disaster in critical infrastructures during the last 20 years, including disasters in industrial plants and critical infrastructures.

### ***5.2 Simulators Requirements Definition***

The goal of this task is to define the requirements of the new simulators.

### ***5.3 Definition of Case studies and Scenarios for Experimentation***

With the support of subject matter experts involved in the research the most interesting case studies for demonstration and experimentation will be defined. These case studies will constitute the main scenarios to be experimented on the new simulators

### ***5.4 Human Reliability Assessment and Error Analysis: Methodologies for human reliability assessment and error analysis***

Starting from both state of art and data input analysis, the approach to address the human factors in the different danger situations will be defined.

### ***5.5 HRA: Methodologies for machines reliability assessment and fault analysis***

For this task the same approach will be applied to machines: Starting from both state of art and machines data, safety allocation techniques which study the machines in different possible danger situations will be defined. The reliability methods will be based on different techniques on reliability analysis.

### ***5.6 Characterization of cognitive model-based human error analysis and reliability assessment***

After the identification of evaluation methods, cognitive models will be created to address human error analysis and reliability assessment.

### ***5.7 Interactions for procedure definition***

Once defined the application methods and cognitive models, the interactions will be model in order to support the identification process of best procedures able to ensure minimization of injuries and victims in accidents.

### ***5.8 Development of global human behaviour models for human reliability assessment and error analysis in critical infrastructures***

Creation of global human behaviour models allowing errors reduction and reliability improvement.

### ***5.9 Analysis of the Optimal Facility Location and Network Design respect industrial Disasters***

Special attention will be paid to examining the most recent industrial disasters respect the methods for facility location and network design. During this activity important aspects will be investigated regarding risk management, medical care management, analysis of the critical issues affecting traffic and supply logistics.

### ***5.10 Identification of most significant parameters and new methodologies definition***

The Preparation of the protocols to be implemented in case of disaster requires identification of key parameters; this achievement leads toward development of the methodologies for transport management in case of emergencies within critical infrastructures and industrial plants.

### ***5.11 Testing and validation of Logistics in Disaster Management***

This phase will be devoted to testing and setting up the related methodologies. The tests will consider both people directly involved in the event and those mobilised for managing the emergency as well as those who indirectly hinder emergency services. The protocols will refer to the specific building or site where the emergency has occurred, to exit flows and evacuation network towards safe places and health care facilities as well as within those facilities themselves. Special attention will also be paid to internal pedestrian mobility.

### ***5.12 Definition of New Emergency Procedures in First Aid Facilities and Hospitals***

During this activity the research team will be in charge of identifying and defining emergency medical scenarios as well as all the situations and consequences of an emergency scenario. Such medical scenarios will be the base for the simulators conceptual models

### ***5.13 Analysis and study of the actual hospital pathway***

This activity will focus on the study and analysis of the actual standard hospital pathway followed by critical ill patients in order to identify critical issues and possible improvements. Interacting with referent professionals involved in critical scenarios management (e.g. surgeons, emergency doctors, anaesthetists), model will developed on the diagnostic resources habitually used by the emergency clinicians.

### ***5.14 Information management in emergency situations***

The aim of this activity is to develop models for information and communication management among professionals, operators and victims involved in an emergency situation.

### ***5.15 Identification and Definition of clinical outcomes for procedures evaluation***

This activity aims at defining clinical performance indicators and outcomes to be used for evaluating the performance of new procedures. Note that the evaluation of the impact of the procedures used on the technical efficiency of the hospital will be evaluated by using the simulators.

### **5.16 New Emergency Procedures and points-of-care programs**

The main idea behind this activity is to define new emergency procedures and points-of-care programs to be used for critical ill patients. This will be a challenging opportunity to create a multitasking Emergency Team made by different professionals involved in the critical ill patient management.

### **5.17 Federation Architecture**

This activity aims at defining and implementing the interoperable simulators conceived. Therefore the main goal of this activity is to implement the architecture of the federation of the interoperable simulators that will be based on the standard for distributed simulation IEEE 1516 HLA (High Level Architecture).

### **5.18 Interoperable Simulators for emergency procedures analysis within First-Aid and Hospital Facilities & for Emergency Management Procedures outside First-Aid and Hospital Facilities**

Both the simulation engines and the interface will be developed as part of this activity based on previously defined conceptual models.

### **5.19 Federation Verification, Validation and Accreditation**

The Verification, Validation and Accreditation (VV&A) of the federation will be done according to the guidelines proposed by the IEEE 1516 HLA standard. In order to assure high quality for the simulators this phase will be on going along the entire development lifecycle.

### **5.20 Federation Development and Integration**

The Federation Development and Integration will be made according to the procedures available in the IEEE standard 1516,3 FEDEP (Federation Development and Execution Process)

### **5.21 Federation Experimentation and Demonstrator**

The Federation will be used to carried out specific experimentations to assess the potentials and validity of the proposed emergency procedures and methodologies on the defined simulation scenarios.

## **6. CONCLUSIONS**

The paper proposes an approach for modelling industrial plant disasters taking into account the human factors as well as the complexity related to the emergency management; special attention is attributed to how to handle and route the injured people along the health care facilities.

This research results from a common effort of multiple Institutions operating in different areas and this paper represents just a preliminary description of the modelling approach and of the technologies under investigation.

Currently the authors are developing the conceptual models while they are also defining the scenario to be studied.

The research is moving forward in development of the new simulators able to create the interoperable federation combining all the different models.

## **ACKNOWLEDGMENTS**

The authors thank MIUR and supporting Institutions (i.e. DIME & DISC University of Genoa, DIMI University of Brescia, MSC-LES University of Calabria, DICAAR University of Cagliari, DICEM University of Cassino, IAC CNR) of DIEM-SSP Project (Disasters and Emergencies Management for Safety and Security in industrial Plants) for their contribution to these researches. The authors thank also Liophant Simulation for promoting the simulation results in the scientific community.

## **REFERENCES**

- Amram O. (2009) "A web - based model to support triage location allocation in mass casualty situations"
- Babrauskas, V. (2003) 'Ignition Handbook', Fire Science and Technology Inc, Issaquah, WA
- Bateman N., Jefferson R., Thomas S., Thompson J., Vale A. (2014) "Toxicology", Oxford University Press, Oxford UK
- Benson M, Koenig KL, Schultz CH. (1996) "Disaster triage: START, then SAVE-a new method of dynamic triage for victims of a catastrophic earthquake", Prehospital Disaster Med, Apr-Jun, 11(2): 117-24
- Bieder, C., P. Le-Bot, E. Desmares, J-L Bonnet, F. Cara (1998) "MERMOS: EDF's New Advanced HRA Method," in Probabilistic Safety Assessment and Management (PSAM 4), A. Mosleh and R.A. Bari (editors), Springer-Verlag, New York
- Bruzzone A.G., Kerckhoffs (1996) "Emergency Management" in the Proceedings of European Simulation Symposium, Simulation in Industry Genoa, Italy, October, Vol. I & II, ISBN 1-56555-099-4
- Bruzzone A.G., Page E., Uhrmacher L. (1999) "Web Based Modeling and Simulation", SCS, San Francisco
- Bruzzone A.G., Tremori A., Massei M. (2008) "HLA and Human Behavior Models", Proceedings of MMS2008, Ottawa, Canada, April 14 - 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., Massei M. (2010) "Intelligent Agents for Modelling Country Reconstruction Operation", Proceedings of Africa MS 2010, Gaborone, Botswana, September 6-8
- Bruzzone A.G. Tremori A., Massei M. (2011) "Adding Smart to the Mix", Modeling Simulation & Training: The International Defense Training Journal, 3, 25-27
- Bruzzone A.G., Novak V., Madeo F. (2012) "Agent based simulation model for obesity epidemic analysis", Proceedings of IWISH 2012, Wien, Austria, pp.195-203



- Bruzzone A.G., Longo F., 2013. An advanced modeling & simulation tool for investigating the behavior of a manufacturing system in the hazelnuts industry sector. *International Journal of Food Engineering*. Volume 9, Issue 3, pp. 241-257
- Chen C. (2010) "A Novel Emergency Vehicle Dispatching System" in : *Vehicular Technology Conference (VTC Spring)*, IEEE 77<sup>th</sup>
- Cossiga F. (1992) "Atto di Indirizzo e Coordinamento alle regioni per la Determinazione dei Livelli di Assistenza Sanitaria di Emergenza", DPR 27/3/1992, Rome
- Crowl D.A. (2003) "Understanding Explosions", American Institute of Chemical Engineers (AIChE), New York, NY.
- CSB (2008) "Sugar Dust Explosion and Fire", Investigation Report, U.S. Chemical Safety and Hazard Investigation Board, February
- CSB (2005) "Combustible Dust Fire and Explosions, Imperial Sugar Company", Investigation Report, U.S. Chemical Safety and Hazard Investigation Board, February
- Dastidar, A.G. et. al., (2005) "Evaluation of Dust and Hybrid Mixture Explosion Potential in Process Plants", *Process Safety Progress*. Vol. 24, No. 4, December
- Frydman C., Longo F., Mekouar K. (2009) "European Modeling and Simulation Symposium Vol.2", La Laguna University Press, ISBN 978-84-692-5415-8
- Hinrichs R., Jones L., Stanely E.M., Kleiner M. (2011) "Report on the 2010 Chilean Earthquake and Tsunami Response", Open File Report 2011-1053 1.1. US Department of the Interior, US Geological Survey
- Higgins S.D. (1987) "Emergency delivery: prehospital care, emergency department delivery, perimortem salvage", *Emergency medicine Clinics of North America*, Vol. 5, N.3, August, pp 539-540
- Jezek B., Vanek J., Antos K. (2012) "Visualization Methods in Emergency Medical Services Readiness Evaluation" *Advances in Visualization, Imaging and Simulation*
- Kuhl, F., Weatherly, R., Dahmann, J., 1999. *Creating Computer Simulation Systems: An Introduction to the High Level Architecture*. Prentice Hall, Upper Saddle River, USA
- Lagi A., Marini F., Cipani S., Sarti A.(2009) "Ecocardiologia per l'Intensivista: Eco FAST nel dipartimento emergenza urgenza", Springer Verlag, Trento, pp 269-274
- Latorre-Biel, J.-I., Jiménez-Macías, E., 2013. Modelling, analysis, and simulation of manufacturing processes with alternative structural configurations by parametric Petri nets. *Advanced Science Letters*, Vol. 19, no. 2, pp. 665-669
- Lin Y., Batta R., Rogerson P., Blatta A., Flanigana M. (2009) "A logistics model for delivery of critical items in a disaster relief operation: heuristic approach" *Socio-Economic Planning Sciences* Vol. 45 132-145
- Longo F., 2013. On the short period production planning in industrial plants: a real case study. *International Journal of Simulation and Process Modeling*, Vol. 8, no. 1, pp. 17-28
- Merkuryeva G., Yuri Merkurjev, Hendrik Vanmaele (2011) "Simulation-Based Planning and Optimization in Multi-Echelon Supply Chains" *Simulation: Transactions of The Society for Modeling and Simulation International*. Vol. 87, No.8, August, pp. 698-713.
- Mujica, M., Piera, M.A.,(2012) "The translation of CPN into NetLogo environment for the modelling of political issues: FUPOL project", *Proceedings of the 24th European Modeling and Simulation Symposium, EMSS, Wien, Austria, September*
- Pereira, D.C., Del Rio Vilas, D., Prado, R.R., Monteil, N.R., Rodriguez, A.L., 2013. Experimental manufacturing system for research and training on human-centred simulation. *Proceedings of the 23rd European Modeling and Simulation Symposium, EMSS 2011, Rome, Italy, pp 400-409*
- Rauner Marion S. (2012). "Resource planning for ambulance services in mass casualty incidents: a DES-based policy model" *Health Care Manag Sci* 15 254-269
- Raybourn, E.M., 2011. Honing Emotional Intelligence with Game-Based Crucible Experiences. *International Journal of Game-Based Learning*, 1(1), 32-44, January-March 2011.
- Raybourn, E.M. 2012. Beyond serious games: Transmedia for more effective training & education.
- Reer, B., V.N. Dang, and S. Hirschberg (2004) "The CESA Method and Its Application in a Plant-Specific Pilot Study on Errors of Commission," *Reliability Engineering & System Safety*, 83: 187-205
- Sell C., Braun I (2009) "Using a Workflow Management System to Manage Emergency Plans Environment" *Department of Geography*
- Shi J. (2008) "Agent-based evacuation model of large public buildings under fire conditions" *Automation in Construction* 18 338-347
- Taboada M., Cabrera E., Iglesias M. L., Epelde F., Luque E. (2011) "An Agent-Based Decision Support System for Hospitals Emergency Departments" *Procedia Computer Science* 4 1870-1879
- Vonolfen S., M. Affenzeller, A. Beham, E. Lengauer, S. Wagner (2011) "Simulation-based evolution of resupply and routing policies in rich vendor-managed inventory scenarios" *Central European Journal of Operations Research*, Vol. 19, No. 4, 2011, pp. 23
- Vuorio, U.M., and J.K. Vaurio (1987) "Advanced Human Reliability Analysis Methodology and Applications." In: *Proceedings of PSA '87: Probabilistic Safety Assessment International Meeting, Zurich, Switzerland, August 3 - September 4, pp.104-109*
- Wybo J.L., Therrien M.C. and Guarnieri F. (1996) "The International Emergency Management and Engineering Conference", Montreal, Canada, May 28-31

# A SYSTEM DYNAMICS APPROACH FOR IMPROVING CONTAINER TERMINAL OPERATIONS

Converso Giuseppe<sup>(a)</sup>, Gallo Mosè<sup>(b)</sup>, Murino Teresa<sup>(c)</sup>

<sup>(a),(b),(c)</sup> D.I.C.Ma.P.I. - Department of Chemical, Materials and Industrial Production Engineering,  
University of Naples Federico II, P.le Tecchio, 80 – 80125 - Naples - Italy

<sup>(a)</sup>[giuseppe.converso@unina.it](mailto:giuseppe.converso@unina.it), <sup>(b)</sup>[mose.gallo@unina.it](mailto:mose.gallo@unina.it), <sup>(c)</sup>[murino@unina.it](mailto:murino@unina.it),

## ABSTRACT

This study provides a simulation based framework for the optimal integration of the various activities within a container terminal. The operation of a container terminal is made possible thanks to the interaction of numerous subsystems, which, in turn, depend on other subsystems. Indeed the performances of the container terminal are affected by many factors such as component subsystems of the whole system, equipment, resources and procedures. In order to represent the relation between these factors, a System Dynamics approach is proposed. This approach tries to explain: 1) the interactions between all subsystems of the system; 2) the effect of any changes of their factors to the global performance.

This study aims to provide management tools to control performance of each subsystem and of the whole system and to anticipate the effect of performance improvement of subsystems on the whole system. The choice of utilizing a SD paradigm derives from the possibility of studying through a casual loop diagram the relationships between all variables involved and the effect of any changes of these factors on the each subsystem's performance and on the whole performance by means of a simulation model.

Keywords: System Dynamics, container terminal operations, simulation

## 1. INTRODUCTION

A container terminal is an open system of material flow with two external interfaces: seaside interface and landside interface. It consists of four functional subsystems connected together:

1. Gate area or landside. This area can be divided into import gate and export gate. Through the export gate, full trucks enter into the yard for unloading the containers. The import gate is dedicated to empty trucks coming into the terminal for picking up cargos.
2. Seaside is a space where the vessel can anchor for unloading/loading cargos. When unloading or loading cargos, an equipment called quay crane is used.

3. Interconnection area: this functional area is composed of equipment to transport containers between the seaside and the stacking area and vice versa.
4. Stacking area: this area can be divided into import and export yard. In order to unloading or pick up containers, trucks enter into export or import yard and drive to the assigned location where reach stackers or rubber tired gantry (RTG) unload or load the cargo. The yard equipments unload the cargo from the quay or handle the containers to load into the ship.

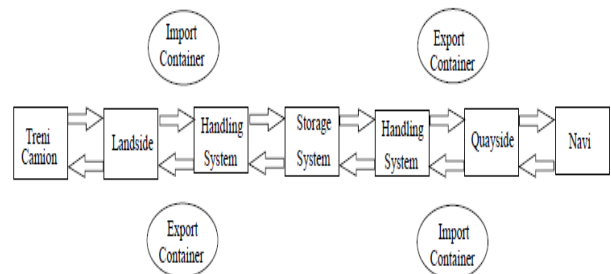


Figure 1: The physical flow of a container terminal.

The physical flow follows these interconnected steps:

- When a container ship reach the port, if a suitable terminal's quay is available, then the container ship enters the berth or waits at the anchorage.
- Container ships wait for service on the berth. If there are idle quay cranes the container ship is unloaded and/or loaded.
- Container trucks transport the container to the yard or to the quay. If there are idle quay cranes and/or yard cranes, empty tugmasters receive containers and move them to corresponding locations.

Therefore, the container terminal operations are:

- Receiving operation
- Delivery operation
- Loading operation
- Unloading operation

The productivity of a container terminal can be measured in terms of productivity of two type of operations. One type is the receiving and delivery operation, in which containers are transferred to/from external trucks. The other one is ship operation, in which the containers are unloaded and loaded from/onto the ships. Both types of operations are affected, in turn, by all internal operations, such as quay crane's operations, transfer operations between quay and yard and vice versa, and the yard operations.

## 2. MODELLISTIC APPROACHES IN LITERATURE

A container terminal is a dynamic and complex system where different operations have to be carried out. In fact it is characterized by highly complex subsystems that involve numerous equipments, operations and handling steps.

In literature different types of study have been proposed to model the operations in a container terminal. Different approaches are focused on modeling single operations or sequences of operations. However, most of the existing literature considers the optimization of a single subsystem of handling equipment, mainly in the scheduling quay cranes, cargo handling systems. These studies emphasized that each single problem, related to an operation, has to be considered as a part of the whole system and it has to be individually optimized in order to improve the whole performance. The operational problems considered are:

- Berth allocation;
- Loading and unloading of the ship: crane scheduling, stowage and unloading plane;
- Transfer of container from quay to storage area and vice versa;
- Stacking of container, scheduling of yard crane;
- Picking up of containers from export trucks;
- Delivery of containers to import trucks.

Many previous research studies on crane's scheduling treat the quay cranes separately from other resources in the terminal. For example, Kim and Kim (1999 a, b) consider the optimization of the routing of a single transfer crane in a port container terminal. The problem is formulated as a mixed integer program, and exploited in an optimization algorithm that uses dynamic programming.

Ascheuer et al. (1999) consider the scheduling of a stacker crane in an automated warehouse. The problem was modeled as an Asymmetric Traveling Salesman Problem (ATSP) and both heuristics and an exact Branch and Cut algorithm are used to solve the problem.

Meersmans and Wagelmans (2001) made a first integrated scheduling for QCs, AGVs, and automated stacker cranes (ASCs) in automated container terminals. They present a branch and bound algorithm and an heuristic beam search algorithm in order to minimize the makespan of the schedule. In a related research by Meersmans and Wagelmans (2001), an optimization-based beam search heuristic is proposed to solve the integrated scheduling problem in a dynamic environment. However, in both static and dynamic cases, only the vessel loading operations are considered. Most of the literature considered the scheduling of a single type of equipment. However, the container terminal subsystems are tightly interrelated, optimizing only one subsystem may not necessarily yield an overall optimal terminal operation.

In the last decade, researchers have started to focus on the simultaneous optimization of quay cranes and other subsystems. Therefore, to improve the productivity of a container terminal some studies focused on scheduling of different types of handling equipment in an integrated way. In this way, the operations performed by the container handling equipment at container terminals are interrelated to each other.

The latest studies (also our study) are focused on modeling the whole system under the point of view of interconnected and synergic subsystems. Considering the whole system, the problem complexity grows up because many problems are linked by shared resources, whose availability are time dependant, and their interconnections have affect on the whole performance. Therefore, the container terminal can be considered a "system of systems" for its complexity, because it is composed of a set of different subsystems with their specific goals. These subsystems, totally or partially, share some resources with fixed capacity in order to ensure the whole system functioning.

The import and export flows are the result of the interaction of the aforesaid four subsystems. Each subsystem is itself a complex system. The individual subsystems are different and can partially operate in an independent way, but their proprieties and interactions type affect the whole system performance.

Furthermore there are different interrelated performance indicators for each functional area of a container terminal, each measuring the productivity and utilization of each resource. The quay crane throughput can be considered the global performance indicator of a container terminal, because it is affected by the productivity of each subsystem. There are many different decisions involved in operating a container terminal and all these affect each other, affecting indirectly the efficiency of quay cranes.

Therefore it becomes clear that the interactions and synergy of these subsystems affect the global performance because the global performance is not only the sum of the performance of each subsystem.

Given the complexity of the "system of systems", as a container terminal, Systems Dynamics is a suitable

approach to analyze the feedback loops, the interactions between subsystems and the temporal delays. Therefore, through a dynamic and experimental approach it is possible to synchronize the subsystems in a global view perspective.

In order to analyze all interactions between subsystems it is necessary to define a conceptual model of the whole container terminal, where quays, yards, gates and transfer systems can be considered as resources. The corresponding simulation model can be used to phase all subsystems and to define their optimal configuration in a system perspective. In fact, considering a container terminal as a whole, the throughput is determined by individual capacity of each component, such as vessel berthing, vessel unloading/loading, container's transfer from quay to yard and yard to quay, yard storage and gate passage. But it is generally equivalent to the lowest capacity of each of these and it is also influenced by their interactions. The occurrence of waiting times in different points throughout system is dependent on the variability of service times of each stage of the process and the different cycle time of the resources involved.

### **2.1. A Casual Loop Diagram for container terminal operations**

The relationships between the various subsystems of a container terminal can be effectively represented through a Causal Loop Diagram (CLD, figure 2). The process of loading (and unloading) of containers from vessel to the container yard and vice versa uses three kinds of equipment: container quay crane, tugmaster and yard equipment (Rubber Tired Gantry for import area and reachstackers for export area) by working in a synergy, so the process of loading and unloading of each container from vessel to the container yard and vice versa depends on harmony of the above mentioned equipment. However, the unevenness of cycle times of these equipments causes often the lack of this harmony and their optimal synergy. Indeed, the speed of loading and unloading process depends on the Berth Occupancy Rate and on the capacity of the resources, but it is also related to the activities of transfer, storage and retrieval that take place into the yard. The lower part of the CLD shows the loops that characterize the subsystems and their interrelations.

In the upper part of the CLD there are several interconnected loops (figure 2). In fact, the waiting time of the ships entering the harbor is influenced by the activity of containers loading and unloading, because a ship can be moored only if a second pier is available or if the loading/unloading operations on the previous ship have been completed. Moreover, a loading and/or unloading process slower and/or a high amount of containers to be handled produces a long ship Turnaround time (TRT) and a greater Pre-Berthing Time for the next vessel, implying a reduction of the ships served. The number of ships served and quay availability are linked by a negative cause-effect

relationship, in fact the increase of quay availability produces the decrease of waiting time for ships. Thus, the number of moored ships increases with the decrease of the waiting time. On the other side, an increase in the number of containers to be handled involves a greater utilization of the resources of terminal, such as quay, transfer system and yard.

The lower part of the CLD (figure 2) shows the internal loops of each subsystem and also the interactions between all subsystems:

#### **1. Quay cranes subsystem.**

The cycle time of quay cranes in loading and unloading process is influenced by many variables, such as human resources, the number of cranes in operation (crane scheduling for single bay of the ship) and also by external variables, like the availability of tugmaster under the crane.

#### **2. Transfer subsystem**

The process of containers transfer from the quay to the yard and from the yard to the quay is affected by internal variables of the transfer system, such as the number of tugmasters (a variable connected to the number of operative crane), the ability of human resources involved, and also by external factors like the level of congestion in the yard for storage and retrieval operations.

#### **3. Yard subsystem**

The cycle time for containers handling (import and export flow) is influenced by the quality of human resources, by the overlapping of quay and gate activities and tugmasters delays.

Therefore, the upper part of the CLD shows the loops connected to the global system, while the lower part reports the loops characteristic of subsystems and their interactions. Considering these loops, the overall optimization of a container terminal does not depend on separated optimization of its subsystems, as the overall performance is not simply the sum of the performance of its components. In fact, increasing capacity of a subsystem may improve its performance but providing a negative effect on those of other components, and even have a negative impact on the overall dynamics of the terminal.

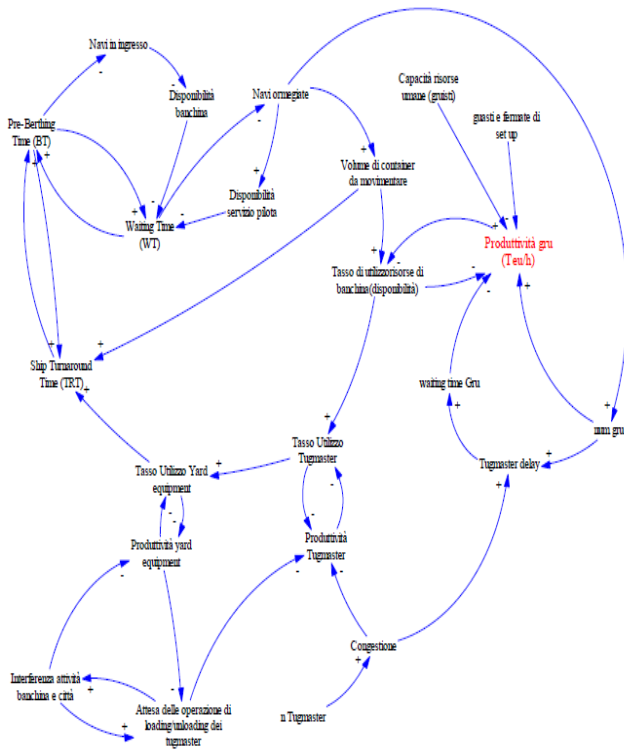


Figure 2: Casual Loop Diagram for terminal container operations.

## 2.2 Case study

The general CLD reported above has been applied to a real case in order to be validated. The container terminal considered handles containers loading –unloading with the wharf lengths of 315 and 295 meters, meaning that maximum two ships can be served at the same time. On the bigger quay there are 4 quay cranes while on the smaller one there are 2 cranes, but the number of operative quay cranes depends on the container stowage plan. When the vessel is anchored, the unloading and loading process can start. In order to transfer container from wharf to yard and vice versa, three tugmasters are assigned for each operative quay crane according to gang modality. The yard area is divided into import and export area according to a static strategy of space allocation. Summing up, the equipment involved in the terminal are:

- 2 quays with different length and draft and each of these is able to serve only one ship at once;
- 6 quay cranes, their function is unloading container from ships to the available corresponding tugmaster or vice versa;
- 10 Rubber Tired Gantries (RTG) in the import yard involved in stocking containers transported by tugmasters and delivering containers to trucks coming from the gate
- 9 Reach Stacker in the export yard. Two of them are assigned to export trucks while the others are assigned one for each operative quay crane;

- 15 tugmasters to transport containers from wharf to container yard or vice versa.

## 3. SCENARIO ANALYSIS

Based on the interconnected mechanism, shown in the CLD, we developed a simulation model of the terminal container considered. First of all, we run some experimental tests in order to validate the model. Then, we used this model to evaluate possible shortages or surplus of capacity for available equipment. In addition, we identified the system's components on which to intervene in order to optimize the capacity of equipment available. Several models have been developed each corresponding to different real system's configurations. The first model describes the current situation and the others allow to analyze possible interventions on elements that have significant influence on the global performance.

Several assumption have been adopted in constructing and implementing the improvement plans for our container terminal. In particular, the equipment configuration scenarios used are:

- Scenario I (SI), existing condition.
- Scenario II (SII), adding 1 tugmaster for each operative quay crane.
- Scenario III (SIII), reducing of 1 unit the number of tugmaster for each operative quay crane.
- Scenario IV (SIV), using also in the import yard the same strategy used in the export yard. Considering one operative RTG corresponding to each unloading quay crane.

As an example of the output from the models, we consider the quay cranes productivity, i.e. the number of container moved per hour. Through an experimental valuation we derived that the optimal number of tugmaster assigned to each loading /unloading crane is the current number, 3, indeed a lower number of tugmasters would cause a reduction in quay cranes productivity, while a higher number would determine the same performance. Basing on these considerations, we could conclude that resources' sizing and composition is already optimal, therefore the problem is the interference between quay and gate services. The Scenario IV, therefore, propose a new resource management strategy able to separate the gate and quay activities.

## 4. CONCLUSION

This study proposes a systemic approach to analyze the operation of a container terminal. The container terminal has been considered as "a system of systems", composed of complex subsystems. By taking into consideration the causal loop relationships between its components, the optimization of each subsystem does not provide the overall optimization of the system, in

fact an increase of subsystem's performance may provide a negative effect on the behavior of other components and on the overall dynamics of the terminal.

Using a systemic approach it is possible to analyze the internal dynamics of the container terminal depending on changes made to the initial conditions. The internal process of the container terminal has been explored through internal causal loops.

This study has also led to develop several simulation models to analyze the operation of a specific terminal container and to point out the dynamic internal changes of the terminal components corresponding to different scenarios for terminal operation. However, these models have been developed to emphasize the causal relationships existing within the system and in which measure these connections contribute to the overall terminal container operation.

## REFERENCES

- Ascheuer, N., Grötschel, M., Abdel-Aziz Abdel-Hamid, A., Order picking in an automatic warehouse: Solving online asymmetric TSPs, 1999, Computational Optimization and Applications, 17 (1), pp. 61-84
- A. Mostafavi, D. Abraham, D. DeLaurentis, J. Sinfield, "Exploring the dimensions of systems of innovation analysis: a System of Systems framework", IEEE Systems Journal, Vo1.5, No.2, 2011.
- Bruzzone, A., Longo, F., Document An application methodology for logistics and transportation scenarios analysis and comparison within the retail supply chain, 2014, Source of the Document European Journal of Industrial Engineering 8 (1), pp. 112-142
- Bruzzone, A., Nicoletti, L., Chiurco, A., Diaz, R., Palaia, E., Document Marine ports environmental sustainability: A state of the art overview, 2012, International Conference on Harbour, Maritime and Multimodal Logistics Modelling and Simulation 1, pp. 188-196
- Chang-Seung Ha, Gyu-Sung Cho, Chun Paik, Developing a Educational System on the Analysis of Transshipment Container Terminal Operation using a Simulation approach, The Korean Society Fisheries and Sciences Educations, 2007.
- C. Bierwirth, F. Meisel, "A survey of berth allocation and quay crane scheduling problems in container terminals", European Journal of Operational Research, 615–627, 202, 2010.
- D. De Laurentis and W. Crossley, "A Taxonomybased perspective for System of Systems design methods", Proc. 2005 IEEE International Conference D.J. Sterman, Business Dynamics: System Thinking and Modeling for a Complex World. McGraw-Hill, 2000.
- F. Meisel, "scheduling seaside resources at container ports", Wiley Encyclopaedia of Operations Research and Management Science, edited by James J. Cochran, 2010.
- Kim H. K., Kim Y.K., An Optimal Routing Algorithm for a Transfer Crane in Port Container Terminals, 1999, February 1, 1999, 17 – 33
- Longo, F., Nicoletti, L., Chiurco, A., Document Cooperative training for ships and tugboats pilots based on interoperable simulation 2013 25th European Modeling and Simulation Symposium, EMSS 2013 pp. 695-703
- Meersmans J, Poot, A., Kant, G., Wagelmans, A.P.M., A savings based method for real-life vehicle routing problems, 2002, Journal of the Operational Research Society 53 (1), pp. 57-68
- N. Karcanias and A. Hessami, "Complexity and the notion of System of Systems: Part(II): defining the notion of System of Systems," Proc. World Automation Congress, pp.I-7, 2010.
- Steenken D.; Voss S.; Stahlbock R. (2005). Container terminal operation and operations research - a classification and literature review In Container Terminals and Automated Transport Systems on Systems, Man and Cybernetics, pp. 86-91, 2005.



# INNOVATION IN HOSPITALS: AN E-PROCUREMENT MODEL IN PHARMACY OPERATIONS IN DAY SURGERY

Piera Centobelli<sup>(a)</sup>, Giuseppe Converso<sup>(b)</sup>, Antonio De Iasi<sup>(c)</sup>, Teresa Murino<sup>(d)</sup>

<sup>(a),(b),(c),(d)</sup> D.I.C.Ma.P.I. - Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, P.le Tecchio, 80 – 80125 - Naples - Italy

<sup>(a)</sup>[piera.centobelli@unina.it](mailto:piera.centobelli@unina.it), <sup>(b)</sup>[giuseppe.converso@unina.it](mailto:giuseppe.converso@unina.it), <sup>(c)</sup>[antonio.deiasi@gmail.com](mailto:antonio.deiasi@gmail.com), <sup>(d)</sup>[murino@unina.it](mailto:murino@unina.it)

## ABSTRACT

The relationship between pharmaceutical companies and hospitals is one of the most important key factors in healthcare system. Hospitals are complex organizational structures that can be improved in order to reduce wastes from both the economic point of view and the resources employed. This leads to a progressive modification of their commercial and technological relationships with pharmaceutical companies. This paper analyses the introduction of e-collaboration models for the supply relationship between hospitals and pharmaceutical companies and presents a new and more streamlined materials purchase procedure based on e-commerce. In order to verify the real benefits and analyze the peculiarities of this approach a simulation model using the System Dynamics has been implemented.

Keywords: lean healthcare, e-procurement, e-sourcing, e-supply chain collaboration

## 1. INTRODUCTION

The main problems of health organizations are bad resources allocation, inefficiency and consequently growing costs. The work presented aims to analyze interconnection between different departments trying to improve information flow through the introduction of e-procurement innovative technology, a web-based platform used for the regulation of materials and services supply. This technology should ensure the reduction of patients' queues, bad resources allocation and subsequent wastes which inevitably increase the total management cost.

In recent years the trend is to manage hospitals as companies, the organization structure passes from a functional structure to a structure focused on processes.

The new healthcare-company concept is based on the unending search for processes improvement and quality control; recently Lean thinking concepts are being applied to healthcare facilities (Habidin, Shazali, Ali, Khaidir, and Jamaludin 2014; Curatolo, Lamouri, Huet, and Rieutord 2014; Bhat, Gijo, and Jnanesh 2014) focused on flows management to reduce deviations and inefficiencies and finally to improve quality of medical treatments.

In order to improve the Day Surgery (DS) process the innovative technology of e-procurement has been identified. The Day Surgery is considered a solution to the problem of a lack of beds in hospitals, queue management and economic resources waste. Past researches have highlighted the economic benefit to perform some operation in DS than the traditional method (Broekelmann 2006).

As regards the e-procurement, there are many authors who have based their studies on the evolution of web-based systems and e-collaboration solutions with a focus on growth, development and expected trends related to implementing e-procurement. In 2002 Rangone and Balocco classify B2B applications focusing in particular on different tools and interactions provided.

The paper is organized into 2 main sections. The first is an analytical section which has the purpose to analyze the context and the new trends in the field of economic management, i.e. waste reduction, improvement of organizations and introduction of innovative models in healthcare organizations thanks to the development of new information technologies. In the second part a simulation models shows improvements achievable thanks to the introduction of new e-procurement tool. The case study has been provided for central pharmacy at CARDARELLI hospital in Naples.

## 2. PURCHASING FUNCTION TRANSFORMATION: E-BUSINESS

The introduction of new technological solutions involves both organisational changes and processes reengineering. The B2B tools implementation in companies allows management to pay more attention to activities with higher added value and strategic value. Therefore, the activities of the purchasing function are modified: as a result of increased access to information and the automation of order management, on the one hand greater control, speed and time available for tactical and strategic activities is achieved; on the other hand increases the dissemination of knowledge and purchasing activities carried out independently by other functions within the organization.

One of the consequences is the apparent loss of 'power' of the Purchasing Department, as it is no longer the only depository of information on contracts, pricing, catalogs and suppliers on the market.

The traditional purchase method changes from centralized to decentralized. This change is a function of the intensity with which the company pursues its buy strategy based on the electronic market.

The production process and effects resulting by the use of certain materials rather than other continues to be important; in this sense the sourcing approach will be strategic in order to manage the most critical direct materials in an appropriate way.

Talking about B2B trading, should be taken into account its application along the logistics chain in two opposite directions: both from customer (downstream, also known as E-commerce) and seller (upstream, where it is known as E-procurement) point of view (Figure 1).

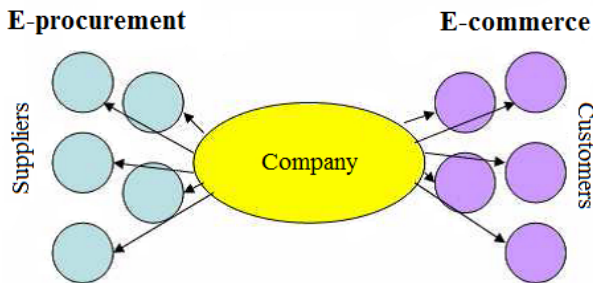


Figure 1: e-business in supply chain

The main aim of e-business, both upstream and downstream in supply chain, is to manage the business transactions in a better way than in the past. This is achieved by exploiting different strategies such as:

- Customized marketing strategy in real time;
- Increase in products and services supply;
- Number of customers reached increased.

The electronic marketplace is the place where online commerce takes place, the so-called e-marketplace is the meeting point between demand and supply on the internet and acts as a connecting point between customers and suppliers.

### 2.1. Implementation of E-procurement in companies

The buying process is vital for every organization; from this process depends success, customer satisfaction and good cost management. Good management of the supply network and the sales process and customer management are the most important goals compared to only attempt to maximize corporate profit; however it's necessary to understand that an excellent supply chain network and demand process management leads to a growth of corporate profit. The purchase process, such as customer management, has been influenced by new information technologies and in particular by the e-economy.

With the introduction of e-Procurement the procurement process is managed online; for this purpose companies use B2B platforms: this tool allows to reduce costs and time of the procurement process, improves warehouse management, inventory levels and the management of all business processes.

This is a necessary solution both for large companies, since it makes easier manage the entire process of purchasing and supply's growing network, and small companies that, adopting e-procurement solutions, take part of a global system with numerous business opportunities. In order to do this the must invest in two basic elements: an electronic catalogue and online payment procedure.

The e-Procurement process implementation involves several steps:

- the first step is to analyze business processes and understand which should be outsourced ("make or buy" policy). Purchase process and suppliers network management should be managed by qualified e-procurement platforms;
- the second step involves the identification of supplier network through an accurate supplier selection process;
- the third step is the definition of laws to be applied at supplier network, as well as suppliers classification and clustering;
- Finally considerations should be done about purchasing policy.

The whole process of e-Procurement is divided into two phases (Figure 2): the e-Sourcing phase (in which define technical and economical products specifications, send request for quotation to suppliers, collect and select best quotations) and the e-Supply Chain Management (in which the order is issued, supplier sends the merchandise that is checked and payment is processed).

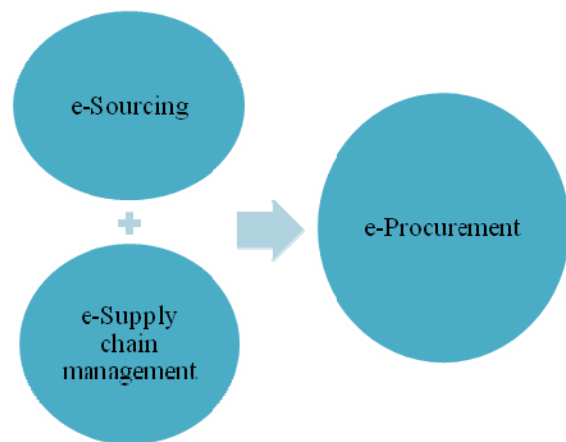


Figure 2: e-Procurement

### 3. E-SUPPLY CHAIN

The term e-Supply Chain is used to define a logistic chain managed in an integrated way thanks to new technological solutions; the effect is the processes and objectives planning with a view to business ethics in an integrated and sharing information way that involves the entire supply chain.

To understand in detail the features of e-Supply Chain, it's necessary analyze two aspects:

- The analysis of processes involved by new technologies (e-Supply Chain Execution and e-Supply Chain Collaboration);
- Company strategies for new technologies implementation.

As regards the first aspect, we can talk, as mentioned above, two types of processes:

- Operational processes (e-Supply Chain Execution): related to the order-delivery-invoice cycle and pre-sale and post-sale processes with the aim to make automated and integrated interface phases in supply chain logistics;
- Collaborative processes (e-Supply Chain Execution): related to planning processes (demand forecasting, inventory management, monitoring and control) as well as new product development and quality management, with the aim to improve the processes efficiency through the know-how sharing.

As regard the analysis of technological solutions, two choices are available:

- EDI (Electronic Data Interchange) solutions: allow data interchange of data between information systems, through a dedicated channel and in a specific format, they are designed ad hoc for company at issue and for the specific data to be exchanged;
- Internet EDI solutions: are more flexible than the previous one because they use internet for information exchange.

#### 3.1. E-supply chain Execution

The e-Supply Chain Collaboration basic stages are summarized in the figure below (Figure 3).

In each stages considered e-SC Execution considers the use of a specific tool. The benefits of e-solutions SC Execution implementation are:

- the chance to reduce operating costs of all interface activities between customer and supplier;
- the chance to improve the service level due to delivery times reduction and to greater accuracy spent in processes management.

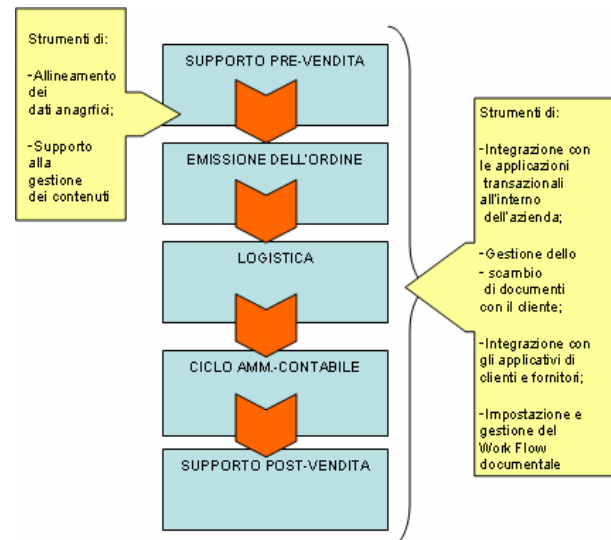


Figure 3: e-Supply chain Execution

#### 3.2. E-supply chain Collaboration

The basic stages of e-Supply Chain Collaboration can be found in the following:

- Supply Chain monitoring and control: involves information sharing (e.g. data sales and availability of production capacity), definition of performance evaluation metrics and the report of critical issues along the supply chain (e.g. stock-out);
- Collaborative management of planning activities: i.e. collaborative demand forecasting, promotions planning and inventory management. It is the application of Collaborative Planning Forecasting and Replenishment (CPFR) model;
- Collaborative management of new product development: provides the sharing of technical documents and project work-flow;
- Collaborative management of marketing communication strategies: involves the collaboration in research and development division as well as the customer relationship management.

An e-Supply Chain Collaboration implementation allows to manage in an integrated way all the phases just highlighted providing significant benefits.

#### 3.3. The benefits and barriers to development

Despite the benefits of collaborative solutions there are numerous barriers that prevent the e-Supply Chain Collaboration. e-Supply Chain Collaboration implementation benefits are longterm effects, arising from difficulties in create, understand and manage the relationship network along the supply chain. Another obstacle is related to the problem of translating the e-collaboration objectives in real company goals that may justify the investment in new technologies.

Furthermore management cultural barriers to change represents an obstacle to innovation.

Benefits are difficult to detect, i.e. is complicated to measure results as the growth of service level. In this paper the main benefits of this solution are highlighted: efficient and effective information management and material flows optimization along the supply chain.

#### 4. THE SIMULATION MODEL

The presented model simulates the evolution of hospital warehouse starting from order emission to the central pharmacy (main warehouse), until required materials arrival. The simulation will follow two phases:

- Definition of time needed for the development of procurement processes and their variability and introduction of events chain in order to obtain a virtual warehouse.
- Quantification of inventory and penalties costs due to a possible stock-out.

Two simulation models have been realized to reproduce the phases described above.

In order to reduce inventory costs materials should be available in warehouse exactly when necessary. In general pick orders from warehouse depend on patient treatment; so in case of change of the previously treatment a new order emission to the central pharmacy would be required.

The challenge is to avoid two common situations:

- Materials in stock before required (increasing inventory cost);
- Materials not available in warehouse (increase penalties cost due to stock-out).

The approach used for modeling relies on setting random upper and lower bounded variables to simulate the variability related to orders emission, demand and inventory level. The resulting model is reported below in Figure 4.

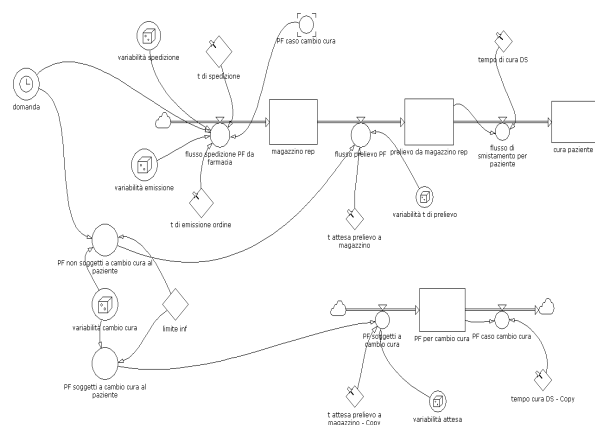


Figure 4: Stock and flow diagram

The second model simulates the total cost, sum of inventory and penalty cost, assessed according to the evolution of virtual warehouse realized in the previous simulation model. The cost diagram is represented in Figure 5.

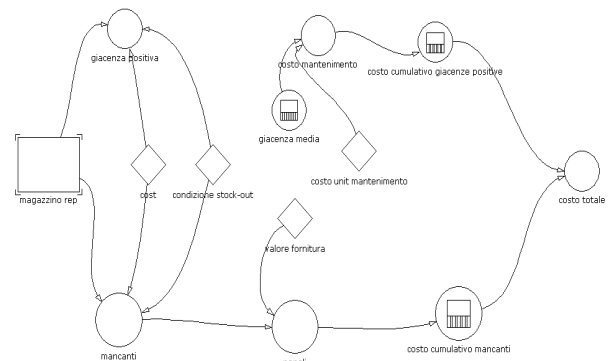


Figure 5: Stock and flow diagram (total cost)

Constants, levels and variables used in system dynamic model development are summarized in the following table (Table 1).

#### 4.1. Case Study: AORN “Antonio Cardarelli” di Napoli.

In order to validate the model a pharmacy department daily demand, derived from the average of Italian hospitals, has been considered and its trend has been represented in Figure 6.

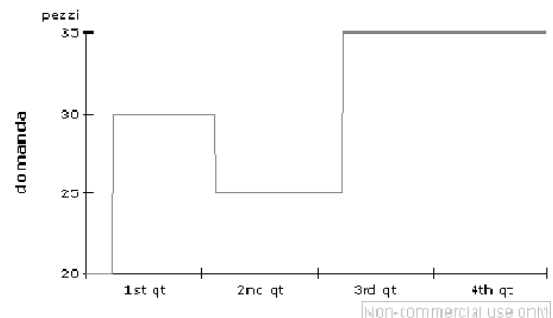


Figure 6: Demand trend

Simulation scenarios, characterized by an observation period of one year, are influenced by the trend of random variables:

- Order variability;
- Shipping time variability;
- Pickup time variability;
- Treatment change.

The warehouse inventory level simulation is based on three scenarios:

- Scenario 1: warehouse evolution and costs analysis without e-procurement;
- Scenario 2: warehouse evolution and costs analysis with e-procurement implementation and orders emitted by the central pharmacy;
- Scenario 3: warehouse evolution and costs analysis with e-procurement implementation and just in time pull strategy.

The following table summarizes parameters used for each simulation scenario (Table 2).

Table 1: Stock and flow diagrams parameters

<b>1<sup>ST</sup> SIMULATION MODEL</b>		
<b>NAME</b>	<b>DESCRIPTION</b>	<b>TYPE</b>
Demand	Products required by the warehouse department of the central pharmacy	Variable
Shipping time	Time required to submit the required product to the warehouse department of the central pharmacy	Constant
Shipping variability	Random variable that changes for each simulation time step	Random Variable
Processing time	Time required to send a product enquiry to the main warehouse (including bureaucratic stops)	Constant
Emission variability	Variable that allows to model the variability of orders emission	Random Variable
Inflow	Materials Flow coming from stock /In stock products	Variable
Warehouse	In stock products	Level
Outflow	Withdrawal rate of products from warehouse	Variable
Waiting time	Time constant that takes into account waiting time to withdraw the necessary materials from the warehouse department	Constant
Time of sampling variability	Variable for the modelling of the uncertainty	Random Variable
Withdrawal from warehouse department	Number of pieces taken from the store's department	Level
Flow clearinghouse for patient	Rate of sorted materials for patient care	Variable
Treatment time in DS	Time spent by patient receiving treatment in day surgery	Constant
Patient care	Materials actually used for the treatment of patient at issue	Level
Condition orders that are subject to treatment change	Condition for which occurs the emission of a new order due to a treatment change compared to the previously drawn .	Variable
PF in case of care exchange	Products must be ordered in case there is a treatment change and new order must be added to the input stream to the warehouse department	Variable
<b>2<sup>nd</sup> SIMULATION MODEL</b>		
<b>NAME</b>	<b>DESCRIPTION</b>	<b>TYPE</b>
Positive stock	Positive stock level	Variable
Stock out condition	Failure condition of the warehouse inventory	Constant
Missing	Negative stock evaluation	
Maintenance cost	Costs due to a positive inventory level.	Variable
Maintenance unit cost	Maintenance cost for a single product in the warehouse department	Constant
Maintenance cumulative cost	Total maintenance cost in the simulated time horizon	Variable
Penalty	Cost related to stock-out. Quantified according to a percentage of the supply request	Variable
Delivery value	Average value of the supply demand	Constant
Penalties cumulative cost	Total penalties cost in the simulated time horizon	Variable
Total cost	Sum of both maintenance and penalties costs calculated in the simulation horizon	Variable

Table 2: Simulation parameters

Parameter	1 <sup>st</sup> scenario	2 <sup>nd</sup> scenario	3 <sup>rd</sup> scenario
Demand	Step function	Step function	Step function
Total simulation time	1 year	1 year	1 year
Shipping time	24 h	24h	24 ore
Shipping variability	0,99-1,20	0,99-1,20	0,99-1,20
Emission time	3h	1 h	1 h
Emission variability	1,00-1,50	1,00-1,50	1,00-1,50
Waiting time	30 h	29 h	25 h
Variability	0,9-1,20	0,9-1,20	0,9-1,20
Treatment time in DS	24 h	24h	24 ore
Treatment change	0,00-1,00	0,00-1,00	0,00-1,00
Time necessary for treatment change in DS	30 h	29 h	25 h
Treatment change variability	0,15-1,00	0,15-1,00	0,15-1,00

#### 4.1.1. Simulation results

The following figures (Figure 7 and Figure 8) show the trend of both drugs inventory level and costs, obtained from an appropriate number of simulations.

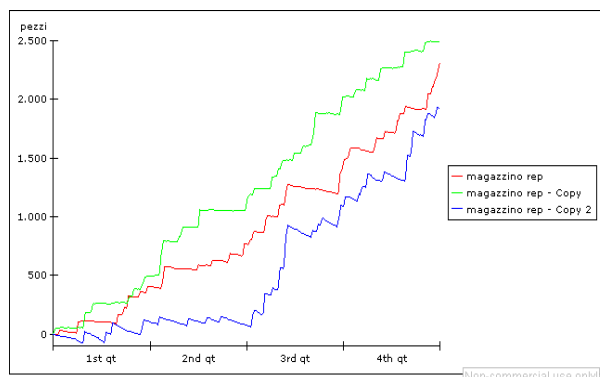


Figure 7: Inventory level trends: 1<sup>st</sup> scenario (green), 2<sup>nd</sup> scenario (red) and 3<sup>rd</sup> scenario (blue)

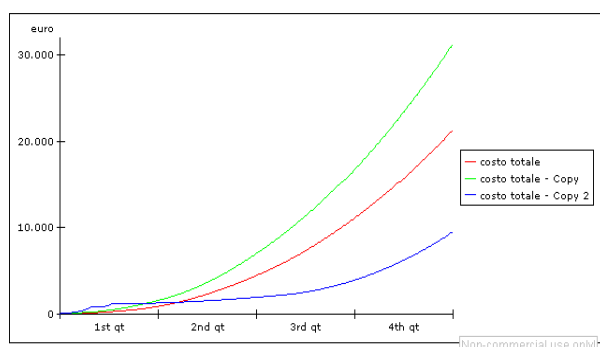


Figure 8: Total cost trend: 1<sup>st</sup> scenario (green), 2<sup>nd</sup> scenario (red) and 3<sup>rd</sup> scenario (blue)

The first and second scenario trends showed an higher inventory level that implies high maintenance costs, while in the third scenario the just in time strategy reduce the inventory level and related costs.

## 5. CONCLUSIONS AND FUTURE DEVELOPMENTS

The aim of the presented model is to show the benefits achievable in a hospital through the implementation of an innovative management model in order to standardize operations and reduce total costs.

An e-collaboration framework has been implemented in Day Surgery where patients spent approximately 24 hours; it is assumed that web-based platforms decrease orders emission time and improve relationships and interactions with provider at issue.

On these two fundamental issues a simulation model has been realized whose results are acceptable and have highlighted the chances to succeed. In the third scenario, based on just-in-time philosophy reached through a lean healthcare management process, the simulation results show that, the lean emission order process allows to achieve a great reduction of inventory cost.

It's important to highlight that the simulation model developed concerns a single hospital department; the model should be extended to all hospital departments in order to connect them with the central pharmacy as well as connect the pharmacy with its supplier and realize an e-collaboration network.

## REFERENCES

- Bhat, S., Gijo, E.V., Jnanesh, N.A., 2014. Application of Lean Six Sigma methodology in the registration process of a hospital. *International Journal of Productivity and Performance Management*, 63 (5), 613-643.
- Broekelmann J., 2006. Ambulatory surgery in Germany 2004 and historical aspects. *Ambulatory Surgery*, Volume 12, Issue 4, pp. 173-176, July 2006.
- Curatolo, N., Lamouri, S., Huet, J.-C., Rieutord, A., 2014. A critical analysis of Lean approach structuring in hospitals. *Business Process Management Journal*, 20 (3), 433-454.
- De Risi, P., 2002. Introduzione alla gestione per processi nelle organizzazioni, Università degli studi di Pisa.
- Habidin, N.F., Shazali, N.A., Ali, N., Khaidir, N.A., Jamaludin, N.H., 2014. Exploring lean healthcare practice and supply chain innovation for Malaysian healthcare industry. *International Journal of Business Excellence*, 7 (3), 394-410.
- Liker, J. K., 2003. *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*. McGraw-Hill.
- Rangone, A., Balocco, R., 2002. Le applicazioni B2b e il ruolo degli e-marketplace. *Mondo digitale*, 3, 2002.



# HEALTH WORKER MONITORING: KALMAN-BASED SOFTWARE DESIGN FOR FAULT ISOLATION IN HUMAN BREATHING

Improta Giovanni<sup>(a)</sup>, Natale Pasquale<sup>(b)</sup>, Santillo Libertina Carmela<sup>(c)</sup>, Triassi Maria<sup>(d)</sup>

<sup>(b),(c)</sup> D.I.C.Ma.P.I. - Department of Chemical, Materials and Industrial Production Engineering,  
University of Naples Federico II, P.le Tecchio, 80 – 80125 - Naples - Italy

<sup>(a),(d)</sup> Department of Public Health Medicine - University of Naples Federico II, Naples - Italy

<sup>(a)</sup>[giovanni.improta@unina.it](mailto:giovanni.improta@unina.it), <sup>(b)</sup>[pasquale.natale@unina.it](mailto:pasquale.natale@unina.it), <sup>(b)</sup>[santillo@unina.it](mailto:santillo@unina.it) <sup>(c)</sup>[triassi@unina.it](mailto:triassi@unina.it)

## ABSTRACT

Breathing is an involuntary action by the human body, and irregularities in breathing suggest possible respiratory disorders. This paper proposes a signal-noise separation algorithm for estimating the Oxygen partial pressure, an important respiratory parameter, by means of Kalman filtering. This technique is considered a powerful method to detect and isolate signal noise and it represents one of the most significant approaches for multisensory data fusion measurements. Our results show that this technique can effectively filter out sensor noise.

Keywords: computer science, software design, sensor information, sensors

## NOMENCLATURE

$\mathbf{x}$  is the state vector;  
 $\mathbf{u}$  is the input vector;  
 $\mathbf{w}$  is system noise (representing model inaccuracies), with zero mean, and known covariance  $\mathbf{Q}$ ;  
 $y \in \mathbf{R}$  is the system output;  
 $v \in \mathbf{R}$  is measurement noise with zero mean and covariance  $\mathbf{R}$ .  
 $y_i$  is the  $i$ -th sensor measure;  
 $\mathbf{R}_i$  is the  $i$ -th measurement noise covariance ;  
 $\mathbf{C}_i$  is the  $i$ -th measurement matrix.

## 1. INTRODUCTION

In the recent years different fusion schemes are introduced in many applications such as communication networks, sensor networks, automotive, aerospace as well as energy systems (Fiengo, Santini, and Glielmo 2008; Carotenuto, Iannelli, Manfredi, and Santini 2005; Brancati, Montanaro, Rocca, Santini, and Timpone 2013; Sarghini, de Felice, and Santini 2003; Manfredi 2014; Manfredi 2014; Valente, Montanaro, Tufo, Salvi, and Santini, 2014; Manfredi, Pagano, and Raimo 2012; Manfredi 2010).

One of the simplest and most intuitive general methods of measurement fusion is to take a weighted average of redundant information provided by multiple sensors and use this as the fused value (Qiang and

Harris 2001; Willisky, Bello, Castanon, Levy, and Verghese 1982; Sderstrm, Mossberg, and Hong 2009; Lee 2003). More complex approaches to measurement fusion use sensor information as well as model information for on line estimation. There are various multi-sensor data fusion approaches of which Kalman filtering is the most significant (Anderson and Moore 2006). Kalman filter is predominantly preferred because it results in estimates for the fused data that are optimal in a statistical sense (Simon 2006).

### 1.1. Medical application of method

Breathing is an involuntary action by the human body, and irregularities in breathing suggest possible respiratory disorders. Oxygen partial pressure is an important parameter for breath analysis and determines how much oxygen can be delivered from the lungs to the blood. This paper proposes a signal-noise separation algorithm for estimating the real Oxygen partial pressure.

The main limitations in the existing oxygen measurement methods suffer from either low sensitivity or low spatial or temporal resolution, or are invasive by nature.

In order to obtain more precise measurement of the parameter and to recognize and isolate the noise from the examined signal, this paper proposes the Kalman filtering as a powerful method to facilitate the estimation of the respiratory parameter. This filter estimates explicitly account for process and measurement noise inherent in clinical data making these estimates more informative than raw observation alone.

The Kalman filter has been recently used in different medical applications. Charleston and Azimi-Sadjadi (1996) suggest the implementation of the Kalman filter to solve the problem of the cancellation of the heart sounds in respiratory signals. Sameni, Shamsollahi, Jutten, and Babaie-Zadeh (2005) proposed the extended version of the Kalman filter for filtering noisy ECG signals and they succeeded in extracting fetal cardiac signals from maternal abdominal signals. In order to eliminate the motion artifact in InfraRed spectroscopy measurement Izzetoglu, Chitrapati, Bunce,

and Onaral (2010) developed a Kalman filter algorithm used in its discrete version.

Moreover, Kalman filtering is one of the most significant approach for multisensory data fusion measurements, which has several biomedical applications with especially regard to critical care monitoring and medical images. Tobergte, Pomarlan, and Hirzinger (200), for example, presented a sensor fusion for pose estimation using optical and inertial data based on extended Kalman filter.

## 2. PROBLEM STATEMENT

Given a generic nonlinear discrete-time process of the form:

$$\begin{cases} \mathbf{x}(k+1) &= f(\mathbf{x}(k), \mathbf{u}(k), \mathbf{w}(k)) \\ y(k) &= h(\mathbf{x}(k), \nu(k)) \end{cases} \quad (1)$$

where:

$\mathbf{x} \in \mathbb{R}^n$  is the state vector;

$\mathbf{u} \in \mathbb{R}^m$  is the input vector;

$\mathbf{w} \in \mathbb{R}^n$  is system noise (representing model inaccuracies), with zero mean, and known covariance  $\mathbf{Q}$ ;

$y \in \mathbb{R}$  is the system output;

$\nu \in \mathbb{R}$  is measurement noise with zero mean and covariance  $\mathbf{R}$ .

Assume that system (1) is observable in absence of disturbances  $w$  and  $\nu$ . In what follows we propose an adaptive measurement fusion strategy, based on Kalman filter, that, on the basis of residuals, adaptively estimates noise covariance of each sensor thus providing a simple fault tolerant data fusion algorithm. The proposed approach is able to cope with sudden and significative faults, like abrupt bias or jump, as well as with sensors slow derives. Further enahcement of the proposed approach will rely on more innovative and sopfisticated adaptive mechanisms.

## 3. ADAPTIVE MEASUREMENT FUSION SCHEME

To solve the above mentioned problem we follow a fusion scheme by exploiting the idea of covariance-matching (Manfredi and Santini 2013), where EKF is Extended Kalman filter given by the following coupled difference equations (Anderson and Moore 1979):

$$\hat{\mathbf{x}}_{k+1|k} = f(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k, 0) \quad (2a)$$

$$\mathbf{P}_{k+1|k} = \mathbf{A}_k \mathbf{P}_{k|k} \mathbf{A}_k^T + \mathbf{W}_k \mathbf{Q}_k \mathbf{W}_k^T \quad (2b)$$

being

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (y_k - h(\hat{\mathbf{x}}_{k|k-1}, 0))$$

$$(3a) \mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{\Omega}_k^{-1}$$

(3b)

$$\mathbf{\Omega}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{V}_k \mathbf{R}_k \mathbf{V}_k^T$$

$$(3c) \mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}$$

(3d)

Where

$$\begin{cases} \mathbf{A}_k &= \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\substack{x = \hat{\mathbf{x}}_{k|k} \\ u = \mathbf{u}_k}} \\ \mathbf{W}_k &= \left. \frac{\partial f}{\partial \mathbf{w}} \right|_{\substack{x = \hat{\mathbf{x}}_{k|k} \\ u = \mathbf{u}_k}} \\ \mathbf{H}_k &= \left. \frac{\partial h}{\partial \mathbf{x}} \right|_{x = \hat{\mathbf{x}}_{k|k-1}} \\ \mathbf{V}_k &= \left. \frac{\partial h}{\partial \nu} \right|_{x = \hat{\mathbf{x}}_{k|k-1}} \end{cases} \quad (4)$$

Let now  $y_i$  be the  $i$ -th sensor measure;  $\mathbf{R}_i$ :  $i$ -th measurement noise covariance and  $\mathbf{C}_i$  be the  $i$ -th measurement matrix. The Measurement Fusion block in Figure 1 provides the fused measurement information  $y$  as [1]:

$$y(k) = \left[ \sum_{j=1}^N \mathbf{R}_j^{-1}(k) \right]^{-1} \sum_{j=1}^N \mathbf{R}_j^{-1}(k) y_j(k), \quad (5a)$$

$$\mathbf{C}(k) = \left[ \sum_{j=1}^N \mathbf{R}_j^{-1}(k) \right]^{-1} \sum_{j=1}^N \mathbf{R}_j^{-1}(k) \mathbf{C}_j(k), \quad (5b)$$

$$\mathbf{R}(k) = \left[ \sum_{j=1}^N \mathbf{R}_j^{-1}(k) \right]^{-1}. \quad (5c)$$

Equations (5) have a clear physical meaning: the information of higher measurement noise covariance sensors ( $\mathbf{R}_i$ ) are weighted less than those of the lower measurement noise covariance sensors. Note that if all the  $N$  sensors have the same measurement noise covariance, equations (5) simply reduce to an arithmetic mean.

Measurement fusion works very well if noise statistics of each sensor are exactly known. Poor "a priori" information about noise statistics may degrade the Kalman filter performance. Moreover, even if we carefully know the noise statistics of each sensor, one of them could suddenly fault (wrong calibration, sudden jump, bias). Measurement fusion schemes, as in Fig.1, is not able to cope with these situations, while an adaptive scheme could recognize and recovery the fault. To this aim a considerable amount of research has been carried out in the adaptive Kalman filtering area (Simon 2006; Mehra 1970; Lee 2003), but in practice an "ad hoc" adaptive scheme has to be developed considering the specific problem characteristics.

Here the sensor fusion based on an adaptive Kalman filter is enhanced by using the covariance matching approach and employing the forgetting factor technique. In Kalman filtering theory, the difference between the current measurement and the predicted one is called innovation sequence, and it is, for an optimal filter, a Gaussian white noise sequence. For analogy with the linear case, residuals

$$c_i(k) = y_i(k) - h(\hat{x}(k|k-1), 0), i = 1, \dots, 3$$

are called the innovation process and are assumed to be well described by a white sequence  $N(0, C_i(k))$ . The innovation sequence plays an important role in the adaptive algorithm design and represents a flag of the actual estimation errors. The monitoring of the innovation process of each sensor can be utilized for fault detection.

The idea of covariance-matching techniques is to make the innovation samples consistent with their theoretical covariance. The theoretical covariance of the innovation sequence is (Simon 2006):

$$C_i(k) = E[c_i(k)c_i(k)^T] = \mathbf{H} \cdot \mathbf{P}(k|k-1) \cdot \mathbf{H}^T + \mathbf{R}_i \quad (6)$$

where the index  $i$  is referred to  $i$ -th sensor.

The actual covariance of  $c_i(k)$  can be approximated by its sampled covariance as:

$$\begin{aligned} \hat{C}_i(k) &= \frac{1}{k} \sum_{j=1}^k \mathbf{c}_i(j) \mathbf{c}_i(j)^T = \\ &= \frac{k-1}{k} \hat{C}_i(k-1) + \frac{1}{k} \mathbf{c}_i(k) \mathbf{c}_i(k)^T \end{aligned} \quad (7)$$

Assuming that (7) would be a good approximation of (6), we could estimate on line noise covariance by (6):

$$\hat{\mathbf{R}}_i(k) = \hat{C}_i(k) - \mathbf{H} \cdot \mathbf{P}(k|k-1) \cdot \mathbf{H}^T \quad (8)$$

We now use  $\hat{\mathbf{R}}_i$  estimation (8) to compute equations (5), thus achieving the following advantages:

1. sensor noise covariances has not to be "a priori" known;
2. if one of the sensors give wrong information it is less weighted in the measurement fusion since its noise covariance estimation is high.

To cope with slowly changing bias we further introduce a forgetting factor in the adaptive scheme. Indeed, estimation in (7) get worse with time  $k$ , thus not allowing the fault detection for slowly sensor faults. To overcome this drawback we introduce the forgetting factor  $\mu$  as:

$$\begin{aligned} \hat{C}_i(k) &= \frac{1}{k} \sum_{j=1}^k \mu^{k-j} \mathbf{c}_i(j) \mathbf{c}_i(j)^T = \\ &= \mu \frac{k-1}{k} \hat{C}_i(k-1) + \frac{1}{k} \mathbf{c}_i(k) \mathbf{c}_i(k)^T \end{aligned} \quad (9)$$

#### 4. VALIDATION OF THE APPROACH

The most common failure mode for sensors is failing to achieve the proper output for a given input signal, resulting in a lower than true reading. Some sensors fail suddenly, stopping all current production, while others give off a burst of energy before ceasing output. The most serious for safe-critical medical applications is the nonlinear sensor saturation to produce the exact output above a given level. This failure taken to extreme results is a sensor generating a fixed high-level output regardless of the physical value of the monitored variable. To avoid this problem we consider different measurements that are fused together to realize a more accurate estimation of oxygen partial pressure. In particular the considered scenario refer to three identical oxygen sensors for measure redundancy.

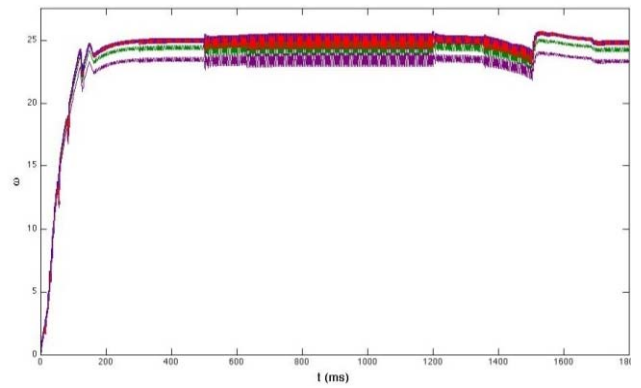


Figure 1: Multiple faults scenario. Oxygen partial pressure (red line) against actual value (blue line). sensor 1 (purple line); sensor 2 (dashed green line); sensor 3 (black line).

Figure 1 shows how the estimated variable (red line) is in perfect agreement with the actual oxygen level (blue line) despite the presence of nonlinear saturated sensors measurements (sensor 1 (purple line) and 2 (black line)).

#### 5. CONCLUSIONS

In this paper a Kalman filter based approach has been used to obtain more accurate estimation of oxygen partial pressure, due to its ability to precisely assess system states in the presence of substantial sensor noise. Our results show that this technique can effectively filter out sensor noise.

Because of the lack of a shared method for signal-noise separation in medical applications, along with the few works present in literature inspecting breath monitoring, it is difficult to compare the method presented in this paper with other techniques. Further studies, after accurate bibliographic researches, will aim to find the best procedure among those proposed in literature.

#### REFERENCES

- Anderson B. D. O., Moore, J. B., 1979. Optimal Filtering, Prentice-Hall, INC., 1979.

- Brancati, R., Montanaro, U., Rocca, E., Santini, S., Timpone, F., 2013. Analysis of bifurcations, rattle and chaos in a gear transmission system. *International Review of Mechanical Engineering* 7 (4), pp. 583-591, ISSN: 19708734, 2013.
- Caputo, G., Gallo, M., Guizzi, G., 2009. Optimization of production plan through simulation techniques. *WSEAS Transactions on Information Science and Applications*, 6 (3), pp. 352-362.
- Carotenuto, S., Iannelli, L., Manfredi, S., Santini, S., 2005. Sensor fusion by using a sliding observer for an underwater breathing system. *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference, CDC-ECC '05 2005*, art. no. 1583399, pp. 7662-7667, ISBN: 0780395689;978-078039568-8, DOI: 10.1109/CDC.2005.1583399, 2005
- Centobelli, P., Cerchione R., Converso, G., Murino, T., 2013. E-procurement and E-supply Chain: Features and Development. *International Conference on Future Software Engineering and Multimedia Engineering IERI Procedia* 6 (2014) 8-14
- Centobelli, P., Converso, G., Murino, T., Santillo, L.C., 2013. Banking operations and strategic marketing plans: A causal loop based model for finance sector. *DocumentInternational Review on Modelling and Simulations*
- Charleston, S., Azimi-Sadjadi, M. R., 1996. Reduced order Kalman filtering for the enhancement of respiratory sounds. *IEEE Transactions on Biomedical Engineering*, 43(4), pp. 421-424.
- Converso, G., De Vito, L., Santillo, L., 2013. *A simulation approach for sustainability of global supply chain network*, 25th European Modeling and Simulation Symposium, EMSS
- Fiengo, G., Santini, S., Glielmo, L., 2008. Emission reduction during TWC warm-up: Control synthesis and hardware-in-the-loop verification. *International Journal of Modelling, Identification and Control*, 3 (3), pp. 233-246, ISSN: 17466172, DOI: 10.1504/IJMIC.2008.020122, 2008.
- Gallo, M., Montella, D.R., Santillo, L.C., Silenzi, E., 2012. Optimization of a condition based maintenance based on costs and safety in a production line. *Frontiers in Artificial Intelligence and Applications*, vol. 246, pp. 457-474.
- Guizzi, G., Murino, T., Romano, E., 2012. An innovative approach to environmental issues: The growth of a green market modeled by system dynamics. *Frontiers in Artificial Intelligence and Applications*, 246, pp. 538-557.
- Izzetoglu, M., Chitrapu, P., Bunce, S., Onaral, B., 2010. Motion artifact cancellation in NIR spectroscopy using discrete Kalman filtering. *Biomed Eng Online*, 9(1), 16.
- Lee, T. G., 2003. Centralized kalman filter with adaptive measurement fusion: its application to a gps/sdins integration system with an additional sensor. *International Journal of Control, Automation and Systems*, vol. 1, no. 4, pp. 444-452, December 2003.
- Manfredi S., Santini, S., 2013. An adaptive measurement fusion scheme for drifted and biased sensors fault isolation. *International Review of Automatic Control*, 6 (5), pp. 552 – 557, ISSN:1746-6172, 2013.
- Manfredi, S., 2010. A reliable cooperative and distributed management for wireless industrial monitoring and control. *International Journal of Robust and Nonlinear Control*, 20 (2), pp. 123-139, doi: 10.1002/rnc.1453, 2010.
- Manfredi, S., 2014. A theoretical analysis of multi-hop consensus algorithms for wireless networks: Trade off among reliability, responsiveness and delay tolerance. *Ad Hoc Networks*, (PART A), pp. 234-244, ISSN:1570-8705, 2014.
- Manfredi, S., 2014. Decentralized queue balancing and differentiated service scheme based on cooperative control concept. *IEEE Transactions on Industrial Informatics*, 10 (1), pp. 586-593, 2014. doi: 10.1109/TII.2013.2265879.
- Manfredi, S., Pagano, M., Raimo, R., 2012. Ultracapacitor-based distributed energy resources to support time-varying smart-grid power flows. *SPEEDAM 2012 - 21st International Symposium on Power Electronics, Electrical Drives, Automation and Motion*, art. no. 6264610, pp. 1148-1153. ISBN: 978-146731299-8 doi: 10.1109/SPEEDAM.2012.6264610, 2012.
- Mehra, R. K., 1970. On the identification of variances and adaptive kalman filtering. *IEEE transaction on automatic control*, vol AC-15, no.2, pp. 175-184, 1970.
- Murino, T., De Carlini, R., Naviglio, G., 2012. An economic order policy assessment model based on a customized ahp. *Frontiers in Artificial Intelligence and Applications*, vol. 246, pp. 445 – 456.
- Murino, T., Naviglio, G., Romano, E., Zoppoli, P., 2009. Single stage multi product kanban system. Optimization and parametric analysis. *Proceedings of the 8th WSEAS, International Conference on System Science and Simulation in Engineering*, ICOSSE '09 pp. 313-318
- Qiang, G., Harris, C.J., 2001. Comparison of two measurement fusion methods for kalman-filter-based multisensory data fusion. *IEEE transaction on aerospace and electronic systems*, vol. 37, no. 1, pp. 273-280, 2001.
- Revetria, R., Catania, A., Cassettari, L., Guizzi, G., Romano, E., Murino, T., Improta, G., Fujita, H., 2012. Improving healthcare using cognitive computing based software: An application in emergency situation. *Advanced Research in Applied Artificial Intelligence, Lecture Notes in Computer Science*, Vol. 7345 LNAI, pp. 477-490.
- Revetria, R., Catania, A., Cassettari, L., Guizzi, G., Romano, E., Murino, T., Improta, G., Fujita, H., 2012. Improving healthcare using cognitive

computing based software: An application in emergency situation. *Advanced Research in Applied Artificial Intelligence, Lecture Notes in Computer Science* Vol. 7345 LNAI pp. 477 – 490.

- Sameni, R., Shamsollahi, M. B., Jutten, C., Babaie-Zadeh, M., 2005. Filtering noisy ECG signals using the extended Kalman filter based on a modified dynamic ECG model. *Computers in Cardiology*, pp. 1017-1020.
- Sarghini, F., de Felice, G., Santini, S., 2003. Neural networks based subgrid scale modeling in large eddy simulations. *Computers & Fluids*, 32 (1) pp. 97-108, ISSN 0045-7930, [http://dx.doi.org/10.1016/S0045-7930\(01\)00098-6](http://dx.doi.org/10.1016/S0045-7930(01)00098-6), 2003.
- Sderström, T., Mossberg, M., Hong, M., 2009. A covariance matching approach for identifying errors-in-variables systems. *Automatica*, vol. 45, no. 9, pp. 2018-2031, 2009.
- Simon, D., 2006. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*, Wiley, July 2006, ISBN: 978-0-471-70858-2.
- Tobergte, A., Pomarlan, M., Hirzinger, G., 2009. Robust multi sensor pose estimation for medical applications. *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on* (pp. 492-497). IEEE.
- Valente, A. S., Montanaro, U., Tufo, M., Salvi, A., Santini, S., 2014. Design of a Platoon Management Strategy and its Hardware-In-the Loop validation. Accepted for publication in *The Second International Workshop on Vehicular Traffic Management for Smart Cities*, 2014.
- Willsky, A. S., Bello, M. G., Castanon, D. A., Levy, B. C., Verghese, G. C., 1982. Combining and updating of local estimates and regional maps along sets of one-dimensional tracks. *IEEE transaction on automatic control*, vol. AC-27, no. 4, pp. 799-813, August 1982.

# THE INDUSTRIAL PLANTS RELOCATION:ISSUES, POLICIES, PROCEDURES AND ALGORITHMS FOR DISASSEMBLY AND REASSEMBLY PHASES

De Falco Massimo<sup>(a)</sup>, Gaita Pellegrino<sup>(b)</sup>, Ilsami Mario<sup>(c)</sup>, Santillo Libertina Carmela<sup>(d)</sup>,

<sup>(a)</sup> DIIN - Department of Industrial Engineering - University of Salerno- Italy

<sup>(b),(c),(d)</sup> D.I.C.Ma.P.I. - Department of Chemical, Materials and Industrial Production Engineering,  
University of Naples Federico II, P.le Tecchio, 80 – 80125 - Naples - Italy

<sup>(a)</sup> [mdefalco@unisa.it](mailto:mdefalco@unisa.it), <sup>(b)</sup> [pellegrino.gaita@unina.it](mailto:pellegrino.gaita@unina.it), <sup>(c)</sup> [mario.ilsami@unina.it](mailto:mario.ilsami@unina.it), <sup>(d)</sup> [santillo@unina.it](mailto:santillo@unina.it),

## ABSTRACT

This work proposes an analytical approach to optimize the transfer project of existing industrial plants in new productive locations. This means to transport in new sites the same functional characteristics that plants have in the original one. The impracticality to move existing plants, shifts the problem into define a transferring project, regarding constraints and opportunities related the execution of two strongly specular phases.

The aim is an algorithmic procedure which, selecting the main components of a plant, determines and optimizes the transfer global cost, as function of the parameters characterizing the three different project phases; disassembling, transporting and reassembling.

The core is constituted by an iterative structure which, starting from an analytical description plant's , integrates the loading and transport phases, estimating at the same time the best level of disassembly sequences and the loading efficiency and plant's reduction. Those evaluations allow to optimize the overall cost of the project.

Keywords: Algorithmic Modeling, Manufacturing, Logistics, Reassembly.

## 1. INTRODUCTION

Offshoring is the relocation by a company of a business process from one country to another.

Relocation of production, whereby a firm transfers all or part of its production facilities from one country to another, has become an important issue in industrial relations across Europe and especially in those western European countries that have seen a number of high-profile cases of outward relocation over recent years.

China and India have emerged as a prominent destinations for production offshoring.

The main reasons why companies resort to offshoring are:

- To increase efficiency;
- Because of the increased competition at the national level;
- to enhance image and prestige of the company.

The main benefits of offshoring are:

- Low costs and taxes;
- Time savings;
- The entry into new market areas.

However there are also risks related to the transfer of production abroad. First and foremost is the reduction in the level of employment and the loss of control and image quality.

In addition, others factors to consider are:

- Political climate in foreign countries;
- Differences in work practices and cultural barriers;
- Hidden costs.

The problem of plant transfer emerges when an already installed and operating plant has to undergo a change of location. The project of plant transfer can be divided in three sequential and temporally distinct phases:

1. Plant disassembly;
2. Transportation to the new production site;
3. Re-assembly.

The small components, obtained from disassembly, are loaded into containers and then they are transported by appropriate means of transport to the new site. Instead the singular transport is provided for large components. In particular small components determine significant problems that will be treated in this paper with a mathematic and algorithmic approach.

## 2. STATE OF ART FOR DISASSEMBLY

All Disassembly is defined as the process of separating products into parts and/or subassemblies with necessary inspections and sorting operations. Generally, disassembly is done with several purposes, i.e., recovering material fractions, isolating hazardous substances, separating reusable parts and/or subassemblies, etc. Note that disassembly is one of core operations in recovering or even disposing of used/end-of-life products.

Disassembly objectives include:

- Recovery of valuable parts or subassemblies,
- Retrieval of parts or subassemblies of discontinued products,



- Removal of hazardous parts,
- Increasing the purity of the remainder of the product,
- Achieving environmentally friendly manufacturing standards.

A first model has been formulated by Gupta and Gungor. It presents a methodology to obtain a near optimal DSP. A DSP is a sequence of disassembly tasks that begins with a product to be disassembled and terminates in a state where all the parts of the product are disconnected. The number of alternative DSPs increases exponentially with the number of parts in the product being disassembled. For example, a product with 10 parts may have up to a whopping  $10! = 3,328,800$  disassembly sequences. Even though the number of feasibility constraints may reduce the number of disassembly sequences, this number is still expected to be large for a complex product.

A second model has been formulated by three Korean engineers, named Han, Yu and Lee.

It considers the selective disassembly sequencing problem under the serial disassembly environment in which only one component is obtained at each disassembly operation. The problem is represented as a disassembly precedence graph and then a new integer programming model is suggested for the objective of minimizing the total disassembly cost. To solve the problem, we suggest a branch and fathoming heuristic.

The disassembly precedence graph, suggested by Lambert (2006), is adopted to represent disassembly operations and their precedence relations. More formally, the disassembly precedence graph can be represented by  $G = (N, A)$ , where  $N \setminus \{0\}$  is the set of nodes that represent disassembly operations, i.e.,  $N = \{0, 1, 2, \dots, n\}$  and  $A$  is the set of arcs that represent the precedence relations between two disassembly operations, i.e.,  $A = \{(i, j) \mid i, j \in N \text{ and } i < j\}$ . Also, node 0 denotes the root node representing the start of disassembly. Note that each disassembly operation is associated with a part or subassembly to be obtained from that operation since we consider the serial disassembly and the number in each node represents the corresponding part or subassembly.

The Container Loading problem (CLP) alludes to the task of packing boxes into containers. More precisely, given the dimensions of the containers and the boxes which need to be loaded, the problem can be defined as to find such an arrangement of boxes that optimizes a given objective function that, in general, is the maximum volume of the loaded boxes. In addition to the geometric constraints, other restrictions can also be considered, such as boxes orientation and cargo stability.

The container loading problem is to pack boxes into containers so as to meet certain objectives. Examples of objectives are:

(a) Minimize the length of the container required for a specified cargo.

(b) Maximize the volume of cargo packed into a container.

(c) Minimize the number of containers needed to pack a specified cargo.

(d) Find a way to pack all given cargo into a container.

### 2.1. George and Robinson's heuristic

J. A. George and D. F. Robinson (1980) present a heuristic for packing boxes into a container. It fills the container layer by layer across its width. A new layer is not commenced until all the previous layers are packed. Criteria are set to determine which box should be used to commence the next layer. When a box type is determined to commence the next layer, as many complete columns of boxes of this box type are filled into this layer. The unfilled space of this layer is cut into cubic spaces that are packed by boxes of other types. Then it will try to amalgamate gaps in this layer with the previous layers to see if it is possible to fill more boxes. One may re-order the priority of choosing the box type to reload the container to see if it is possible to fill the cargo specified.

### 2.2. The new heuristic

The new approach described in this paper is based on George and Robinson's heuristic. It is a "wall building" heuristic and the pack is performed along the depth direction. The container is open in the front and boxes are packed through this opening, from the back along its length. One of the modifications to the George and Robinson's heuristic is related with the container. We consider a finite length to the container. With this modification we can eliminate the unsuccessful packing and the automatic repacking procedures of the packing algorithm. Another modification concerns the packing in the end of the container. The George and Robinson's heuristic uses a minimal length parameter that prohibits the packing of new layers in the end of the packing process. This causes sub-optimal layers. To avoid this in the GRMod heuristic the layer depth dimension is dependent of the volume of unpacked boxes. This way in the end of the container layers could have small depth but with better volume utilization.

Following the GRASP (Greedy Randomised Adaptive Search Procedure) paradigm the approach discussed in this paper is divided in two different steps. In the first step solution is built and in the second step this solution is improved with a local search algorithm. In the construction phase the container is loaded until one of the following three conditions is met: there are no more free spaces in the container; there are no more boxes to be packed; or the dimensions of the remaining free spaces are smaller than the dimensions of the boxes still available to pack. Afterwards a local search phase is run to improve this solution. Alike George and Robinson's heuristic, this constructive heuristic deals with empty spaces in two different ways. If an empty space has the same height and width than the container, then this space is treated by the heuristic as a new layer. In this case the layer's depth dimension is defined by the depth

dimension of the type of box chosen to start the layer. Otherwise the space is treated like a free space. When a new layer is started the boxes are placed in vertical columns along the width of the container. In the other cases the algorithm tries to pack the boxes in the free spaces left by the boxes packed in the current layer and by previously built layers. When the unpacked boxes do not fit in the free spaces then those spaces are temporarily marked as “rejected”. The mark is only temporary because, if a new adjacent layer is built, this marked space can be amalgamated, as in the George and Robinson’s heuristic. The new heuristic is based on the “wall building” procedure. The container is filled with transversal walls and the depth of the layer is determined by the first box placed in the layer. A local search for the best box/orientation is performed to open a new layer and to fill a space. For all the boxes available this procedure computes the best arrangement when considering all possible orientations for each type of box. The best arrangement is found by simulating all the choices of boxes types and possible orientations and computing the correspondent volume utilization. If more than one arrangement yields the best volume utilization one of them is randomly chosen to become the definitive packing. After that the list of unpacked boxes and the list of free spaces are updated. The layer depth is equal to the box dimension placed along that direction. The number of boxes placed along the width and the height is limited by the container dimensions and the availability of that type of box. First the height of the container is filled as much as possible with an integer number of boxes and then these columns are replicated along the width of the container. An incomplete column is permitted. If there is some free space left between the layer and the container height or width, new spaces are generated (Figure 1) so that remaining boxes can be later on packed there.

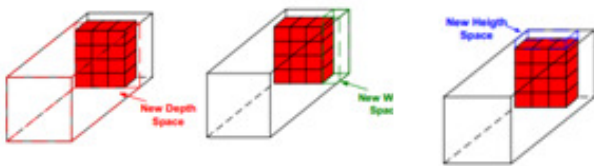
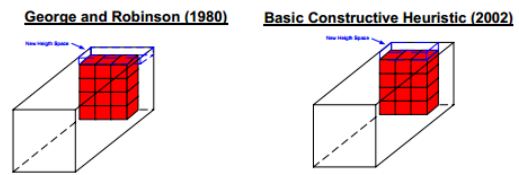


Figure 1. Example of space generation

Their generation follows a fixed order. The first space to be created is the depth space that corresponds to the frontal free space. This space is always created until the front of the container is reached. The next space to be generated is the width space and finally the height space is created. It should be noticed that, if the arrangement of boxes fits perfectly in the container along one of these dimensions, these spaces may have null dimensions, i.e. do not exist. In George and Robinson’s heuristic at the time when the width space is created if the dimension is smaller than the minimum box dimension then the new width space is not accepted. In this case the height space assumes the width of the original space (Figure 2).

Figure 2. Width of the original space

This results in no fully supported boxes. To improve the



cargo stability, in this situation the GRMod heuristic assumes that: If one of the dimensions of the newly created spaces is smaller than the smallest dimension of the boxes not yet packed, then this particular space is marked as “rejected”. This way the approach guarantees that all boxes are fully supported.

All the generated spaces are placed in a list of spaces in the order by which they are generated. Later on, when free spaces are considered to pack boxes, they are used following a first-in-last-out strategy, favouring the full support of the packed boxes and increasing the cargo stability.

When a new space is marked as “rejected” the algorithm tries to increase its size by amalgamating it with contiguous spaces belonging to the previous layer and also marked as “rejected”. By this procedure a new useful space may be generated (figure 3). If no amalgamation can be performed with spaces of the previous layer then the “rejected” space is kept in the list hoping that, in the next layer, any new “rejected” space can be amalgamated with it.

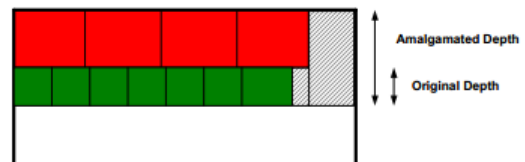


Figure 3. New useful spaces

It should be noticed that, by this process, an efficient and dense packing may be achieved. A direct consequence is that boxes with depth dimensions larger than the depth of the layer can now be packed there. This generates intersected walls.

As stated before, very small spaces may be rejected as they have not been amalgamated with contiguous spaces and cannot be used in the future to pack any boxes. To avoid this space fragmentation the original G&R heuristic proposed the concept of flexible width. This parameter bounds the number of columns that can be placed along the width in a new layer. Its value is propagated from the previous layer and is equal to the width of the arrangement of boxes that started the previous layer. For instance, if the previous layer was started with a box with a width of 30 cm and 4 columns were placed along the width, then the flexible width for the next layer would be 120 cm (Figure 4). While George and Robinson’s heuristic bounds the number of columns in the new layer by taking the smallest integer

that contains the flexible width, in the present algorithm the largest integer smallest than the flexible width will be taken. Taking the example presented in Figure 4, George and Robinson's heuristic would place an additional column in the new layer.

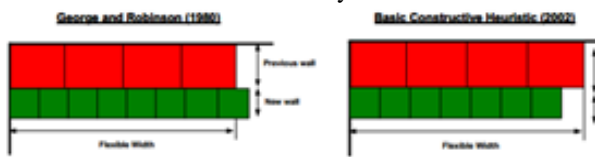


Figure 4. Flexible width for next layer

The construction of a layer ends by filling the free spaces that were generated in the first step of the layer construction. The first space to be filled is the height space, following the previously mentioned last-in-first-out strategy (the height space was the last one to be created). Only the boxes that have smaller dimensions than the space dimensions are considered. For each type of box, the procedure computes all possible arrangements (number of columns in depth and width directions and number of boxes per column) and selects the one that yields the best volume utilization. Then for all best volume utilization arrangements one is randomly chosen, following the GRASP paradigm, and the free space is filled with that box type and arrangement. When no feasible arrangement of boxes is found, this space is marked as "rejected". Then the algorithm tries to amalgamate this space with any other previously marked spaces.

After filling a space new depth, width and height spaces are generated, processed and inserted in the spaces list. The last one to be inserted will be the first one to be used. This filling spaces procedure is recursively applied until no more free spaces, different from the container front space, are available. In that case the new layer procedure is started applied to the container front space.

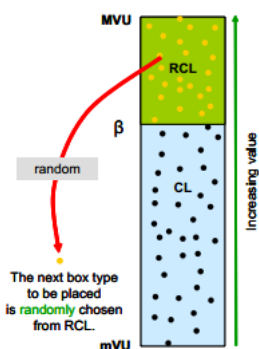


Figure 5. RCL list

Following the GRASP paradigm, randomization is used in this approach. In each iteration the choice of the next type of box to pack is made over a candidate list that contains the several alternatives of box types, ordered by volume utilization. A totally greedy strategy would lead to the choice of the best type of box (the first

element of the candidate list) and a completely random strategy would draw from the entire list.

However, a restricted candidate list (RCL) is built with the best candidates. Then a random choice is made from this RCL list (Figure 5).

The volume utilization tries to measure the benefit of selecting each type of box for a new layer or for a free space.

To define which candidates will belong to the RCL list a parameter  $\alpha$  is used, which will control the level of greediness of the algorithm. This parameter can vary between [0,1].

After computing the volume utilization for all candidates (types of boxes) the RCL list is filled according to the following threshold:

$$\beta = MVU + \alpha * (mVU - MVU)$$

where:

- $\beta$  is the volume utilization threshold;
- MVU is the maximum volume utilization computed for all possible arrangements;
- mVU is the minimum volume utilization computed for all possible arrangements;

If the volume utilization for one arrangement is bigger or equal to the  $\beta$  parameter, then the arrangement is added to the RCL. It is easy to see that when  $\alpha=1$ ,  $\beta$  is minimum and the basic heuristic is random; if  $\alpha=0$ ,  $\beta$  is maximum and the basic heuristic is greedy.

In the local search phase the algorithm starts with the solution built in the construction phase. Then a neighborhood of this solution is built. If a better solution is founded in the neighborhood, then it becomes the new current solution and a new neighborhood is built around this new better solution. The local search procedure stops when no better solutions in the neighborhood are found.

In order to build a neighborhood a disturbance to the solutions must be defined (figure 6).

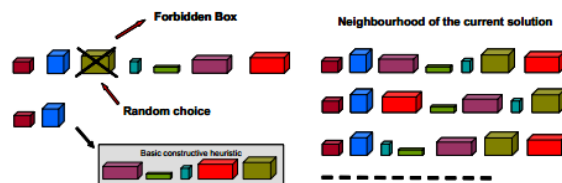


Figure 6. Example of defining disturbance

In this approach a position in the sequence by which the boxes were placed is randomly selected. Then all the boxes placed from that position until the end of the sequence are removed from the list of placed boxes and inserted in the list of unpacked boxes. The type of box that corresponds to the random position becomes "forbidden" and is temporary removed from both lists.

Then, for all boxes belonging to the unpacked boxes list the heuristic applies the constructive heuristic, but now without any randomness ( $\alpha=0$ ). After the packing the first type of boxes the "forbidden" type of box is reinserted in the list of unpacked boxes. By this the box type that previously occupied the disturbed position will not retake that place. The constructive heuristic, in its

greedy flavor, continues until no more boxes can be packed and a new solution is obtained.

### 3. REASSEMBLY

Reassembly is the last phase of the transport process of an industrial plant.

All major components, which were previously disassembled and then transported to the new site, are cleaned to remove any residue of paint, rust, dirt, grease, oils, deposits of coal, coke and other contaminants. Each component is thoroughly inspected and the company determines whether it can be recovered and reused, if it must be repaired to go back to proper working condition or if it is sufficiently worn out or obsolete and it must be replaced with a new component. All recovered, refurbished and new core components are finally reassembled to create a remanufactured product whose performance is equal to that of a new product. A remanufactured product is expected to match current factory specifications, meet or exceed the original expected lifetime, and may even need to comply with the latest engineering specifications and environmental requirements.

#### 3.1. Two types of reassembly

It is possible to distinguish two different types of reassembly:

- Symmetric reassembly;
- Asymmetrical reassembly.

In general, if our industrial plant is newly built and the aim is simply to transfer it to a new location so the reassembly phase is, in many instances, perfectly symmetrical with the disassembly phase. In other words, the reassembly sequence is exactly the reverse of the disassembly sequence. Instead, in the case of the transport of older generation systems, it is usually necessary to replace obsolete machines and components and to adopt through engineering studies a new and better layout configurations. In this case, therefore, the reassembly sequence may be asymmetrical with the disassembly sequence. The asymmetric reassembly is adopted for the following objectives:

- To optimize the space available;
- To speed up the transfer of material and/or persons from the various departments of the plant;
- To ensure, as far as possible, the proximity between departments where there is an important relationship.

### 4. THE PROPOSED MODEL

The proposed approach aims to overcome the traditional logic scheme, characterized by two sequential and independent phases: the separated determination of the disassembly sequence and the items loading into containers (Figure 7). A more deepened analysis would evidence that the cost connected the these two phases is function of the chosen level of plant disassembly, and that this will impact the disassembly and local re-assembly costs. Thus is evident that they would be lower if the plant remains assembled during the transport. The transport cost, on the contrary, would

decrease with the raise of the disassembly level; it is widely known in literature that a reduction of the average dimensions of the items produces a more effective loading, with possible optimization of the number of containers and then transportation cost. These considerations suggest the possibility to find an optimum disassembly level which minimizes the sum of disassembly, transport and re-assembly costs. Therefore, a new logical scheme for the problem approaching is configured as shown in Figure 1-b. The phases of disassembly and loading are no longer independent: a further connection "area" will take into account all the possible disassembly modalities and evaluates them on the overall generated costs basis.

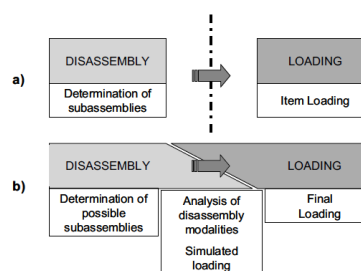


Figure 7. Determination of sequence

The structured procedure proposed for determining the optimum modality of plant transfer, according to the total cost minimization criterion, requires the following activities:

- analysis of the items subject to the procedure;
- disassembly/assembly sequences identification;
- identification of the optimum level and global costs calculation;
- determination of the Shipping List.

In Figure 2 the general flow chart of the procedure is illustrated. It is easy to notice the presence of an iterative block justified by the mutual dependence of the disassembly and loading/transportation phases.

Logistics Planning and Control Models

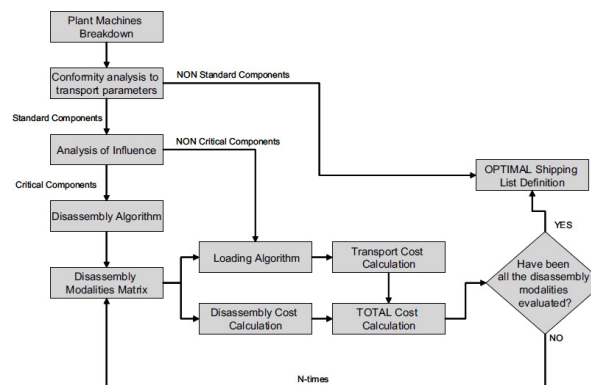


Figure 8. Logistics planning and control models

#### 4.1. The algorithm

Given the dependence of the transport costs from the disassembly phase, the algorithm starts determining all the possible disassembly modalities, loading modalities

and classifies them according to the resultant global cost.

In practice, the algorithm produces a  $M(N \times R)$  matrix, in which  $N$  is the number of possible disassembly modalities and  $R$  is the number of disassembly actions present in one modality at least. Will be:

- $M_{ij} = 1$  if the  $j$ th action occurs in the with disassembly modality;
- $M_{ij} = 0$  in the contrary case.

		Actions							
		1	2	...	j	j+1	j+2	...	R
Modalities	1	0	0	0	0	0	0	0	0
	...								
	i	1	1	0	1	0	0	0	0
	i+1								
	...								
	N								

Figure 9. Matrix of disassembly modalities

The first line corresponds to the fully-assembled component and therefore is characterized by the 0 value for every action. Once obtained the matrix, the procedure goes through two paths:

- calculation of the disassembly/re-assembly cost;
- implementation of the container loading algorithm and subsequent calculation of the transport cost.

The disassembly/re-assembly cost relative to the with modality  $C_{dis(i)}$  can be expressed as:

$$C_{DIS(i)} = (2 \cdot C_{man} \cdot T_{OP(i)} + C_{ext} \cdot T_{ext}) \cdot N_{wor} + C_{eq}$$

in which:

- $C_{man}$  = manpower hourly cost [€/h];
- $T_{OP(i)} = \sum_{j=1}^R M_{ij} \cdot T_j$  = operation time associated with the  $i$ th disassembly modality [h]; the factor 2 means that the same time is assumed necessary also for the reassembly phase;
- $T_j$  = time necessary for the  $j$ th action [h];
- $N_{wor}$  = number of necessary workers;
- $C_{ext}$  = travelling daily allowance [€/day];
- $T_{ext}$  = number of necessary transfer days.
- $C_{eq}$  = equipment cost.

The implementation of the container loading algorithm requires, for each disassembly modality, the list of components with relative dimensions. The result is the total number of containers necessary to contain the entire disassembled plant, which influences directly the transport cost.

At this stage will be possible to identify the disassembly/assembly level-sequence corresponding to the minimum total cost and, therefore, to define the optimal shipping list.

## 5. CONCLUSIONS

This thesis proposes a structured approach to industrial plant transfer problem; the main result is a analytical procedure which allows to correlate all possible

disassembling-loading modalities of the standard-critical components with the relative global transfer cost, identifying, then, the optimum. The procedure consists of an iterative cycle in which, for each disassembly modality, a run of the container loading algorithm is performed. The sum of the consequent disassembly and transport costs determines the best disassembly sequence for the industrial plant.

Future developments of the research will be targeted to an increase of the procedure efficiency. In fact the number  $N$  of the disassembly sequences, being the industrial plant a complex object, can be extremely large; this means that the algorithm of container loading has to be run  $N$  times with the risk of not acceptable computational times. An interesting starting point could be the definition of a method to pre-select the “more promising” disassembly sequences in order to reach the best solution in a limited number of interactions. The risk of achieving only a local optimum solution should be compared with the advantages of the computational time saving.

## REFERENCES

- Converso, G., De Vito, L., Santillo, L., 2013. A simulation approach for sustainability of global supply chain network, 25th European Modeling and Simulation Symposium, EMSS
- Converso, G., Gallo, M., Aveta, P., Santillo, L.C. Planning of supply chain risks in a make-to-stock context through a system dynamics approach, 2012
- De Falco M., S. Miranda, 2007. “Analytical model for the optimization of industrial plant transfer projects”, Proceedings of the 12th International Symposium on Logistics, 8-10 July 2007, Budapest.
- Hee-Jong Han, Jae-Min Yu, Dong-Ho Lee, 2012. “Selective Disassembly Sequencing with Sequence-dependent Setups for Multiple Target Components”, Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference 2012.
- Gallo, M., Murino, T., Pensato, V., Santillo, L.C., 2013. A proposal for an exact algorithm for solving a SALBP-1 problem, Proceedings - 19th ISSAT International Conference on Reliability and Quality in Design, RQD 2013, pp. 383-388.
- Gungor A., Gupta S.M., 1998. “Disassembly sequence planning for complete disassembly in product recovery”, Proceedings of the 1998 Northeast Decision Sciences Institute Conference, Boston, Massachusetts, March 25-27.
- Longo, F., Nicoletti, L., Chiurco, A., Document Cooperative training for ships and tugboats pilots based on interoperable simulation 2013 25th European Modeling and Simulation Symposium, EMSS 2013 pp. 695-703
- Moura A., José Fernando Oliveira, 2005. “A GRASP Approach to the Container Loading Problem”, IEEE Computer Society.



## Author's Index

Ablazevica,	557
AbouRizk,	153 191 325 334
Affenzeller	40 418
Aguilar Zaragoza,	238
Aizstrauts	357 523 536 557
Almeida,	220
Alonso-Vicario	402
Arana-Cuena,	469
Armenzoni,	543
Asmuss,	444
Astalos	384
Atanasijevic Kunc	254
Axelsson,	144
Baltruks,	523 557
Beham,	418
Blanco-Fernández,	576 582 588 620
Boenzi	504
Bogdan,	608
Borges,	402
Bottani,	543
Boutin,	112
Brozek,	46 451
Bruzzone,	649
Buil	357 536 557
Burkhardt,	523
Camañes,	517
Camci,	438
Capel Tuñón,	30
Capocchi,	563
Carotenuto	649
Carvajal-Fals	576 582
Castaño,	40
Castelo Branco	229
Cavalheiro,	24
Centobelli,	662
Cerbaso,	424 432
Çetinkaya,	438
Chacón,	582



Champliaud,	295
Chandrakash,	480
Chiurco,	649
Choi	184
Cirulis,	530
Converso,	657 662
Crespo-Pereira,	498
Crouseilles,	112
Cruz L.P.M.,	269
Cruz M. A.,	229
Cui,	171
da Cunha Santos,	229
da Cunha,	229 263 269
Danzer,	164
Darrigrand,	112
Davey	474 480 490
Davoli,	136
De Falco	673
De Felice	280 649
De Iasi	662
de Prada	56
D'Elia,	280
Di Bona	424
Divis,	307
Dorfer,	79
Douglas,	229
Dreija,	557
Drinovec	254
Duraccio	424 432
Ekyalimpa,	153
El Boukili	65
Elduque,	517
El-Laithy,	608
Escamilla-Serna	463
Faboya,	245
Fadda,	649
Falcone,	432 649
Fancello,	649
Faria,	220

Feng,	295
Fernandes,	263
Ferreiro-Cabello,	626
Fikejz,	371
Flores	87 120 238 570
Forcina,	424 432
Fraile-García,	626
Franch	517
Frascio,	649
Gabriel,	164
Gago	451
Gaita	673
Gallo M.	657
Gallo S.A.	136 199
García,	613
García-Alcaraz,	594
Ginters	357 364 523 530 536 557
Glasa,	384
Glavan,	392
Gomes	212
Govoni,	136
Gradisar,	392
Guadalupe Serna-Díaz,	469
Guimarans	570
Guo,	17
Gusev	557
Gutierrez	523
Hague,	191
Halada,	384
Harris,	175
Hathurusingha	490
Hennerbichler,	79
Hernandez-Romero	463 469
Hou	171
Ilsami	673
Improta	668
Jacak,	93
Jacob,	212 220
Jakes,	451

Javierre,	517
Jimenez-Gonzalez,	469
Jiménez-Macias	576 588 594 601 608 620
Joseph,	334
Jungles,	24
Justo,	212
Kalapyshina,	106 128
Karas	72
Karder,	418
Kartak,	343
Kavicka	46 307 319 343
Kerr,	334
Kim	184
Kocyan,	301
Kommenda,	418
Kraujalis,	98
Kronberger,	40 418
Kwon	184
Lakhdar	378
Lamas-Rodriguez,	498
Landau,	263 269
Lara de Paz,	120
Latorre-Biel,	588 594 601 608
Lauks	444
Leça Coelho	220
Leiva-Lázaro,	620
León-Samaniego,	613
Li B.	171
Li D.	334
Li H.,	17
Li W.	289
Liu T.	313
Liu Y.,	1
Liu Z.	289
Loimayr,	79
Longo	136 199 641 649
Lopes,	263 269
Lopez-Rojas,	144
Luengo,	40

Mahé,	112
Mariano,	280
Martin	402
Martínez-Cámara	582 588 620 626
Martinovic,	301
Mazaeda,	56
Medina-Marin,	463 469
Melechovsky	410
Melloni,	136 199
Mercado-Flores,	469
Merrett	175
Mittal	632
Mohamed	191 325 334
Montanari,	543
Mosconi,	199
Mujica Mota	570
Murino,	136 657 662
Musalimov,	106
Music,	392
Muzitano Reis	229
Muzy,	550
Natale	668
Nazemi,	523
Newstead,	191
Nicoletti,	641
Nutaro	350
Nuzhdin,	106 128
Oberbauer,	164
Obritzberger,	79
Oliveira J. F.,	212
O'Neill	480
Onggo,	46
Orive,	402
Ou,	171
Padovano	641
Penzes,	319
Perechesova,	106 128
Pérez-Parte	588 594 601 613
Petre,	334

Petrillo	280 649
Piera Eroles	357 536 557
Pilkauskas	98
Pioletti,	112
Podhoranyi	301
Pranevicius H.,	98
Pranevicius M.,	98
Pranevicius O.,	98
Pröll	93
Rakotomanana	112
Ramos,	212
Raska,	205
Ren Lei,	171
Ren Li,	17
Rezanina	371
Rinaldi,	543
Rios Prado	498
Rodrigues,	212
Rossainz López,	30
Rossetti,	220
Ruscino	280
Sáenz-Díez,	588 594 613
Sakne,	557
Sánchez,	87
Sanchez-Orozco,	576
Sanchez-Roca,	576 582
Santatriniaina,	112
Santillo	668 673
Santucci	563
Schaller,	40 79 164
Scheibenpflug,	418
Seck-Tuoh-Mora	463 469
Sentocnik,	254
Shen,	1
Siebers,	245
Siemann Santos Pereira	325
Silvestri,	424 432
Sklenar,	11
Snipas,	98

Soares Freire,	24
Solari	543
Song,	1
Sonntagbauer,	523
Sonntagbaure,	523
Sorrosal,	402
Stepucev,	557
Strmcnik	392
Tao,	1
Tejerina,	56
Tellez-Jurado,	469
Triassi	668
Ulrych	205
Velkoski	557
Vena,	641
Ville,	563
Voracek,	319
Wagner,	418
Wang,	313
Weinberger,	164
Weisenpacher,	384
Wendt	457
Wiedemann,	457
Winkler	40 79 164
Wu	313
Yang	289
Zake,	364
Zamoruev	106
Zanoni,	649
Zavanella,	649
Zeigler	350 550
Zennir	378
Zhang L.,	1
Zhang X.,	1
Zia,	175
Zizlavsky,	72
Zou J.	191
Zou W.	474