11th INTERNATIONAL INDUSTRIAL SIMULATION CONFERENCE 2013

ISC'2013

EDITED BY

Veronique Limère

and

El-Houssaine Aghezzaf

MAY 22-24, 2013

GHENT, BELGIUM

A Publication of EUROSIS-ETI

11th Industrial Simulation Conference 2013

GHENT, BELGIUM

MAY 22-24, 2013

Organised by

ETI- The European Technology Institute

Sponsored by

EUROSIS - The European Simulation Society

In Cooperation
With

Ghent University

Co-Sponsored by

EK

Incontrol Simulation Solutions

LMS International

University of Skovde

UPV

Hosted by

Ghent University

Department of Industrial Management

Ghent, Belgium

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

EDITORS

General Conference Chairs

Veronique Limère, Ghent University, Ghent, Belgium
El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium

Local Programme Committee

Hendrik Van Landeghem, Ghent University, Ghent, Belgium
Stijn De Vuyst, Ghent University, Ghent, Belgium
Birger Raa, Ghent University, Ghent, Belgium
Karel Bauters, Ghent University, Ghent, Belgium
Derek Verleye, Ghent University, Ghent, Belgium
Cédric Verbeeck, Ghent University, Ghent, Belgium

Post Conference Chair

Vladimír Janoušek, Brno University of Technology, Brno, Czech Republic

Publications Chair

Yan Luo, NIST, Gaithersburg, USA

INTERNATIONAL PROGRAMME COMMITTEE

Discrete Simulation Methodology, Languages and Tools
Matthias Becker, University Hannover, Hannover, Germany
Jonathan Berglund, Chalmers University of Technology, Gothenburg, Sweden
Luis M.S.Dias, University of Minho, Braga, Portugal
Helge Hagenauer, Universitaet Salzburg, Salzburg, Austria
Sophie Hennequin, ENIM, Metz Cedex, France
Mauro Iacono, SUN II, Universita di Napoli, Naples, Italy
Bjorn Johansson, Chalmers University of Technology, Gothenburg, Sweden.
Sarka Kvetonova, Brno University of Technology, Brno, Czech Republic
Erik Lindskog, Chalmers University of Technology, Gothenburg, Sweden
Stefano Marrone, Seconda Universita di Napoli. Naples, Italy
Paulo Novais, Universidade do Minho, Braga, Portugal
Jiri Safarik, University of West Bohemia, Plzen, Czech Republic
Antonella Zanzi, Universita' degli Studi dell'Insubria, Varese, Italy

Ambient Intelligence and Simulation
Selwyn Piramuthu, University of Florida, Gainesville, USA

Simulation in Manufacturing
Pascal Berruet, Universite Bretagne Sud, Lorient, France
Peter Byrne, Dublin City University, Dublin, Ireland
Ana Camacho, UNED, Madrid, Spain
Luis Camarinha-Matos, Universidade Nova de Lisboa, Lisbon, Portugal
Eduardo Castellano, IKERLAN Technol. Res. Centre, Mondragon-Arrasate, Spain
Remy Dupas, Universite Bordeaux 1, Bordeaux, France
Alexander Felfernig, University of Klagensfurt, Klagensfurt, Austria
Michel Gourgand, Universite Blaise Pascal, Clermont-Ferrand, France
Imed Kacem, Université Paul Verlaine Metz, Metz, France

INTERNATIONAL PROGRAMME COMMITTEE

Carlo Meloni, Politecnico di Bari - DEE, Bari, Italy
Pascal Meyer, Forschungszentrum Karlsruhe, Karlsruhe, Germany .
José A V Oliveira, University of Minho, Braga, Portugal
Marina Valles, UPV, Valencia, Spain
Markus Vorderwinkler, PROFACTOR GmbH, Steyr, Austria
Roland Wischnewski, (RIF) e.V., Dortmund, Germany
Farouk Yalaoui, Universite de Technologie de Troyes, Troyes Cedex, France

Simulation in Steel Manufacturing
Brian Hollocks, Bournemouth University, Bournemouth, United Kingdom

Simulation in Automotive Systems
Naoufel Cheikhrouhou, EPFL, Lausanne, Switzerland
Julien Richert, Daimler AG, GR/PAA, Sindelfingen, Böblingen, Germany
Rob Walker, Anglia Ruskin University, Chelmsford, United Kingdom

Simulation in Robotics
Leopoldo Arnesto Angel, UPV, Valencia, Spain
A. Chatzinikolaou, Athens, Greece
Andrzej Dzielinski, Warsaw University of Technology, Warsaw, Poland
Markus Koch, Orga Systems GmbH, Paderborn, Germany
George L.Kovacs, Hungarian Academy of Sciences, Budapest, Hungary
Alexandru Marin, Polytechnical Inst.of Bucharest, Bucharest, Romania
Martin Mellado, UPV, Valencia, Spain
Bogdan Raducanu, Computer Vision Centre, UAB, Barcelona, Spain
Krzysztof Skrzypczyk, Silesian University of Technology, Gliwice, Poland
Sergiu-Dan Stan, Technical University of Cluj-Napoca, Cluj-Napoca, Romania

Simulation in Electronics, Computer and Telecommunications
Teresa Alvarez, University of Valladolid, Valladolid Spain
Christos Bouras, University of Patras, Patras, Greece
Tom Dhaene, Ghent University, INTEC, Ghent, Belgium
Silvia Mirri, University of Bologna, Bologna, Italy
Maurizio Palesi,Universita di Catania, Catania, Italy
Marco Roccetti, University of Bologna, Bologna, Italy
Fernando Boronat Segui, UPV, Valencia, Spain
Renate Sitte, Griffith University, Gold Coast, Australia

Simulation in Electronics Manufacturing
Theresa Roeder, San Francisco State University, USA
Gerald Weigert, Dresden University of Technology, Germany

Simulation of Complex Multiprocessor Systems
Orhan Gemikonakli, Middlesex University, London, United Kingdom

Simulation in Computer Science
Ernst Kesseler, NLR, Amsterdam, The Netherlands
Wolfgang Kreutzer, University of Canterbury, Christchurch, New Zealand

Simulation in Logistics, Transport and Harbour Simulation
El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium .
Christian Almeder, European University Viadrina, Frankfurt (Oder), Germany
Maria Sameiro Carvalho, University of Minho, Guimaraes, Portugal
Isabel Garcia Guttierez, Univ. Carlos III de Madrid, Madrid, Spain .
Thomas Hanne, Univ. of Applied Sciences Northwestern Switzerland, Olten, Switzerland .
Peter Lawrence, Swinburne University, Lilydale, Australia
Marie-Ange Manier, UTBM, Belfort, France
Herve Manier, UTBM, Belfort, France
Roberto Montemanni, IDSIA, Manno-Lugano, Switzerland

INTERNATIONAL PROGRAMME COMMITTEE

Jaap Ottjes, TU Delft, Delft, The Netherlands
Roberto Razzoli, University of Genova, Genova, Italy
Rosaldo Rossetti, University of Porto, Porto, Portugal
Rik van Landeghem, Ghent University, Ghent, Belgium
Hans Veeke, TU Delft, Delft, The Netherlands
Pengjun Zheng, University of Southampton, Southampton, United Kingdom

Hospital Logistics Simulation
Track Chair: Giorgio Romanin-Jacur, University of Padova, Vicenza, Italy
Jose Machado, University of Minho, Braga, Portugal
Peter Summons, University of Newcastle, Australia

Complex Systems Modelling
Track Chair: Igor N Litvine, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa
Frantisek Capkovic, Slovak Academy of Sciences, Bratislava, Slovak Republic
Eugene Kindler, Ostrava University, Ostrava, Czech Republic
Alexandre Nketsa, LAAS, Toulouse, France
Michal Pechoucek, Czech Technical University, Prague, Czech Republic
Miguel Rocha, University do Minho, Braga, Portugal
Alfonso Urquia, UNED, Madrid, Spain

Simulation in Aerospace
Reza Azadegan, Urmia University, Urmia, Iran
Wolfgang Kuehn, University of Wuppertal, Wuppertal, Germany
Martin Spieck, DLR, Goettingen, Germany

Marine Simulation
Sergeij Kalashnkikow, DANFOSS, Austria

Simulation in Industrial Design and Product Design
Sergio Baragetti, University of Bergamo, Dalmine, Italy
Chiara Catalano, IMATI-CNR, Genoa, Italy
Yan Luo, NIST, Gaithersburg, USA
Jorge Perez Mateos, AOS, Madrid, Spain
Catarina Rizzi, University of Bergamo, Bergamo, Italy

Simulation in Engineering Processes
Chrissanti Angeli, Technological Institute of Piraeus, Athens, Greece
Alejandra Gomez Padilla, University of Guadalajara, Mexico
Jan Studzinski, Polish Academy of Sciences, Warsaw, Poland
Joao Tavares, University of Porto, Porto, Portugal

Civil and Building Engineering
Alistair Borthwick, Oxford University, Oxford, United Kingdom
Graham Saunders, Loughborough University, Loughborough, United Kingdom

Simulation in Energy and Power Systems
Manuel Cordeiro, CETAV/UTAD, Vila Real, Portugal .
Sergeij Kalashnkikow, DANFOSS, Austria
Janos-Sebestyen Janosy, KFKI Atomic Energy Research Institute, Budapest, Hungary

Simulation in Multibody Systems
Track Chair: Bernd Schaefer, DLR, Wessling, Germany
Ignacio Garcia-Fernandez, University of Valencia, Valencia, Spain

Simulation in Chemical, Petroleum and Mining Engineering
Diganta Das, Loughborough University, Loughborough, United Kingdom
Mohamad R. Riazi, Kuwait University, Kuwait

INTERNATIONAL PROGRAMME COMMITTEE

Simulation in Military and Defense
Roberto de Beauclair Seixas, IMPA, Rio de Janeiro, Brazil
Carlos Palau, UPV, Valencia, Spain
Matthias Reuter, CUTEC Gmbh, TU-Clausthal, Clausthal, Germany

Virtual Reality and Graphical Simulation in Industrial Simulation
Track Chair: Guodong Shao, NIST, Gaithersburg, USA
Emilio Camahort, Universidad Politecnica de Valencia, Valencia, Spain
Anders Hast, University of Gavle, Gavle, Sweden
Fabrizio Lamberti, Politecnico di Torino, Turin, Italy
Sudhir Mudur, Concordia University, Montreal, Canada
Marcos A.Rodrigues, Sheffield Hallam University, Sheffield, United Kingdom

Simulation of Complex Multiprocessor Systems
Peter Kvasnica, Alexander Dubcek University of Trencin, Trencin, Slovakia

Verification, Validation and Accreditation
Mokhtar Beldjehem, University of Halifax, Canada
Agustin Yague, Technical University of Madrid, Madrid, Spain

Simulation and Training
Manuel Alfonseca, Universidad Autonoma de Madrid, Madrid, Spain
Wenji Mao, Chinese Academy of Sciences, Beijing, P.R. China
Eshan Rajabally, Loughborough University, Loughborough, United Kingdom
Gerhard Schreck, Fraunhofer IPK, Berlin, Germany

Workshops

Workshop on Intelligent Transport Systems
Track Chair: Anna Syberfeldt, University of Skovde, Skovde, Sweden
Paul Davidsson, Blekinge Institute of Technology, Ronneby, Sweden
Petr Hanacek, Brno University of Technology, Brno, Czech Republic
Jairo Montoya Torres, Universidad de la Sabana, Chia, Columbia

NANOSIM
Clemens Heitzinger, Cambridge University, Cambridge, United Kingdom
Yong K. Kim, University of Massachusetts Dartmouth, Dartmouth, USA
Javier Marin, ETSI, University of Malaga, Malaga, Spain

Workshop Augmented Reality and Pervasive Systems in Simulation
Alessandro Genco, University of Palermo, Palermo, Italy

Workshop on Simulation in Lean Manufacturing
Track Chair: Leif Pehrsson, Volvo Car Corporation, Skovde, Sweden
El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium
Hendrik Van Landeghem, Ghent University, Ghent, Belgium

Simulation-based Optimization in Industry
Amos H.C. Ng, University of Skovde, Skovde, Sweden
Anna Syberfeldt, University of Skovde, Skovde, Sweden

OPENSIM
Philippe Geril, ETI Bvba, Ostend, Belgium

INDUSTRIAL SIMULATION 2013

© 2013 EUROSIS-ETI

Responsibility for the accuracy of all statements in each peer-referenced paper rests solely with the author(s). Statements are not necessarily representative of nor endorsed by the European Simulation Society. Permission is granted to photocopy portions of the publication for personal use and for the use of students providing credit is given to the conference 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 and the conference.

All author contact information provided in this Proceedings falls under the European Privacy Law and may not be used in any form, written or electronic, without the written permission of the author and the publisher. Infringements of any of the above rights will be liable to prosecution under Belgian civil or criminal law.

All articles published in these Proceedings have been peer reviewed

EUROSIS-ETI Publications are ISI-Thomson and IET referenced
Legal Repository: Koninklijke Bibliotheek van België, Keizerslaan 4, 1000 Brussels, Belgium
CIP 12.620 D/2011/12.620/1

Selected papers of this conference are published in scientific journals.

For permission to publish a complete paper write EUROSIS, c/o Philippe Geril, ETI Executive Director, Greenbridge NV, Ghent University Ostend Campus, Wetenschapspark 1, Plassendale 1, B-8400 Ostend, Belgium.

EUROSIS is a Division of ETI Bvba, The European Technology Institute, Torhoutsesteenweg 162, Box 4, B-8400 Ostend, Belgium

Printed and bound in Belgium by Reproduct NV, Ghent, Belgium
Cover Design by Grafisch Bedrijf Lammaing, Ostend, Belgium

EUROSIS-ETI Publication
ISBN: 978-90-77381-76-2
EAN: **978-90-77381-76-2**

PREFACE

Dear Participants,

On behalf of the Organizing Committee, we would like to welcome each of you to the 11th Annual Industrial Simulation Conference (ISC 2013), hosted by the Department of Industrial Management of Ghent University, in Belgium from May 22nd-24th, 2013.

During two and a half days, we offer a very high-level scientific program, spanning a wide array of topics in simulation, from theory to applications, over a large spectrum of industrial and service domains. Three keynote speakers, from academia and industry, will be sharing their professional expertise.

To encourage the exchange of techniques and ideas, which is the basic premise at the heart of the Industrial Simulation Conference, the timetable is organized to allow for ample opportunity to meet and discuss. We are convinced that this will permit you to elaborate on some issues and go home with a collection of bright new ideas for your future research.

We are indebted to all of you, without whom, the conference would not be possible. First of all, thanks go to all reviewers for critically evaluating the papers and keeping an eye on the scientific quality of the conference.

Moreover, we are grateful to the keynote speakers for accepting our invitation and for presenting their latest contemplations on significant simulation themes.

Furthermore, we would like to express our thanks to Mr. Philippe Geril, Secretary General of EUROSIS, who has taken the responsibility for most of the organizational matters. The Industrial Simulation Conference is only possible thanks to his hard work and never-ending enthusiasm.

Finally, our warm thanks go to all of you for submitting your research results, attending sessions and actively participating and discussing. We wish you all a stimulating and productive meeting.

General Conference Chairs
Veronique Limère
El-Houssaine Aghezzaf

CONTENTS

Preface ..... **XI**
Scientific Programme **1**
Author Listing **245**

KEYNOTE

Knowledge Simulation and Return on Education in the Logistics Industry
Matthias Klumpp **5**

MANAGEMENT SIMULATION

Three Eight Dimensional Correlation Studies
William Conley **13**

Simulation Gaming and Natural Language Processing for Modelling Stakeholder Behavior in Energy Investments
A. Ittoo, N.B. Szirbik G.B. Huitema and J.C. Wortmann **23**

Hybrid Simulation of Mobile Business Processes
Stefan Hellfeld **28**

How Two Management Engineers and a Simulation Software can support the Organization of a New Surgical Pre-Admission Unit
Giorgio Romanin-Jacur and Federico Segala **34**

WASTEWATER MANAGEMENT SIMULATION

Planning Revitalization Sewerage Systems using Decision Rules of Fuzzy Logic
Agnieszka Służalec and Andrzej Ziółkowski **39**

Using Optimization Approaches for Modeling, Control and Management of Water Networks
Jan Studzinski **44**

SIMULATION IN ENGINEERING

Development and Evaluation of an Order-N Formulation for Multi-Flexible Body Space System
Tushar Ghosh and Leslie J.Quiocho **51**

CONTENTS

Industrial Simulation in Fluid Power Systems by SIMULINK and AMESIM
Daniela Vasiliu, Georgiana Claudia Vasiliu, Nicolae Vasiliu and
Petru-Cristinel Irimia **59**

Thermal Modeling of a Battery Package used on Road Vehicle Powertrain
Valerian Croitorescu, Jan Anthonis and Nicolae Vasiliu **67**

TURBINE SIMULATION

Numerical Simulation and Experimental Validation of the transients from High Head Hydropower Plants
Radu Puhalschi, Marius Mitroi, Constantin Călinoiu and Nicolae Vasiliu **75**

Lifetime Evaluation of the Mechanical Elements for Hydraulic Turbines by Numerical Simulation
Constantin Drăgoi, Vlad Florin Pîrăianu and Petru-Cristinel Irimia.....**83**

Modelling Building Effects on Atmospheric Dispersion Comparison between ADMS and ANSYS CFD
Claudia Gheorghe (Nicolici), Ilie Prisecaru, Stefan Nicolici and
Radu M.Bilegan **88**

ELECTRONIC NETWORKS SIMULATION

Improving Performance Prediction of Parallel and Distributed Discrete Event Simulation: A Rough Sets Based Approach
László Muka and István Derka.....**95**

Testing The Speed up of Parallel Discrete Event Simulation in Heterogeneous Execution Environments
Gábor Lencse and István Derka **101**

Algorithms for Accelerated Rare Event Simulation with Markov Chain Modelling in Communication Networks
Izabella Lokshina and Evan Schiele **108**

Supervising Cathodic Protected Nets with CI-based Methods
Matthias Reuter.....**114**

Time Response of Operating Systems simulated by Load Balancing in Gridsim
Peter Kvasnica, Igor Kvasnica and Rozmarína Dubovská.....**119**

CONTENTS

MANUFACTURING SIMULATION

Simulation of a Winding Production Unit for the Supply Automation of Bobbins
Brahmadeep and Sébastien Thomassey **127**

Impact of Industrial Sorting on various Pistachio Categories in link with Aflatoxin Contamination thanks to a Monte Carlo Simulation
Nathalie Wesolek, Alejandra Ramirez-Martinez and Alain-Claude Roudot **132**

PRODUCTION OPTIMIZATION

Using Shewhart Control Charts for Continuous Model Optimization
Rodrigo Rezende Amaral, Peter Ottoy and Veronique Limère **139**

Database-Driven 3D Simulation: a Method Specification using the UML Metamodel
Martin Hoppen, Michael Schluse and Juergen Rossmann **147**

Simulation System for Linear Cutting Plans
Ovidiu-Dorin Alupei-Cojocariu, Nicolae Vasiliu and Szilard Feher **155**

Automotive Production Location Cost Optimization with Productivity and Sustainability Constraints
Matthias Klumpp and Viktor Kuchshaus **160**

AGV SIMULATION

Coupling of a Discrete Event Multi-Agent Simulation with a Real Time Multi-Agent Simulation
Christoph Schwarz, Jurij Schachmanow, Jürgen Sauer, Ludger Overmeyer and Georg Ullmann **165**

Decision-Making in Conceptual AGV Systems Design using Simulation Based Optimization
Leif Pehrsson and Amos H.C. Ng **171**

A Simulation Model using AMR Sensors Applied to Vehicles Classification and Identification
Ramón V. Cirilo Gimeno, Marta Pla Castells, Antonío García Celda and Javier Martínez Plumé **177**

CONTENTS

SUPPLY CHAIN SIMULATION

Sectioned Walking Worker Lines with Loop Balancing
Leif Pehrsson, Amos H.C. Ng, Jacob Bernedixen and David Stockton **187**

Generic Simulation Model for Assembly Line Supply
Veronique Limère, Kurt Decock and El-Houssaine Aghezzaf **192**

A Scalable Agent Based Decision Support System: Case Manufacturing Supply Chain
Lauri Lättilä and Samuli Kortelainen **198**

Estimating the Cost Structure of Wood Chip Supply Chains: A Simulation Approach
Lauri Lättilä, Kalle Karttunen, Olli-Jussi Korpinen, Jarno Föhr and Tapio Ranta **203**

Multivariate Inventory and Selected Problems using Multi-Stage Monte Carlo Optimization
William Conley **208**

LOGISTICS SIMULATION

A Framework for Discrete Event Simulation of Pick Up and Delivery Problems
Anna Syberfeldt **219**

Simulation Optimization of Waste Collection Based on Ultrasonic Level Measurements
Anna Syberfeldt **227**

TRAFFIC SIMULATION

Analysis of Possible Speedup of the Microscopic Road Traffic Simulation using Aggregate Movement of Vehicles
Tomas Potuzak **235**

Microscopic Simulation of Multimodal Urban Traffic Network
Ivana Ćavar, Marjana Petrović and Miroslav Vujić **240**

SCIENTIFIC PROGRAMME

KEYNOTE

Knowledge Simulation and Return on Education in the Logistics Industry

Matthias Klumpp
Institute for Logistics and Service Management
FOM University of Applied Sciences
Leimkugelstraße 6, D-45141 Essen, Germany
E-mail: matthias.klumpp@fom-ild.de, www.fom-ild.de

KEYWORDS
Knowledge Management, Berufswertigkeit, Logistics Education, Return on Education, DEA

ABSTRACT
With 2.8 million employees the logistics industry is the third largest in Germany. Challenged with increasing technology and sustainability requirements, especially the knowledge base of these employees is crucial for further success and value creation in this special industry. The contribution outlines simulation concepts based on a 2012 field survey regarding competences of 1.068 employees in the German logistics industry. This enables companies within the industry also to calculated a prognosis regarding possible returns of education with specific education measures as e.g. trainings and study programs for their employees.

PROBLEM DESCRIPTION
Many industries are facing increasing knowledge levels and requirements as for example the health care, financial services, automotive and electrical industries as well as the logistics industry (Klumpp/Krol/Zelewski 2006; Lin et al. 2011; Roth 2012; Klumpp/Clausen/ten Hompel 2013). This is the reason why several publications and quthors are also using the terms "knowledge economy" or "knowledge society" (Böhme/Stehr 1986; Kim/Mauborgne 1998; David/Foray 2002; Dunning 2002).
This development will even increase in terms of economic impacts in countries with the challenge of sharp demographic changes like in Europe or Japan (Klumpp/Bioly/Abidi 2013). Therefore the business practice question of an *impact evaluation of knowledge and education management* plays an important role in such industries, namely the logistics industry. Until today research regarding education impact has mainly focused on the *return on education investments* – also mainly on the individual personal level (see figure 1).

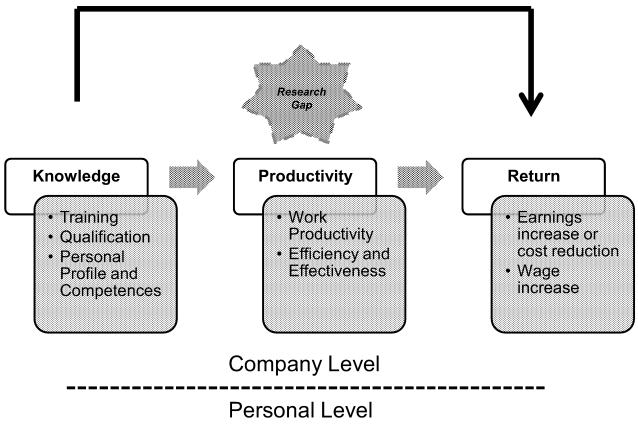


Figure 1: Knowledge Acquisition and Return on Education

As outlined in figure 1, the existing research results may leave a causal research gap in the question of transmission of increased knowledge into an increased return – and this also on the individual level of one person as well as on the company level of corporate processes and performances.

In order to contribute to the closure of this research gap this article proposes two interdependent steps in establishing a quantitative relation between knowledge and return on this knowledge or knowledge increase (education and training) by first (a) modeling competence evaluation and measurement – this is exemplified with the measurement concept of "Berufswertigkeit", applied in a large-scale survey in the German logistics industry in 2012 (Klumpp et al. 2013). And second (b) a competence impact measurement and simulation – this is ecxemplified here with a DEA measurement concept for training inputs and logistics performance outputs. Both steps are depicted in figure 2 below. With this two-step approach a further evaluation and also management concepts can be established in order to enhance the knowledge and education management results in the logistics industry. This concept can also be transferred to other knowledge-intensive industries.

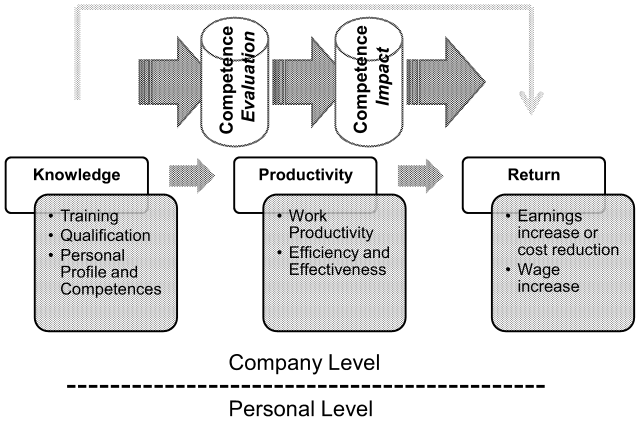


Figure 2: Modelling and Simulation Concept Return on Education

EXISTING RESEARCH
Starting point for many research concepts regarding the impact and value of education was the systems level and the question of a return on education in the increasing number and share of higher education in the form of university graduates in all modern societies in the last five decades (Teichler 2007, Teichler 2009).
Further evaluations have then concentrated on the individual return on any education investment by one person, mainly on higher education as this has been seen as the main individual

decision investing further in training after a minimum training level (scholl or vocational training levels depending on the specific society and economy) has been completed (Carroll/Ihnen 1967; Blaug 1972; MacMahon/Wagner 1981). Further analysis has been applied towards specific countries and further influencing sub-factors as e.g. gender and social beackground as determining factors (Ilon 1992; Judson 1998; Walton/Graves/Sexton 2005). Further research studies moving in the direction of modeling and simulation have established analytic and prognostic of these connections (e.g. Camgoz-Akdaga/Zaimb 2012; Walker/Florea 2013).
All these research results are directed towards the direct link of education and training towards the economic return of such measures (cp. Figure 1 and 2).

KNOWLEDGE MANAGEMENT AND COMPETENCE EVALUATION IN THE LOGISTICS INDUSTRY

The establishment of knowledge management concepts is an important research focus within logistics research today as through technological change and global developments the knowledge requirements for effective and efficient logistics processes are rising continuously (cp. Klumpp et al. 2011; Meyer/Reidel/Schmidt 2012; Kowalski/Zelewski/Bergenrodt 2012; Klumpp/Bioly/Zelewski 2013; Klumpp/Kandel/ Wirsing 2013; Robles/Wei/Noche 2013). The following figure 3 depicts an exemplary knowledge management circle that can be applied to logistics companies.

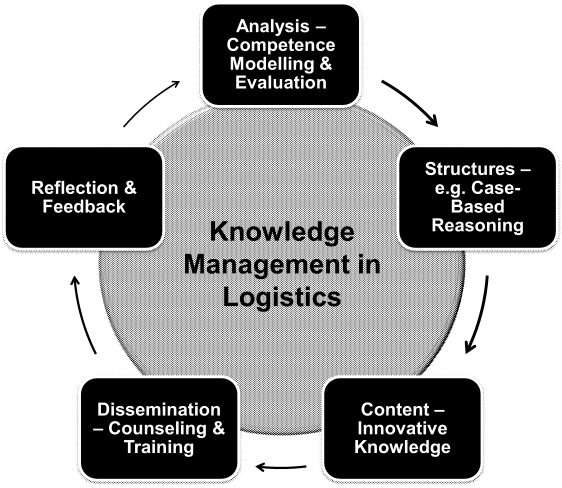


Figure 3: Knowledge Management Cycle in Logistics

In order to contribute to the before outlined research gap this mainly qualitative approach has to be complemented with a quantitative evaluation concept. Since 2007, for exactly this purpose the output-oriented measuring concept "Berufswertigkeit" has been developed and serves also as an industry evaluation concept for the development of an European Qualifications Framework for the the logistics industry (EU ESCO reference group "transportation and storage") and integrates the necessary investigation of competences (Klumpp and Schaumann, 2007). The 36 evaluation criteria of this concept are listed within figure 4 below.
Since the end of June 2011 and into 2012 a field survey with 1,068 persons from the logistics industry within the German research project "Wissenschaftliche Weiterbildung in der Logistik" (EffizienzCluster LogistikRuhr) was started. There are three types of survey instruments which could be executed such as written form, telephone and internet. Due to the experiences in the two Berufswertigkeit studies of 2007 and 2009 the project consortium decided for a telephone survey. The survey was executed in North Rhine-Westphalia and Hesse as both states depict a very good representation of the whole country of Germany (economic and population structures). Since January 2011 the survey was in preparation by development of the survey instrument, addresses of the logistic industries in both states, distribution of qualification in the logistic in-dustry, questionnaire, pretests and selection of the institute for questioning. The telephone survey of more than 1,000 persons had been begun on the end of June and took place till September 2012. 808 persons in North-Rhine Westphalia (NRW) and 259 persons in Hesse with different education levels were questionned.
Therein existing skills and competences of persons in the logistics industry will be described and the above mentioned draft for an industry qualifications framework logistics will be reworked. Also traditional formal degrees in vocational and aca-demic education will be classified according to evaluated practical competence levels. In total 1,068 persons from the logistics industry were questioned and these represent 379 female and 689 male respondents.

Altogether 75.7% of the respondents are from North Rhine-Westphalia and the remaining 24.3% are from Hesse. The respondents represent various professional levels at their current working place. 88.6% are employed as white-collar worker in the different levels like branch manager, team leader and office clerk in their company and 11.4% work in the warehouse or as truck drivers, namely blue-collar workers. The figure 2 shows the Berufswertigkeitsindex which significantly presents that 80% of the competences of 21.8% blue-collar as well as white-collar workers in Hesse are higher evaluated than blue-collar as well as white-collar workers in North-Rhine Westphalia, hereby it is to assume that the Ger-man federal state of Hesse indicates a very specialized logistics area with "hotspots" such as e.g. Frankfurt Airport.
The following figure 4 to 6 show results from this survey in the specific shape of group distribution functions, allowing for a comparison of groups of different sizes as on the y-axis percentage shares of the group total in percent are depicted regarding the different levels of people meeting the "Berufswertigkeitsindex" (BWI) competence requirement criteria (36 criteria normalized onto an 100 percent scale).
With this application it can be shown that *specific characteristics* such as the school degree and therefore a specific amount of education have *specific, positive effects* on the level at wich people meet the Berufswertigkeit criteria (figure 5). This can be interpreted as competence level regarding specific groups or persons. Also in a further analysis non-formal education aspects as for example the traning-on-the-job effect ("experience") can be measured as shown in figure 6 – where persons with more business experience state higher levels in the BWI measurement concept than others. This could enable companies even to simulate the prognostic competences of employees in the future or after specific education and training measures (prognosis on BWI increase in the future).

1. Efficiency
2. Independence and own initiative
3. Flexibility and adaptability
4. Work virtues
5. Stress resistance
6. Motivation and ability to lifelong learning and maintain to own competence profile
7. Coordinate the work- and lifetimes
8. Creativity
9. Loyalty
10. Risk-taking
11. Charisma
12. Ability to write and speak in German
13. Knowledge of a foreign language
14. Staff requirements and staff mission planning / staff development
15. Team, staff and leadership
16. Improving responsible care
17. Legal knowledge
18. Ability to apply modern information- and communication technologies (work place)
19. Communication and rhetoric
20. Assertiveness
21. International and intercultural competence
22. Costumer focus
23. Skills in mathematics and statistics
24. Preparation of cost estimates and quotations
25. Planning, implementation and documentation of orders and projects
26. Negotiations capacity
27. Analytical problem-oriented work
28. Quality management (optimization of processes and products or service quality)
29. Conceptual and strategic implementation of industry-specific knowledge and experience
30. Identification with the company
31. Strategic orientation, determine / control the complete company
32. Understanding of solutions for complex technical problems
33. Basic knowledge of business administration
34. Perception of functions of management and organization
35. Conceptual working in immediate workplace
36. Planning, control procurement and logistics processes

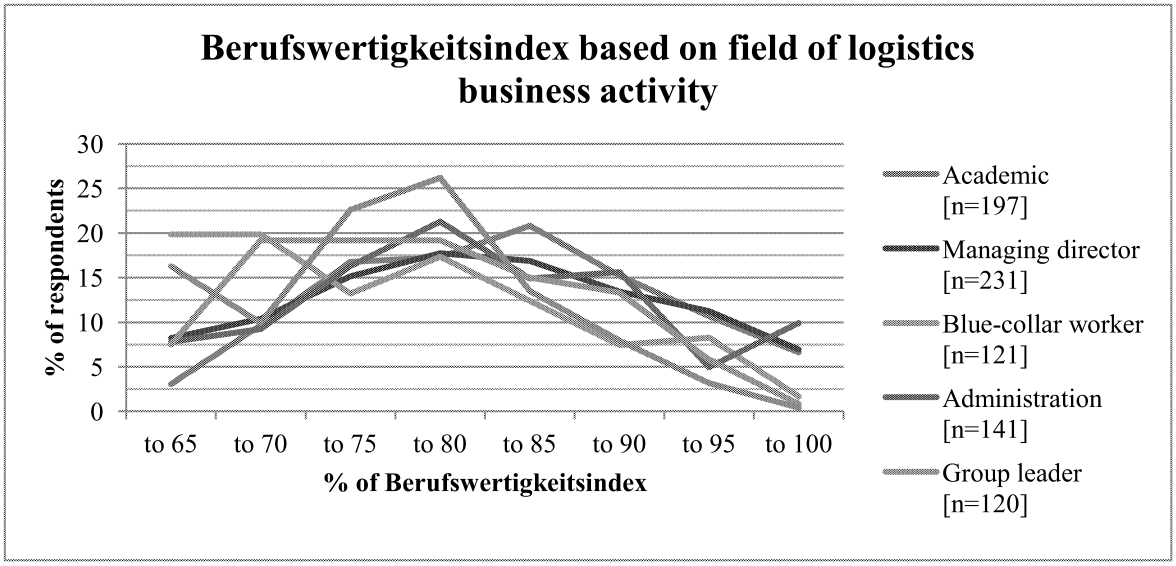


Figure 4: BWI Criteria and Total BWI Distribution among all Persons

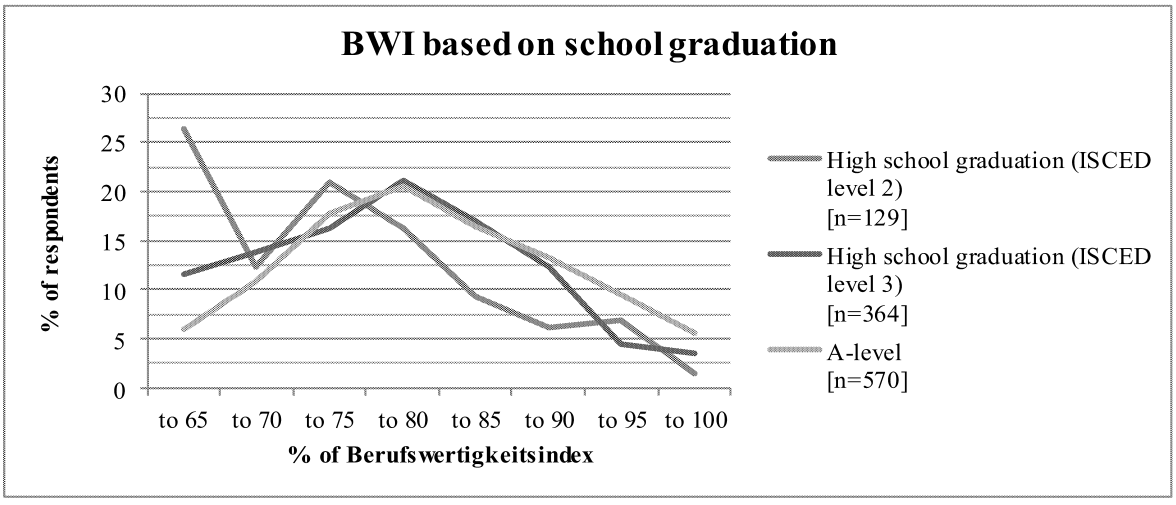


Figure 5: Total BWI Distribution According to School Degree

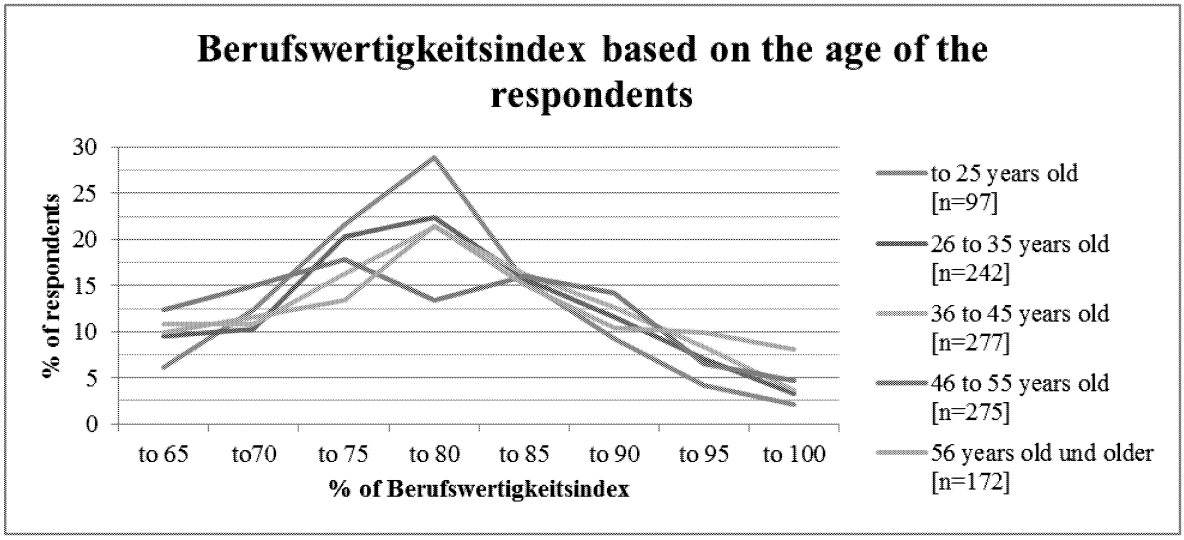


Figure 6: Total BWI Distribution According to Age

COMPETENCE IMPACT EVALUATION: DEA

For an evaluation of the competence impact – especially in the logistics industry a data envelopment analysis (DEA; Charnes/Cooper/Rhodes 1978) could be applied, e.g. with the input indicators training hours (or better: competence levels) and the output indicators logistics performance (tonkilmetres, timeliness, quality of service etc.).

The connectivity of such a modelling setup has to be checked in terms of existing research e.g. proof of impact of training hours towards competence leves on the one hand and also prrof of impact of logistics performance indicators towards business performance (earnings, profit).

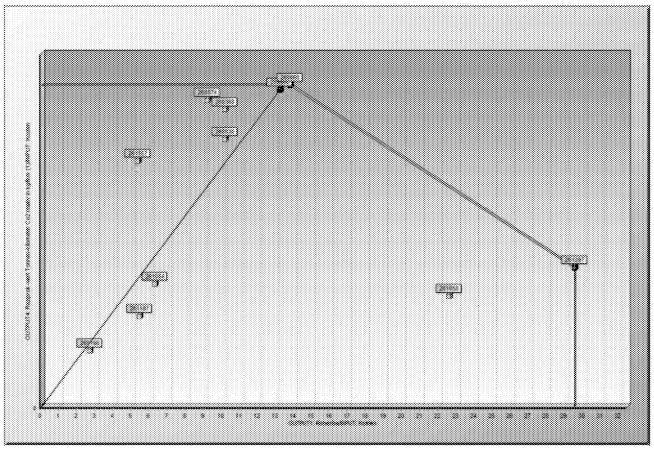


Figure 7: Example DEA for Knowledge Productivity in Logistics (Input Traing Costs, Outputs TKM Earnings and Service Level)

CONCLUSIONS

This approach towards knowledge modelling and simulation in the logistics industry has shown an opening for further research that could contribute to a large extend towards business improvement and impact in terms of increased effectiveness and efficiency. This should also enhance the today largely qualitative discourses in management and research regarding the role of knowledge and education.

Further research should be applied with company data from the logistics and also other industries to broaden the knowledge and insight in this important research and business management field.

Acknowledgement:This research is connected to the national excellence research cluster EffizienzCluster LogistikRuhr (www.effizienzcluster.de), funded by the German Federal Ministry of Education and Research (BMBF) in the project funding line 01|C10L19D.

REFERENCES

Blaug, M. (1972), "An Introduction to the Economics of Education", Penguin, Baltimore and Middlesex.

Böhme, G., Stehr, N. (1986), "The Knowledge Society", Sociology of the Sciences Yearbook 1986, Reidel Publishing, Dordrecht.

Camgoz-Akdaga, H., Zaimb, S. (2012), "Education: a comparative structural equation modeling study", in: Procedia - Social and Behavioral Sciences, Vol. 47, pp. 874- 880.

Carroll, A.B., Ihnen, L.A. (1967), "Costs and Returns For Two Years of Postsecondary Technical Schooling: A Pilot Study," in: Journal of Political Economy, Vol. 75, pp. 862-873.

Charnes, A., Cooper, W.W., Rhodes, E. (1978), "Measuring the efficiency of decision making units", in: European Journal of Operational Research, Vol. 2, No. 6, pp. 429-444.

David, P.A., Foray, D. (2002), "An introduction to the economy of the knowledge society", in: International Social Science Journal, Vol. 54, No. 171, pp. 9-23.

Dunning, J.H. (2002), "Regions, Globalization, and the Knowledge-Based Economy", Oxford University Press, Oxford.

Ilon, L. (1992), "Fitting Girls Schooling into Existing Economic Paradigms: Confronting the Complexities", in: Educational Development, Vol. 12, No. 2, pp.147-159.

Judson, R. (1998): "Economic Growth and Investment in Education: How Allocation Matters", in: Journal of Economic Growth, Vol. 3, pp. 337-359.

Kim, W.C., Mauborgne, R. (1998), "Procedural justice, strategic decision making, and the knowledge economy", in: Strategic Management Journal, Vol. 19, No. 4, pp. 323-338.

Klumpp, M. (2013): "Herausforderung Logistikqualifikation: Berufswertigkeitsanalyse und Industrie-Qualifikations-rahmen Logistik", in: Wolf-Kluthausen, H. (Hrsg.): Jahrbuch Logistik 2013, Korschenbroich, www.jahrbuchlogistik.de, pp. 106-109.

Klumpp, M., Abidi, H., Krol, B., Stender, T., Bioly, S. (2013), "Berufswertigkeit und Logistikqualifikation", Logis, Berlin.

Klumpp, M., Bioly, S., Abidi, H. (2013), "Zur Interdependenz demografischer Entwicklungen, Urbanisierung und Logistiksystemen", in: Göke, M., Heupel, T. (eds.), Wirtschaftliche Implikationen des demographischen Wandels, Wiesbaden (Gabler), pp. 161-175.

Klumpp, M., Bioly, S., Zelewski, S. (2013), "Integrating sustainability and technology innovation in logistics management", in: McIntyre, J.R.; Ivanaj, S.;Ivanaj, V. (eds.): Strategies for Sustainable Technologies and Innovations, Cheltenham (E. Elgar Publishing), in print.

Klumpp, M., Clausen, U., ten Hompel, M. (2013), "Logistics Research and the Logistics World of 2050", in: Clausen, U., ten Hompel, M., Klumpp, M. (eds.), Efficiency in Logistics, Lecture Notes in Logistics, Heidelberg (Springer), pp. 1-6.

Klumpp, M., Kandel, C., Bioly, S., Rodermund, S. (2011), "Potenziale eines GPS-Trackingsystems", in: Zeitschrift für wirtschaftlichen Fabrikbetrieb, ZWF106 (5), pp. 2-5.

Klumpp, M., Kandel, C., Wirsing, E. (2013), "Cargo Telematics for Operational Transport Excellence and Strategic Knowledge Management", in: Kreowski, H.-J., Scholz-Reiter, B., Thoben, K.-D. (eds.): Dynamics in Logistics, Lecture Notes in Logistics Series, Heidelberg (Springer), pp. 71-84.

Klumpp, M., Krol, B., Zelewski, S. (2006), "Knowledge Intensive Service Production Analysis (DEA)", in: Conference Proceedings 14th International Working Seminar on Production Economics, Innsbruck,, Vol. 1, pp. 193-206.

Klumpp, M. and Schaumann, U. (2007), "Requirements for Leadership Personnel and the Concept Berufswertigkeit", Kölner Blätter für Wirtschaftspädagogik, Vol. 12/2007, pp. 3-50.

Kowalski, M., Zelewski, S. and Bergenrodt, D. (2012), "Applying of an Ontology-driven Case-based Reasoning System in Logistics", International Journal of Computers & Technology, Vol. 3 No. 2, pp. 347-350.

Lin, L.F., Zhang, W.Y., Öou, Y.C., Chu, C.Y. and Cai, M. (2011), "Developing manufacturing ontologies for knowledge reuse in distributed manufacturing environment", International Journal of Production Research, Vol. 49 No. 2, pp. 343-359.

MacMahon, W.F., Wagner, A.P. (1981), "Expected Returns to Investment in Higher Education," in: Journal of Human Resources, Vol. 16, pp. 274-85.

Meyer, N., Reidel, J., Schmidt, I. (2012), "Corporate social responsibility-management in the management sector – natural conditions and stumbling blocks", paper presented at the X. International Logistics and Supply Chain Congress, 8-9 November, Istanbul.

Roth, A. (2012), "Analyse von Anforderungen in der Logistikbranche – Erste Implikationen für die Berufsbildung", Zeitschrift für Berufs- und Wirtschaftspädagogik, Vol. 108 No. 4, pp. 604-626.

Robles, M., Wei, F. and Noche, B. (2013), "Challenges in the Planning, Organization, Execution and Control of International Supply Chains", in Clausen, U., ten Hompel, M. and Klumpp, M. (eds.), Efficiency and Logistics, Springer, Berlin Heidelberg, pp. 245-252.

Teichler, U. (2007), "Graduate Employment and Work: Various Iddues in a Comparatice Perspective", in: Teichler, U. (ed.), Careers of University Graduates, Springer, Dordrecht, pp. 1-14.

Teichler, U. (2009), "Higher Education and the World of Work", Sense, Rotterdam, Taipeh.

Walker, R., Florea, L. (2013), "Easy-Come-Easy-Go: Moral Hazard in the Context of Return to Education", in: Journal of Business Ethics, DOI 10.1007/s10551-013-1656-8.

Walton, S.M., Graves, P.E., Sexton, R.L. (2005), "New Entry and the Rate of Return to Education: The Case of Registered Nurses", in: Atlantic Economic Journal, Vol. 33, pp. 325-336.

BIOGRAPHY

MATTHIAS KLUMPP studied economics and business administration after a vocational degree in logistics (Speditionskaufmann IHK, 1996) at University Leipzig and the IECS Strasbourg from 1995 to 1998 (Diplom-Kaufmann, Diplom-Volkswirt). Parallel to his professional consultant career in strategy management, logistics and education (zeb/, conoscope) he also obtained a PhD at University Leipzig in 2007 and started as a professor at FOM University of Applied Sciences. He founded the Institute for Logistics and Service Management (ild) at FOM in 2009.

He is member of the scientific committee for the German national excellence cluster EffizienzCluster LogistikRuhr funded by the German national ministry for research (BMBF) and has done research or teaching stays in Australia, China, Finland, France, Netherlands, Norway, Spain, Switzerland, Turkey and the USA.

MANAGEMENT SIMULATION

THREE EIGHT DIMENSIONAL CORRELATION STUDIES

William Conley
Austin E. Cofrin School of Business
280B Wood Hall
University of Wisconsin-Green Bay
Green Bay, Wisconsin 54311-7001
U.S.A.
Conleyw@uwgb.edu

KEYWORDS
CTSP correlation statistic, multivariate analysis, complex statistics

ABSTRACT

Personal computers are so affordable, ubiquitous and fast in our 21^{s} century that it is now possible to carry out massive simulations with them in seconds or a few minutes of computer run time. This allows the opportunity to develop new more sophisticated multivariate statistics for analyzing large spreadsheets of data to see what, if any, relationships may be hidden in their *k* dimensional space (for *k* variables). One such statistic is the new CTSP correlation statistic (discovered in 2002) which is featured here to analyze three multivariate spreadsheets of data involving sets of eight variables which may be correlated. The analysis of the first two spreadsheets shows a correlation between the variables in question. However, the third spreadsheet of data involves variables that were thought to be correlated, but the CTSP calculations prove this assumption to be false.

INTRODUCTION

The idea behind the new CTSP class of statistics is dealt with in (Conleya 2007 and Conleyb 2007). However, because the three large scale eight dimensional examples presented here involve too many dimensions to picture the idea geometrically, a brief discussion of the approach is mentioned here.

The CTSP acronym is short for correlation using the travelling salesman problem mathematical optimization approach. The classic travelling salesman problem (TSP) involves a salesperson leaving the company's home office and travelling to n – 1 clients or customers and returning to the home office while travelling on the shortest route , saving time, money and fuel.

The TSP problem has defied all attempts at a theoretical mathematics solution for large *n*. It is easy to state, but difficult to solve. Therefore (Conley 1991) used the multi stage Monte Carlo optimization (MSMCO) simulation technique on two famous test problems (problem number 30 and problem number 32) from the mathematics literature with 150 and 200 points or destinations. The MSMCO simulation approach produced hundreds of shorter routes than the best (shortest) ever found trying other methods. What is being advocated here is to find the shortest TSP routes through spreadsheets of data points (that represent variables).

Then to compare these shortest routes with shortest routes through similar size random data. If the shortest routes through the real data are shorter than the shortest routes through the corresponding random data, then the real data is following a pattern. Hence, the variables represented by the data are correlated in some fashion. Note that we are using a much more general purpose all-encompassing idea of correlation than the statistical theorists who developed linear correlation analysis (before the computer age) were able to do. Let us look at three eight dimensional examples (eight variables) with some accompanying data.

A NONLINEAR CORRELATION

Consider the following $n = 99$ lines of eight columns (representing eight variables) of data. It is hypothesized by engineers (Hayter, 2002) that the X_8 variable (the percentage increase in yield of the compound being manufactured) is a function of the number of units of seven additional catalysts and chemicals (variables one through seven) added to the long established production formula. The $n = 99$ lines of data represent 99 trial runs of the process. Therefore, the engineers wish to test the hypothesis

H_0: no correlation between the eight variables X_1, X_2, X_3, X_4, X_5, X_6, X_7 and X_8
Versus
H_A: a correlation between the variables

Therefore, the spreadsheet of eight columns of $n = 99$ lines of data is read into the MSMCO CTSP algorithm computer program adjusted for 99 lines of data (99 trial runs) and eight variables. The program starts by calculating the 99 x 99 distances from each point to each other data point (using an eight dimensional generalization of the Pythagorean theorem distance measure). Then this 99 x 99 array is column ranked from smallest to largest distances. As an example, the first column top entry (upper left hand corner) of the ranked array would represent the distance from point number 1 to its nearest point. The second number down would represent the distance from the point number 1 to its second nearest point and so on. The value in the lower left hand corner would represent the distance from the point number 1 to the eight dimensional point that is furthest away from it. Therefore, each column in the ranked array would have the distances to the nearest points at the top of the array. This is where the MSMCO CTSP algorithm concentrates its simulation efforts because many of these relatively short sub-routes are likely to be on the optimal shortest route or a near approximation.

A few minutes of computer run time on a desktop PC produced a shortest route of total distance A = 5759.427 through the 99 points. The route read left to right, top to bottom is in Table 1.

Table 1: The route

d =A =5759.427

49	97	61	43	66	58	89	63	99	62
96	71	87	40	68	88	64	19	21	29
33	91	18	57	69	59	95	9	81	15
79	74	42	34	45	67	56	93	82	50
27	53	17	72	52	16	31	26	36	46
98	14	84	77	11	*65*	38	23	76	73
51	10	90	85	37	12	28	*60*	80	20
1	92	13	32	25	48	3	47	75	39
8	7	2	4	86	6	94	83	30	78
44	55	22	54	24	35	5	70	41	49

Table 2 contains the data.

Table 2: The data

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	68	86	6	1	55	0	32	60
2	83	92	14	99	45	3	36	65
3	2	66	63	98	13	38	6	58
4	60	77	24	75	28	13	39	82
5	32	21	50	5	100	40	52	75
6	25	57	10	77	50	29	41	88
7	43	100	7	73	63	29	31	78
8	23	72	15	73	93	4	29	75
9	76	23	86	97	22	100	81	53
10	100	12	95	33	83	90	2	47
11	68	95	67	67	12	99	58	60
12	47	71	4	16	56	87	15	70
13	64	53	52	46	17	0	53	79
14	89	63	100	58	41	35	50	74
15	77	2	52	84	35	82	67	74
16	70	44	56	43	73	45	33	95
17	14	2	95	24	23	23	31	54
18	19	55	53	36	52	87	56	83
19	27	56	90	32	10	87	12	56
20	6	100	32	7	65	11	4	50
21	2	99	98	3	17	96	1	8
22	87	98	7	6	4	74	90	53
23	51	77	5	67	41	63	6	78
24	57	59	56	11	83	82	66	79
25	57	20	22	56	35	19	6	80
26	69	93	41	33	41	55	63	83
27	6	10	100	6	49	30	54	53
28	46	75	27	18	68	34	15	82
29	42	76	80	3	35	80	55	66
30	7	12	12	61	47	49	84	75
31	90	74	48	44	44	75	50	83
32	56	21	39	44	15	8	13	73
33	31	79	86	16	54	76	70	69
34	8	22	52	11	16	96	11	49
35	37	52	81	16	100	61	57	76
36	82	46	59	46	57	31	91	84
37	57	70	8	46	100	90	38	76
38	48	96	35	81	62	62	18	81
39	14	78	23	68	93	52	4	74
40	43	62	89	82	100	98	73	65
41	38	67	56	46	86	1	67	82
42	33	9	69	16	3	84	49	57
43	100	33	98	80	21	30	96	53
44	77	13	11	59	40	22	82	78
45	11	23	37	11	51	100	33	64
46	83	67	76	59	95	32	78	79
47	11	26	55	91	50	6	4	68
48	30	14	16	79	6	62	2	62
49	31	50	25	8	36	33	85	75
50	26	18	86	36	77	26	94	70
51	87	10	12	85	85	63	26	75
52	77	8	89	40	70	59	42	77
53	1	18	100	14	22	5	44	44
54	52	66	31	6	74	67	53	83
55	57	53	20	58	61	54	92	89

56	0	30	73	56	98	84	43	69
57	14	45	41	47	16	91	84	63
58	61	64	44	100	17	4	77	68
59	14	71	84	78	23	70	100	59
60	33	95	60	31	67	12	0	66
61	96	52	82	21	18	27	100	54
62	55	74	50	53	75	78	100	79
63	61	100	64	90	72	38	100	68
64	39	57	92	39	39	86	60	75
65	39	100	46	91	24	61	47	74
66	72	51	72	79	33	10	81	77
67	6	49	28	40	93	93	28	72
68	44	45	81	80	65	88	48	82
69	13	14	71	90	5	60	100	52
70	11	59	39	22	93	44	55	81
71	23	64	70	50	88	77	85	78
72	56	31	100	2	61	30	50	67
73	94	76	26	98	42	100	24	62
74	40	23	98	60	2	100	55	52
75	4	85	53	77	84	21	0	65
76	75	67	43	70	32	83	1	74
77	59	88	94	90	7	71	47	58
78	61	23	9	51	15	53	72	79
79	76	21	75	44	8	76	67	69
80	27	90	84	14	60	22	12	62
81	94	28	57	91	30	93	73	67
82	76	7	93	65	83	12	65	69
83	28	27	20	81	95	49	73	84
84	94	67	79	86	8	15	14	57
85	73	100	54	46	100	83	9	64
86	23	53	3	62	30	35	48	84
87	5	69	48	37	91	79	68	74
88	25	34	82	78	45	92	40	75
89	85	99	40	82	24	35	97	63
90	94	54	100	24	65	62	1	59
91	10	95	60	29	56	74	44	72
92	100	55	19	34	35	6	32	72
93	45	31	94	70	96	48	53	80
94	3	45	27	71	75	18	66	82
95	34	32	98	87	56	82	86	67
96	47	61	52	73	71	85	83	85
97	30	72	56	1	25	37	100	60
98	99	73	72	88	84	28	40	75
99	53	70	46	67	92	46	79	89

Then four sets of n=99 lines of eight columns of random data (in similar ranges as the real data) are read into the MSMCO CTSP algorithm computer program. Their shortest routes connecting each of the n=99 lines of 8 dimensional points in a complete closed loop tour are 6454.494, 6288.674, 6285.939 and 6589.618. Then CTSP =A/B =5759.427/ (6454.494+6288.674/2)=.904 where B is the median of the four shortest routes from the random data. Now taking the 3x4=12 A_i/B_i quotients from the four random data shortest routes, one sees that the vast majority of the sampling

6454.494/6288.674 = 1.020
6285.939/6589.618 = .954
6288.674/6454.494 = .974
6589.618/6285.939 = 1.048
6454.494/6285.939 = 1.027
6285.939/6454.494 = .974
6454.494/6589.618 = .979
6589.618/644.494 = 1.021
6288.674/6285.939 = 1.000
6285.939/6288.674 = 1.000
6288.674/6589.618 = .954
6589.618/6288.674 = 1.048

distribution of the CTSP statistic (under the hypothesis of no correlation) will be in the range 6285.939/6589.618 = .954 to 6589.618/6285.939 = 1.048. Our CTSP = .904 is comfortably below this range; therefore H_0 can be rejected. The variables are correlated because they are following a pattern and hence the real data has a shorter shortest route than similar random data would have. It turns out that the equation

$$Y = X_8 = 100 - .0045535 (X_1^2 - 104X_1 + X_2^2 - 94X_2 + X_3^2 - 82X_3 + X_4^2 - 118X_4 + X_5^2 - 126X_5 + X_6^2 - 92X_6 + X_7^2 - 98X_7 + 18{,}561)$$

fits that data quite well. The engineers additionally wish to maximize this function subject to the safety constraints of $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 \leq 300$ and $2X_1 + 4X_2 + 9X_3 + 6X_4 + X_5 + 12X_6 + X_7 \leq 1400$. A short multi stage Monte Carlo optimization (MSMCO) program is run several times producing three usable solutions of

$X_1=43$ $X_2=39$ $X_3=33$ $X_4=51$ $X_5=54$ $X_6=38$ $X_7=41$
$Y = X_8 = 97.805$

$X_1=44$ $X_2=39$ $X_3=32$ $X_4=50$ $X_5=55$ $X_6=38$ $X_7=41$
$Y=X_8 = 97.805$

$X_1=44$ $X_2=39$ $X_3=32$ $X_4=51$ $X_5=55$ $X_6=38$ $X_7=40$
$Y=X_8= 97.805$

AN UNUSUAL LINEAR CORRELATION

Consider the following n=129 lines of eight columns (representing eight variables) of data in Table 3.

Table 3: Eight variables

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	67	82	94	37	5	45	16	62
2	45	40	48	25	34	39	93	54
3	65	44	93	58	36	62	34	46
4	4	42	1	47	58	33	2	23
5	100	35	2	11	47	24	10	44
6	29	35	50	92	99	78	9	27
7	0	58	32	88	25	35	34	39
8	61	49	56	0	22	33	37	49
9	69	100	83	19	3	38	62	83
10	46	72	15	68	66	45	86	70
11	23	64	72	69	76	74	8	41
12	59	6	32	28	93	60	28	24
13	58	98	66	42	78	69	27	71
14	19	80	90	32	12	46	55	60
15	60	100	79	69	24	52	45	77
16	55	33	24	49	79	53	16	34
17	98	56	86	6	100	83	72	70
18	100	39	62	26	11	34	59	58
19	20	61	74	2	82	69	70	54
20	35	38	16	69	55	41	55	42
21	14	28	36	69	83	62	100	43
22	63	61	4	2	79	39	35	55
23	63	44	11	90	21	25	17	42
24	25	43	26	83	53	46	7	30
25	1	15	82	37	11	43	86	30
26	63	33	49	100	63	62	68	49
27	9	98	77	6	74	67	87	75
28	64	3	76	5	83	71	68	33
29	49	46	44	63	85	66	17	39
30	76	72	91	56	46	66	82	60
31	67	53	30	17	41	34	56	57
32	92	10	71	23	22	42	84	47
33	79	40	53	60	92	72	75	58
34	74	54	75	73	76	75	80	65
35	64	79	8	67	15	18	36	65
36	31	87	32	40	55	44	23	58
37	88	39	71	74	15	45	100	66
38	33	51	8	10	93	48	41	44
39	98	100	51	16	51	47	88	97
40	85	60	12	58	25	24	69	68
41	46	7	46	5	29	33	100	39
42	51	64	32	32	12	23	48	57
43	17	36	60	33	78	65	8	25
44	74	68	62	72	82	73	85	74
45	26	5	5	85	40	31	62	24
46	88	28	2	77	16	18	68	52
47	100	35	25	41	24	26	80	61
48	44	99	91	70	27	58	77	81
49	84	42	20	62	79	53	68	58
50	24	52	91	76	57	73	82	53
51	51	89	73	52	77	72	4	59
52	19	60	83	65	56	68	12	39
53	47	74	21	91	37	37	62	65
54	75	100	33	29	85	57	65	86
55	10	59	48	72	100	75	45	44
56	76	28	10	15	100	53	12	35
57	9	16	10	20	32	22	82	31
58	53	88	44	100	17	38	2	58
59	15	100	62	66	48	56	48	67
60	88	95	91	2	38	55	15	73
61	100	100	4	64	25	21	65	91
62	46	9	57	53	12	35	75	34
63	1	82	56	19	24	37	45	54
64	9	67	9	78	72	47	6	38
65	17	92	81	78	60	71	2	52
66	33	69	4	35	68	38	38	53
67	14	76	17	43	100	59	88	65
68	23	29	70	54	4	37	46	32
69	24	6	31	23	66	47	98	33
70	86	6	100	13	24	54	33	31
71	40	100	42	67	24	37	98	86
72	94	24	37	7	19	25	19	39
73	54	5	72	71	89	80	13	18
74	33	94	58	25	45	48	61	72
75	43	100	11	98	67	48	1	62
76	100	71	20	3	45	30	24	66
77	61	5	9	7	60	33	100	42
78	16	78	11	93	41	35	53	58
79	37	66	56	59	55	56	76	62
80	26	38	93	4	26	51	66	42
81	37	33	80	70	19	50	58	40
82	100	100	24	50	88	57	76	94
83	51	24	100	96	57	79	23	30
84	64	56	74	40	100	82	68	61
85	37	100	50	34	68	57	87	82
86	40	1	65	64	7	38	79	30
87	20	49	12	44	50	34	27	37
88	72	61	30	52	44	39	31	56
89	70	67	89	36	46	62	2	51
90	79	39	45	21	35	37	87	60

91	100	36	35	91	23	36	17	46
92	7	80	100	89	50	75	93	67
93	100	99	3	75	100	57	0	75
94	34	31	88	31	25	52	58	38
95	79	13	100	84	94	95	79	48
96	83	72	0	82	71	43	100	82
97	29	50	39	100	36	45	9	35
98	40	14	39	100	67	60	60	32
99	15	16	64	29	16	37	8	14
100	54	89	40	74	79	62	42	69
101	94	61	43	19	46	42	78	73
102	89	20	49	12	49	45	18	35
103	11	84	85	68	59	71	23	52
104	100	89	51	26	12	30	64	85
105	95	0	89	2	59	65	70	39
106	96	52	76	97	100	90	100	74
107	54	18	41	84	32	42	82	42
108	100	91	27	83	38	39	13	74
109	92	100	92	59	52	69	82	94
110	4	100	37	59	17	30	32	61
111	10	100	6	40	83	46	56	68
112	3	64	42	92	69	61	41	44
113	26	30	47	14	77	57	68	39
114	0	43	40	66	30	38	4	23
115	39	30	35	48	77	56	76	44
116	8	63	97	10	26	53	64	51
117	88	96	96	30	48	66	83	91
118	42	12	9	44	81	47	83	37
119	92	17	11	17	2	8	54	44
120	1	100	83	84	76	79	2	53
121	1	48	63	53	61	61	43	36
122	8	14	9	36	29	22	62	25
123	39	47	32	86	13	30	11	36
124	65	61	97	31	48	66	23	52
125	3	18	81	7	51	58	0	10
126	7	5	77	97	94	87	25	11
127	49	44	32	1	34	29	100	59
128	80	55	28	93	92	66	90	70
129	0	97	53	79	97	77	43	62

The scientists believe that the eight variables may be correlated in some fashion and would like to test that hypothesis. Therefore, they test

H_0: no correlation between X_1, X_2, X_3, X_4, X_5, X_6, X_7 and X_8

H_A: a correlation between the variables.

Similarly to the previous example (of 99 x 8), the n = 129 lines of eight columns of data (Table 3) is inserted into the MSMCO CTSP algorithm and a shortest route through these 129 eight dimensional points is found to be A=6209.184 (see Table 4 for the complete point to point route read left to right, top to bottom).

Table 4: The route

d =A =6209.184

44	34	30	37	107	26	98	21	118	115
113	19	27	85	74	59	110	63	14	116
80	94	81	68	25	86	62	2	127	90
101	31	42	8	72	119	47	46	40	35
88	91	23	123	97	24	64	4	87	66
36	111	67	10	96	128	106	95	84	17
28	105	70	102	5	76	22	*38*	56	12
16	29	6	126	73	83	3	124	89	60
1	15	48	109	117	39	104	9	18	32
41	77	69	57	122	45	20	53	78	7
114	99	125	43	121	112	55	129	120	65
103	52	11	51	13	100	79	50	92	71
61	108	58	75	93	82	54	33	49	44

Then six sets of n=129 lines of eight columns of random data in the same 0 to 100 ranges of the original data are read into the MSMCO CTSP algorithm. Their shortest routes were calculated to be 8224.152, 8161.167, 7966.983, 8281.473, 7801.320 and 7726.985. Now, taking their 5 x 6 = 30 A_i/B_i quotients, the range of these quotients is 7762.985/8281.473 = .937 to 8281.473/7762.985 = 1.067. Then again using the median for B in the CTSP calculation, CTSP = 6209.184/(7966.983+8161.167)/2) = .770 which is significantly below the .937 to 1.067 range where the vast majority of the sampling distribution of CTSP must be if H_0 is true. Therefore, H_0 is rejected and the variables are correlated. The multivariate piece wise linear functions $X_6 = .41X_3 + 12X_4 + 47X_5$ and $X_8 = .23X_1 + .52X_2 + .25X_7$ fit the data quite well.

NO CORRELATION

Business researchers are considering the following n=60 lines of eight columns (representing eight variables) of data for possible correlation (Mendenhall and Reinmuth 2010) between the variables. Please see Table 5 for the data.

Table 5: The 60 x 8 data

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	71	100	84	100	68	5	49	35
2	96	7	22	67	92	58	43	49
3	32	11	88	12	12	68	90	26
4	50	78	20	47	100	70	45	58
5	71	11	57	31	91	53	28	2
6	5	100	41	60	93	13	80	51
7	80	32	100	8	73	42	19	51
8	7	17	40	85	45	31	100	100
9	1	45	46	79	59	64	58	89
10	94	60	34	86	100	59	44	51
11	7	79	93	59	40	20	75	52
12	97	47	37	11	45	61	62	34
13	95	62	42	12	6	30	14	27
14	31	40	63	19	66	17	26	3
15	50	20	92	47	86	95	5	13
16	85	90	95	72	97	41	3	68
17	58	22	34	3	27	24	9	47
18	19	22	65	83	23	15	97	20
19	59	36	3	5	100	55	17	30
20	86	76	100	41	55	63	88	65
21	2	99	98	3	1	96	1	8
22	91	0	23	25	52	82	24	34
23	50	71	84	69	78	57	59	49
24	39	29	35	54	4	61	15	67
25	49	94	16	92	94	48	71	99
26	97	50	55	91	52	14	40	6
27	98	35	62	57	93	46	66	91
28	22	81	74	23	100	100	93	55
29	62	54	80	98	75	88	66	22
30	62	36	77	28	100	58	44	53
31	90	8	75	62	7	16	24	4
32	54	73	54	87	37	66	62	95
33	25	95	25	27	41	0	49	87
34	82	37	2	100	61	85	68	61
35	76	33	40	48	33	75	45	21
36	100	19	23	73	17	75	83	31
37	24	52	92	81	73	23	6	38
38	66	90	52	31	28	99	19	69
39	59	19	100	38	19	89	45	99
40	11	71	16	84	39	4	93	99
41	56	21	44	72	27	43	11	100
42	59	23	25	50	71	81	55	20
43	75	71	59	31	77	76	73	70
44	40	77	99	1	64	55	69	20
45	48	100	73	47	48	74	5	100
46	5	85	36	85	78	15	19	78
47	22	92	32	100	9	70	26	48
48	75	100	52	10	14	97	36	76
49	2	90	97	85	51	69	56	26
50	84	75	11	17	6	86	43	9
51	90	22	54	24	72	7	21	65
52	76	36	60	32	78	68	21	36
53	57	83	86	87	61	100	70	17
54	56	18	20	60	73	4	100	49
55	2	82	52	100	31	19	41	72
56	28	40	92	83	87	91	76	56
57	34	73	24	9	46	85	93	6
58	44	65	99	2	71	88	99	16
59	9	79	79	78	68	19	30	7
60	100	93	80	57	47	77	39	70

Therefore, they test

H_0: no correlation between X_1, X_2, X_3, X_4, X_5, X_6, X_7 and X_8
Versus
H_A: a correlation between the variables

The n = 60 lines of eight columns of numbers are read into the MSMCO CTSP algorithm. A short run on a desktop PC produces a shortest route closed loop tour connecting the 60 eight dimensional points with a total distance of A = 4317.431. Please see Table 6 for the point to point route read left to right and top to bottom.

Table 6: The route

d =A = 4317.431

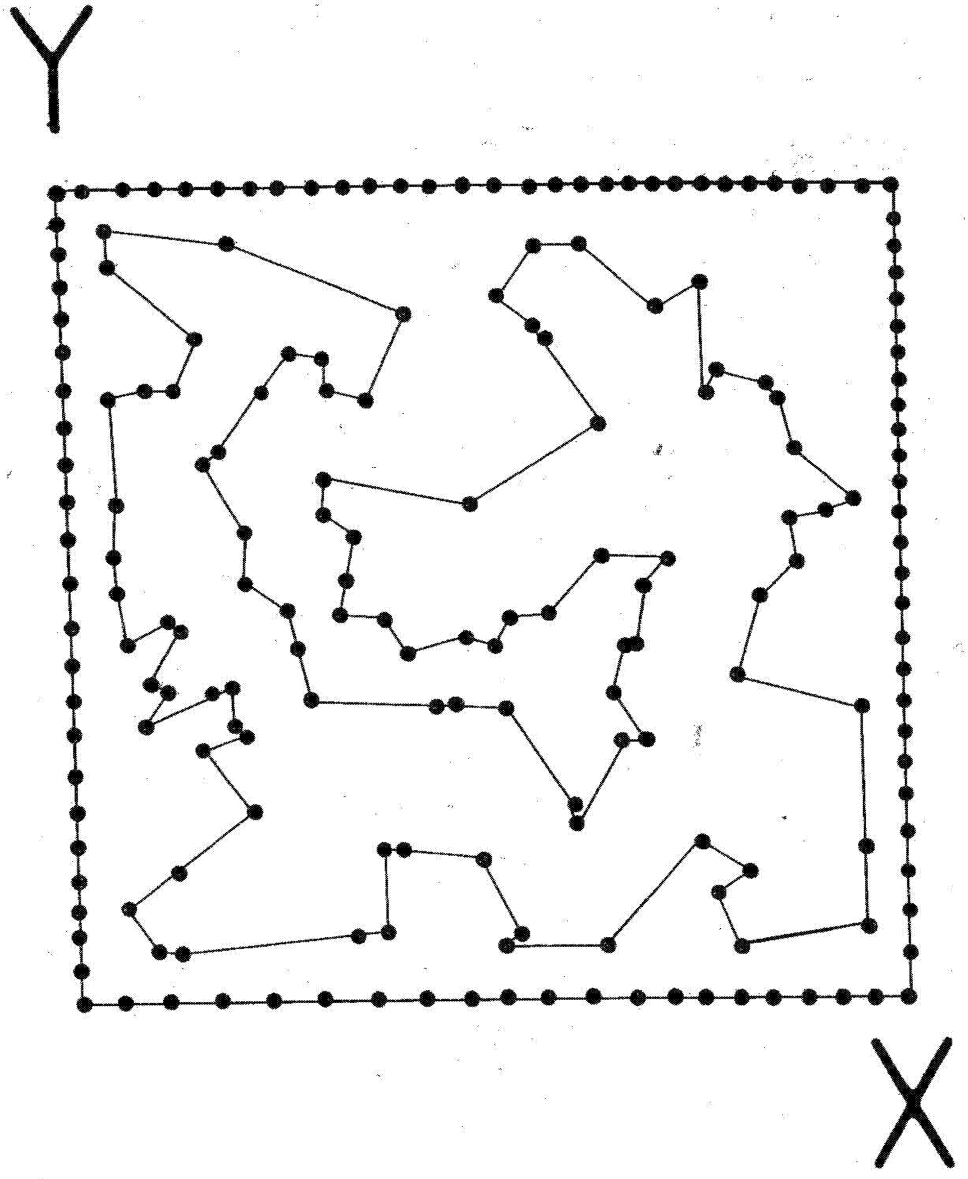
11	49	59	37	16	1	26	31	13	17
51	7	30	52	15	5	14	19	4	25
32	9	8	40	33	6	46	55	47	24
41	39	3	18	54	34	36	35	12	50
57	42	22	2	10	27	43	20	60	45
38	48	21	44	58	28	56	29	53	23

Then four sets of n=60 lines of eight columns of random numbers (in the same 0-100 ranges as the original data) are sequentially entered into the MSMCO CTSP algorithm and run producing shortest routes of total distances 4472.419, 4170.986, 4224.071 and 4371.717. Taking the 3 x 4 = 12 A_i/B_i quotients as before, the range of these is 4170.986/4472.419 = .933 to 4472.419/4170.986 = 1.072. Now CTSP = A/B = 4317.431/((4224.071 + 4371.717)/2) = 1.005 is very near one and right in the middle of the sampling distribution range if H_0 of no correlation is true. Therefore, the null hypothesis of no correlation between the variables cannot be rejected.

THE GEOMETRY CORRELATION

The basic premise behind the CTSP general correlation idea is that points that are correlated are more compact because they are following a pattern dictated by the relationship that governs them. Therefore, the shortest route connecting them will be relatively shorter than the shortest route for the same number of random points in similar ranges. This is difficult to picture in an eight dimensional space. However, Figures 1, 2, and 3 illustrate it for two dimensional correlations of X Y variables. Figure 1 shows n = 98 X Y points (the points on the border) following a precise rectangular pattern because X Y are correlated in a piece wise linear fashion (four line segments.)

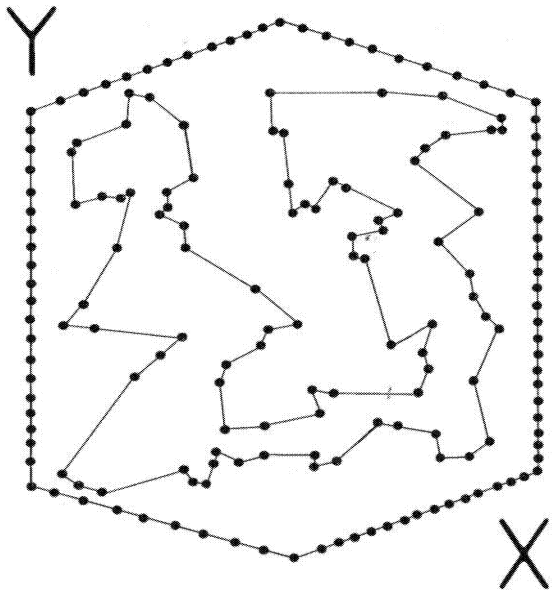
Figure 1. Border Correlated. Inside Points No Correlation



However, the n = 98 points inside the border are indicative of no relationship or correlation between the X and Y variables they were taken from because they do not follow a pattern. Notice that the shortest route connecting them is much longer than the shortest route connecting the n = 98 border points.

Figure 2 similarly shows a hexagonal relationship from the n X Y points on the border, and no relationship with the random n X Y points inside the border.

Figure 2: Border points represent X Y correlation



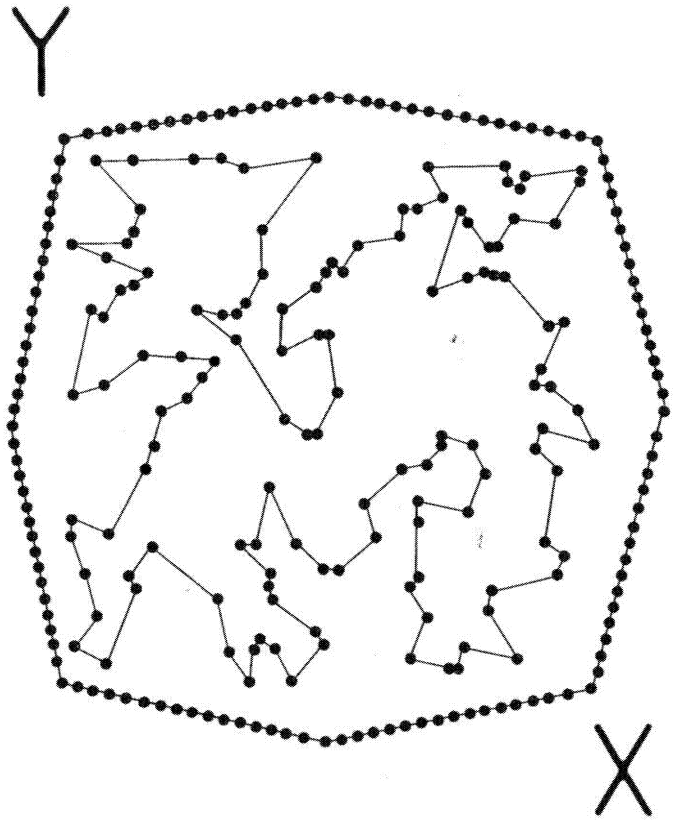
Also, Figure 3 similarly represents an octagonal relationship (with a shorter shortest route) on the border, but no correlation is found between X and Y with the data set inside the border.

NORMALITY ASSUMPTIONS

The standard linear correlation analysis involving the r correlation coefficient for X Y data and the larger multivariate R for three or more variables has its testing theory rooted in underlying normality assumptions (Anderson 2003) for its distributions.

Consider the n = 50 X Y points arranged in a "random" sample from a bivariate normal distribution in Figure 4.

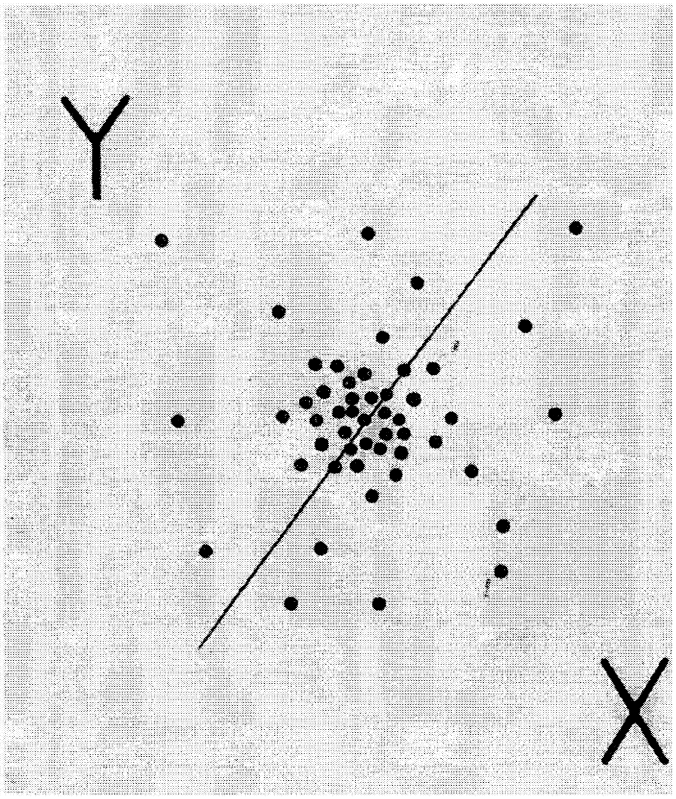
Figure 3: Points inside the border represent no X Y correlation



The shortest route connecting them will be much longer than the shortest route connecting 50 points along and near the straight line bisecting the points in Figure 4. Therefore, if one wants to assume underlying normal distributions (or any other distribution for that matter), the CTSP analysis should be able to find more restrictive correlations, be they linear and/or nonlinear.

Additionally, the new DTSP statistic (Conley 2005) uses the standard normal approximation to the runs test for randomness along with shortest route analysis to test what multivariate distribution the data may have come from, if there are doubts about the underlying distribution.

Figure 4: Multivariate normal distribution



SOFTWARE FOR CTSP

The first example presented here showed a correlation between eight variables where the first seven were driving the eighth variable in a nonlinear fashion. Least squares and/or other curve fitting techniques could be reasonably expected to produce a yield equation similar to the one that was stated and optimized here.

The third example showed no correlation, so fitting an equation was not appropriate. However, the second example showed a strong overall correlation between the eight variables represented by the 129 x 8 data array) but it was really divided into two separate linear relationships between X_1, X_2, X_7, X_8 and X_3, X_4, X_5, X_6. Therefore, what is needed now is a CTSP software package that will not only do the overall correlation test on the k variables in question, but will follow it up (where appropriate) by automatically doing correlation tests on all the possibly hundreds of subsets of the variables. There is enough computing power available on inexpensive desktop PC computers to use a software package of that type.

CONCLUSION

Three eight variable correlation studies were presented here, where the new CTSP correlation statistic was used to test whether the variables were correlated or not. The first one showed a correlation that proved to be nonlinear. The second one had a piece wise linear correlation. The third one was clearly shown to have no correlation at all between the variables.

The CTSP correlation analysis was compared and contrasted with the standard linear correlation analysis (based on normality assumptions) (Anderson 2003) which has been available and in use for some time.

The geometry of the CTSP idea was discussed and partially illustrated. The case for a CTSP software package that not only can do the initial correlation analysis, but can automatically test all relevant subsets of the variables for correlation was mentioned.

It also should be noted that multi stage Monte Carlo optimization (MSMCO) is a general purpose simulation optimization technique that can be applied across a wide spectrum of application areas of interest to engineers, scientists and people in business. Some examples in applied mathematics (Wong 1996), engineering (Szarkowicz 1995), ecology (Conley, 2008) and transportation (Conley 2012) are mentioned in the references.

REFERENCES

Anderson, T. W. 2003. Multivariate Statistical Analysis, 3rd edition, Wiley and Sons, New York.

Conley, W. C. 2012. "A million variable nonlinear transportation problem." In *Proceedings from of 2012 Industrial Simulation Conference*. ISC2012 EUROSIS, Ghent, Belgium, pp. 113-121.

Conley, W. C. 2008. "Ecological optimization of pollution control equipment and planning from a simulation perspective," *International Journal of Systems Science*, Vol. 39, No. 2, pp. 1-7.

Conley, W. C. 2007a. "Finding relationships that are functions." In *Proceedings of the 2007 Industrial Simulation Conference*. ISC2007 EUROSIS, Ghent, Belgium, pp. 61-64.

Conley, W.C. 2007b. "Simulation optimization and correlation with multi stage Monte Carlo optimization," *International Journal of Systems Science*, vol. 38, No. 12, pp. 1013-1019.

Conley, W.C. 2005. "A new multivariate distribution statistic applied to health care research." In *Proceedings of Western Simulation Multiconference*, January 2005, New Orleans, Society for Computer Simulation, San Diego, pp. 55-60.

Conley, W. C. 1991. "Programming an automated punch or drill," *International Journal of Systems Science*, vol. 22, No. 11, pp. 2039-2056.

Hayter, A. J. 2002. Probability and Statistics for Engineers and Scientists, 2nd Edition, Duxbury Press, Pacific Grove, California.

Mendenhall, J. R. and Reinmuth, A. H. 2010. Statistics for Business and Economics, 10th Edition, Duxbury Press, Pacific Grove, California.

Szarkowicz, D. 1995. "Investigating the brachistochrone with a multi stage Monte Carlo optimization," *International Journal of Systems Science*, Vol. 26, pp. 233-243.

Wong, J. Y. 1996. "A note of optimization in integers." *International Journal of Mathematics Education in Science and Technology*, Vo. 27, pp. 865-874.

BIOGRAPHY

WILLIAM CONLEY received a B.A. in mathematics (with honors) from Albion College in 1970, an M.A. in mathematics from Western Michigan University in 1971, an M.Sc. in statistics in 1973 and a Ph.D. in mathematics - computer statistics from the University of Windsor in 1976. He has taught mathematics, statistics, and computer programming in universities for over 30 years. He is currently a professor emeritus of Business Administration and Statistics at the University of Wisconsin at Green Bay. The developer of multi stage Monte Carlo optimization and the CTSP multivariate correlation statistics, he is the author of five books and more than 200 publications world wide. He is a member of the American Chemical Society, a fellow in the Institution of Electronic and Telecommunication Engineers and a senior member of the Society for Computer Simulation.

SIMULATION GAMING AND NATURAL LANGUAGE PROCESSING FOR MODELLING STAKEHOLDER BEHAVIOR IN ENERGY INVESTMENTS

A. Ittoo
N.B. Szirbik
G.B. Huitema, J.C. Wortmann
Department of Operations
Faculty of Economics and Business
University of Groningen
The Netherlands
{r.a.ittoo, n.b.szirbik, g.b.huitema, j.c.wortmann}@rug.nl

KEYWORDS
Simulation, gaming, energy, investments, natural language processing .

ABSTRACT

Investment decisions in energy infrastructure are multi-stakeholder decisions. Stakeholders are not only Transportation Systems Operators (TSOs) but also banks, energy traders, producers, consumers, governments and regulators. Classical approaches for modeling investment decisions are often inadequate to support decision makers in energy infrastructure investments. In the classical approaches, a single party's decision problem is modeled and the optimal course of action is selected. In a multi-stakeholder situation, however, each party has to acknowledge other stakeholders' autonomous decisions. This introduces uncertainty which cannot be modeled mathematically, but has to be handled by negotiations or in scenarios. Accordingly, this paper describes an approach to study such decision making situations by simulation gaming with (representatives of) real stakeholders. The games aim to capture the considerations taken into account by the players in order to learn about their behavior. The paper describes an innovative way of capturing these considerations empirically by using natural language processing technology to analyze written conversations and negotiations. The game is currently designed for decisions on expansion or renewal of the gas transportation infrastructure, but the technology could be applied to study multi-stakeholder decision making in more general settings.

INTRODUCTION

Governments around the world are seeking massive levels of investment in new energy infrastructure in order to satisfy the growing energy demands. For the same reason, they aim to modernize and upgrade the existing infrastructures. In order to entice investors, several energy-related policies have been enacted in many countries, including feed-in and transport tariffs, R&D subsidies, and tax benefits (Holburn 2012).

However, despite the financial appeal of these policies, long-term energy infrastructure investments (referred in the rest of the text as energy investments) are still considered as high-risk endeavors (Holburn 2012; Pelletier and Wortmann 2009; Barradale 2010). One of the biggest risks for investment firms is that the regulatory policies exhibit a high degree of uncertainty since they are susceptible to a wide range of unpredictable factors. In order to understand the situation it should be noted that the increased liberalization of energy markets goes together with increased regulation of the infrastructure business. Accordingly, the costs of transportation, storage, and conversion of energy is not determined by market conditions but by a regulatory regime responsible for low prices. It should also be emphasized that in many European and North-American countries, the regulator is deliberately independent and not subject to political control by the parliament. Accordingly, governmental policies and regulatory policies in a country are not necessarily in harmony, i.e. aligned, since there is no harmonization force or mechanism. Any apparent alignment is merely spurious or accidental.

In particular, a major deterrent to energy investments are the constantly shifting fiscal and political priorities of governments and regulatory agencies, which may lead to the reversal or repeal of favorable policies. There are no systematic approaches to determine or forecast when such policy changes will materialize and when will they be implemented. In most cases, they tend to occur haphazardly, on an ad-hoc basis (Holburn 2012).

According to previous research (Holburn 2012; Pelletier and Wortmann 2009; Barradale 2010; Fabrizio, 2012) regulatory uncertainties exert considerably more influence on energy investments than other social, economic and environmental factors (Holburn 2012; Barradale 2010). They have been found to exacerbate the already complex process

of contract negotiation and decision making between investment stakeholders. As described in (Barradale 2010) stakeholders often make diametrically opposing assumptions about each other's objectives, and, to a certain extent, tend to behave irrationally in the face of regulatory uncertainties. As a result, contract negotiations may be stalled, which in turn, hinder the necessary energy market growth.

Several optimization and mathematical programming techniques have been proposed for modeling energy investments under uncertainties. However, they suffer from various limitations, as we will discuss in the section "Overview of Existing Techniques". In this article, we propose a novel simulation gaming approach to accurately and efficiently analyze real-life multi-stakeholder negotiation and behavior under regulatory uncertainties. We conjecture that a sound understanding of the negotiations and decision making dynamics will foster the energy market growth. The main novelty in our approach is that it is augmented with Natural Language Processing (NLP) techniques for analyzing the interactions between the stakeholders during the game. As surrogate for the interactions, we use the conversations that the stakeholders exchange via dedicated chat boxes when they negotiate during the game. The extracted information can be used at a later stage, in real-life situations, to help stakeholders better understand each other and align their conflicting objectives and assumptions.

This article is organized as follows. In the next section, we provide a real-life example of the impact of regulatory uncertainties on investment decisions. Then, we review some related research, and highlights their limitations. Next, our proposed augmented simulation gaming (ASG) framework is described, focusing on its Knowledge and Behavior Capture component (natural language processing). Then, we briefly present some technical/implementational aspects of our proposed framework. The last 2 sections respectively conclude the article, and presents areas for future work.

REAL-LIFE EXAMPLE OF UNPREDICTBILITY

Example: Energy Transport Investments

In the gas market, regulators typically impose transport tariffs, and expect transmission system operators (TSOs) to self-finance their investments for gas transports. The fluctuation of the levels of the imposed tariffs, however, has a high degree of uncertainty. For example, they can be repealed at any time by the regulators, without prior warning. As can be expected in such an environment, TSOs are reluctant to invest due to the high risks involved. In addition, some TSOs consider the current tariffs to be so small that hardly any returns on investment (ROI) is possible (Pelletier and Wortmann 2009), and they demand higher tariffs. However, regulators deliberately maintain the gas transport tariff rates as low possible to ensure competitiveness in the trading and shipping areas of the market, and to guarantee the availability of gas at the lowest possible cost to the society.

OVERVIEW OF EXISTING METHODS

Operational Research Techniques

Despite their application for modeling energy investments under different uncertainties, classical Operational Research (OR) methods, particularly, "inexact optimization techniques" (Zeng, et al., 2011), are inadequate and present several shortcomings for handling regulatory uncertainties. First, these techniques are restricted to uncertainties in market parameters, which can be straightforwardly estimated from historical/forecast data, for e.g. supply-demand. On the other hand, regulatory uncertainties are completely decoupled from market mechanisms (Pelletier and Wortmann 2009). Instead, they are intrinsically linked to the fiscal/political priorities/agendas of policy-makers. As such, they are of a more implicit and intricate nature. They cannot be explicitly identified, and estimated from historical/forecast data. Second, classical techniques assume that the uncertainties (parameters) have known probability distributions, and can be formally expressed, for e.g. using fuzzy sets (FMP), stochastically/probabilistically (SMP) or within known intervals (IMP) (Verderame, et al., 2010). However, in most real-life scenarios, the information required for establishing the parameters' distribution is often incomplete (Li, et al., 2010). Furthermore, regulatory uncertainties are completely unpredictable, and cannot be formulated using traditional models, such as fuzzy sets, probability distributions or bounded intervals. Third, existing techniques are limited to only 1 stakeholder, while real-life strategic investment decisions typically involve more than 1 party. Thus, these classical OR techniques fail to take into account the interactions, for e.g. negotiations, between the stakeholders. A promising alternative to the aforementioned techniques is simulation gaming (also known as serious gaming).

SIMULATION GAMING (SERIOUS GAMING)

Simulation gaming is a solution to unlock complex and time-consuming socio-technical processes – like investment strategy– which tend to deadlock easily due to human factors, linked to stakeholders' personal preference, their different worldviews, and even ideological standpoints. Simulation games are played typically for edutainment, with the desired effect to improve the decision making skills of the playing actors that have to take these decisions in the real world (Duke and Geurts 2004), and also "to overcome a

designated problem or deficiency, and provide appropriate feedback to the user about their efforts" (Susi et al. 2007).

Simulation gaming requires a domain definition, i.e. an abstraction of the real world subset under investigation. In our case, the domain is that of strategic investment process in the energy infrastructure. In simulation gaming, the role of the experimenters is to become observers and eventually controllers. Players are the actual actors from the investigated domain who will take investment decisions in real world. To mitigate conflicts and game deadlocks, we introduce a special game participant, called the Deus ex Machina (DEM) (Roest and Szirbik 2009).

By playing a simulation game of the kind presented above, decision makers (players/actors) can learn, explore, and express principled ideas and spontaneous emotions in the safe environment of the game. This enables the actors to acquire a better understanding of the human-related reasons underlying success or failure of strategic negotiations and investments. We posit that, in real-life, a strategic investment process with actors that have opposed worldviews has a greater chance of success if the actors play beforehand the similar process in the controlled environment of a game emulating the real process.

PROPOSED SIMULATION GAMING PLATFORM

The platform is envisaged as a modular architecture, with 3 main functional areas: modeling, game execution and knowledge and behavior capture.

The Modeling Functions

The modeling components are used to set-up the games, starting from a set of investment scenarios. These scenarios define the set of agents that represent the typical actors in the investment processes. To play a game, each player is first assigned to one or more agents. The agents have limited autonomy in that they can make trivial and routine decisions. Otherwise, if made by the actors', these decisions would drastically slow down the game. However, complex decisions that are difficult to model in sufficient details (non-trivial and based on human factors) are still taken by the stakeholders who control the agents. Agents' decisions can, at all times, be overturned by the controlling stakeholders.

Also modeled are the physical elements and their characteristics, for e.g. territory (map where gas infrastructure is to be placed), production, storage and transport equipment, as well as the economic elements, for e.g. price elasticity. These elements are obtained apriori from investment scenarios and domain experts. The games are enacted by using an interaction model, defined in the agent interaction modeling language TALL (Stuit and Szirbik 2009). A set of predefined interactions and typical patterns of hierarchies of interactions are to be modeled apriori as a library of interactions of the platform. In the introduction of this paper much attention was given to the role of the regulator and the government in investment decisions. Accordingly, the game has built-in capabilities to support these stakeholders playing their roles.

The Game Execution Functions

The game interaction model allows for a high degree of flexibility, and the situation on the map evolves as the decisions are made by the agents and their assigned players. The realism of evolving the physical and financial situation is continuously checked by appropriate models (which are not indicated in the architecture explicitly, but are part of the domain models). The "game instance execution" functionality allows the DEM and the stakeholder to reach a (positive or negative) conclusion in the game. It also offers a virtual game clock (which for gas investment games covers huge time spans, like 20 to 40 years), and a real-time clock. A game, if played synchronously, can have a real duration of 3 to 5 hours (limited by player fatigue). Asynchronously, a game can span over weeks, depending on the players commitment and available time. The game can be organized in a centralistic way, where stakeholder teams are in the same room (so called LAN party - synchronous), and can see the map situation on a large screen. Each player has an individual computer allowing him/her to participate in playing the game. Alternatively, the game can be organized in a distributed setting, where players access remotely the platform via web-browser based interfaces (asynchronous).

Knowledge and Behavior Capture Functions

During the game, players communicate and negotiate with each other via chat-boxes. Communication can be private (peer-to-peer) or public (broadcast). The players express themselves in (written) English, including informal constructs, such as emoticons, and abbreviations like "IMHO". These communications are saved during a game session, and analyzed using natural language processing (NLP) techniques.

Our NLP algorithms for conversation analysis are hinged on the notion of speech acts (Searle 1969). Speech acts are an idiosyncratic type of utterances that implicitly perform an action, express different attitudes and decisiveness levels, for e.g. "I regret….", "I want…", "I wish…" or "I suggest…". Thus, they convey useful information about the intent of the stakeholders during negotiations, which enable us to better understand the decision making dynamics underlying energy investments.

Example Application

A chat-box, for players to negotiate, implemented as part of our simulation gaming platform is depicted in Figure 3. It illustrates a negotiation between 2 parties, namely a consumer (consumer1) and a grid operator. Both parties are negotiating in private (i.e. peer-to-peer), and they can see the messages sent by each other.

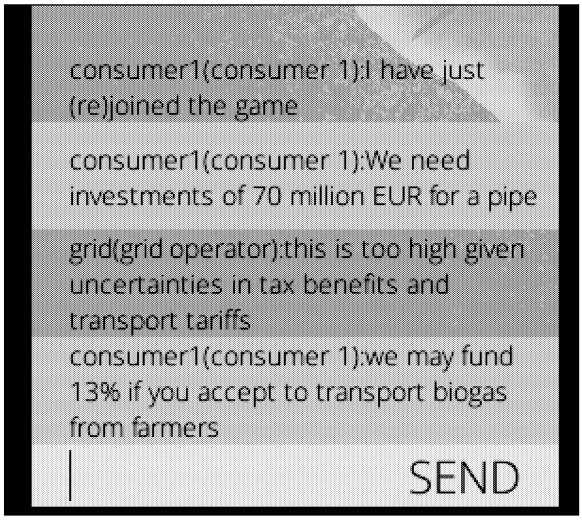


Figure 1: Negotiation via Chat-Boxes

In the 1st utterance, the consumer states a fact about a particular amount of investment. This utterance is classified by our algorithms as an Assertive, and will be awarded a relatively high degree of decisiveness. In the 2nd utterance, the grid operator reacts to the consumer. It is classified both as an Assertive and an Expressive. In the 3rd utterance, the consumer is willing to lower the amount under a condition. Thus, this utterance is classified as Commissive.

As depicted in the above example, our NLP algorithms automatically determine the types intentions conveyed by the utterances, and identify relevant linguistic markers, such as "may offer" or "must have", that can either thwart negotiations or enable them to succeed. This information can then be used in real-life scenarios to enable stakeholders to align their conflicting assumptions and expectations, which facilitate the investment negotiations in the face of uncertainty, and lead to market growth. Details on the NLP techniques are briefly described in the next section.

SIMULATION GAMING PLATFORM IMPLEMENTATION

The proposed simulation gaming platform is developed by a consortium of industrial and academic partners. The game user-interface and 3D components are implemented in Unity 3D FX with SmartFoxServer as the backend server. The AIMMS ("Advanced Interactive Multidimensional Modeling System") mathematical optimization package is employed for the financial/economic models. A partial example of the "game board" (and chat-box) is depicted in Figure 2. The elements on the board correspond to pipes, factories and digesters that are placed in the board by the players during the game.

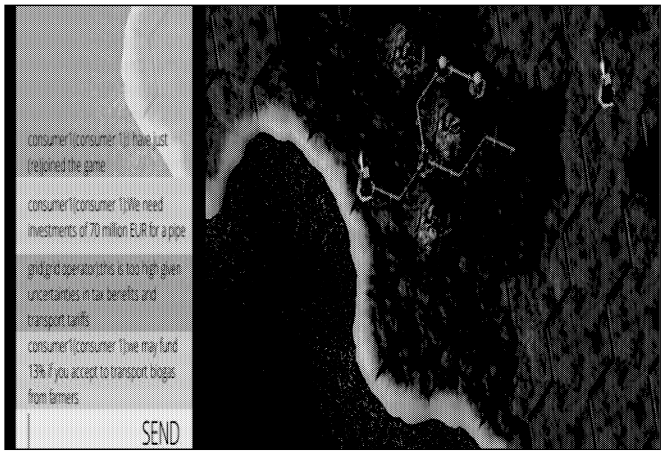


Figure 2: Sample Game Board

To learn about speech acts and their intentions and to generate a corresponding computational model, our NLP algorithms rely on various linguistic and statistical techniques. As input, we use 2 well-known text collections that contain utterances (conversations). Each utterance is further annotated with the intentions that it conveys. The text collections are, namely, the Switchboard corpus (Godfrey, Holliman and McDaniel 1992) and Meeting Recorder Dialog Act corpus (Janin, et al. 2004). Then, we linguistically process the utterances, for e.g. by determining the part-of-speech (POS) of words in the utterances, using the Stanford Part-of-Speech tagger (Toutanova, et al. 2003). Next, for each intention type, we extract as features, sequences of n-words (known as n-grams; currently n=2) together with their POS-tag, for e.g.
<"I|PRP" , "WANT|V">, where PRP is the POS-tag for pronouns (e.g. "I") and "V" is for verbs (e.g. "want").

We then identify the most promising features per intention type by computing their pointwise mutual information (PMI) scores, according to equation (1). It measures the association strength between an intention type, c, and a given, feature, f,. The value p(c|f) is the conditional probability of c given f, and is estimated based on the frequency counts obtained from the Switchboard and MRDA collections.

$$PMI(f,c)=\log\ (p(c \mid f))/p(c) \qquad (1)$$

Finally, a computational model that maps utterances to the intentions they convey is generated using Support Vector Machines (Joachims 2002). Given a new utterance made by a player during a game, the model extracts its features and predicts the utterance type (e.g. Assertive, Commissive).

CONCLUSION

In order to cope with the necessary enhancements and transformations of the current energy infrastructures huge investments have to be done. In times where regulatory and government policies shift constantly, negotiations between related stakeholders – all with different objectives – are hampered by uncertainties and mostly end-up in unwanted results. In this paper, we explained that the classical approaches of Operational Research have severe limitations in analyzing regulatory uncertainties. Thus, we proposed a simulation gaming approach to analyze multi-stakeholder negotiations under uncertainties; the key novelty being the application of Natural Language Processing techniques. We expect that the information extracted from the negotiations during the games can be used in real-life for stakeholders to better understand each other and to align their conflicting objectives. Another added value of our approach is that the process of capturing and analyzing stakeholder interactions is completely automated, which provides scalability and speed. This can be beneficial in intensive, time pressing negotiation situations. Furthermore, because of its formal/computational groundings, our approach can be considered as being more accurate, objective (i.e. less subjective), and traceable than manual analyses. Overall we expect that our approach will foster the conditions for the necessary energy market growth.

FUTURE WORK

Initially, we will focus on the energy market, as described in this article. In future work, we will investigate the portability and generalizability of our platform. Since the proposed platform and the underlying simulation gaming methodology is generic, it can, in principle, be applied to support decision making for infrastructure transformation or expansion in other domains. Practical applications can be found, for example, in investment decisions in public transport, such as the discussion about the introduction of tram infrastructure in Dutch cities like Nijmegen and Leiden, the enhancement of European high speed trains or the upgrading of the broadband telecommunications networks.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of Wilrik Mook in developing the serious gaming platform and in providing valuable feedback that have improved the quality of this article.

This research has been financed by a grant of the Energy Delta Gas Research (EDGaR) program. EDGaR is co-financed by the Northern Netherlands Provinces, the European Fund for Regional Development, the Ministry of Economic Affairs, Agriculture and Innovation and the Province of Groningen

REFERENCES

Barradale, M.J. 2010. "Impact of public policy uncertainty on renewable energy investment: Wind power and the production tax credit." *Energy Policy* 38, no. 12, 7698--7709.

Duke, R.D., and J Geurts. 2004. *Policy games for strategic management.* Dutch University Press.

Fabrizio, K.R. 2012. "The Effect of Regulatory Uncertainty on Investment: Evidence from Renewable Energy Generation." *Journal of Law, Economics, and Organization* (Oxford Univ Press).

Godfrey, J. J., E.C. Holliman, and J. McDaniel. 1992. "SWITCHBOARD: Telephone speech corpus for research and development." *IEEE International Conference on Acoustics, Speech, and Signal Processing*.

Holburn, G.L.F. 2012. "Assessing and managing regulatory risk in renewable energy: Contrasts between Canada and the United States." *Energy Policy*, 45, 654-665.

Janin, A., J. Ang, S. Bhagat., R. Dhillon, J. Edwards, J. Macias-Guarasa, N. Morgan, B. Peskin, E. Shriberg, A. Stolcke, and others. 2004. "The ICSI meeting project: Resources and research." *Proceedings of the 2004 ICASSP NIST Meeting Recognition Workshop.*

Joachims, T. 2002. *Learning to Classify Text using Support Vector Machines*. Springer.

Li, YF, GH Huang, YP Li, Y. Xu, and W Chen. 2010. "Regional-scale electric power system planning under uncertainty—A multistage interval-stochastic integer linear programming approach." *Energy Policy* 38, no. 1, 475--490.

Pelletier, C. and Wortmann, JC. 2009. "A risk analysis for gas transport network planning expansion under regulatory uncertainty in Western Europe." *Energy Policy* 37, no. 2, 721--732.

Roest, G.B., and N.B. Szirbik. 2009. "Escape and Intervention in Multi-Agent Systems." *Journal of AI & Society* (Springer) 24, no. 1, 25--34.

Searle, J.R. 1969. *Speech acts: An essay in the philosophy of language.* Cambridge Univ Press.

Stuit, M., and N.B. Szirbik. 2009. "Towards agent-based modelling and verification of collaborative business processes: An approach centred on interactions and behaviours." *Intl. Journal of Cooperative Information Systems* (IJCIS-World Scientific Publishing) 18, no. 3-4, 423--479.

Susi, T., M. Johannesson, and P. Backlund. 2007. "Serious games--An overview." Technical Report, University of Skövde.

Toutanova, K., D. Klein, C. D. Manning, and Y Singer. 2003. "Feature-rich part-of-speech tagging with a cyclic dependency network." *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*.

Verderame, P.M. and Elia, J.A. and Li, J. and Floudas, C.A. 2010. "Planning and scheduling under uncertainty: a review across multiple sectors." *Industrial \& engineering chemistry research* 49, no. 9, 3993--4017.

Zeng, Y., Y. Cai, G. Huang, and J. Dai. (2010). "A Review on Optimization Modeling of Energy Systems Planning and GHG Emission Mitigation under Uncertainty." *Energies* 4, no. 10, 1624--1656.

HYBRID SIMULATION OF MOBILE BUSINESS PROCESSES

Stefan Hellfeld
FZI Research Center for Information Technology,
Manager of the FZI House of Living Labs,
Haid-und-Neu-Str.10-14,
76131 Karlsruhe,
Germany
E-mail: hellfeld@fzi.de

KEYWORDS
Simulation, mobile, business process, Petri-net

ABSTRACT

A sustainable integration of mobile devices, especially smartphones and tablet pc, into day-to-day business requires an adaption or a redesign of the underlying business processes. The redesign should focus the efficient and effective usage of the mobile devices used by the mobile executing authorities in the business processes. The dynamically changing situations of the executing authorities and the impact on the usage of the mobile devices have to be considered in the redesign of the business processes. Methods for the analysis of the efficient and effective usage of mobile devices in business processes are missing. The presented approach enhances the existing conventional simulative analysis of a business process by the possibility to simulate specific task executions of mobile executing authorities in reality. This improves the analysis of business processes, in which mobile executing authorities are supported by mobile devices in their task execution.

INTRODUCTION

According to a current forecast, 350 million people will use their smartphone for executing operational tasks in 2016 (Schadler and McCarthy 2012). As a result for the enterprises, the usage of mobile devices has to be considered in the underlying business processes. Therefore the business processes have to be adapted or redesigned. For analysing the adapted or redesigned business processes, concerning the usage of mobile devices and the identification of improvements, the business process models can be software-based simulated (Oberweis 1996). However to consider every environmental factor, which could influence the task execution and thus the use of the mobile device in the business process, results in a big overhead in the software-based simulation. Therefore it needs a method for simulating the relevant sub-processes of the business processes, in which the executing authorities are mobile. This simulation should have the opportunity to test the task execution supported by a mobile device in reality for gathering all contingencies, which could influence the efficient and effective usage of the mobile devices. The presented approach in this paper proposes a solution for this problem, in form of an extension of the software-based simulation of business processes.
The paper is structured as follows: In the second part, the analysis of mobile business processes, respectively the models of mobile business processes, is introduced.. In the third part of the paper the new approach of the hybrid simulation of mobile business processes is explained. In the fifth part a short summary of the tool-supported usage of the method in practice is given. The sixth part concludes the paper.

ANALYSIS OF MOBILE BUSINESS PROCESSES

The approach presented in this paper focusses business processes modeled as Petri-nets. A Petri-net is a directed bipartite graph, in which the two disjoint node sets of places and transitions are connected by arcs with each other. The places represent local states in the business process. The transitions represent state changes or activities. The directed arcs connect places and transitions and represent the flow relation in the business process. A place can only be connected with a transition and vice versa (Desel et al. 2004; Reisig 2005).
In literature, different methods for analyzing Petri-net based business process models are mentioned. Methods for the qualitative analysis of a business process, respectively the business process model, can analyze the validation and verification of the structure of the business process model. Methods for the quantitative analysis can validate the achievement of objectives of the business process, related to predefined parameters. Examples for the quantitative analysis are markov-chains or the queueing theory (van der Aalst and van Hee 2004; Neumann and Morlock 2002). Another example is the method of the computerised simulation, which allows a "virtual play through" of the execution of single activities in the business process (Oberweis 1996).
Business processes, in which executing authorities are mobile (so called mobile business processes), can also be analysed by a computerised simulation as well as "normal" business processes. Mobile business processes have the characteristic, that the situations of the executing authorities can dynamically change because of their mobility. These changes can affect the execution of specific tasks in the business process (Mourão and Antunes 2004; Trunko 2011). The identification of the entirety of the environmental

factors, influencing the task execution, is either not practicable or only realizable with a significant effort. To reduce the effort and to improve the simulation of mobile business processes, the method of hybrid simulation of mobile business processes can be used.

HYBRID SIMULATION OF MOBILE BUSINESS PROCESSES

The hybrid simulation of mobile business processes is a new form of simulation and is not comparable with the hybrid simulation known as a combination of the continuous and discrete-event simulation of business processes. The method of hybrid simulation extends the conventional simulation of a business process model. With hybrid simulation the usage of a mobile device by an executing authority in reality is simulateable. The executing authorities have the possibility, to test the usage and the support of a specific mobile device in their task executions (for example using the mobile device for preventing a media disruption). These "tests" are realized within hybrid simulation. The method of hybrid simulation is defined as follows:

The hybrid simulation is a specific extension of the computerised simulation of Petri-net based business process models. Parts of the simulation of the business process were executed in reality. The realistic simulated activities are realized as mobile human-computer interactions between a human executing authority in a role in the business process and a mobile device.

So, the method of hybrid simulation enables a virtual simulation as well as a simulation in reality. The virtual simulation is realized by an existing business process engineering software, extended by a tool, which implements the hybrid simulation (described in detail in the following section “Use in practice”).
The aimed improvements in the business process, which should be realized by the usage of mobile devices, are specific task executions. These task executions represent in hybrid simulation human-computer-interactions among executing authorities and their mobile devices in the business process. These interactions are simulated in reality during hybrid simulation. The necessary steps for realizing hybrid simulation are splitted into three stages, which will be described in the following sections.

Stage One: Interaction identification

The interaction identification is the first stage of the method of hybrid simulation. In this stage the business process model is prepared for hybrid simulation. The user of the hybrid simulation, respectively the modeller, has to identify tasks in the business process model, which could be supported by the usage of a mobile device. There exists two types of interactions with mobile devices: Implicit and explicit interactions (Schmidt 2000). An implicit interaction describes the sensor data detection in a mobile device. An explicit interaction describes the interaction of an human with the device, for example a text input. These interactions can be influenced by environmental factors and should be consequently simulated in reality.
An identified task in a Petri-net-based business process model, which can be identified by the modeller, is either a single transition (activity) or a summary of transitions, a transition-edged subnet (see figure 1). If a transition-edged subnet is identified, the atomar activities are mapped onto one single interaction with a mobile device (supported by one single user interface for realizing an explicit interaction).

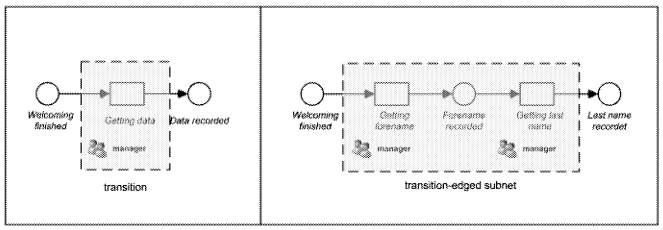


Figure 1: Identified transitions in a Petri-net based business process model

The identified tasks are called *hybrid scheme components*. These hybrid scheme components encapsulate the information of the type of interaction. During the conventional simulative analysis of the business process model, the hybrid scheme component are sent to the corresponding mobile devices and interpreted by a specific application installed on the device. So the interactions can be simulated in reality. A hybrid scheme component is defined as follows:

A hybrid scheme component is a document, formatted in the extensible markup language, XML (World Wide Web Consortium 2002), which represents a mobile human-computer interaction among an human executing authority and a specific mobile device.
It can contain either the definition of an user interface for enabling the interaction or the definition of the hardware component especially the sensor, from which data should be detected. A hybrid scheme component is interpreted by specific apps on mobile devices. The result of the interaction is saved in the hybrid scheme component.
The hybrid scheme component is represented in the business process model by a single transition or a transition-edged subnet. The hybrid scheme component is assigned to a single executing authority in its specific task execution.

In case of mapping the identified tasks onto interactions, the user of the method of hybrid simulation have to identify the type of interaction. Choosing an implicit interaction requires the assignment of the sensor, from which data will be detected. Choosing an explicit interaction requires the definition of an user interface.
The method of hybrid simulation assists the user in choosing user interfaces for the corresponding interaction by giving recommendations. These recommendations are based on studies communicated in literature, which analysed the day-to-day usage of mobile devices and the used modalities/user interfaces. The interaction of text input

(MarketingCharts 2011) and the interaction of selecting objectives are executed most often in day-to-day usage (Little 2010; Comscore 2011; Butts and Cockburn 2002; James and Reischel 2001). Using interactions for task execution, which are already known from day-to-day usage of a mobile device, can significantly reduce the training period of an human learning a specific interaction (Roto 2010; Wellman 2007). Therefore the method of hybrid simulation recommends these interactions. More user support is given by the possibility to choose the user interface out of an amount of already designed user interfaces for enabling a text input or an objective selection interaction. Two examples of these designed user interfaces are shown in figure 2 (others are shown in figure 5).

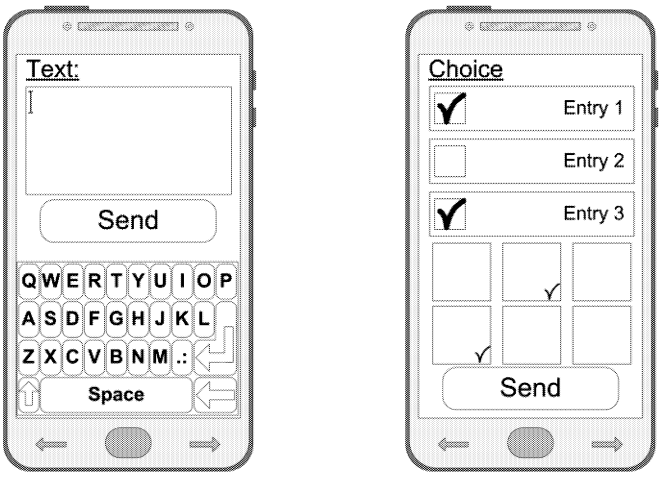


Figure 2: Already designed user interface for use in the hybrid simulation

If the user wants to realize an interaction, which is not enabled by the already designed user interfaces, he/she has the possibility to design an individual user interface. The user is also supported in designing that interface by the method of hybrid simulation. This support is realized in stage two of the method, the interaction support.

Stage Two: Interaction support

In the stage of interaction support the user of the method of hybrid simulation gets support in generating a hybrid scheme component, especially the user interfaces for realizing an explicit interaction, if needed.
For supporting more than one mobile device with a preferably minimum of effort, the method of hybrid simulation uses the following specific provisions:

- A Client-Server Architecture: The tool, in which the method is implemented, works with a client-server architecture for task sharing. As a consequence a maximum of independency of the different mobile device specifications is reached.
- Reuse by means of a declarative markup language: The reuse of the interfaces on different mobile devices is realized by transforming the interaction-enabling user interfaces in specific documents formatted in a declarative markup language. So the user interfaces have not to be designed in detail for each specific mobile device.
- Simplified representation: The process of designing the user interfaces is significantly simplified by choosing controls out of an amount of already designed controls, with which the user can built up an user interface without having knowledge of how to programm the interface in the device or operating system specific programming language.

Figure 3 shows an overview over the different provisions for designing user interfaces in the stage of interaction support.

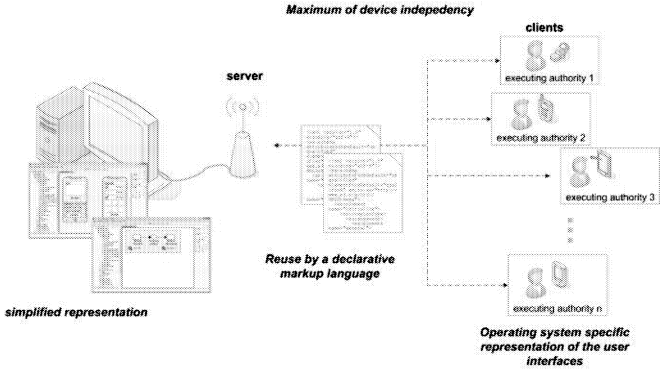


Figure 3: Overview of the provisions in the stage of interaction support.

As written in literature, a mobile device can be used more efficiently and effectively, if the user interfaces of the applications are designed and layouted simliar to the user interfaces of the underlying operating system running on the mobile device (Roto 2010; Wellman 2007; Olson et al. 2012; Verclas and Linnhoff-Popien 2012; Puscher 2010; Woodcock 2011). Therefore in hybrid simulation the designed user interfaces supporting the explicit interactions have the design and layout of the underlying operating system. As a consequence a single user interface have to be designed for every identified explicit interaction and for every device with its operating system. This would lead to an overhead in designing user interfaces. That's the reason, why the method of hybrid simulation uses different mechanisms/provisions for enabling the reuse of the user interfaces and reducing the effort of designing and layouting user interfaces by the user.
The user of the method of hybrid simulation has not to be aware of the different specific designs and layouts of the operating system controls, because the tool of the method of hybrid simulation supports designing user interfaces in a generic representation. This means, that the different controls have all a generic design.. Figure 4 shows an example how the design and layout of an user interface with generic controls is distinguished from an user interface in the design and layout of the Google® Android operating system in version 4.2.
Beyond that, the complexity of designing the user interfaces is also reduce by the fact, that the amount of suitable controls in the method of hybrid simulation is an intersection of controls of different mobile operating systems. This intersection includes controls of the Apple®

operating system iOS 5.0.1, the Android operating system in version 4.0 from Google®, the Windows Phone operating system from Microsoft®, the Blackberry OS 7 from Research in Motion® and the operating system Bada 2.0.4 from Samsung®. These selected operating systems cover more than 99 percent of the used operating systems on mobile devices world wide (Cozza and Milanesi 2012). So designed user interfaces with controls taken out of this intersection are displayable on more than 99 percent of the used operating systems on mobile devices.

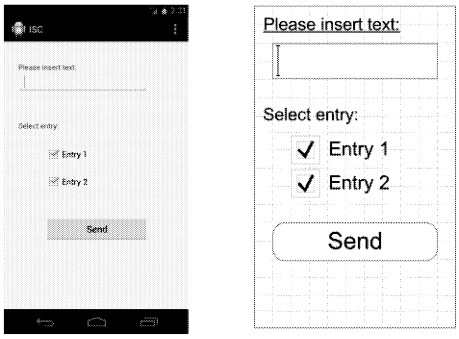


Figure 4: An user interface in operating system specific and generic design.

Based on the recommendations for the "right" interactions (as already explained above), the method of hybrid simulation supports the user with a preselection of already designed user interfaces. Figure 5 shows an extract of already designed user interfaces, supporting different text input and objective selection interactions.
For designing individual user interfaces the method of hybrid simulation provides an extended approach of prototyping (Frank, Naugler, and Traina 2005; Gomaa and Scott 1981; Tudhope, Beynon-Davies, and Mackay 2000). A simple editor based on the "What you see is what you get"-principal (Preim and Dachselt 2010). enables the designing of user interfaces without any knowledge about their programming in the operating system specific programming language.



Figure 5: Already designed user interfaces for hybrid simulation

For reusing the designed user interface on more than one mobile device in the operating system specific design and layout, the interface is transformed in a representation formatted in a declarative markup language (Reid 1980). This language was developed as a simple XML derivative. The whole hybrid scheme component, including the description of the graphical design and layout of the user interface supporting the explicit interaction is transformed in a document, formatted in this declarative markup language. So the interfaces only have to be interpreted on the mobile devices.
By using a client-server-architecture in the tool a high independency of the specific devices is reached. Specific operating system- and device-dependent applications interpret the hybrid scheme components containing the presentation of the user interfaces in the operating system-specific design and layout and the sensor-based data detecting of a hardware sensor in the mobile device. This interpretation is executed in the last stage of the method of hybrid simulation.

Stage Three: Interaction interpretation

The interaction interpretation is the last stage of the method of hybrid simulation, in which the hybrid scheme components are interpreted by specific applications on the mobile devices. These applications have access to the different hardware components on each mobile device. Depending on the content of an hybrid scheme component, the applications are executed in various ways. If an implicit interaction is defined, the application detects data from the appropriate sensor and integrates the result into the hybrid scheme component. If an explicit interaction is defined in the hybrid scheme component, the controls, which were described in the declarative markup language, are mapped onto operating system specific controls. These controls are displayed to the executing authority in manner of supporting its task execution.
To collect all possible data about a situation of an executing authority, the method of hybrid simulation uses so called *reports*. Reports are recorded sensor data of the actual situation the executing authority is in during its task execution. Data from all sensors is detected by the application running on the mobile device. The user of the method of hybrid simulation/the modeller can use this data for characterization of the situation, and to optionally adapt the usage of the mobile device in terms of an efficiency and effectiveness (for example is the usage of a sensor in an activity of the business process senseless, if it cannot detect valid data in the specific situation).
The identification interpretation is the last stage of the method of hybrid simulation. The different stages are executed tool-supported during and as an extension of the conventional computerised simulation of the business process model.

USE IN PRACTICE

The method of hybrid simulation was implemented as a plugin for the business process engineering software *horus*

business modeler. In this software the conventional simulative analysis is combined with the hybrid simulation described in this paper. The tool and the underlying the method could be evaluated in different projects with industrial partners. The result was an enhancement in the efficience and effectiveness of the usage of the mobile devices in each business: For example the usage of sensors for automation of working steps could be simulated in reality and the usage could be improved if the sensors didn't detect valid data. This was realized in a project with an industrial partner. In the business process the execution authorities should use the data of a smartphone sensor for locating themselves near a client address. Because of environmental factors in form of buildings around the position of the executing authority, the sensor was not detecting valid data, so the smartphone could not be used in the business process for an improvement of the business process. So the usage of the mobile device could be optimized. The efficience and effectiveness of the usage of the mobile device was enhanced by hybrid simulation.
Another example was the restructuring of working steps of executing authorities in a business responsible for the garbage collection in a city. Mobile devices should be used to give the executing authorities a hint of the place of the dustbins. With hybrid simulation, it was possible, that the the executing authorities can give feedback, that the initial installation of the device at the garbage collection car requires a new working step of walking to the car after every clearance of the dustbins. This feedback was giving after hybrid simulation. So the business process with its working steps could to be redesigned and the mobile device was installed on the arm of the executing authority. So the working step of walking to the garbage collection car was no longer necessary.
These examples illustrate that the method of hybrid simulation improves the integration of executing authorities in redesign and adaption of business processes.

CONCLUSION

The method of hybrid simulation is an extension of the conventional computerized simulative analysis of business processes modeled as Petri-nets. The method allows an analysis of the dynamic changing situations, in which the executing authorities perform their task execution and in which they use the mobile devices for supporting their task execution. With the method of hybrid simulation the executing authorities getting a possibility to participate in the redesign and adaptation of the business processes in regard of the efficient and effective usage of mobile devices.

REFERENCES

van der Aalst, W. and van Hee, K., 2004: *Workflow management. Models, methods and systems*. MIT Press, Cambridge, Mass..

Butts, L. and Cockburn, A. 2002: An Evaluation of Mobile Phone Text Input Methods. In: *Grundy, J.; Calder, P.: Third Australasian User Interface Conference (AUIC2002)*.

Cozza, R. and Milanesi, C. 2012: *Market Share. Mobile Devices by Region and Country*.

Desel, J.; Reisig, W.; and Rozenberg, G. 2004. *Lectures on concurrency and Petri nets: Advances in Petri nets*. Springer, Berlin.

Frank, C. E.; Naugler, D.; and Traina, M. 2005: Teaching user interface prototyping. In: *Journal of Computing Sciences in Colleges*, Vol. 20; Page 66-73.

Gomaa, H. and Scott, D. B. H. 1981: Prototyping as a tool in the specification of user requirements. In: *Proceedings of the 5th international conference on Software engineering*; Page 333-342.

James, C. L. and Reischel, K. M. 2001: Text input for mobile devices: comparing model prediction to actual performance. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*; Page 365-371.

Mourão, H. and Antunes, P., 2004: Exception Handling Through a Workflow. In: *Meersman, R. and Tari, Z.: On the Move to Meaningful Internet*. Systems 2004: CoopIS, DOA, and ODBASE; Page 37–54.

Neumann, K. and Morlock, M, 2002.: *Operations Research*. Hanser.

Oberweis, A., 1996: *Modellierung und Ausführung von Workflows mit Petri-Netzen*. Universität Karlsruhe, Habilitation. Teubner, Stuttgart. [in German]

Olson, S.; Hunter, J.; Horgen, T. H.; Horgen, B.; and Goers, K. 2012: Professional Cross-Platform Mobile Development in C#. John Wiley & Sons, 2012.

Preim, B. and Dachselt, R. 2010: *Interaktive Systeme: Grundlagen, Graphical User Interfaces, Informationsvisualisierung*. Springer. [in German]

Reid, B. K. 1980: A high-level approach to computer document formatting. In: *Proceedings of the 7th ACM SIGPLAN-SIGACT symposium on Principles of programming languages;* Page 24-31.

Reisig, W. 2005. *Petri Nets: An Introduction*, Springer-Verlag, Berlin.

Roto, V. 2010: *Web Browsing on Mobile Phones*. Lambert Academic Publishing.

Schmidt, A. 2000: Implicit human computer interaction through context. In: *Personal Technologies*, Vol. 4(2-3); Page 191–199.

Trunko, R. 2011: *Kontextsensitive Ausnahmebehandlung in Geschäftsprozessen*. Dissertation. Karlsruhe, 2011. Dr. Hut, München. [in German]

Tudhope, D.; Beynon-Davies, P.; and Mackay, H. 2000: Prototyping praxis: constructing computer systems and building belief. In: *Human-Computer Interaction, Vol. 15(4)*; Page 353-383.

Verclas, S. and Linnhoff-Popien, C. 2012: *Smart Mobile Apps: Mit Business-apps ins Zeitalter Mobiler Geschäftsprozesse*. Springer-Verlag New York Inc.. [in German]

Woodcock, J. 2011: *App Design: How to Conceive, Design, and Market Bestselling Apps*. Book Sales, Inc..

WEB REFERENCES

Comscore 2011: *The comScore 2010 Mobile Year in Review*. http://www.comscore.com/Insights/Presentations_and_Whitepapers/2011/2010_Mobile_Year_in_Review, [2013-01-21]

Little, A. D. 2010: *Mobile Applications in Germany*. http://www.adlittle.de/uploads/tx_extthoughtleadership/ADL_TIME_2010_German_Apps_Study.pdf. [2013-01-14]

MarketingCharts 2011: *Texting Top Smartphone Activity*. http://www.marketingcharts.com/direct/texting-top-smartphoneactivity-18152/, [2012-01-15]

Puscher, F. 2010: *App-Design für Touchcomputing: Gesten und Symbole für gängige Aktionen*. http://www.drweb.de/magazin/app-design-fuer-

touchcomputing-gesten-und-symbole-fuer-gaengige-aktionen/, [2013-02-01] [in German]

Schadler T. and McCarthy J. C. 2012: *Mobile Is The New Face Of Engagement* http://www.forrester.com/Mobile+Is+The+New+Face+Of+Engagement/fulltext/-/E-RES60544?objectid=RES60544. [2013-01-20] [in German]

Wellman, S. 2007: *Google Lays Out Its Mobile User Experience Strategy*. http://www.informationweek.com/mobility/business/google-lays-out-its-mobile-user-experien/229216268. [2013-02-20]

World Wide Web Consortium, 2002: Extensible Markup Language (XML) 1.1 (Second Edition), In: *W3C Recommendation*. http://www.w3.org/TR/xml11/. [2013-02-23]

HOW TWO MANAGEMENT ENGINEERS AND A SIMULATION SOFTWARE CAN SUPPORT THE ORGANIZATION OF A NEW SURGICAL PRE-ADMISSION UNIT

Giorgio Romanin-Jacur and Federico Segala
Department of Management and Engineering
University of Padova
I 36100 Vicenza
Italy
E-mail: romjac@dei.unipd.it

KEYWORDS
Health care, decision support, discrete simulation, personal computers.

ABSTRACT

We report a practical experience in a working hospital. Every surgical operation is preceded by a pre-admission process, destined to analyze patients conditions, which involves many different departments beside the one executing the operation. We started from an organization which provided independent processes for every surgical department, which asked elsewhere for inspections and consulting. A transformation was effected, by designing a new unit which manages a centralized process, utilized by all surgical departments, where all other interested departments, services or units work together in synchrony. That permits to save time and movements of patients inside the hospital, and to reduce total pre-admission times, thus obtaining a better utilization of hospital resources. In order to rule a transformation which required the collaboration of many different groups independently operating before, an accurate design and management control was operated by simulation. After a quantitative analysis of process phases, a new model was built and coded in language Arena. The model evidenced critical aspects suggested useful modifications and adaptations which permitted to get a better working. We think the experience from a specific case may be easily extended to other similar ones.

INTRODUCTION

In a situation where large amounts of resources and money were available, and the amount and complexity of surgical operations were smaller than today, every department executing surgical operations was equipped with its own operating plate (or plates) and managed it (or them) and the interested patients in a totally autonomous way. As reported for instance in (Holt et al. 2007, Pollard et al 1997, Pollard et al 1999, Van Klei et al. 2002) the "preoperative process", consisting in a sequence of exams, visits, interviews with the scope of an accurate knowledge of patient's clinical conditions, is absolutely necessary and permits to ameliorate surgical operation and to reduce length of stay. Some of the visits, interviews and exams are performed inside the same department which will effect the surgical operation, but many of them are to be performed outside the same department, take for instance cardiological visits, or by personnel of another hospital service, take for instance x-ray and anaesthesiology visit.
A first interesting change took place in many hospitals when some, and more often all, operating plates were shared among many or all departments; such a change permitted a remarkable saving for what concerns both technical and human costs, due to increased efficiency.
More recently efforts were made to perform all pre-admission processes interesting surgical patients in an only unit serving the whole hospital, with consequent increase in the connected efficiency. In the present paper the story of the building and management of a new pre-admission process unit is reported, evidencing critical phases and interesting results.

GENERAL PROBLEM

Before the above pre-admission process transformation, every department or department set executed all visits or exams connected with its own operations in a totally independent way, both from the clinical and from the organizational point of view. As a consequence we could remark a dramatically irregular work schedule, due both to the normal variability of surgical operations' schedule, and to the fact that the majority of exams were executed out of the concerned department; see for instance ematochimic exams, executed in a specialized laboratory, x-ray, performed in radiology department, electrocardiogram, made inside cardiology department, anaesthetic visit, performed in the specialized anaesthesiology ambulatory; obviously every exam or visit performed outside had to follow all rules and times of the related department, service or unit, and that caused a lot of time to wait, move patients and transmit results in an ordered way.
According to the new conception, all pre-admission visits, exams and interviews are managed by a new independent unit for all hospital surgical patients. Patients are managed in a dedicated continuous place, where they come only once to get all needed services in sequence in a comparatively short time, permitting to save time and money. Obviously such a new unit needs to be properly organized, in order to arrange many different operators, coming from different departments or units, to synchronize all operations and to schedule interested patients, so to save time and to minimize queues; in the same time, an accurate planning of operator

turns of duty has to be performed to minimize employed personnel and technical resources cost.

OLD AND NEW MODELS

We started from the old organization in a public hospital placed in North-East Italy, characterized by six different departments or department sets executing surgical operations: 1)Obstetrics and Gynecology, 2)General Surgery, 3)Urology and Otolaryngology, 4)Orthopedics and Traumatology, 5)Oftalmology, 6)Day surgery.
For every one above we accurately analyzed the list of pre-admission examinations and visits; because of space limits, we report only the one related to general surgery: 1)general interview (anamnesis) with writing of nursery board and pre-admission scheduling, 2)ematochemical exams, 3)thorax x-ray (for every patient over 30, on demand for younger ones), electrocardiogram, anaesthetic visit, possible further required visits or consulting; for every one time probability distribution was obtained from the previous year registers. By examining all lists, we realized a total of twenty visits or exams common to almost all departments; Obstetrics and Gynecology require in addition two different sets of specific exams or inspections; Day surgery requires only the anaesthetic visits, with addition of other common exams whenever required by the anaesthesiologist; all surgical operations require the anaesthetic visit. The second analysis concerned every department work charge, i.e., number of operated patients, which revealed that their number presents small variations during the year, for all departments.
A model of old organization appeared extremely difficult to build and totally useless, given the absolute managerial independence of every department, service or unit effecting required operations; patients were asked to be present at 8 and wait until the first operation was over, then, after the first examination, move to another place for the second one and wait until scheduled, and so on till the end of the day, to return the following day if all visits had not been effected, depending on operators' turns of duty. We could estimate that in the most favourable situation, corresponding to minimum queueing time, the average patient stay time was one day (10 hours) and 6 hours: apart from the necessity of coming two times, a lot of unpleasant movements inside the hospital was required.
With the scope of process centralization the hospital instituted the new unit, to perform all examination and visits, and gave it a dedicated place, with a first room for the acceptance and compilation of the nursery board (two desks), a second room for electrocardiogram and withdrawals (four beds), a third room for anaesthetic visit, placed beside a waiting corridor; the only outer exams remained the x-ray and the eco ones, performed in the radiology department, which is placed very close to the unit location and directly reachable. The next task concerned the construction of patients and operators scheduling to reach the following objectives: minimizing patients' total waiting time and returns to already crossed ambulatories, maximizing operators' and technical resource efficiency. The structural decisions concerned dimensioning of ambulatories and desks, the functional decisions concerned patients' arrival times and operators' turns of duty and assignment to ambulatories and desks. To make some decisions easier, a new operation was performed: all patients were no more grouped according to the concerned department, but to the pre-admission process "path", i.e., the sequence of visits or exams to be performed; they were grouped into seven classes characterized by probability; for every class we listed the visits' sequence, with related time distributions and involved human and technological resources' list. A first special class is added, concerning Obstetrics and Gynecology patients, getting specialized examination inside the interested department and joining other patients for other more general visits or exams; a second special class concerns Day surgery patients, for whom the list is reduced.
In order to support structural and functional decisions, a simulation model, coded in Arena, particularly easy to be understood by non experts (doctors and nurses) and implementable on Personal Computers, was built up, made up of about ninety blocks and reported in the figure below. The model, highly detailed, is characterized by independent sections, which simulate single parts of the pre-admission process. Sections are listed from left to right ad from high down:
-Withdrawal results;
-Patient creation: independent parallel blocks for every day of the week for normal patients and separate blocks for patients of Obstetrics and Gynecology department and of Day surgery; created patients are assigned the class according to probability defined by the week plan; the assigned class is associated with the sequence of exams and visits to be performed;
-Electrocardiogram, with possible exit for DS surgery;
-Ematochemical and Urine examinations; it includes blocks representing sample withdrawal and decision about next visit for Day surgery patients;
-Thorax x-ray;
-Anesthetic visit, which is the most complex one, as: 1)Day Surgery patients arrive directly there, without any previous examination, 2) the anaesthesiologist may ask for further examinations or visits after the ones already made, which may cause a wait and return in a short time, when possible, or in the following days, if not;
-Other visits or exams, usually required by the anaesthesiologist
-Nursing board compilation
-Abdomen eco
-Spinal column x-ray
-Neck x-ray and neck eco
-Abdomen computer assisted tomography
-Other visits or exams, usually required by the anaesthesiologist
-Pre-admission process end, with patients exit
Every operation asks for time and resources; paths of taken samples from the withdrawal room to the laboratory and path back of results are simulated; paths of patients requiring specialized external visits are simulated.
Every created patient crosses some or all sections according to the related sequence, depending on the class.

The adopted decisions define below model parameters:
-patient arrival numbers and times, and possible priorities
-human resources' turns of duty and possible movements among the unit ambulatories and desks
-technological sets at disposition in time.
For every adopted decision set the model supplies the following results: -patient's total working time, total waiting time, total transfer time, total spent time in the hospital; patient's waiting time and patients' queue length before every operation; work load for every resource and related efficiency. For every result the minimum, the maximum and the average value related to one simulated year is reported.
On the basis of results, the critical operations are evidenced and consequently all decisions adjusted.
The final distribution of patients arrival and nurses' presence and movements were found in a comparatively easy way. For every day there was decided to treat 21 orderly patients and 10 patients from Day Surgery and from the Obstetrics and Gynecology department; orderly patient arrival was divided into three turns, at the distance of half an hour by starting at 7.30; arrivals of Day Surgery patients and patients from the Obstetrics and Gynecology department was stated together with the third turn, with priority for the anaesthetic visit.
In the unit there are four nurses, of which one works with the electrocardiograph, one at the nursing board compilation and two begin with the ematochemical exams and then gradually, one by one at different times, move to the nursing board compilation.
The electrocardiograph revealed to be a mild critical resource, as an only machine is at disposition.
A highly critical resource appeared the anaesthesiologist. Supported by simulation, a new schedule could be arranged, with increased presence of the anaesthesiologist between 8.30 and 13.00 every day.
Such a way the total mean total time for patients varied between 102 and 386 minutes, depending on the class, with waiting times under 30 minutes for all operations except for the anaesthetic visit with a mean waiting time of an hour. Working loads revealed satisfactory for all operators, except for the anaesthesiologist, where they resulted comparatively high, but were accepted by interested operators. The efficiency of the unit was considered high, with total satisfaction of all interested departments' directors and of the hospital planning and management unit.
In a gradual way, in order to avoid or reduce unexpected difficulties, all departments transferred their pre-admission process to the new unit. Every step was preceded and followed by simulation. Now the new pre-admission unit has been entirely activated, and the model performance validated.

REFERENCES

Holt N. F.; Silverman D. G.; Ravindra P.; Dziura J.; Ruskin K. J. 2007. "Preanesthesia Clinics, Information Management, and Operation Room Delays: Results of a Survey of practising Anesthesiologists." *Anesth Analg* 104, No. 3, 615-618.

Pollard, J. B.; Garnerin P.; Dalman R. L. 1997. "Use of Outpatient Preoperative Evaluation to Evaluation to Decrease Length of Stay for Vascular Surgery." *Anesth Analg* 85, 1307-1311.

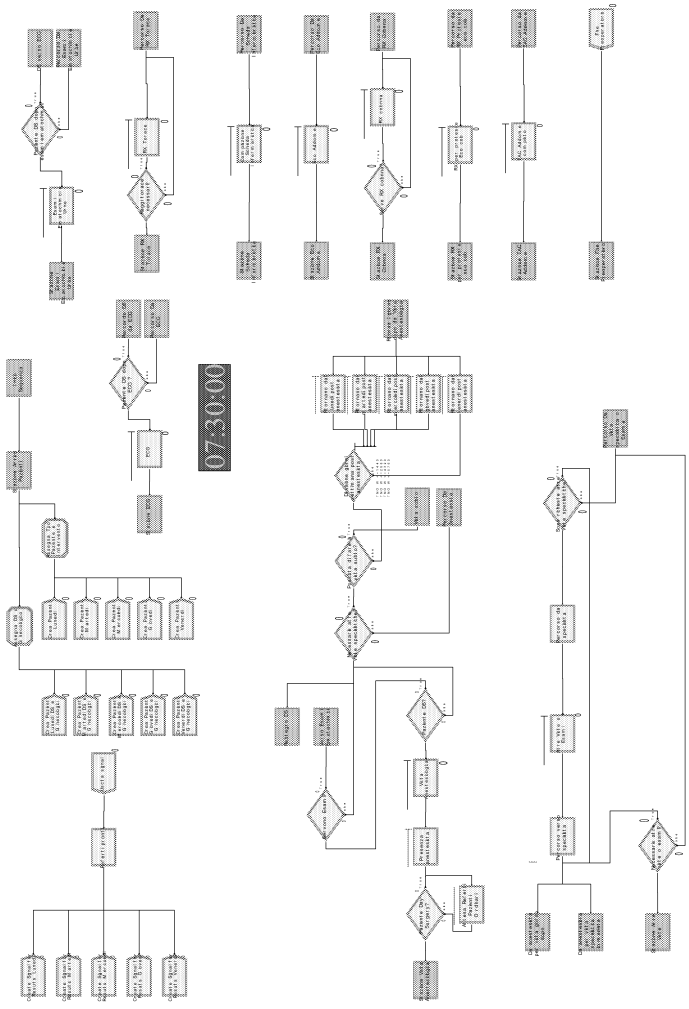
Pollard, J. B.; Olson L. 1999. "Early Outpatient Preoperative Anesthesia Assessment: Does it Help to Reducw Operating Room Cancellations?" *Anesth Analg* 89, 502-505.

Van Klei W. A.; Moons K. G. M.; Rutten C. L. G.; Schuurhuis A.; Knape T. A.; Kalkman C.; Grobbee D. E. 2002. "The Effect of Outpatient Preoperative Evaluation of Hospital Inpatients on Cancellation of Surgery and Length of Hospital Stay." *Anesth Analg* 94, 644-649.

BIOGRAPHIES

GIORGIO ROMANIN-JACUR went to the University of Padova, Italy, where he studied electronical engineering and obtained his degree in 1970. After doing his military service, he worked for ten years in the National Council of Researches, then he become first associate and then full professor of Operations Research in the Schoool of Engineering of the University of Padova, where he currently operates.

FEDERICO SEGALA went to the University of Padova, Italy, where he studied managementl engineering and obtained his degree in 2012. Now he is working in a private industry.



WASTE WATER MANAGEMENT SIMULATION

PLANNING REVITALIZATION SEWERAGE SYSTEMS USING DECISION RULES OF FUZZY LOGIC

Agnieszka Służalec
Polish Academy of Sciences
Systems Research Institute
Newelska 6, 01-447 Warsaw
Poland
e-mail: sluzalec.a@gmail.com

Andrzej Ziółkowski
Warsaw Information Technology
Newelska 6, 01-447 Warsaw
Poland
e-mail: a.ziolkowski@wit.edu.pl

KEYWORDS
Revitalization of sewerage systems, decision support systems, fuzzy logic, simulation.

ABSTRACT
In the paper the structural and functional unit supporting revitalization of sewerage systems, using decision rules of fuzzy logic is discussed. The main advantage of using of fuzzy logic rules compared to classical logic rules is the possibility of ordering results according to the membership grade to categories. This approach allows to estimate the condition of the sewerage system and outlines what steps to take to ensure its continued proper operation. The rightness of decision is verified by simulations of the network model. In typical situation when limited financial resources must be considered the development of work plans is facilitated. This approach has been implemented in a web application MOSKAN (Modeling and Simulation of Sewerage Systems) created at the Systems Research Institute of the Polish Academy of Sciences[1].

INTRODUCTION

An important element of the management of the sewer system in addition to the expansion, due to the adaptation to the changing nature of land use, is the revitalization of the existing system. After some period of operation, damages and irregularities appears inside the sewers, which may in consequence lead to reduction of their functions.
Revitalization includes actions aimed at ensuring the proper operation of the system, enabling the collection of sewage from customers on a continuous basis without negative impact on the environment. The scope includes: maintenance (including washing, cleaning of mechanical and hydrodynamic clears), local repair (including sealing point), renovations, which cause depending on the applied technology, improved hydraulic conditions, strengthen the construction, pipe and seal replacement (rebuilding) existing network e.g. due to the necessity to increase the capacity by increasing the cross-section.
The factors forcing revitalization are:

- Increase the amount of sewage caused by changes in the nature of land use and the connection new invested areas,
- Insufficient of capacity sewer structure, arising from the increase in loads of vehicle fleet or caused by corrosion or abrasion sewer bottom,
- Difficult operating conditions resulting from inadequate sewer decreases, causing the silting and the emergence of aggressive compounds that affect the design,
- Collisions with projected underground communication structures,
- Errors made at the building stage e.g. incorrect sewer slope, faulty made connection of sewers to the wells,
- Errors made at the design stage, e.g. erroneous design assumptions regarding the amount of sewage, bad choice of construction materials and seals the joints,
- The quality of materials, such as embedded leaky pipe joints,
- Defect caused by long-term operation of the network (the process of wear and aging structures),
- Tightening the rules on the protection of the environment. Factors may be more, the specific for a particular region or city, such as mining damage.

The factors forcing the revitalization of existing systems mean that need of revitalization do not raise concerns. An important aspect, however, is a correct planning of the revitalization. At present decisions are often made ad hoc and, in the best case, are based on practical experience after failures appear. Whereas, it is important to react at the right time, before serious failures occur. This is a problem that must be measured by companies authorized to operate in the public water supply or collective sewage (water supply and sewerage companies) and communal organizational units without legal personality carrying out such activities.
The need to support this type of activity is especially visible, when limited financial resources must be considered. Automation of creating revitalization's plans can facilitate not only the development of investment plans by companies possessing a limited budget, but also to prevent the occurrence of accidents.
In order to carry out these tasks, at the Systems Research Institute revitalization planning support system of sewerage systems was created. It is included in the structure of an integrated information system MOSKAN, whose general idea was presented in (Służalec et al. 2012a, 2012b, 2013).
In this system we propose using decision rules of fuzzy logic. The main advantage of using of fuzzy logic rules compared to classical logic rules is the possibility of ordering results according to the membership grade to categories. This approach allows to estimate the condition of the sewerage system and outlines what steps to take to ensure its continued proper operation. The rightness of decision is verified by hydraulic simulations of the network model. For this

[1] The paper describes the results of the research project No. N N519 6521 40 sponsored by the Polish National Center of Science (NCN).

purpose, a hydraulic model SWMM (Rossmann 2012) is used, which is one of the components of MOSKAN.
Application of fuzzy logic rules sets the support module for planning the work of revitalization, which is implemented in the MOSKAN, apart from Care-S (Computer Aided Rehabilitation for Sewer and Storm Water Networks (Saegrov 2005).
Presented in this paper system is currently in the testing phase with the real data coming from some Polish water and sewage company.

PLANNING REVITALIZATION OF SEWERAGE NETWORKS IN THE MOSKAN SYSTEM

System for supporting of revitalization planning should suggest, when and what action should be taken to ensure their continued reliable operation. Algorithms for decision-making is not an easy one. In the process of decision-making, data easy to determine is used e.g. time of using sewers, diameter, material, but we must also take into account information about the purity level of the sewer, the scale of damage and deformation. This information is obtained from network inspection (usually based on results from CCTV inspection) and it must be entered into the database system.
Due to the fact, that the degree of detail of information from the inspection is contingent upon the advancement of equipment used for inspection, and further evaluation of the results is often subjective in nature, it was decided in the development of the module on the use of fuzzy logic decision rules.
Decision support system includes the following elements:
- a database for collecting data about sewage network state,
- a tool to specify definition of fuzzy decision rules and definition of the fuzzy values,
- algorithms for decision-making.

Apart from it benefits from hydraulic model to check results after revitalization.

Database

Decision support system requires, as it was mentioned before, the information about the network condition. This information is taken from the system database, which collects data on:
- Time use of the channel,
- The material of which it is made
- Cross-section,
- Inspections,
- Carried out repairs and incurred the costs.

Individual information is contained in three catalogues: *Conduits State, Repairs (Rehab)* and *Inspections*, available from the *Graph* level (Fig. 1). Data entry is done interactively, by selecting the appropriate tab (Fig. 2).
The year of building the sewer, the material from which it was made and the irregularities found there are introduced in the *Conduits State* tab, which is an important element of the system. In fact, these are the source of information used by the module support the revitalization plan to determine the indicators of decision rules, that show if the sewer only requires current maintenance, local repair, renovation or rebuilding (replacement).
On the other hand, information on the repairs carried out on different parts of the sewerage system including: the date of repair, the type of repair, the costs incurred to correct irregularities is entered in *Repairs* tab. *The type of repair* is chosen from included *The Dictionary of damages*. The proposed scheme of the damages can be extended and tailored to the needs of the user. The aim of standardizing names was enable the user to make further analysis, e.g. what damages appeared mostly. Entering the cost helps to estimate, how much effort is spent on ensuring reliable operation of a section of the sewerage system.
An important element of the system is also information from the sewer inspection, as reports and photo documentation. Based on it, parameters (e.g. roots penetration, moving pipe joint, infiltration of groundwater, incorrect slope, sewer corrosion) characterizing the condition of the sewer are defined.

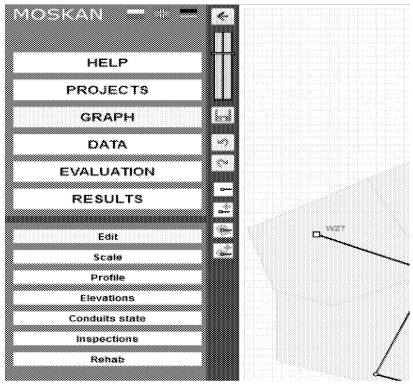


Figure 1. Database catalogues.

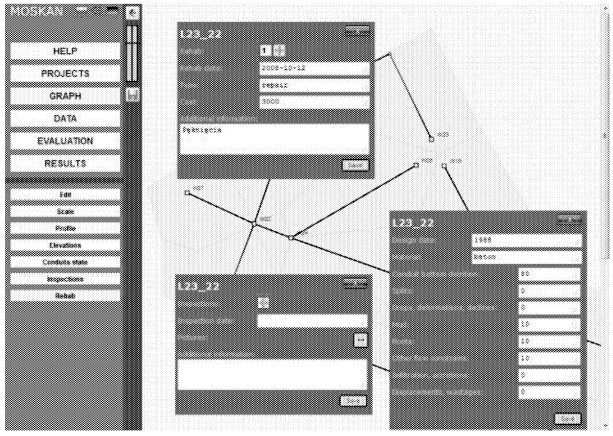


Figure 2. Interactively introduction of data.

Application of fuzzy logic decision rules

To define the revitalization plan of sewerage systems we use the calculation of fuzzy decision rules proposed in 1965 by L.A. Zadeh, as an extension of classical logic, based on two core values: "1" and "0" (true and false).
The main objective of this theory is to represent and process imprecise information contained in some natural language definitions. This linguistic variables are one of the fundamental concepts related to fuzzy sets. They translate human language daily formulation to the language of mathematics. Their values are not numbers, but words or sentences, called linguistic values. For example, the linguistic variable temperature can have linguistic values: cold, heat, hot (Zadeh 1965, Kacprzyk 1986).

The use of fuzzy logic relies on using fuzzy function for determining the membership grade to a collection of elements.
According to the definition: if X is a collection of object denoted generically by x, then a fuzzy set A in a non-empty space X, is a set of ordered pairs, described by equation (1):

$$A=\{(x,\ \mu_A(x))\ x \in X\} \qquad (1)$$

where: μ_A:X→[0,1] is called membership function of fuzzy set A (Zadeh 1965, Zimmermann 1996).

This function assigns to each element $x \in X$ the grade of membership to the fuzzy set *A*. Each element can fully belong to the fuzzy set *A* ($\mu_A = 1$), does not belong to this set ($\mu_A = 0$) or partially belongs ($0 < \mu_A < 1$).
Going away the dichotomy, the finding that the sentence can only be true or false, the object can belong to a set or not, but there is no other solution. In the case of a fuzzy set object may belong to a set to some degree or in the same way do not fall, which in turn leads to the fact that the sentence may be true in part at the same time to the same extent partially false.
The most commonly used range of values of membership functions is the unit interval [0,1]. In this case each membership function maps elements of a given universal set X into real numbers in [0, 1] (Klir and Yuan 1995).
In practice, only a few types of functions are used, e.g (Rutkowska et al. 1999):
- singleton - described by equation (2) and shown in Fig. 3,

$$\mu_A(x) = \begin{cases} 1, & x = x_0 \\ 0, & x \neq x_0 \end{cases} \qquad (2)$$

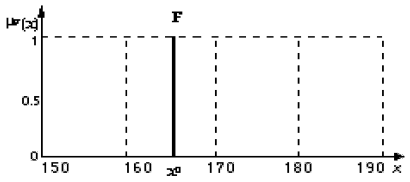


Figure 3. Example of membership function described equation (2) above.

- triangular - described by equation (3) and shown in Fig.4,

$$\mu_A(x) = \begin{cases} \dfrac{x-a}{b-a}, & a \leq x \leq b \\ \dfrac{x-c}{b-c}, & b < x \leq c \\ 0, & x < a \vee x > c \end{cases} \qquad (3)$$

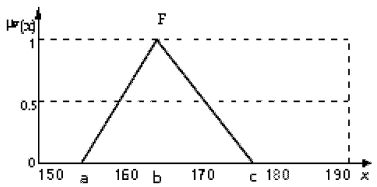


Figure 4. Example of membership function described equation (3) above.

- trapezoidal - described by equation (4) and shown in Fig. 5.

$$\mu_A(x) = \begin{cases} \dfrac{x-a}{b-a}, & a \leq x \leq b \\ \dfrac{x-d}{c-d}, & c < x \leq d \\ 1, & b < x \leq c \\ 0, & x < a \vee x > d \end{cases} \qquad (4)$$

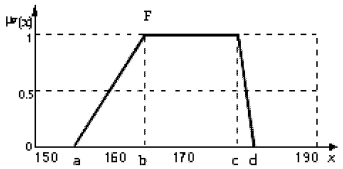


Figure 5. Example of membership function described equation (4) above.

The use of fuzzy sets allows to create the fuzzy model of the system, representing the essential features with fuzzy sets theory. The most important feature of such systems is that they are based on the concept of fuzzy coding information.
As mentioned in the introduction, the module uses a fuzzy logic decision rules. The advantage of its use is that as a result of calculation we get a membership grade of sewer to a certain category (recommended maintenance, repair local recommended, recommended renovation or reconstruction - replacement) as a number from 0 ... 1. Some actions can be strongly recommended (membership grade 1.0), but some of them may be recommended less or more strongly (membership grade <1.0). This form of the results makes it very easy to decide when funds earmarked for regeneration is limited. The fact that some recommendations are stronger and the other weaker allows in rational way to determine the order of the work of regeneration.

Specifying definition of fuzzy decision rules and definition of the fuzzy values

Due to the fact that the evaluation of the results of the inspection is conducted often subjective nature, it was decided to apply the fuzzy logic decision rules in two layers: both to determine the conditions of the selection and indicators.
In the System MOSKAN, four groups of revitalization work were defined, described in the application as a condition of selection, namely:

- *Maintenance,*
- *Local repair,*
- *Renovation,*
- *Rebuilding.*

To define the rules of decision by which the sewers are classified to the appropriate category of work revitalization (recommended maintenance, recommended local repair, recommended renovation or reconstruction – Fig.6), the following indicators are introduced:

- The status of the channel structure – *Conduit State*,
- The degree of flow limitation – *Flow constraints*
- The degree of leakage - *Leaks.*



Figure 6. Decision rules defining the conditions of selection.

In the above examples, such as *Conduit State* and *Flow constraints* are the names of indicators used to define the rules of decision-making. Adjectives such as *good*, *acceptable*, are the names of fuzzy values. Conjunctions *AND* and *OR* allow to create logical expressions that determine when to perform *Maintenance, Local repair, Renovations or Rebuilding*. The word *END* is the tag end of the fuzzy logic expressions. Has been added to allow full interactive editing of fuzzy logic expressions.
Due to the aforementioned aspect, each of the indicators is based on pre-determined decision rules peculiar to it (Fig. 7).

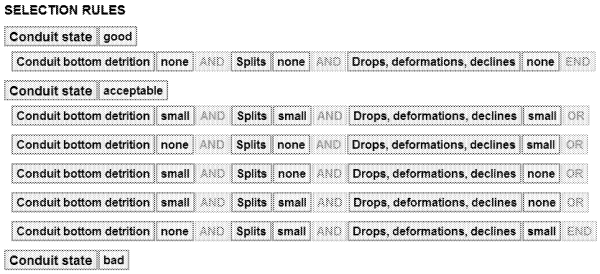


Figure 7. Decision rules defining the conditions for selection of indicators.

For indicators (linguistic variables e.g. *Conduit bottom detrition; Splits; Drops, deformation, declines*, etc.) defined fuzzy values presented graphically as colored bars (Fig.8).
Attributes of these values e.g. *significant, small, none*, are placed on the right side of the bars.
Green colour bar defines the scope of fuzzy. Fuzzy values in the module are represented by trapezoids (Fig. 9). Because of the height of trapezoid is always 1.0 (membership grade 0...1,0) to determine the trapezoid is enough to give four coordinates on the x-axis. They are stored as a number from 0…1000 (a value of 1 corresponds to 0.1% of full scale value of the indicator).
Both the range of values, as well as fuzzy the lower and upper limit of the range of values can be freely configured in the system in interactive mode.

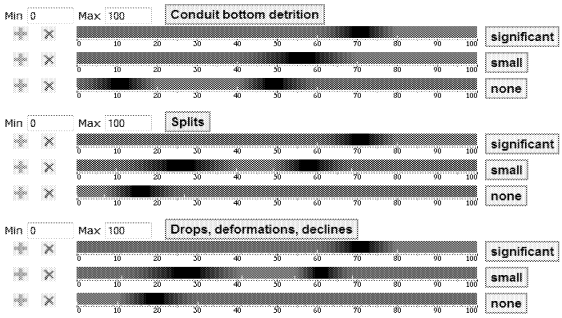


Figure 8. Defining of fuzzy values.

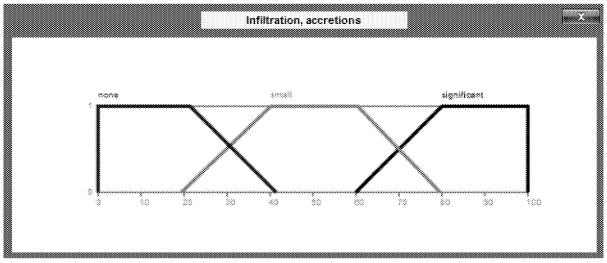


Figure 9. Trapezoids determining the fuzzy values.

Phases of decision making
In the whole process of decision making related to revitalization supported by MOSKAN we can separate some phases:
1. Collecting data about sewage network state. Collected data is available also for other purposes.
2. Specifying definition of fuzzy decision rules and definition of the fuzzy values. Participation of experienced decision makers in this activity is essential. Available data about sewage network state must be translated to fuzzy value names e.g. splits in 10-20% of conduit length could be considered “small” and in more than 40% “significant”. Splits in 20-40% are both “small” and “significant” with the appropriate membership grade depending on the exact value.
3. Calculations. Fuzzy decision rules are applied to data about sewage network state. In result, for each conduit, the membership grade to each category (*Maintenance, Local repair, Renovation, Rebuilding*) is calculated.
4. Workout of revitalization plan. All decisions are made by decision maker who takes into account available resources and presented results from calculations.

Presentation of the results
In the system MOSKAN *Revitalization options* are available from the *Evaluation* level.
The results are presented in the form of colored graph network, in which color corresponds to the membership of the recommendations (Fig. 10) and tabular (Fig. 11).
Presentation of the results in the form of a graph facilitates finding areas where revitalization is needed. Results in tabular are ordered according to the membership grade to the selected category making it easier to make decisions about the order of work.

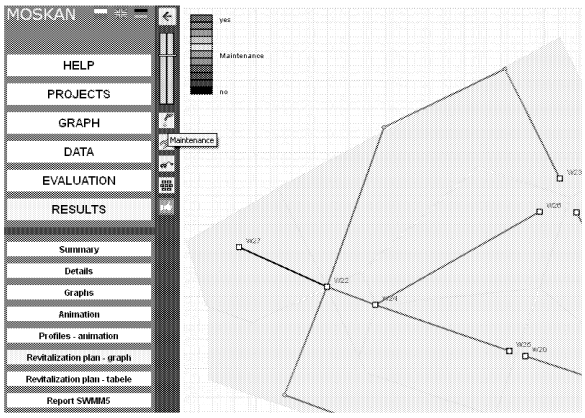


Figure 10. Presentation of the results in the form of graph.

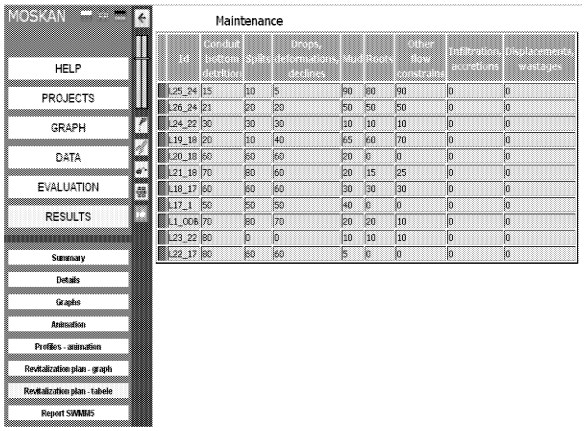


Figure 11. Presentation of the results in the tabular form.

THE MOSKAN SYSTEM - IMPLEMENTATION

Based on the experiences of information of water networks and considering how an important matter, in addition to the design of new sewer systems and their optimization, is also the revitalization of existing systems, in the Systems Research Institute of the Polish Academy of Sciences, it has been attempted to work out an integrated information system MOSKAN, which will be realize all of above tasks.

The system is based on the hydraulic model SWMM (http://www.epa.gov/nrmrl/wswrd/wq/models/swmm/). However, it has its own, accessible interface.

MOSKAN system was implemented as a web application in order to provide convenient access to the results of a number of potential users and to facilitate the entry and update the data required for its operation.

For the implementation of the interface MOSKAN system used languages JavaScript, PHP and HTML. To present Sewerage systems as a graphical object used canvas with HTML-5. Sewage systems are drawn on the background of the site map as image files in PNG or JPG format. Data describing the network are stored in XML format and on their basis are generated data in the format required by the SWMM program. It runs on a server when it is necessary to make simulation calculations.

SUMMARY

The need of revitalization sewers system and the need of improving revitalization plans do not raise concerns. Revitalization planning support system, presented in the paper, allows the identification of the technical condition of the sewerage system, determine failure trends, generate scenarios for revitalization work on which it is possible to create investment plans tailored to business investment opportunities.

The use of fuzzy logic facilitates the development of work plans and decision making with limited financial resources, because the received proposals can be ordered according to the membership grade to categories (e.g., need renovation). Decision rules are written in a simple manner similar to the natural language and the system also makes it easy to follow their activities.

Presented in this paper decision support system is currently in the testing phase with the real data coming from some Polish water and sewage company is currently in the testing phase.

REFERENCES

Kacprzyk, J. 1986. „Zbiory rozmyte w analizie systemowej", PWN, Warszawa.

Klir, G.J. and Bo Yuan. 1995. „Fuzzy sets and fuzzy logic. Theory and Applications". Prentice Hall PTR, New Jersey, USA, 11-29.

Rossman, L. 2012. „Storm Water Management Model (SWMM) – User's manual", Version 5.0.022", 2012. Online http://www.epa.gov/nrmrl/wswrd/wq/models/swmm.

Rutkowska, D., Piliński, M. and L. Rutkowski. 1999. „Sieci neuronowe, algorytmy genetyczne i systemy rozmyte", Wydawnictwo Naukowe PWN, Warszawa, Łódź.

Saegrov, S. 2005. „Care-S – Computer Aided Rehabilitation for Sewer and Storm Water Networks", IWA Publishing, Alliance House, London.

Służalec, A., Studziński, J. and A. Ziolkowski. 2012. „MOSKAN – eine WWW-Anwendung in der HTML-5 Technologie zur Abbildung und Simulation von kommunalen Abwassernetzen", In: *Simulation in Umwelt- und Geowissenschaften,* Workshop Hamburg 2012 (J. Wittmann, B. Page, Hrsg), 85-95.

Służalec, A., Studziński, J. and A. Ziolkowski. 2012. „Rechnerunterstützte Planung von kommunalen Abwassernetzen mittels des hydraulischen Models und statischer Optimierung ", In: *Modellierung und Simulation von Ökosystemen,* Workshop Kölpinsee 2012 (A. Gnauck, Hrsg), in printing

Służalec, A., Studziński, J. and A. Ziolkowski. 2013. „Optimization of sewege structure in the integrated system for design, management and revitalization - MOSKAN", In: *Simulation in Umwelt- und Geowissenschaften,* Workshop Hamburg 2012 (J. Wittmann, B. Page, Hrsg), in printing.

Zadeh, L.A. 1965. „Fuzzy sets", Information and Control No.8 338-353.

Zadeh, L.A. 1975. „The concept of a linguistic variable and its application to approximate reasoning", Part I, Information Sci. No.8 199-249, Part II, Information Sci. No.8 301-357, Part III, Information Sci. No.9 43-80.

Zimmermann, H.J. 1996. „Fuzzy set theory and its aaplications (3rd ed.)", Kluwer Academic Publishers, Norwell Massachusetts, USA, 11-20, 53-66.

USING OPTIMIZATION APPROACHES FOR MODELING, CONTROL AND MANAGEMENT OF WATER NETWORKS

Jan Studzinski
Polish Academy of Science
Systems Research Institute
Newelska 6, 01-447 Warsaw
e-mail: jan.studzinski@ibspan.waw.pl

KEYWORDS
Decision making support, water distribution system, modeling, optimization, control, management.

ABSTRACT
In the paper some numerical algorithms supporting the management of water networks are described. The algorithms are included as software modules into the structure of an integrated IT system whose general idea was presented in (Stachura et al., 2012). The system uses mathematical models and optimization algorithms for solving several tasks of operation and maintenance realized in municipal drinking water supply systems with the goal, to make their execution easier, faster and more exact.

The research was financed by the Polish National Center of Research and Development (NCBiR) in frame of the project no. NR14-0011-10/2010

INTRODUCTION

Municipal water supply system

Municipal water supply system is a key element in each waterworks for it affects the work of the subsequent systems connecting with it which are wastewater network and sewage treatment plant. A correct management and operation of the water network decides on the quality and quantity of drink water supplied to the city customers and also on the proper removal and processing of the wastewater operation outgoing the water net. The main tasks of the water net management that can be automated by means of mathematical models and optimization algorithms are as follows:

- planning SCADA systems for water nets,
- calibration of water net hydraulic model,
- loads forecasting for the whole water net as well as for its individual nodes,
- supply and distribution of water under sufficient pressure,
- planning modernization and expansion of the network,
- location of water leaks,
- energy saving control of pump stations installed on the water net.

Computer algorithms supporting the execution of the above tasks will be presented in the following. They have been successfully tested with the real data coming from some Polish waterworks.

IT system for water network management

At the Systems Research Institute an integrated IT system for complex management of municipal water networks has been developed (Fig. 1). It consists of several programs cooperating each other via special data files and its main modules providing all programs with the input data are GIS and SCADA systems and a hydraulic model of the water net.

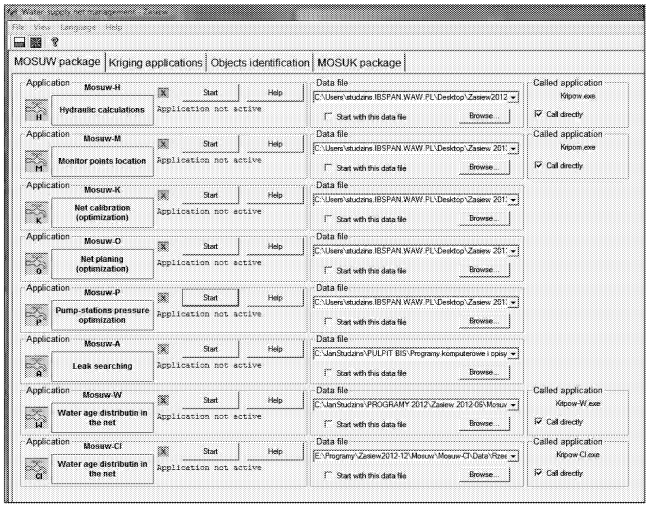


Fig. 1. View on the interface screen of the IT system.

The advantages of this IT system are the close integration of its programs and the use of mathematical models and optimization algorithms for the computer aided solution of the water net management tasks. It stands in strong contrast to the usual practice of the waterworks operation that bases mostly on individual experience and intuition of water net operators and if there a waterworks computerization occurs then it relies usually on the implementation of computer programs that run autonomic realizing their tasks without any cooperation with another software. In this way the synergy effect resulting from collective work of several programs is lost and handling of many separated programs in the waterworks becomes more and more complicated after the number of programs oversteps some critical quantity.

In the following some selected optimization approaches implemented in the IT system are presented.

MATHEMATICAL MODELS AND OPTIMIZATION ALGORITHMS HELPING WITH THE WATER NET MANAGEMENT

SCADA system planning

SCADA systems provide the water net operator with the data of water flows and pressures in the network giving him the

current information about the state and quality of its operation. The systems can be used either to calibrate and recalibrate the water net hydraulic models. A very important task by SCADA planning is a selection of correct measurement points for monitoring the network in such the way that a possibly minimal number of the points can deliver a possibly maximal amount of information. The points which are chosen properly are called *characteristic points* and their attribute is a great sensivity against the flow or pressure changes in the water net in relatively big surrounding of them. In the algorithm for calculation of the water net sensivity given by (Straubel and Holznagel, 1999) the simulation of water leaks in each node of the water net is done by means of the hydraulic model and the following formulas are used:

$$S_{pm} = \frac{\sum_{k \neq m} (\Delta p_m / p_m) L_{km}}{\sum_{k \neq m} L_{km}} \qquad S_{qm} = \frac{\sum_{k \neq m} (\Delta q_m / q_m) L_{km}}{\sum_{k \neq m} L_{km}}$$

where: k – node with the water leak simulated, m – measurement point considered, p – water pressure, q –water flow, Δp_m and Δq_m – differences in measurements by standard and emergency states of the water net, L –distance between points k and m.

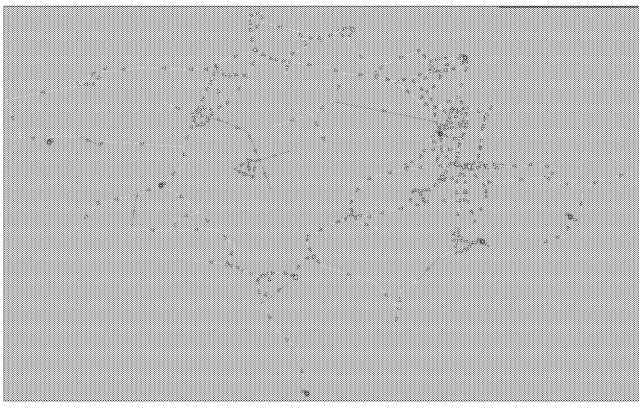


Fig. 2. Water net investigated.

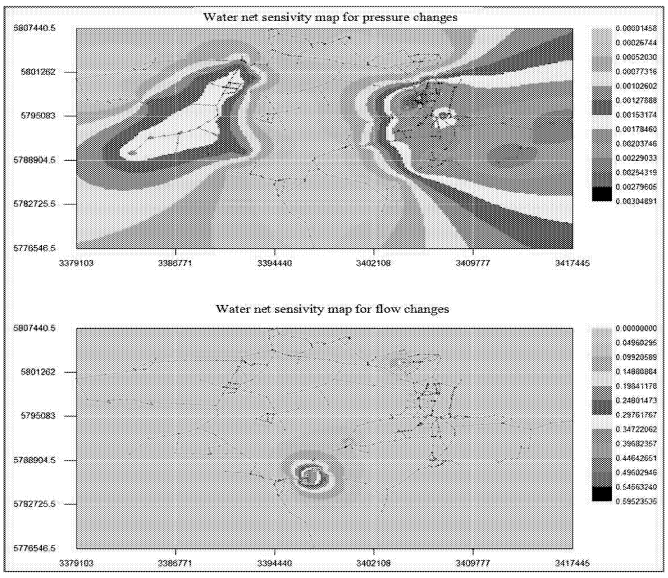


Fig. 3. Visualization of the water net sensivities.

As the result of calculation the lists of water net nodes sorted regarding their sensivity against pressure and flow channges are got and the sensivity distributions can be visualized by use of kriging approximation (Bogdan and Studzinski, 2007) (Fig. 2 and 3). Subsequent a multi criteria optimization for determining the number and location of the selected measurement points is applied with the following quality functions:

- number of measurement points – to be minimized,
- sensitivity on the water leaks – to be maximized,
- installation cost of measurement points – to be minimized.

There is to notice that the SCADA system planning is an iterative process realized step by step for for the water leaks simulation a properly calibrated water net hydraulic model is needed and to calibrate the model an already rightly designed SCADA system has to be used.

Hydraulic model calibration

Beside SCADA and GIS systems a hydraulic model is third of the main elements of the IT system and at the same time it is the basic computational tool supporting the water net management. Its mathematical description in form of a set of linear and nonlinear algebraic equations results from the first and second Kirchhoff laws known from electrical engineering (Todini and Pilati, 1987) and its structure consists of pipes, nodes (pipe junctions), pump stations, valves and storage tanks or reservoirs. Hydraulic model calculates the water flow in each pipe, the pressure at each node, the height of water in each tank, and with it the water age in the water net and the concentration of chemical species passing the network can be calculated (Rossman and Boulos, 1996). Because of the importance of the hydraulic model for the water net management a fundamental problem is the model calibration, i.e. its fitting to the network. The calibration is usually made *per hand* what means that the model operator changes its parameter values intuitively and checks his decisions by following model simulation runs; the calibration is finished when the simulation results match the measurement data. This process is very time consuming and it can be automated by use of a multi criteria algorithm with the following quality functions (Fig. 4 and 5):

- deviation of calculated pressures from the measured ones in the water net nodes – to be minimized,
- deviation of calculated flows from the measured ones in the water net pipes – to be minimized.

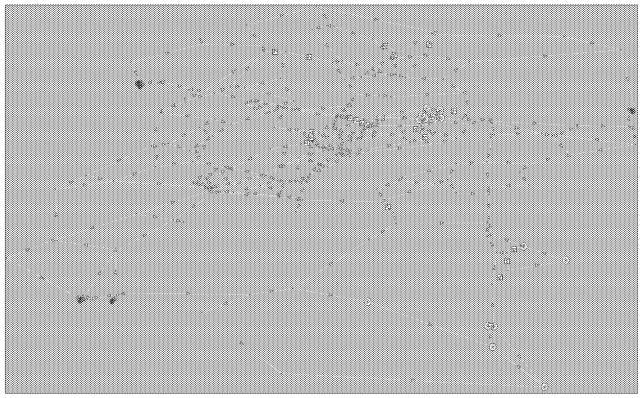


Fig. 4. Water net investigated with the monitoring points marked.

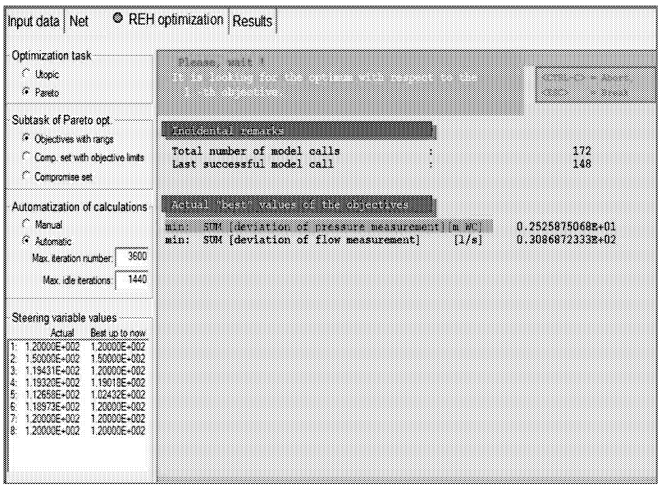


Fig. 5. Hydraulic model calibration with two quality functions.

To optimize the quality functions the roughness values in the model equations are selected automatically and the whole optimization process is effective only when the monitoring system installed on the water net has been developed in accordance with the requirements regarding the network characteristic points. The multi criteria optimization algorithm used to solve this task and also the following ones has been developed by (Straubel, 2005).

Water net planning

After the water net hydraulic model has been calibrated with the monitoring system properly designed then one can use the model to assess the work quality of the water net by calculating the pressure in all nodes and the flow values in all pipes of the network. An exemplary water flows distribution calculated by use of a hydraulic model and mapped by means of kriging approximation is shown in Fig. 6. The network investigated works incorrect for the water speed in many pipes is to small what results in worsening the water quality.

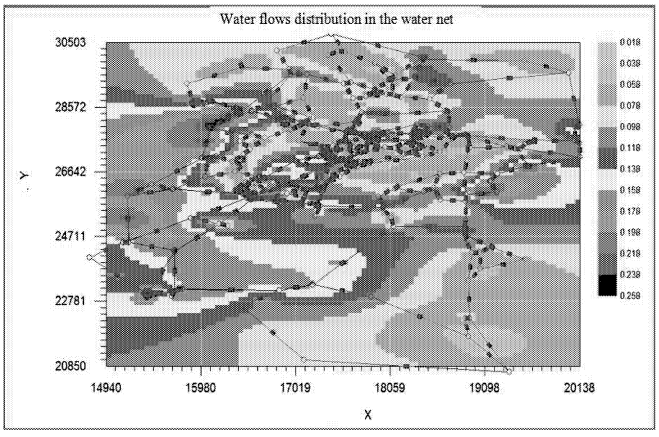


Fig. 6. Visualisation of the water net work.

With the multi criteria optimization enclosed in the IT system the water net functioning can be improved by changing the pipes diameter or the control scenario of the pump stations or retention tanks or by the change of state of valves installed on the network. The appointment of values of controlling parameters which are the input data for the optimization algorithm is shown in Fig. 7. While running the optimization algorithm the following quality functions are calculated (Fig. 7):

- deviation of calculated pressures from the measured ones in the water net nodes – to be minimized,
- efficiency of energy consuming by the network pumps to be maximized,
- sum of output pressures of all pump stations – to be minimized,
- speed values of the water passing the network pipes – to be minimized,
- investment costs of all operations – to be minimized,
- price of 1 m^3 of water produced – to be minimized.

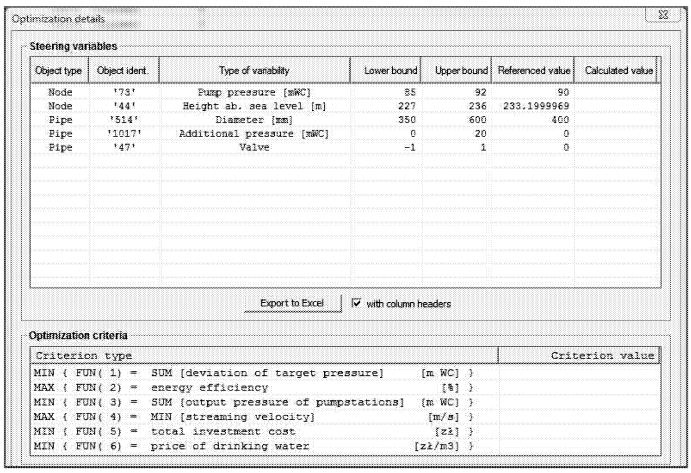


Fig. 7. Data preparation, quality functions and the calculation results regarding the water net optimization.

Energy saving control of pump stations

The pumps installed on the water net at the water intake stations, in the retention tanks and inside of the network for increasing the pressure in remote parts of the water net are devices which are consuming a lot of energy and a reduction of their energy costs of only few per cent in a year time could bring essential financial savings for the waterworks. Energy saving control of pump stations has been investigated e.g. by (Waterworth, 2002) who proposed an algorithm to minimize the energy consumption of pumps by their time distributed switching on and off and under consideration of their wear time and their exchange against the new devices.

An another algorithm is included into the IT system that uses the multi criteria optimization with the following quality functions:

- deviation of calculated pressures from the measured ones in the water net nodes – to be minimized,
- total power value of all pump stations installed on the network – to be minimized,
- sum of output pressures of all pump stations – to be minimized,
- total cost of the produced drinking water – to be minimized.

In Figures 8 and 9 the water net investigated with 6 pump stations installed on it and the appointment of input data for the optimization algorithm with its quality functions calculated are shown.

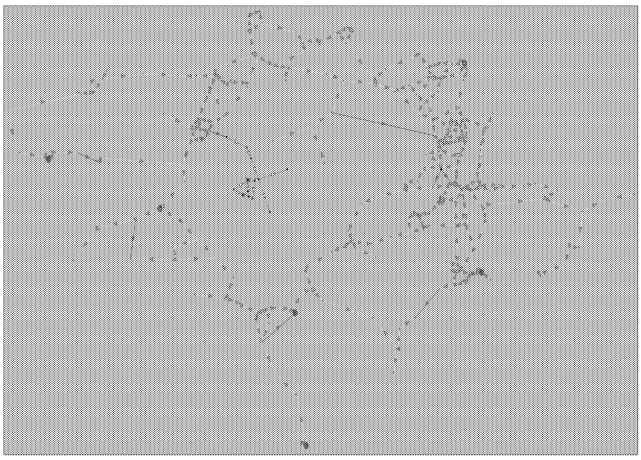


Fig. 8. Water net investigated with 6 pump stations to be controlled.

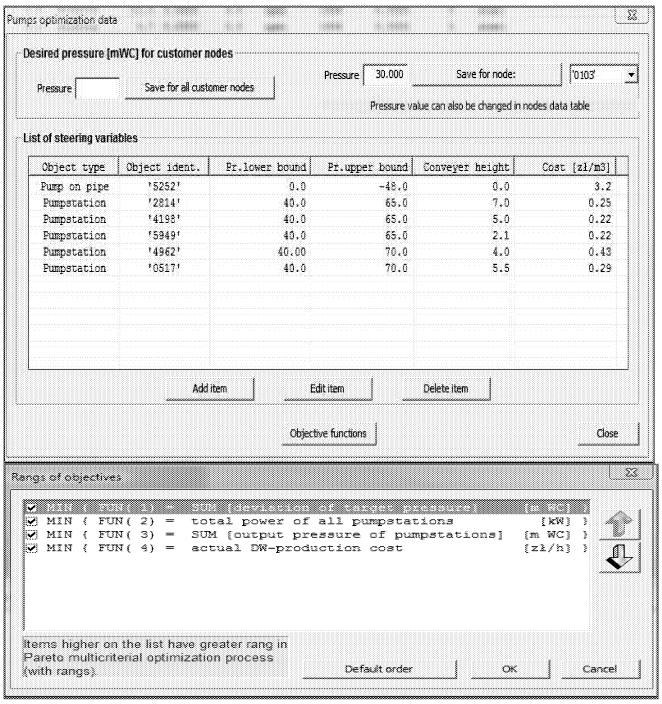


Fig. 9. Data preparation for the pump stations control and the quality functions optimized.

Water net load forecasting

The knowledge of the network hydraulic load as a whole as well as of the loads of its individual nodes is needful while calculating the hydraulic model of the water net and for the generation of control scenarios for its pump stations and retention tanks. The forecasting models for these loads can be developed by means of the fuzzy sets, neuronal nets and time series methods which are most simple, fast and easy in operation.

In the IT system the time series algorithms of the classical least square method to model and forecast the hydraulic load changes are applied (Bartkiewicz and Studzinski, 2010). The process and model equations and the quality function formulated as difference between the process and model outputs are shown in the following:

$$y_n = -A(z^{-1})y_n - \sum_{i=1}^{M} B(z^{-1})x_{in} + v_n \qquad \hat{y}_n = -\hat{A}(z^{-1})y_n - \sum_{i=1}^{M} \hat{B}(z^{-1})x_{in}$$

$$S_v(c) = \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 = \sum_{n=1}^{N} \hat{v}_n^2$$

where: y_n and $\hat{y}_n$ –process and model outputs, respectively, A, B and $\hat{A}, \hat{B}$ – operators with process and model parameters. The forecasting models received by means of the time series methods are very accurate (Fig. 10) and in many cases they are more exact than the models calculated by means of fuzzy sets or neuronal nets; such the comparative calculations have been executed (Studzinski and Stachura, 2011).

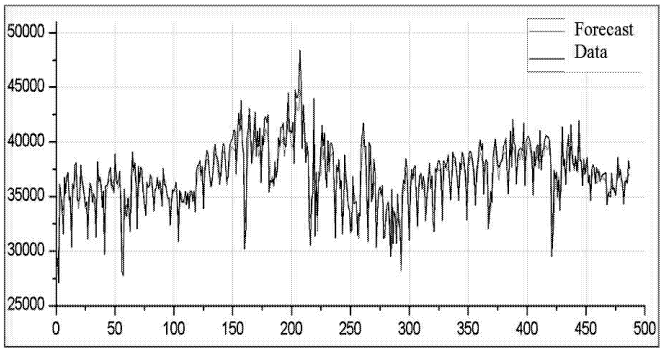


Fig. 10. Exemplary results of network load forecasting with the time series method of Clarke.

Water leaks location

Water leaks in water nets caused by rusting of pipes, tenseness changes of ground or incorrect operation of the network induce essential losses of water in waterworks reaching often up to 20 % of the whole water production. As a result the exploitation costs of water nets are growing what raises the water price for the individual end users of the network. Therefore an early detection and location of water leaks and especially of these hidden ones is of great importance for the improvement of the waterworks functioning.

In the IT system there are a couple of algorithms that have been developed to solve the problem of water leaks in the water net pipes. In case of normal operation of the water net there are typical pressure and flow distributions on the nodes and pipes of network. A failure occurring on the water net and causing the water leak is an unusual situation that results in unusual variation in the nodal pressures and pipeline flows. One of the algorithms enclosed in the IT system detects the faults by comparing the actual pressures and flows in the water net which are on line monitored at the measurement points of the SCADA system with these standard ones that have been recorded before. To locate the water leaks the neuronal nets in form of MLP models are used. For teaching the neuronal nets the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm is used to perform the optimization computing for which the quality function in form of the squares sum of residuals is formulated.

The algorithm mentioned consists of the following steps (Studzinski and Rojek, 2012):

- On-line registration of current pressures and flows from the SCADA measurement points and signalization of a water leak in case when the current measured values differ from the standard ones.
- Having the alarm signal received, calculation of pressure and flow distributions determined by the data requested from the measurement points of SCADA after the water leaks simulations have been made subsequently in each node of the network (Fig. 11) by means of the hydraulic model.

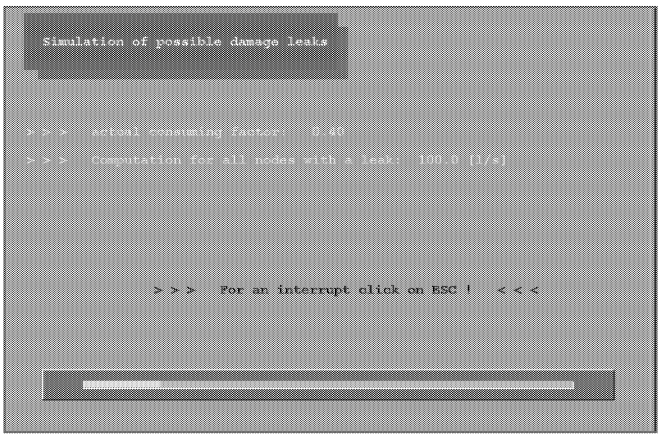


Fig. 11. Simulation of water leaks with the hydraulic model.

- Development of models for failures localization in form of MLP neuronal nets (Table).

Table. Neuronal models for water leaks location.

No.	Neural network name	Learning quality	Testing quality	Validation quality
1	MLP 44-25-21	97,2222	100,0000	98,1481
2	**MLP 44-18-21**	**100,0000**	**100,0000**	**100,0000**
3	MLP 44-21-21	98,0159	96,2963	98,1481
4	MLP 44-18-21	87,6984	85,1852	83,3333
5	MLP 44-19-21	90,0794	88,8889	83,3333
6	MLP 44-24-21	80,5556	75,9259	81,4815
7	MLP 44-8-21	93,6508	90,7407	94,4444

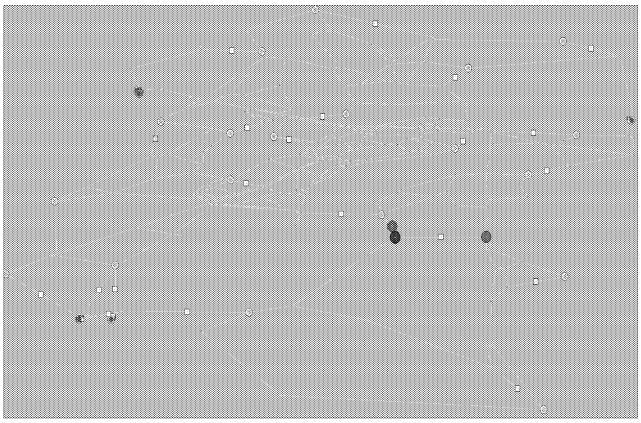


Fig. 12. View on the water net investigated with results of the water leak finding with 3 network nodes in which a failure is guessed.

- Localization of the network node with the water leak guessed by means of the best MLP model (Fig. 12).

CONCLUSIONS

The use of mathematical models and optimization algorithms can help to solve a lot of problems concerning the management of complex water networks. These models and algorithms cooperating each other in form of an united IT system could became a standard information tool supporting the running of communal water supply systems. The benefits from the use of such the systems are the shortening of time of individual management operations, making the operational work easier, faster and better, and the reduction of exploitation costs of the waterworks.

The algorithms described above are already implemented into the IT system being still all the time under development while some new algorithms for solving other water net management tasks are under investigation.

REFERENCES

Bartkiewicz, L. and Studzinski, J., 2010. *Mathematical modeling of the hydraulic load of communal wastewater networks*. In: Modeling and Simulation 2010, ESM'2010 (G.K.Janssens, K.Ramakers, A.Caris, eds.) EUROSIS-ETI, Hasselt Belgium 2010, 156-160.

Bogdan, L. and Studzinski, J., 2007. *Modeling of water pressure distribution in water nets using the kriging algorithms.* In: Industrial Simulation Conference ISC'2007 (J. Ottjes and H. Vecke, eds.), June 11-13, 2007, Delft, TU Delft Netherlands, 52-56.

Rossman, L.A. and Boulos, P.F., 1996. *Numerical methods for modeling water quality in distribution systems: A comparison*. J. Water Resour. Plng. and Mgmt, Vol. 122, No. 2, 137-146.

Stachura, M., Fajdek, B. and Studzinski J., 2012. *Model based decision support system for communal water networks*. ISC 2012, Brno, June 4-6, 2012, 19-23.

Straubel, R., 2007. *REH – Ein Programm für Rechnergestützte Entscheidungshilfe*. Forschungsbericht d. Ingenieurbüros Dr. Straubel, Berlin 2005.

Straubel, R. and Holznagel, B., 1999. *Mehrkriteriale Optimierung fuer Planung und Steuerung von Trink- und Abwasser-Verbundsystemen*. In: Wasser•Abwasser, 140, No. 3, 1999, pp. 191-196.

Studzinski, J. and Rojek, I., 2012. *Failures localization within water networks by means of neuronal nets*. ESM 2012, Essen, October 22-24, 2012, 64-68.

Studzinski, J. and Stachura, M., 2011. *Fuzzysatz- und Neuronalmodelle zur Vorhersage der Wassernetzbelastung*. ASIM Workshop über Simulation in Umwelt- und Geowissenschaften, Berlin.

Todini, E. and Pilati, S., 1987. *A gradient method for the analysis of pipe networks*. In: International Conference on Computer Applications for Water Supply and Distribution, Leicester Polytechnic, UK, September 8-10.

Waterworth G., 2002. *Efficiency modeling in the cost reduction of water pump maintenance*. In: Proceedings of QRM'2002, Ed. G. J. McNulty, PEP London, pp. 141-144.

SIMULATION IN ENGINEERING

DEVELOPMENT AND EVALUATION OF AN ORDER-N FORMULATION FOR MULTI-FLEXIBLE BODY SPACE SYSTEM

Tushar K. Ghosh
L-3 Communications, STRATIS
1002, Gemini Avenue, Suite 200
Houston, TX 77058, USA
E-mail: tushar.ghosh@l-3com.com

Leslie J. Quiocho
Mail Code ER7
NASA Lyndon B. Johnson Space Center
2101 NASA Parkway
Houston, TX, 77058, USA
E-mail: leslie.j.quiocho@nasa.gov

KEYWORDS

Multibody dynamics, Rigid and flexible bodies, Tree topology, Closed loops, Order-N.

ABSTRACT

This paper presents development of a generic recursive Order-N algorithm for systems with rigid and flexible bodies, in tree or closed-loop topology, with N being the number of bodies of the system. Simulation results are presented for several test cases to verify and evaluate the performance of the code compared to an existing efficient dense mass matrix-based code. The comparison brought out situations where Order-N or mass matrix-based algorithms could be useful.

INTRODUCTION

The Software, Robotics and Simulation Division (SRSD) of NASA Lyndon B. Johnson Space Center provides mathematical modeling and simulation in support of engineering analyses and crew training activities for the center. The division currently has an efficient generic multibody dynamics code based on a dense mass-matrix formulation, which is used for simulating systems involving on-orbit robotic manipulators such as the Canadian Space Agency-built Space Station Remote Manipulator System (SSRMS). It is generally known that Order-N (O(N)) algorithms, which involve arithmetic operation counts of the order N, where N is the number of bodies (actually, the operation count is true when all bodies have the same number of joint and flexible degrees of freedom and there are no loops), perform more efficiently for systems with large degrees of freedom, compared to mass matrix-based algorithms with operations of order N^3. It was therefore decided to develop an O(N) simulation for SRSD to investigate applications where they may perform better. This development was performed in-house to allow maximum flexibility and control for different simulations. It should be mentioned that SRSD is also evaluating the Divide-and-Conquer algorithm (Featherstone, 1999) for possible use in its space systems simulations. This paper reports findings related to the O(N) algorithm.

ALGORITHM DEVELOPMENT

The earliest work on the O(N) algorithm was reported by (Armstrong, 1979) in a formulation of dynamic equations for the SRMS (Shuttle Remote Manipulator System built by the National Research Council of Canada) modeled as an open chain of rigid links with three degrees of freedom (DOF) joints. Other early works, all for systems with rigid bodies, include (Featherstone, 1983) on chain of bodies with revolute and prismatic joints, (Brandl, Johanni and Otter, 1987) on bodies in tree topology with arbitrary joints and (Brandl, Johanni and Otter, 1987) for systems with arbitrary joints and kinematic loops. More recent works on O(N) with flexible bodies and kinematic loops include (Banerjee, 2003) which considers motion-induced stiffness of flexible bodies and (Banerjee and Lemak, 2007) which uses an efficient variable formulation of individual flexible links. The references in the last two papers include a broad list of work on the O(N) formulation.

The formulation presented in this work uses a modular approach by algebraically putting together the following sets of equations: (1) kinematic equations relating motions between consecutive joints, (2) equations of motion of a single body, rigid or flexible, (3) equations relating the total spatial force and active forces and moments at a joint and (4) equations for loop closures.

Derivation of the Order-N algorithm is presented in the Appendix. The effect of motion-induced stiffness (Banerjee, 1993) in flexible bodies of the system has not yet been incorporated. Inter-body forces may cause this effect to be important in some cases even for slow motions typical of space systems.

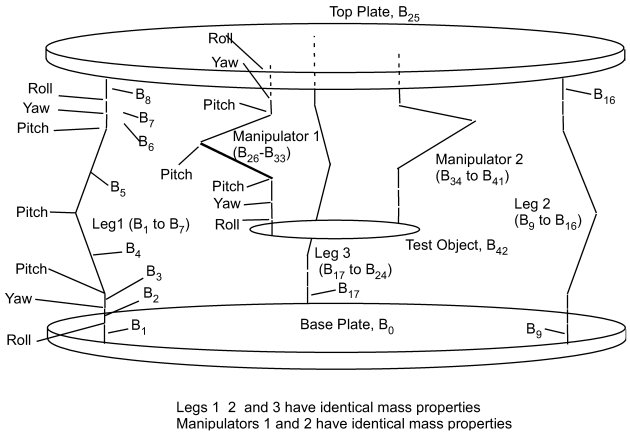
CODE VERIFICATION

The O(N) code was verified against the existing mass matrix-based code for several test cases, which itself was verified against other simulations in the industry, including TREETOPS (Singh et. al.,1985).The mass-matrix algorithm has been in use for many facilities at the Johnson Space Center for many years. Results from the two implementations matched with high accuracy (i.e., within 1.0e-10 or better).

SIMULATION TEST CASES

The test cases are based on variations of the system shown in Figure 1. The base plate B_0 is a rigid circular body. Three identical articulating legs and two identical manipulator arms are rigidly attached to the top plate. The other ends of the legs are rigidly connected to the base plate and the other ends of the manipulators hold a

test object rigidly. The system can be used to represent a space vehicle with three legs visiting a planet, where the top plate is replaced by the crew module modeled as a rigid body and the base plate is fixed to the planet instead of floating. The legs have articulating joints to control the position and orientation of the crew module and are attached to the surface of the planet. During testing of the algorithm the base plate has been allowed to float in order to add additional degrees of freedom to the system.



Figures 1: Test Model Description

All links of the legs and manipulators are modeled as rigid or flexible cylindrical rods. The two long links (called booms) in the middle of a leg or a manipulator are modeled as flexible. Other links are short and modeled as rigid. Each flexible link has four flexible bending modes. In several configurations the boom elements of the legs have been split in the middle into two flexible links of equal length with four bending modes each, and then joined rigidly, for adding additional bodies and flexible degrees of freedom to the system. All joints of the legs and manipulators are single axis rotational joints.

System Configuration	Legs			Manipulators		Total Number of Bodies
	Number of Bodies	Number of Joints	Boom Construction	Number of bodies	Number of Joints	
1	6	6	One rod	0	0	20
2	8	6	Two rods	0	0	26
3	10	7	Two rods	0	0	32
4	10	7	Two rods	8	7	49

Table 1: System Test Case Configurations

Two configurations of the legs and manipulators are considered, one with six joints, the other with seven. The axes of the seven jointed manipulators and legs are in the order roll, yaw, pitch, pitch, pitch, yaw and roll. The axes for ones with six joints are in the order roll, yaw, pitch, pitch, yaw and roll. The system was driven by forces and moments on the top plate and the test object, and moments on joints of the first leg. The forces and moments were held constant for every 20 seconds and then switched in sign. Table 1 shows the configurations. Open loop cases are obtained by freeing the joints between legs 2, 3 and top plate and between manipulator 2 and the test object. For closed loop simulations constraint forces were determined at these same points.

RESULTS

The responses of mass-matrix and Order-N simulations matched very closely, to within numerical accuracies of computations. This is expected because the two methods solve the same equations in two different ways.

Performance of the O(N) code is measured by the CPU time for a 100 second simulation and comparing it with the existing mass-matrix code. The results of the comparison are listed below for rigid and flexible models separately for both open-loop and closed-loop scenarios. Rigid cases were run with 0.001 second and flexible cases were run with 0.0001 second integration time step. An Euler-Cromer integration scheme was used.

Table 2: Timing Results

Case	Simulation	Configuration	Degrees of Freedom			CPU Time (sec)	
			Rigid	Flexible	Total	Order-N	Mass Matrix
Without Loops							
1	R	1	24	0	24	10.1	6.7
2	R	2	24	0	24	12.4	7.9
3	R	3	27	0	27	15.0	10.1
4	R	4	42	0	42	22.3	18.6
5	F	1	24	24	48	166.8	174.0
6	F	2	24	48	72	277.9	423.6
7	F	3	27	48	75	275.9	478.0
8	F	4	42	80	112	408.6	1127.9
Two Loops							
9	R	1	24	0	24	40.7	17.1
10	R	2	24	0	24	48.0	18.4
11	R	3	27	0	27	59.4	22.7
12	F	1	24	24	48	613.4	567.6
13	F	2	24	48	72	937.2	1402.0
14	F	3	27	48	75	1037.9	1551.2
Three Loops							
15	R	4	42	0	42	133.8	57.4
16	F	4	42	80	112	2257.3	4063.9

Simulation: R = Rigid, F = Flexible;
Configuration: As per Table 1

The simulations were run with inputs that produced slow motion. All nonlinear terms in flexible variables and their rates, and products of rigid rates and flexible variables and their rates were neglected. Products of rigid rates

were retained. The derivation in the Appendix however, does not make such assumptions.

DISCUSSION AND CONCLUSIONS

The timing results confirm that the dense mass matrix formulation is faster than the O(N) formulation for smaller degrees of freedom (DOFs) and O(N) is faster otherwise. In our cases for all rigid systems, mass-matrix take less CPU time than O(N) for both open-loop and closed-loop systems because they have fewer DOFs. Flexible body systems have more DOFs and O(N) runs faster, except for Case 12 involving two loops and fewer DOFs than the other flexible cases. Loops add large amount of computation and does so more for O(N). For systems with loops the advantage of O(N) is seen to be less for three loops than for two loops. With more loops the gain in CPU time is less partly because the algorithm requires formation ($O(n_c^2)$ operations) and solution ($O(n_c^3)$ operations) of the constraint equation (Appendix, Equation (4.20)), where n_c is the total number of constrained degrees of freedom. Generation of terms relating acceleration of constrained points in terms of constraint forces, needed by the constraint equation is computation-intensive, even though it is of the order of the number of bodies in the loops.

ACKNOWLEDGEMENT

The authors acknowledge the support provided by An Huynh and Thomas Brain of METECS Corporation, Houston Texas, USA during coding of the simulation, and to Arun Banerjee, formerly of Lockheed Martin Palo Alto Research Laboratory, California, USA for review of the manuscript.

REFERENCES

Armstrong, W.W. 1979. "Recursive Solution to the Equations of Motion of an n-link manipulator," *Proc. 5th World Congress on Theory of machines and Mechanisms,* Vol. 2, Montreal: ASME.

Banerjee, A. K. 1993. "Block-Diagonal Equations for Multibody Elastodynamics with Geometric Stiffness and Constraints," *Journal of Guidance, Control and Dynamics,* Vol. 16, No. 6.

Banerjee, A. K. and M. E. Lemak. 2007. "Recursive Algorithm with Efficient Variables for Flexible Multibody Dynamics with Multiloop Constraints." *Journal of Guidance, Control and Dynamics,* Vol. 30, No, 3.

Baumgarte, J. 1972. "Stabilization of Constraints and Integrals of Motion in Dynamical Systems," *Computational Methods in Applied Mechanics and Engineering,* Vol.1.

Brandl, H.; R. Johanni; and M. Otter. 1986. "A Very Efficient Algorithm for the Simulation of Robots and Similar Multibody Systems Without Inversion of the Mass Matrix." *Proceedings of the IFAC/IFIP/IMACS International Symposium on Theory of Robots,* Vienna, Dec 3-5.

Brandl, H.; R. Johanni; and M. Otter. 1987. "An Algorithm for the Simulation of Multibody Systems with *Kinematic* Loops," *Proceedings of the 7th World Congress on the Theory of Machines and Mechanisms*, (IFToMM), Sevilla.

Featherstone, R. 1983. "The Calculation of Robot Dynamics Using Articulated Body Inertias," *The International Journal of Robotics Research,* Vol. 2, No. 1.

Featherstone, R. 1999. "A Divide-and-ConquerArticulated Body Algorithm for parallel O(log(n)) Calculation for Rigid Body Dynamics, Parts 1 and 2, *The International Journal of Robotics Research*, Vol. 18, No. 9.

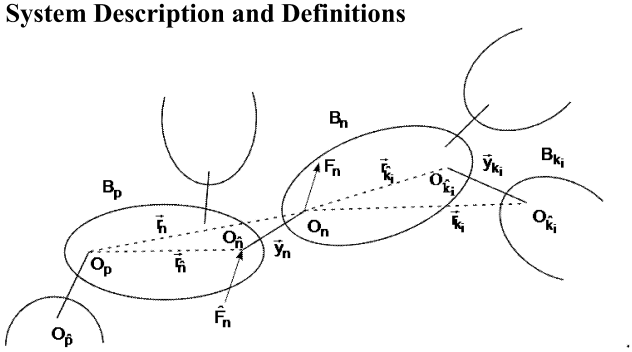
Quicho, L,J.; T. Ghosh; D. Frenkel; and A. Huynh. 2010. "Mode Selection Techniques in Variable Mass Flexible Modeling," In *AIAA Meeting Papers, AIAA Modeling and Simulation Technologies Conference, Aug 2010, Toronto, Canada, AIAA-2010-7606.*

Sincarsin, G. B. and P. C. Hughes. 1989. "Dynamics of an Elastic Multibody Chain: Part A – Body Motion Equations," *Dynamics and Stability of Systems,* Vol. 4 Nos. 3 & 4.

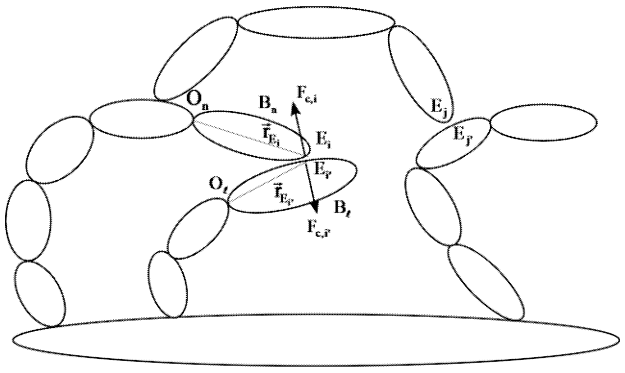
Singh, R.P.; R. J. van der Voort; and P. W. Likins, P. 1985. "Dynamics of Flexible Bodies in Tree Topology- Computer Oriented Approach," *Journal of Guidance, Control and Dynamics*, Vol.26, No. 3.

APPENDIX: DERIVATION OF O(N) EQUATIONS

System Description and Definitions



Figures A1: Multibody System Definitions



Figures A2: Constrained Motion Definitions

Consider the set of rigid and/or flexible bodies, in Figure A1, with the associated labeling of forces and points. $\Re_n$ and $\Re_{\hat{k}_i}$ are frames attached to body n (B_n) at its inboard and outboard joints O_n and $O_{\hat{k}_i}$ respectively. Figure A2 shows the labeling for constrained systems. Throughout this derivation it is assumed that all the vectors and inertia matrices are either generated in or are converted to a single reference frame, which is the structural reference frame $\Re_0$ of the base body of the multibody system. The conversion is performed every time the system states are updated.

Kinematics of Motion between O_p and O_n

Let $\vec{v}_n, \vec{a}_n$ represent the inertial velocity and acceleration respectively, of O_n, $\vec{\omega}_n$ represent the inertial angular velocity of $\Re_n$ and η_p represent the flexible variables of the body B_p. Kinematic equations relating position, velocity and acceleration of O_p and $O_{\hat{n}}$, and angular velocity and angular acceleration of frames $\Re_p$ and $\Re_{\hat{n}}$ are given by

$\vec{r}_{\hat{n}} = \vec{\rho}_{\hat{n}} + \phi_{\hat{n}}\eta_p$, $v_{\hat{n}} = \vec{v}_p + \tilde{\omega}_p r_n + \phi_{\hat{n}}\dot{\eta}_p$,

$\vec{\omega}_{\hat{n}} = \vec{\omega}_p + \psi_{\hat{n}}\dot{\eta}_p$

$\dot{\vec{\omega}}_{\hat{n}} = \dot{\vec{\omega}}_p + \psi_{\hat{n}}\ddot{\eta}_p + \tilde{\omega}_p\psi_{\hat{n}}\dot{\eta}_p$

$\vec{a}_{\hat{n}} = \vec{a}_p - \tilde{r}_n\dot{\vec{\omega}}_p + \phi_{\hat{n}}\ddot{\eta}_p + \tilde{\omega}_p\tilde{\omega}_p\vec{r}_n + 2\tilde{\omega}_p\phi_{\hat{n}}\dot{\eta}_p$

$\vec{\rho}_p$ is the undeformed value of $\vec{r}_p$, the vector from O_p to $O_{\hat{n}}$. Combining the last two equations we can write

$$A_{\hat{n}} = \mathcal{F}_{\hat{n},p}A_p + S_{\hat{n}}\ddot{\eta}_p + A_{\hat{n},r} \qquad (1.1)$$

where, $A_{\hat{n}} = \begin{Bmatrix} \vec{a}_{\hat{n}} \\ \dot{\vec{\omega}}_{\hat{n}} \end{Bmatrix}$, $\mathcal{F}_{\hat{n},p} = \begin{bmatrix} 1 & -\tilde{r}_{\hat{n}} \\ 0 & 1 \end{bmatrix}$, $S_{\hat{n}} = \begin{bmatrix} \phi_{\hat{n}} \\ \psi_{\hat{n}} \end{bmatrix}$

and $A_{\hat{n},r} = \begin{Bmatrix} \tilde{\omega}_p\tilde{\omega}_p\vec{r}_{\hat{n}} + 2\tilde{\omega}_p\phi_{\hat{n}}\dot{\eta}_p \\ \tilde{\omega}_p\psi_{\hat{n}}\dot{\eta}_p \end{Bmatrix}$.

$\mathcal{F}$ denotes a shift operator corresponding to an offset between two points identified by the subscripts. In the case of $\mathcal{F}_{\hat{n},p}$ the offset is $\vec{r}_{\hat{n}}$, the vector from O_p to $O_{\hat{n}}$. The tilde (~) is the usual cross product operator on a vector.

The joint n may have up to six DOFs allowing both translation and rotation. The translation of O_n with respect to $O_{\hat{n}}$ is given by $\vec{y}_n = \sum_{i=1}^{NT_n} \hat{g}_{i,n}\delta_{i,n} = G_n\delta_n$ where $\hat{g}_{i,n}$ are linearly independent unit vectors fixed in $\Re_{\hat{n}}$, $NT_n \le 3$ is the number of translational degrees of freedom of joint n and $\delta_{i,n}$ are scalar quantities, representing joint translations in the directions of the unit vectors. The inertial acceleration of O_n is:

$$\vec{a}_n = \vec{a}_{\hat{n}} - \tilde{y}_n\dot{\vec{\omega}}_{\hat{n}} + G_n\ddot{\delta}_n + \vec{a}_n^{ry} \qquad (1.2)$$

where, $\vec{a}_n^{ry} = \tilde{\omega}_{\hat{n}}\tilde{\omega}_{\hat{n}}\vec{y}_n + 2\tilde{\omega}_{\hat{n}}G_n\dot{\delta}_n$

Inertial angular velocity of $\Re_n$ is given by

$$\vec{\omega}_n = \vec{\omega}_{\hat{n}} + \sum_{i=1}^{NR_n} \hat{l}_{i,n}\dot{\theta}_{i,n} = \vec{\omega}_{\hat{n}} + L_n\dot{\theta}_n$$

where $\hat{l}_{i,n}$ is the unit vector in the direction of the rotation $\theta_{i,n}$ and NR_n is the number of rotational degrees of freedom of the joint. L_n is a matrix whose columns are the unit vectors $\hat{l}_{i,n}$ and θ_n is a column matrix containing the rotations $\theta_{i,n}$. Differentiating $\vec{\omega}_n$ in the inertial frame we get

$$\dot{\vec{\omega}}_n = \dot{\vec{\omega}}_{\hat{n}} + L_n\ddot{\theta}_n + \vec{\alpha}_n^{r\theta} \qquad (1.3)$$

$\vec{\alpha}_n^{r\theta} = \dot{L}_n\dot{\theta}_n$

Combining Equations (1.2) and (1.3),

$$A_n = \mathcal{F}_{n,\hat{n}}A_{\hat{n}} + P_n\ddot{\gamma}_n + A_n^{rj} \qquad (1.4)$$

$A_n = \begin{Bmatrix} \vec{a}_n \\ \dot{\vec{\omega}}_n \end{Bmatrix}$, $\ddot{\gamma}_n = \begin{Bmatrix} \ddot{\delta}_n \\ \ddot{\theta}_n \end{Bmatrix}$, $A_n^{rj} = \begin{Bmatrix} \vec{a}_n^{ry} \\ \vec{\alpha}_n^{r\theta} \end{Bmatrix}$

$$\mathcal{F}_{n,\hat{n}} = \begin{bmatrix} 1 & -\tilde{y}_n \\ 0 & 1 \end{bmatrix} \text{ and } P_n = \begin{bmatrix} G_n & 0 \\ 0 & L_n \end{bmatrix} \qquad (1.5)$$

Finally, combining Equations (1.1) and (1.4) we get

$$A_n = \mathcal{F}_{n,p}A_p + S_n\ddot{\eta}_p + P_n\ddot{\gamma}_n + A_n^r \qquad (1.6)$$

where, using the relationship $\vec{r}_n = \vec{r}_{\hat{n}} + \vec{y}_n$,

$\mathcal{F}_{n,p} = \mathcal{F}_{n,\hat{n}}\mathcal{F}_{\hat{n},p}$, $S_n = \mathcal{F}_{n,\hat{n}}S_{\hat{n}}$, $A_n^r = \mathcal{F}_{n,\hat{n}}A_{\hat{n}}^r + A_n^{rj}$

Inter-Body and Joint Actuator Forces

Let F_n be the spatial forces exerted by B_p on B_n at O_n and $\hat{F}_{\hat{n}}$ be the spatial force exerted by B_n on B_p at $O_{\hat{n}}$. Then, from equilibrium considerations,

$$\hat{F}_{\hat{n}} = -\mathcal{F}_{n,\hat{n}}^T F_n \qquad (2.1)$$

$\mathcal{F}_{n,\hat{n}}$ is given by Equation (1.5). Let $\mu_{i,n}$ and $\nu_{j,n}$ be the i-th actuator force $(i \le 3)$ and j-th actuator torque $(j \le 3)$ on body n at O_n, respectively. They act in the directions of the degrees of freedom $\hat{g}_{i,n}$ and $\hat{l}_{j,n}$ of the joint. The array of actuator forces and moments at joint n can be defined as $\sigma_n = \begin{bmatrix} \mu_n \\ \nu_n \end{bmatrix}$. Using the principle of virtual work it is straight forward to show that

$$\sigma_n = P_n^T F_n \qquad (2.2)$$

where P_n is given by Equation (1.5).

Equations of Motion of Body n

A minor modification of equations of a flexible body in Quiocho, et. al (Quiocho, 2010), produces the equations of a body B_n in rigid and flexible coordinates as

$$M_{rr,n}A_n + M_{re,n}\ddot{\eta}_n = \sum_i \mathcal{F}_{i,n}^T F_{e,i,n} + B_{r,n}$$

$$M_{er,n}A_n + M_{ee,n}\ddot{\eta}_n + K_{ee,n}\eta_n + D_{ee,n}\dot{\eta}_n = \sum_i S_{i,n}^T F_{e,i,n} + B_{e,n}$$

In the above equations, which use a convenient notation of Sincarsin and Hughes (Sincarsin, 1989), the following variables have been defined for body n: $M_{rr,n}$, $M_{re,n}$, $M_{er,n}$ $(= M_{re,n}^T)$, $M_{ee,n}$ are the submatrices of the symmetric positive definite mass matrix, $B_{r,n}$ and $B_{e,n}$ are parts of rigid and flexible inertia forces that are not

dependent upon A_n and $\ddot{\eta}_n$, $K_{ee,n}$ and $D_{ee,n}$ are the stiffness and damping matrices, $F_{e,i,n}$ is the i-th external spatial force, $\mathfrak{F}_{i,n}$ and $S_{i,n}$ are the shift and shape/slope function respectively, at the application point of $F_{e,i,n}$. The subscripts r and e correspond to rigid and flexible degrees of freedom. The first equation follows from force and torque equations and the second equation follows from virtual work equation for the flexible coordinates. We shall write the equations of motion of the body n after separating the external spatial forces $F_{e,i,n}$ on B_n as (i) from the previous body acting at the inboard joint (F_n), (ii) from child bodies B_{k_i} acting at outboard joints ($\hat{F}_{\hat{k}_i}$), (iii) forces at constrained points ($F_{c,j}$) when there are such points on the body, and (iv) external forces ($F_{ext,i,n}$) acting on the body. Because we chose the reference frame of the body to be at the inboard joint, the shift function at the inboard joint is an identity matrix and the shape/slope function at the inboard joint is a null matrix. Equations for body n are:

$$M_{rr,n}A_n + M_{re,n}\ddot{\eta}_n = F_n + \sum_i \mathfrak{F}_{k_i,n}^T \hat{F}_{\hat{k}_i} + \sum_{j\in Z_c(n)} \mathfrak{F}_{c,j}^T F_{c,j} + \sum_i \mathfrak{F}_{i,n}^T F_{ext,i,n} + B_{r,n}$$

$$M_{er,n}A_n + M_{ee,n}\ddot{\eta}_n + K_{ee,n}\eta_n + D_{ee,n}\dot{\eta}_n = \sum_i S_{\hat{k}_i}^T \hat{F}_{\hat{k}_i} + \sum_{j\in Z_c(n)} S_{c,j}^T F_{c,j} + \sum_i S_{i,n}^T F_{ext,i,n} + B_{e,n}$$

In the above equations $F_{c,j}$ is the spatial force at the j-th constrained point, $\mathfrak{F}_{c,j}$ is the shift function for the offset of the constrained point from O_n and $S_{c,j}$ is the shape/slope function of body n at the constrained point. Let us define

$$G_{r,n} = \sum_i \mathfrak{F}_{i,n}^T F_{ext,i,n} + B_{r,n} \qquad (3.1)$$

$$G_{e,n} = \sum_i S_{i,n}^T F_{ext,i,n} + B_{e,n} - K_{ee,n}\eta_n - D_{ee,n}\dot{\eta}_n \qquad (3.2)$$

Using Equations (2.1) for $\hat{F}_{\hat{k}_i}$ and then Equations (3.1) and (3.2) we get the equations of motion of the body n as

$$M_{rr,n}A_n + M_{re,n}\ddot{\eta}_n = F_n - \sum_i \mathfrak{F}_{k_i,n}^T F_{k_i} + \sum_{j\in Z_c(n)} \mathfrak{F}_{c,j}^T F_{c,j} + G_{r,n} \qquad (3.3)$$

$$M_{er,n}A_n + M_{ee,n}\ddot{\eta}_n = -\sum_i S_{k_i}^T F_{k_i} + \sum_{j\in Z_c(n)} S_{c,j}^T F_{c,j} + G_{e,n} \qquad (3.4)$$

Here, $Z_c(n)$ is the set of indices for constrained points on body n.

Recursive Solution of Equations of Motion

The dynamical equations of motion of the system are solved recursively in steps as follows.

Step 1. Forward Pass Kinematics: In this step all position and velocity states are generated starting with the base body using equations derived in Kinematics section.

Step 2. Inward Pass Dynamics Equations:

Observing the equations of motion it is reasonable to expect that for any n F_n can be expressed as linear functions of A_p, $\ddot{\eta}_p$ and constraint forces on itself and on bodies on outer branches:

$$F_n = D_{A,n}A_p + D_{q,n}\ddot{\eta}_p + d_n + \sum_{j\in Y_c(n)} D_{c,n,j}F_{c,j} \qquad (4.1)$$

$Y_c(n)$ is the set of indices of all constrained points on body n and on all bodies of its outer branches.

We shall determine the coefficients in the above equation recursively. In Equation (4.1) using k_i in place of n and n in place of p, and using the result in Equation (3.4) we get

$$\ddot{\eta}_n = \hat{Q}_{A,n}A_n + \hat{q}_n + \sum_{j\in Y_c(n)} \hat{Q}_{c,n,j}F_{c,j} \qquad (4.2)$$

where,

$$\hat{Q}_{A,n} = -\hat{M}_{ee,n}^{-1}\hat{M}_{er,n}, \quad \hat{M}_{ee,n} = M_{ee,n} + \sum_{i\in Z(n)} S_{k_i}^T D_{q,k_i}$$

$$\hat{M}_{er,n} = M_{er,n} + \sum_{i\in Z(n)} S_{k_i}^T D_{A,k_i}$$

$$\hat{q}_n = \hat{M}_{ee,n}^{-1}[G_{e,n} - \sum_{i\in Z(n)} S_{k_i}^T d_{k_i}]$$

$$\hat{Q}_{c,n,j} = -\hat{M}_{ee,n}^{-1}S_{k_i}^T D_{c,k_i,j} \quad \text{for } i\in Z(n),\ j\in Y_c(k_i)$$

$$\hat{Q}_{c,n,j} = \hat{M}_{ee,n}^{-1}S_{c,j}^T \quad \text{for } j\in Z_c(n)$$

$Z(n)$ is the set of indices of child bodies of B_n. In the same manner, substituting for F_{k_i} in Eq. (3.3) and rearranging we get,

$$\overline{M}_{rr,n}A_n + \overline{M}_{re,n}\ddot{\eta}_n = F_n + \overline{G}_{r,n} + \sum_{j\in Y_c(n)} \overline{D}_{c,n,j}F_{c,j} \qquad (4.3)$$

where,

$$\overline{M}_{rr,n} = M_{rr,n} + \sum_{i\in Z(n)} \mathfrak{F}_{k_i,n}^T D_{A,k_i},$$

$$\overline{M}_{re,n} = M_{re,n} + \sum_{i\in Z(n)} \mathfrak{F}_{k_i,n}^T D_{q,k_i},$$

$$\overline{G}_{r,n} = G_{r,n} - \sum_{i\in Z(n)} \mathfrak{F}_{k_i,n}^T d_{k_i}`$$

$$\overline{D}_{c,n,j} = -\sum_i \mathfrak{F}_{k_i,n}^T D_{c,k_i,j} \quad \text{for } i\in Z(n),\ j\in Y_c(k_i)$$

$$\overline{D}_{c,n,j} = \mathfrak{F}_{c,j}^T \quad \text{for } j\in Z_c(n)$$

Using Equation (4.2) for $\ddot{\eta}_n$ in Equation (4.3) we get

$$\hat{M}_{rr,n}A_n = F_n + \hat{G}_{r,n} + \sum_{j\in Y_c(n)} \hat{D}_{c,n,j}F_{c,j} \qquad (4.4)$$

where,

$$\hat{M}_{rr,n} = \overline{M}_{rr,n} + \overline{M}_{re,n}\hat{Q}_{A,n}, \quad \hat{G}_{r,n} = \overline{G}_{r,n} - \overline{M}_{re,n}\hat{q}_n$$

$$\hat{D}_{c,n,j} = \overline{D}_{c,n,j} - \overline{M}_{re,n}\hat{Q}_{c,n,j}$$

Using Equation (1.6) for A_n in Equation (4.4) and rearranging, we get

$$\hat{M}_{rr,n}\mathfrak{F}_{n,p}A_p + \hat{M}_{rr,n}S_n\ddot{\eta}_p + \hat{M}_{rr,n}P_n\ddot{\gamma}_n = F_n + \hat{G}_{r,n} - \hat{M}_{rr,n}A_n^r + \sum_{j\in Y_c(n)}\hat{D}_{c,n,j}F_{c,j}$$

Pre-multiplying this equation by P_n^T, using $\sigma_n = P_n^T F_n$ from Equation (2.2) and solving the resulting equation for $\ddot{\gamma}_n$ we get

$$\ddot{\gamma}_n = B_{A,n}A_p + B_{q,n}\ddot{\eta}_p + b_n + \sum_{j\in Y_c(n)}B_{c,n,j}F_{c,j} \qquad (4.5)$$

where,

$$M_{\gamma\gamma,n} = P_n^T\hat{M}_{rr,n}P_n, \qquad B_{A,n} = -M_{\gamma\gamma,n}^{-1}P_n^T\hat{M}_{rr,n}\mathfrak{F}_{n,p},$$

$$B_{q,n} = -M_{\gamma\gamma,n}^{-1}P_n^T\hat{M}_{rr,n}S_n,$$

$$b_n = M_{\gamma\gamma,n}^{-1}\left[\sigma_n + P_n^T(\hat{G}_{r,n} - \hat{M}_{rr,n}A_n^r)\right],$$

$$B_{c,n,j} = M_{\gamma\gamma,n}^{-1}P_n^T\hat{D}_{c,n,j}$$

Equation (4.5) for $\ddot{\gamma}_n$ in Equation (1.6) yields for A_n

$$A_n = W_{A,n}A_p + W_{q,n}\ddot{\eta}_p + w_n + \sum_{j\in Y_c(n)}W_{c,n,j}F_{c,j} \qquad (4.6)$$

where,

$$W_{A,n} = \mathfrak{F}_{n,p} + P_nB_{A,n},\ W_{q,n} = S_n + P_nB_{q,n},$$

$$w_n = P_nb_n + A_n^r, \text{ and } W_{c,n,j} = P_nB_{c,n,j}$$

Equation (4.6) for A_n in Equation (4.2) gives

$$\ddot{\eta}_n = Q_{A,n}A_p + Q_{q,n}\ddot{\eta}_p + q_n + \sum_{j\in Y_c(n)}Q_{c,n,j}F_{c,j} \qquad (4.7)$$

where,

$$Q_{A,n} = \hat{Q}_{A,n}W_{A,n},\ Q_{q,n} = \hat{Q}_{A,n}W_{q,n},$$

$$q_n = \hat{Q}_{A,n}w_n + \hat{q}_n, \text{ and } Q_{c,n,j} = \hat{Q}_{A,n}W_{c,n,j} + \hat{Q}_{c,n,j}$$

Finally, using Equation (4.6) in Equation (4.4) we get

$$F_n = D_{A,n}A_p + D_{q,n}\ddot{\eta}_p + d_n + \sum_{j\in Y_c(n)}D_{c,n,j}F_{c,j} \qquad (4.8)$$

where,

$$D_{A,n} = \hat{M}_{rr,n}W_{A,n},\ D_{q,n} = \hat{M}_{rr,n}W_{q,n}$$

$$d_n = \hat{M}_{rr,n}w_n - \hat{G}_{r,n}, \text{ and}$$

$$D_{c,n,j} = \hat{M}_{rr,n}W_{c,n,j} - \hat{D}_{c,n,j}$$

We can see that Equation (4.8) is in the same form as Equation (4.1) we started with, confirming that if the latter is true for child bodies, it would be true for the current body also. Following the same steps as above for any outermost body for which $F_{\hat{k}_i} = 0$ it is easy to show that Equation (4.1) is true for such bodies. Using the induction logic it therefore follows that Equation (4.1) is true for all bodies of the system.

Computations for systems with multiple branches need to start from an outermost body, move inward till a body with multiple child bodies is reached. Inward computation from a body with multiple child bodies should continue only after computations for all of its child bodies are completed.

A_0 and F_0 for the Base Body B_0 in terms of the constraint forces are obtained from Equation (4.4). When B_0 is fixed in inertial frame $A_0 = 0$ and we then have

$$F_0 = -\hat{G}_{r,0} - \sum_{j\in Y_c(0)}\hat{D}_{c,0.j}F_{c,j}$$

When B_0 is free in the inertial frame $F_0 = 0$ and

$$A_0 = \hat{M}_{rr,0}^{-1}[\hat{G}_{r,0} + \sum_{j\in Y_c(0)}\hat{D}_{c,0.j}F_{c,j}] \qquad (4.9)$$

The flexible coordinate acceleration $\ddot{\eta}_0$ for the base body is obtained from Equation (4.2)

$$\ddot{\eta}_0 = \hat{Q}_{A,0}A_0 + \hat{q}_0 + \sum_{j\in Y_c(0)}\hat{Q}_{c,0,j}F_{c,j} \qquad (4.10)$$

For systems without closed loops the $F_{c,j}$ term drops out and the equations obtained above are sufficient for determining A_0 and $\ddot{\eta}_0$, and then the system accelerations, by successive computation of $\ddot{\gamma}_n$, $\ddot{\eta}_n$ and A_n in an outward sweep.

Step 3. Accelerations of Constrained Points in terms of Forces at Constrained Points:

For the determination of acceleration of constrained points in terms of forces at these points we seek to express the accelerations A_n and $\ddot{q}_n$ for bodies in the path from base body to the constrained points in the form

$$A_n = h_n^A + \sum H_{n,j}^A F_{c,j} \quad \text{and} \quad \ddot{\eta}_n = h_n^q + \sum H_{n,j}^q F_{c,j} \qquad (4.11)$$

Here the summation index j covers all constrained points. It follows from Equations (4.9) and (4.11) that when the base is free in inertial frame

$$h_0^A = \hat{M}_{rr,0}^{-1}\hat{G}_{r,0},\ H_{0,j}^A = \hat{M}_{rr,0}^{-1}\hat{D}_{c,0,j}, \quad h_0^q = \hat{Q}_{A,0}h_0^A + \hat{q}_0$$

and $H_0^q = \hat{Q}_{c,0,j}$.

When the base is fixed, $h_0^A = 0$, $H_{0,j}^A = 0$, $h_0^q = \hat{q}_0$ and $H_0^q = \hat{Q}_{c,0,j}$.

Using Equations (4.7) and (4.11) we get the recursive equations for h_n^q and $H_{n,j}^q$:

$$h_n^q = Q_{A,n}h_p^A + Q_{q,n}h_p^q + q_n \text{ and}$$

$$H_{n,j}^q = Q_{A,n}H_{p,j}^A + Q_{q,n}H_{p,j}^q + Q_{c,n,j}$$

Using Equation (4.6) and (4.11) we get the recursive equations for h_n^A and $H_{n,j}^A$:

$$h_n^A = W_{A,n}h_p^A + W_{q,n}h_p^q + w_n, \text{ and}$$

$$H_{n,j}^A = W_{A,n}H_{p,j}^A + W_{q,n}H_{p,j}^q + W_{c,n,j}.$$

Let E_i be the point in constraint i located on body n and $E_{i'}$ be the mating point of the same constraint, located on body ℓ. The spatial acceleration of E_i is

$$A_{E_i} = \mathfrak{F}_{E_i,n}A_n + S_{E_i}\ddot{\eta}_n + A_{E_i,r} \qquad (4.12)$$

where $\mathfrak{F}_{E_i,n}$ is the standard shift operator for the offset $\vec{r}_{E_i}$ of E_i with respect to O_n $S_{E_i} = \begin{bmatrix} \phi_{E_i} \\ \psi_{E_i} \end{bmatrix}$ is the shape/slope function of B_n at E_i and,

$$A_{E_i,r} = \begin{bmatrix} \tilde{\omega}_n(\tilde{\omega}_n \vec{r}_{E_i} + 2\phi_{E_i}\dot{\eta}_n) \\ \tilde{\omega}_n \psi_{E_i} \dot{\eta}_n \end{bmatrix}.$$

For each constraint k (k = 1, 2, …, n_c) the constraint forces $F_{c,k}$ and $F_{c,k'}$ at the two constrained points E_k and $E_{k'}$ respectively are equal and opposite, i.e, $F_{c,k'} = -F_{c,k}$. Using this and Equation (4.11) for A_n and $\ddot{\eta}_n$ in Equation (4.12) we get,

$$A_{E_i} = \mathfrak{F}_{E_i,n} h_n^A + S_{E_i} h_n^q + A_{E_i,r} + \sum_{k=1}^{n_c} [\mathfrak{F}_{E_i,n}(H_{n,k}^A - H_{n,k'}^A) + S_{E_i}(H_{n,k}^q - H_{n,k'}^q)] F_{c,k} \quad (4.13)$$

Similarly, spatial acceleration of $E_{i'}$ is given by

$$A_{E_{i'}} = \mathfrak{F}_{E_{i'},\ell} h_\ell^A + S_{E_{i'}} h_\ell^q + A_{E_{i'},r} + \sum_{k=1}^{n_c} [\mathfrak{F}_{E_{i'},\ell}(H_{\ell,k}^A - H_{\ell,k'}^A) + S_{E_{i'}}(H_{\ell,k}^q - H_{\ell,k'}^q)] F_{c,k} \quad (4.14)$$

Step 4. Determination of Constraint Forces**:**

Using Equations (4.13) and (4.14) the difference in the accelerations at the constraint i may be written as

$$\Delta A_{E_i} = A_{E_{i'}} - A_{E_i} = \Delta h_{E_i} + \sum_k \Delta H_{E_i,k} F_{c,k} \quad (4.15)$$

where,

$$\Delta h_{E_i} = \mathfrak{F}_{E_{i'},\ell} h_\ell^A + S_{E_{i'}} h_\ell^q + A_{E_{i'},r} - \mathfrak{F}_{E_i,n} h_n^A - S_{E_i} h_n^q - A_{E_i,r}$$

$$\Delta H_{E_i,k} = \mathfrak{F}_{E_{i'},\ell}(H_{\ell,k}^A - H_{\ell,k'}^A) - \mathfrak{F}_{E_i,n}(H_{n,k}^A - H_{n,k'}^A) + S_{E_{i'}}(H_{\ell,k}^q - H_{\ell,k'}^q) - S_{E_i}(H_{n,k}^q - H_{n,k'}^q)$$

The summation range for k is all the constraints.

The linear and angular constraints at E_i are

$$\hat{g}_{c,i,j} \bullet (\vec{v}_{E_{i'}} - \vec{v}_{E_i}) = 0 \text{ for } 0 < j \le n_{t,i}$$

$$\hat{\ell}_{c,i,j} \bullet (\vec{\omega}_{E_{i'}} - \vec{\omega}_{E_i}) = 0 \quad \text{for } 0 < j \le n_{r,i}$$

where $\vec{v}$ and $\vec{\omega}$ represent the inertial velocity and angular velocity of points corresponding to the subscripts, $\hat{g}_{c,i,j}$ and $\hat{\ell}_{c,i,j}$ are unit vectors in the direction of translational and rotational constraints respectively, and $n_{t,i}(\le 3)$ and $n_{r,i}(\le 3)$ are the number of these constraints, respectively. $\vec{v}_{E_i}$, $\vec{v}_{E_{i'}}$, $\vec{\omega}_{E_i}$ and $\vec{\omega}_{E_i}$ are determined in the manner used for the point $O_{\hat{n}}$ in the Kinematics section. Defining $\Delta V_{E_i} = \begin{bmatrix} \vec{v}_{E_{i'}} - \vec{v}_{E_i} \\ \vec{\omega}_{E_{i'}} - \vec{\omega}_{E_i} \end{bmatrix}$ and $P_{c,i} = \begin{bmatrix} G_{c,i} & 0 \\ 0 & L_{c,i} \end{bmatrix}$ where $G_{c,i}$ is a $3 \times n_{t,i}$ matrix whose columns are the unit vectors $\hat{g}_{c,i,j}$ and $L_{c,i}$ is a $3 \times n_{r,i}$ matrix whose columns are the unit vectors $\hat{\ell}_{c,i,j}$, the difference in the spatial velocities of the constrained points at constraint i may be written as $P_{c,i}^T \Delta V_{E_i}$. The difference in the spatial acceleration in the constraint directions at the constraint point should be zero, giving

$$P_{c,i}^T \Delta A_{E_i} + \dot{P}_{c,i}^T \Delta V_{E_i} = 0$$

Using Baumgarte's stabilization (Baumgarte, 1972) scheme to limit constraint violation caused by numerical errors, this equation is modified to

$$P_{c,i}^T \Delta A_{E_i} + \dot{P}_{c,i}^T \Delta V_{E_i} + K_i P_{c,i}^T \Delta X_{E_i} + C_i P_{c,i}^T \Delta V_{E_i} = 0$$

K_i and C_i are positive constants to provide constraint stabilization. Using Equation (4.15) for ΔA_{E_i} we have

$$P_{c,i}^T \sum_k \Delta H_{E_i,k} F_{c,k} = -\dot{P}_{c,i}^T \Delta h_{E_i} - \dot{P}_{c,i}^T \Delta V_{E_i} - K_i P_{c,i}^T \Delta X_{E_i} - C_i P_{c,i}^T \Delta V_{E_i} \quad (4.16)$$

Let us define $f_{c,k,j}$ to be the constraint force in the direction $\hat{g}_{c,k,j}$ and $\tau_{c,k,j}$ the constraint torque in the direction $\hat{\ell}_{c,k,j}$ and $\hat{F}_{c,k} = \begin{bmatrix} f_{c,k} \\ \tau_{c,k} \end{bmatrix}$. The force $F_{c,k}$ at the constraint point E_k due to forces and moments in the constraint directions is then given by $P_{c,k}\hat{F}_{c,k}$. Let $P_{f,k} = \begin{bmatrix} G_{f,k} & 0 \\ 0 & L_{f,k} \end{bmatrix}$ where $G_{f,k}$ and $L_{f,k}$ are made of the unit vectors normal to the constraint directions for translation and rotation and $\hat{F}_{f,k}$ be the array of forces and moments in these directions. Net spatial force at E_k is

$$F_{c,k} = P_{c,k}\hat{F}_{c,k} + P_{f,k}\hat{F}_{f,k} \quad (4.17)$$

Restricting to cases where $\hat{F}_{f,k}$ is fully known, we use Equation (4.17) in Equation (4.16) to get

$$P_{c,i}^T \sum_k \Delta H_{E_i,k} P_{c,k} \hat{F}_{c,k} = Z_{c,i} \quad (4.18)$$

where,

$$Z_{c,i} = -P_{c,i}^T [\sum_k \Delta H_{E_i,k} P_{f,k} \hat{F}_{f,k} + \Delta h_{E_i}] - \dot{P}_{c,i}^T \Delta V_{E_i} - K_i P_{c,i}^T \Delta X_{E_i} - C_i P_{c,i}^T \Delta V_{E_i} \quad (4.19)$$

Stacking Equation (4.18) for all constraints we get

$$M_c \hat{F}_c = Z_c \quad (4.20)$$

where the $[i,k]$ submatrix of matrix M_c is given by

$$M_{c,i,k} = P_{c,i}^T \Delta H_{E_i,k} P_{c,k}$$

and Z_c is made of arrays $Z_{c,i}$ given by Equation (4.19). Equation (4.20) is solved for $\hat{F}_c$ and the spatial forces $F_{c,k}$ at constraint points are found using Equation (4.17).

Step 5. Computation of System Accelerations**:**

After determination of forces at the constrained points, A_0 and $\ddot{\eta}_0$ are determined using Equations (4.9) and (4.10). $\ddot{\gamma}_n$, $\ddot{\eta}_n$, and A_n are determined recursively in a forward pass using Equations (4.5), (4.7) and (1.6) respectively.

Order of Computations

For systems without loops, the equations obtained for the first forward kinematics pass of Step 1, inward dynamics pass of Step 2 and the final forward pass of Step 5 are executed once for each body. If the joint and flexible degrees of freedom are the same for all bodies, the computations are of the order of the number of bodies (N) of the system. However, for systems with loops additional computations for constraint forces at loop closure points are needed and they involve only the bodies in the loops and, formation and solution of the linear Equation (4.20). Computations for inward and outward passes for loops are performed once for each body in the loop. Subsequent generation of variables in the arrays M_c and Z_c of Equation (4.20) is proportional to the square of number of constraints at the loop closure points. Solution of Equation (4.20) which is of dimension of the total number of constraints, involves arithmetic operations of the order of the cube of the number of constraints. The amount of computation for systems with loops is therefore a linear combination of number of bodies in the system, number of bodies in the loops, square of number of constraints at the loops and, cube of the number of constraints at the loops.

Implementation Notes

It can be shown that in the equations presented above $\overline{M}_{re,n} = \hat{M}^T_{er,n}$ and the matrices $\hat{M}_{ee,n}$, $\hat{M}_{rr,n}$, $\overline{M}_{rr,n}$, $M_{\gamma\gamma,n}$ and M_c are symmetric and positive definite. This can be taken advantage of during coding.

For the test cases presented in this paper, the Baumgarte constraint stabilization constants K_i and C_i in Equation (4.19) for all the constrained points were chosen to be diagonal matrices with the same constants k and c respectively, appearing at the diagonal elements, where $k = [2\pi/10\Delta T]^2$, $c = 2\sqrt{k}$ and ΔT is the integration step. This choice corresponds to a nominal oscillation of the constraint error with a natural time period of ten integration time steps and an added critical damping (damping factor = 1.0). With this choice the constraint errors remained very small (< 1.0e-10) throughout the simulation.

AUTHOR BIOGRAPHIES

TUSHAR K GHOSH received his B.S. and M.S. degrees from Indian Institute of Technology, Kharagpur in 1967 and 1969 respectively and M.E. and Ph.D. degrees from the University of Florida, Gainesville, Florida, USA in 1970 and 1974 respectively. He worked as a machine designer in India from 1974 to 1984. He has been working as a dynamics model developer for different defense and aerospace companies in the United States since 1985. He has been developing dynamics models of on-orbit space robotic manipulators and space vehicles at NASA Lyndon B. Johnson Space Center since 1990.

LESLIE J QUIOCHO received his B.S. in Mechanical Engineering from the University of Texas at El Paso in 1986 and his M.S. in Mechanical Engineering from the University of Houston in 1989 while under a NASA Graduate Research Fellowship. He has worked for the NASA Lyndon B. Johnson Space Center since 1989, primarily focusing on dynamics and controls modeling for on-orbit robotic manipulator systems. He has also been extensively involved in pre and post flight mission analyses involving these robotic systems, and in particular, their interaction with vehicle docking and berthing mechanisms. His current focus is on the development of robotics simulator facilities for crew training.

INDUSTRIAL SIMULATION IN FLUID POWER SYSTEMS BY SIMULINK AND AMESIM

Daniela Vasiliu, Georgiana Claudia Vasiliu, Nicolae Vasiliu
University POLITEHNICA of Bucharest
313, Splaiul Independentei, Sector 6
RO 060042 Bucharest, Romania
E-mail: vasiliu@fluid-power.pub.ro

Petru-Cristinel Irimia
LMS Test Division - CAE Division
15A, I. Slavici Street
RO 500398 Brasov, Romania
E-mail: cristi.irimia@lmsintl.com

KEYWORDS
Numerical simulation, electro hydraulic control systems, fine tuning

ABSTRACT

The paper presents the strategy of finding a good control algorithm for the bent axis electro hydraulic servomotors through numerical simulation. Two simulation environments - SIMULINK and AMESIM are used in order to point out the utility of the numerical simulation in any industrial design process. At the same time, the advantages and the drawbacks of each language are identified. The theoretical results regarding the stability and the accuracy of the force feedback servo mechanisms are found in good agreement with the experimental ones. A final numerical study on the ways to improve stability improvement ways shows practical interesting results concerning the real effectiveness of the small metering orifices, widely used as damping components in hydraulic control systems. These hydraulic resistances can simply improve the stability.

INTRODUCTION

The modern hydrostatic transmissions requiring a continuously variable motor shaft speed are based on the electro hydraulic proportional control of a bent axis axial pistons motor displacement (Fig.1). The servo valve is governed by a DC solenoid attached to the control cover. When the solenoid current increases above the threshold current, the servo piston starts to move from the max towards the minimum displacement position. The shaft speed vs. current relation is a non-linear one. In order to minimize hysteresis, a pulse-width modulated control signal of 70 to 90 Hz is to be utilized. This type of hydraulic motor has gained wide spread acceptance in closed loops industrial systems because they do not need an auxiliary pump: a three-way selector takes the small control oil flow from the motor power ports. The overall motor size is small enough to fit the driving wheels, winches etc. The swash plate variable displacement motors are heavier and expensive because they need an external servo valve, and an additional medium pressure gear pump. They are promoted especially in high performance applications like aerospace and military ones. The main problem of these servomotors dynamics is the control stability generated by the low stiffness feedback spring.

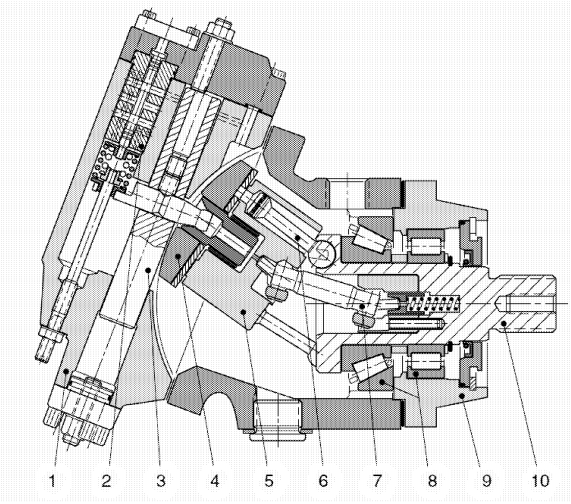


Figure 1: Bent axis electro hydraulic servomotor (PARKER): 1 - servomechanism body; 2 - flow control valve; stroking piston; 4 - sliding valve plate.

Industrial hydraulic power systems need different types of control systems for servomotor displacement, but the hydro mechanic and the electro hydraulic ones are widely used because their full compatibility with the internal force feedback servomechanism. Figure 2 presents the hydraulic diagram for both types.

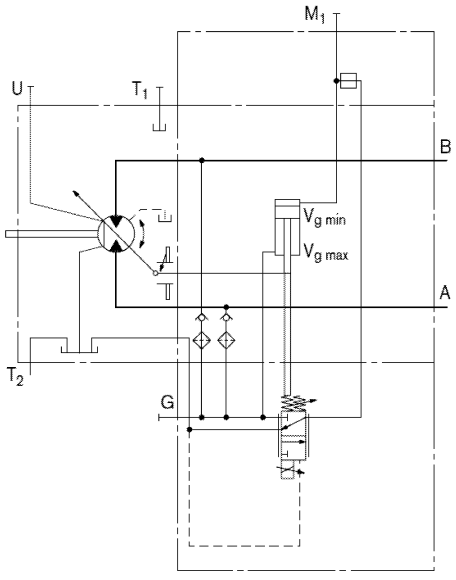


Figure 2: Structure of an electro hydraulic bent axis servomotor controlled by a proportional solenoid

The use of a differential setting piston allows the use of a simple three way flow control valve with round ports and slight positive overlap (see ANNEX A). The high performances applications require a four-way flow control valve with critical center and a symmetric piston (Fig. 3). The dynamic stability requires two metering orifices placed on the valve external ports.

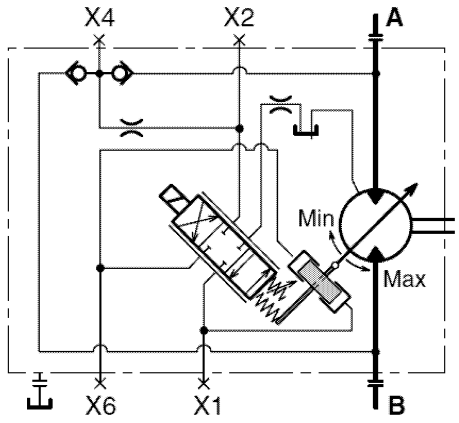


Figure 3: High dynamics electro hydraulic bent axis servomotor (V14-PARKER)

The high accuracy closed loop control systems need an additional electric feedback supplied by an immersed LVDT attached to the setting piston (Fig. 4).The electrical feedback is also useful for the real time survey of the overall servo motor state.

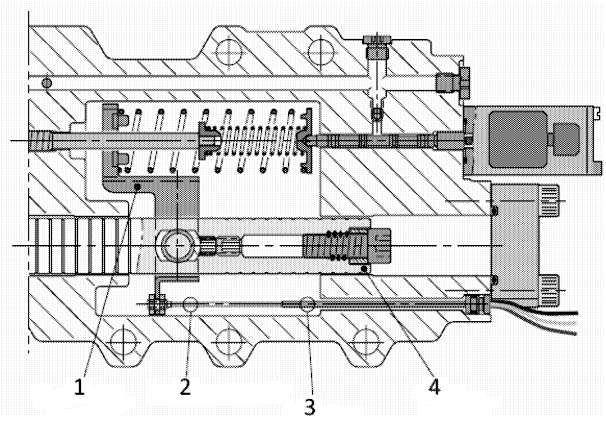


Figure 4: Electro hydraulic bent axis servomotor with additional electric feedback (V14-PARKER): 1-feedback arm; 2-moble armature; 3-position sensor; 4-setting piston.

The force feedback servomechanism included in a bent axis electro hydraulic servo motor has a quasi-linear steady-state displacement-solenoid current characteristics, (Fig. 5), depending on the threshold current. The external steady state characteristics of the electro hydraulic force feedback servomotor, representing the relation between the shaft speed and the solenoid current has a parabolic shape (Fig. 6), corresponding to the typical option of a fine speed control at the start of a power train hydrostatic transmission. The bent angle wide range field (6.5°…35°) allows a wide speed range (100…3000 rev/min) with a good overall efficiency in a large pressure range (210 bar…420 bar); the average specific power reaches 3.0 KW/kg, and the specific torque overcomes 10 Nm/kg.

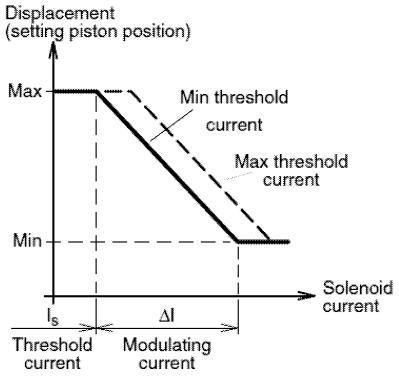


Figure 5: Steady state characteristics $V_g(I)$ of an electro hydraulic force feedback servomotor (PARKER)

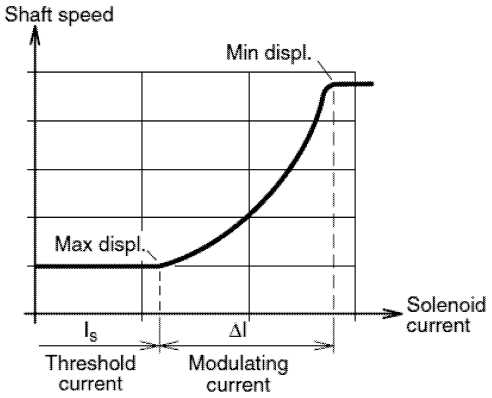


Figure 6: External steady state characteristics of an electro hydraulic force feedback servomotor (PARKER)

This paper is devoted to the mathematical modeling and numerical simulation of the basic servo motor design from ANNEX A. The system dynamics is studied by the aid of both SIMULINK and AMESIM simulation languages in order to point out their benefits and the drawbacks.

MATHEMATICAL MODELING

A typical force feedback servomechanism included in a bent axis electro hydraulic servomotor, presented in the ANNEX A includes a three-port valve, shown in Figure 7. The important advantage of single critical axial dimension requires bias pressure acting on one side of an asymmetrical cylinder for direction reversal. Usually the head-side piston area is twice the rod-side area and supply pressure acts on the smaller area to provide the bias force for reversal. The simplest spool control system is a hydraulic one. It is used in heavy vehicle hydrostatic transmissions in close connection with the mail pump flow control pressure. The valve is supplied by the high

pressure motor line by a three way selector. The same pressure is applied on the rod-side area.

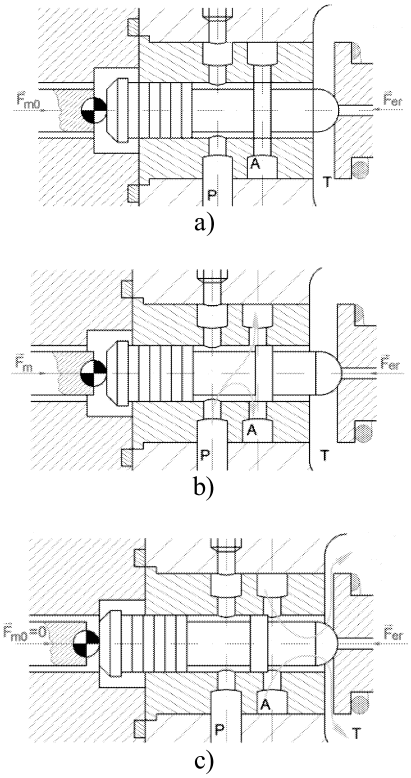


Figure 7: Control valve connections: a) all the ports closed; b) P→A and T closed; c) A→T and P closed.

The modern hydrostatic transmissions are controlled by industrial process computers, which need an electromagnetic solenoid to actuate the control valve. In this last case, the mathematical model of the studied system includes the following equations.

a) Motion equation of the solenoid plunger and the spool:

$$m_x\ddot{x} = F_m - F_{em} - F_{er} - F_h - F_{fv} - F_{fs} \tag{1}$$

Here

$$F_m = K_m(i - i_0) \tag{2}$$

is the electromagnetic force corresponding to the difference between the input current and the threshold one, denoted by

$$I = i - i_0 \tag{3}$$

In practical computation the dynamics of the solenoid plunger can be regarded as a first order lag with a time constant of about 35 ms (Muraru et al. 2000).The slope of the solenoid static characteristics is

$$K_m = \frac{F_m(I_{\max})}{I_{\max}} \tag{4}$$

The solenoid plunger is retired inside the coil by the spring force

$$F_{em} = K_{em}(x + x_{0em}) \tag{5}$$

which depends on the solenoid spring stiffness, K_{em}, the preload x_{0em}, and the spool travel from the neutral position, x defined by the geometrical closure of all the three valve orifices, P, A and T. The feedback force,

$$F_{er} = K_{er}y \tag{6}$$

depends on the setting piston stroke, y and the feedback spring stiffness, K_{er}.The maximum setting piston stroke can be approximated by the relation

$$y_{\max} \cong R_d(\alpha_{\max} - \alpha_{\min}) \tag{7}$$

Here R_d is the radius of the cylindrical surface of the servomechanism body in contact with the motor sliding valve plate. The flow force depends on the spool travel direction. For x > 0,

$$F_h^+ = \rho Q^+ v^+ \cos\theta \tag{8}$$

Here

$$Q^+ = 2c_d A^+(x)\sqrt{\frac{2(p_s - p_1)}{\rho}} \tag{9}$$

and

$$v^+ = c_v\sqrt{\frac{2(p_s - p_1)}{\rho}} \tag{10}$$

Here θ is the jet angle regarding the spool axis. The valve sleeve has two round orifices with a diameter of d_0. Finally,

$$F_h^+ = K_h \cdot A(x)\cdot(p_s - p_1) \tag{11}$$

Here p_s - input motor line pressure; p_1 - pressure in the hydraulic cylinder "passive" chamber. The flow force constant,

$$K_h = 2c_d c_v \cos\theta \tag{12}$$

has no dimensions. This force trends to close the valve.

If x < 0,

$$F_h^- = \rho Q^- v^- \cos\theta \tag{13}$$

Here,

$$Q^- = 2c_d A^-(|x|)\sqrt{\frac{2p_1}{\rho}}$$

(14)

and

$$v^- = c_v\sqrt{\frac{2p_1}{\rho}} \tag{15}$$

The flow force F_h^- trends also to close the valve:

$$F_h^- = K_h A^-(|x|)\,p_1 \tag{16}$$

The viscous friction force,

$$F_{fv} = K_{fv} \cdot \dot{x} \tag{17}$$

includes a coefficient which has to obtained through experiments, K_{fv}. The same problem arises by the Coulomb static friction force:

$$F_{fs} = K_{fs} \cdot sign\,\dot{x} \tag{18}$$

The moving valve assembly includes the equivalent plunger mass, m_m the spool mass, m_s and a part of the feedback spring mass m_{er}:

$$m_x = m_m + m_s + 0{,}35\, m_{er} \tag{19}$$

b) The setting piston motion equation is

$$m_y \ddot{y} = F_{p1} - F_{p2} - F_{er} + F_b - F_{fvp} - F_{fsp} \tag{20}$$

where

$$m_y = m_p + m_t + m_{bc} \tag{21}$$

is the stroking assembly equivalent mass, composed by the stroking piston mass m_p, feedback arm mass, m_t and the barrel equivalent mass, m_{bc}. The other forces can be computed as follows:

$$F_{p1} = p_1 A_1 \tag{22}$$

$$F_{p2} = p_2 A_2 \tag{23}$$

$$F_{er} = K_{er} y \tag{24}$$

$$F_{fvp} = K_{fvp} \dot{y} \tag{25}$$

$$F_{fsp} = K_{fsp} sign\,\dot{y} \tag{26}$$

$$F_b = p_s K_b - y K_y \tag{27}$$

The barrel tilting force coefficients K_b and K_y have to be found by experiments.
The systematic experimental research carried out in RWTH from Aachen has shown (Haas 1989) that the component depending on the stroking piston trends to increase the motor displacement.
This means a positive feedback in the displacement control loop!
However, the value of the coefficient K_b is usually much more than the value of the coefficient K_y. For example, for the hydraulic motor HYDROMATIK A6V, K_b= 2,6 N/bar and K_y = 1,35 N/mm.
c) The flow equation corresponding to the "passive" chamber of the stroking piston depends on the spool motion direction from the neutral position.

If $0 \le x \le x_{max}$,

$$Q_d^{+} - \dot{y} A_1 - Q_{l1} = \frac{V_1}{\varepsilon} \cdot \frac{dp_1}{dt} \tag{28}$$

The control flow valve can be computed by the relation

$$Q_d^{+} = 2 c_d A(|x|) \sqrt{\frac{2(p_s - p_1)}{\rho}} \ge 0 \tag{29}$$

Usually, the leakage flow has a basic laminar component:

$$Q_{l1} = K_{l1} p_1 > 0 \tag{30}$$

in which

$$K_{l1} \cong 2{,}5 \frac{\pi d_1 j_1^{3}}{12 \eta l_1} \tag{31}$$

and

$$V_1 = V_{10} + A_1 y \tag{32}$$

is the "passive" chamber volume. For y = 0, $V_1 = V_{10}$ (the initial volume). The flow equation becomes:

$$\dot{p}_1 = \frac{\varepsilon}{V_1} \left(Q_d^{+} - \dot{y} A_1 - Q_{l1} \right) \tag{33}$$

If $x_{min} \le x < 0$, the flow equation becomes

$$\dot{p}_1 = -\frac{\varepsilon}{v_1} \left(Q_d^{-} + \dot{y} A_1 + Q_{l1} \right) \tag{34}$$

where

$$Q_d^{-} = 2 c_d A(|x|) \sqrt{\frac{2 p_1}{\rho}} \ge 0 \tag{35}$$

The metering orifices can be computed by the relation

$$A(x_r) = \frac{A_0}{\pi} \left[\arccos(1 - 2x_r) - 2(1 - 2x_r) \sqrt{x_r - x_r^2} \right] \tag{36}$$

where x_r is the relative opening of the valve, $x_r = |x|/d_0$ and $A_0 = \pi d_0^2 / 4$ is the maximum orifices area.

NUMERICAL SIMULATION BY SIMULINK

The above mathematical model was integrated by SIMULINK-MATLAB using the "compressed" network presented in Figure 8. The basic model parameters are the following: i_0=0.09934 A; K_m=129 N/m; $1/m_x$ =7.0922 kg^{-1} ; $x_{0em} + x_0 = 1 \cdot 10^{-3}$ m; K_{er} = 1620 N/m; K_h = 0.428 m; $A_0/\pi = 1.5625 \cdot 10^{-6}$ m^2; K_{fv}=200 N/m; K_{fs}=100 N/m; $A_1 = 3.14 \cdot 10^{-4}$ m^2; $A_2 = 1.57 \cdot 10^{4}$ m^2; K_b=2.6 N/bar; K_{fvp} = 4000 N/m; K_{fsp} = 2000 N/m; $1/m_y$ = 0.2717 kg^{-1}; c_d = 0,61; ρ= 900 kg/m^3; $K_{l1} = 1.924 \cdot 10^{-13}$ N/bar; $A_d = 0.196 \cdot 10^{-6}$ m^2; $\varepsilon/V_1 = 4 \cdot 10^{13}$ N/m^2; K_y = 1350 N/m. The figures 9, 10 and 11 present the variation of the main servomotor parameters for a linear input signal of 2 s and p_s=100 bar.

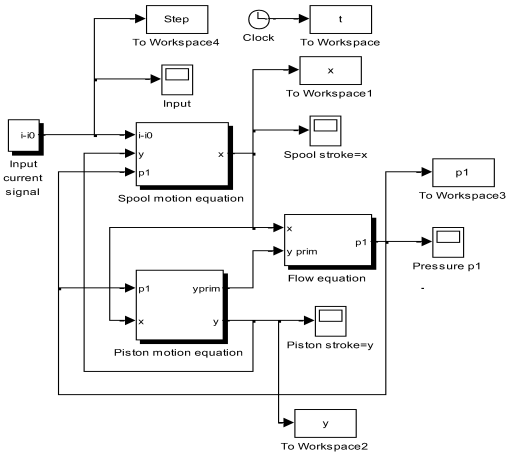


Figure 8: Simulation network of the servomotor dynamics for a linear input

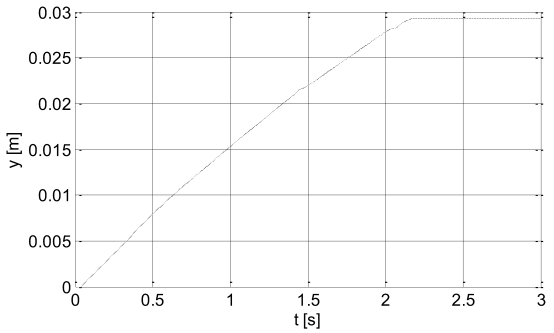


Figure 9: The servo mechanism response for a 75% ramp input signal

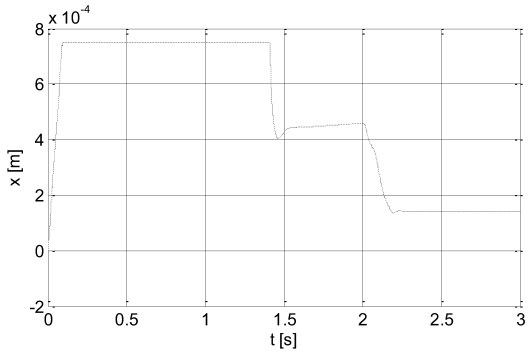


Figure 10: The spool travel for a 75% ramp input signal

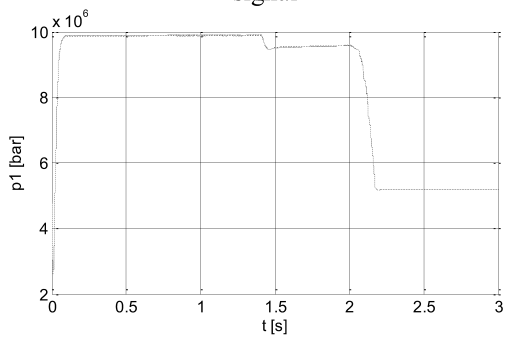


Figure 11: Pressure variation in the passive chamber of the stroking piston for a typical ramp input signal

Some significant parameters of the mathematical model were optimized step by step, in order to facilitate the wide use of this type of servomotor in practical applications.

NUMERICAL SIMULATION BY AMESIM

The above simulations performed by the aid of SIMULINK have shown a poor dumping factor of the displacement control system. A significant improvement can be achieved by introducing a small metering orifice between the three-way non-linear flow control valve and the pressure control chamber of the piston.

The study of the influence of the metering orifice size, d_a was performed by the mathematical model of the same servomotor built by the aid of general-purpose symbols from the libraries of AMESIM simulation environment (ANNEX B). Figure 12 presents the influence of the orifice diameter on the servomechanism response for a step input current signal applied to the control solenoid.

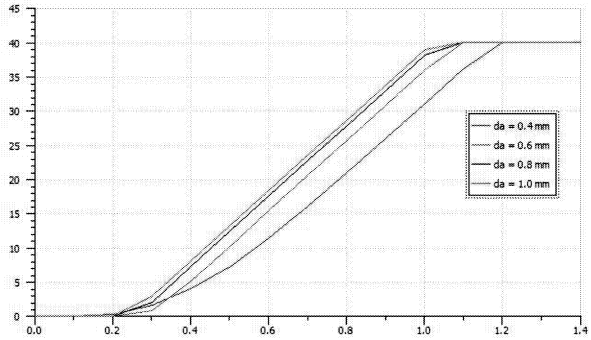


Figure 12: The servo mechanism response for a step input signal (p_s=300bar, d_a=0.4/0.6/0.8/1.0 mm)

The leakages between the supply port and the piston control chamber can initiate periodical small disturbances. After performing the travel for increasing the motor speed, (Figure 13), the pressure force on the control piston is slowly increasing and decreasing around the level of the steady state pressure, according the force feedback effect on the valve spool actuated by the solenoid.

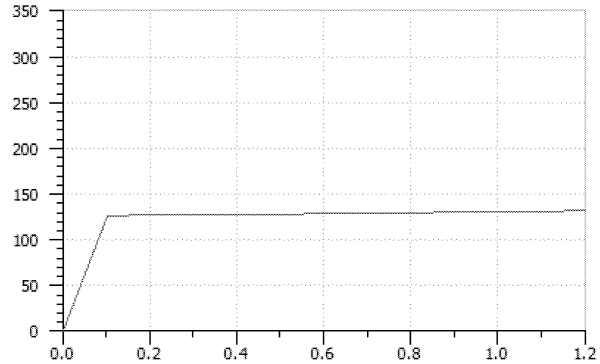


Figure 13: Typical variations of the pressure in the control chamber of the hydraulic cylinder (p_s=300 bar, d_a=0.4 mm)

At a first sight, the tilting force components have small values with small influence on the system behavior, but any change in the operating parameters generate changes in the motor displacement (Figure 14).

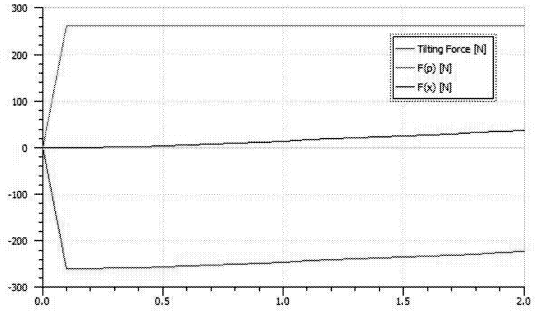


Figure 14: Tilting force components variations during a typical control process (p_s=100 bar, d_a = 0.4 mm)

The system response for different supply pressure (Figure 15) has a small delay, and is a near linear one. The attempt to use critical lap three-way control valve instead of round holes control valves leads to unstable servo system because of the very small dumping factor of the force feedback system with relative small stiffness feedback spring.

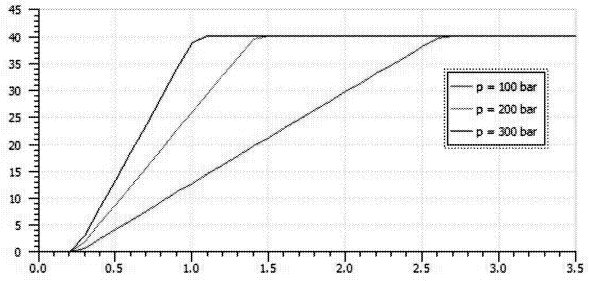


Figure 15: The influence of the supply pressure of the servomechanism on response time (p_s=100/200/300 bar)

Two attempts to eliminate the great amplitude oscillations generated by any step input were performed:
a) Introducing metering orifices on the input port of the control valve;
b) Introducing metering orifices on the input port of the passive chamber of the stroking piston.
Both of them are increasing the response time of the step input response of the servomechanism, but they avoid the oscillations.
The only feasible explanation of this situation regards the combination between the small fluid volumes involved in the control process, the small forces demanded by the displacement control, and especially - the high level of the pressure in the input port of the servomechanism.
The above remarks are sustaining the idea of avoiding a low stiffness force feedback, and the utility of an additional electrical position feedback.

CONCLUSIONS

The numerical results supplied by the two simulation environments are nearly identical. The simple manner of building practical simulation models and the strong integration algorithms used by AMESIM are explaining the wide spread of this new programming manner.
The big difference between the two languages comes from the time needed to build the simulation networks. Both languages allow the full control of the integration process, but some variable step algorithms developed for AMESIM are more accurate in the high-speed phenomenon description. However, the variable step integration is difficult to implement on a general purpose real-time simulation platform like dSPACE without the conversion in SIMULINK.
A new generation of AMESIM simulation models is now prepared for running on PXI and CompactRIO platforms (Vasiliu C. 2011) in different fields like speed governors for power units (Calinoiu C. et al. 2011), automotive engineering (Irimia P.C. 2012), mobile hydraulics (Popescu T.C. 2011) etc. Other analysis and synthesis software like AUTOMATION STUDIO are developed now in the same way (Vasiliu D. et al. 2012).
It seems also that the high-speed development of the modern industry is sustained by Hardware-In-The Loop Simulation experiments using different hardware platforms (Croitorescu V. and Anthonis J., 2012).

REFERENCES

Calinoiu, C., Negoiț ă, G. C., Vasiliu, D., Vasiliu, N. 2011. "Simulation as a Tool for Tuning Hydropower Speed Governors". ISC'2011, Venice, Italy.

Haas, H.J. 1989. "Sekundargeregelte Hydrostatische Antriebe im Drehzah und Drehwinkelregelkreis". Ph.D. Thesis, RWTH Aachen.

Lebrun M., Vasiliu D., Vasiliu N. 2009. "Numerical simulation of the Fluid Control Systems by AMESim". *Studies in Informatics and Control with Emphasis on Useful Applications of Advanced Technology*, Volume 18, Issue 2, p.111-118.

Muraru, V., Vasiliu, N., Vasiliu, D. 2000. "Structural Analysis of the Proportional Solenoids for Control Systems using ANSYS Software". World ANSYS Conference, Pittsburgh, U.S.A.

Negoita, G.C. 2011. "Researches on the Dynamics of the Hydrostatic Transmission". Ph.D. Thesis, University POLITEHNICA of Bucharest.

Popescu, T.C. et al. 2011. "Numerical Simulation - a Design Tool for Electro Hydraulic Servo Systems", in *"Numerical Simulations, Applications, Examples and Theory"*, INTECH PRESS, Zieglergasse 14, 1070 Vienna, Austria.

Vasiliu D., Vasiliu G.C. and Vasiliu N. 2012. "Using Automation Studio for the Design of the Fluid Control Systems". POLITEHNICA PRESS, ISBN 978-606-515-397-4.

Vasiliu N., Vasiliu G.C.,Vasiliu D. and Irimia P.C.2012."Modern Developments of the Hydraulic Transmissions in Automotive Technology". Romanian Academy of the Technical Sciences and AGIR PRESS HOUSE, ISSN 2066-6586.

Vasiliu C. 2011. Real-Time Simulation of the Electric Powertrains. Ph.D. Thesis, University POLITEHNICA of Bucharest.

***Imagine.2001. Numerical Challanges posed by Modeling Hydraulic Systems. Technical Bulletin 114.
***BOSCH-Automation Technology, 1999. *"Servo Solenoid Valves. Technical Specification 13/2"*. Stuttgart.
***LMS INTERNATIONAL, 2011. *"Advanced Modeling and Simulation Environment"*, Release 10 User Manual, Leuven.
***Parker Hannifin. 2008. Catalogue HY30-8223/UK.1.5 08/2008 PC. Hydraulic Motors Series V12, V14 and T12.

WEB REFERENCES

http://www.atos.com/
http://www.boschrexroth.com
http://www.frydenbo-marine.no/
http://www.eaton.com/
http://www.sauer-danfoss.com/
http://www.moog.com/
http://www.hydac.com/
http://www.parker.com
http://www.hydromatic.com
http://www.famictech.com/
http://www.automationstudio.com/
http://www.ni.com/
http://www.lmsintl.com/
http://www.mathworks.com/products/simulink/
www.acslX.com
http://www.nhancetech.com/
http://www.ni.com/matrixx/
http://www.dspace.com/

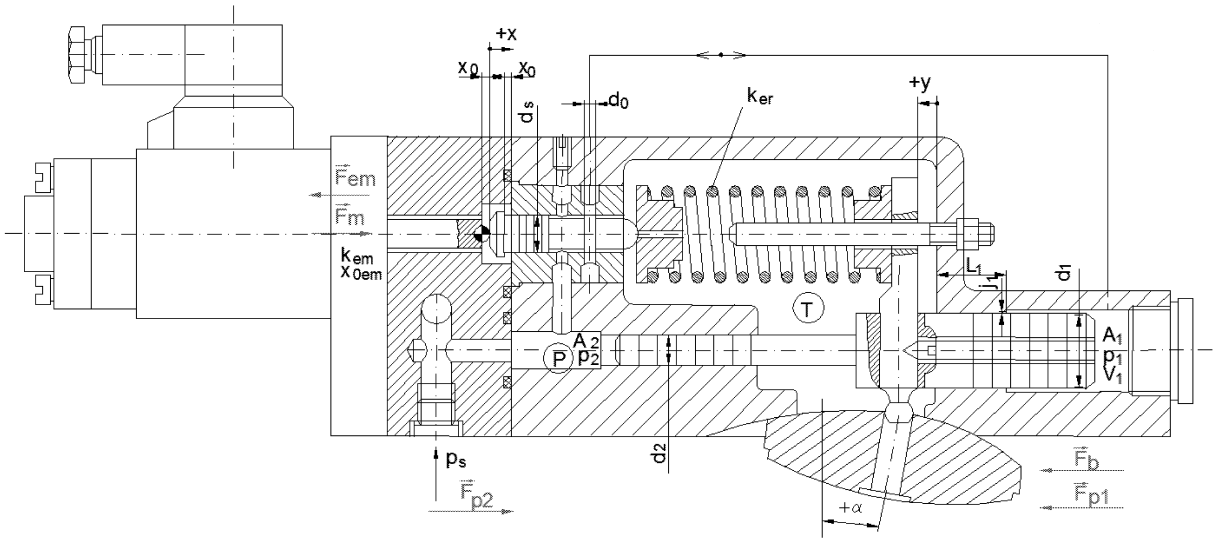
BIOGRAPHIES

DANIELA VASILIU graduated in mechanical engineering in 1981 and prepared the Ph.D. thesis in the field of the dynamics of the hydrostatic transmissions. She is currently professor in the Department of Hydraulics, Hydraulic Machines and Environmental Engineering, head of Fluid Power Laboratory of the University POLITEHNICA of Bucharest with works in the field of modeling, simulation, and experimental identification of the electro hydraulic control systems; member of EUROSIS, FPNI, SIA, FLUIDAS etc.

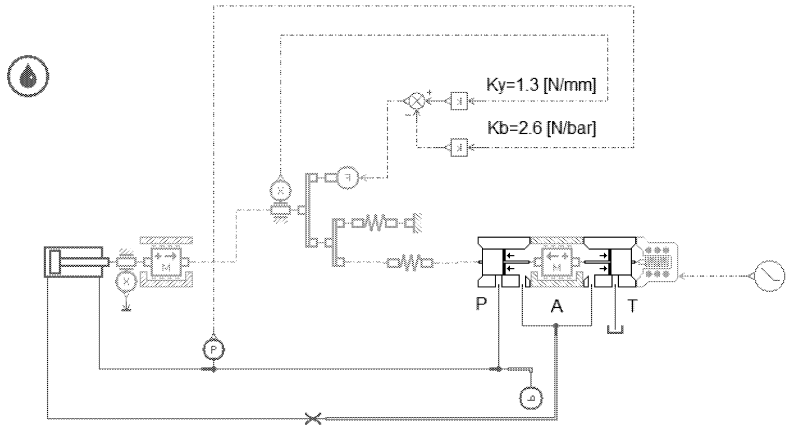
GEORGIANA CLAUDIA VASILIU graduated in mechanical engineering in Bucharest, in 2008. She prepared a Ph.D. thesis in the field of modeling, simulation and experimental identification of the hydrostatic transmissions in 2011. She is currently lecturer in the frame of the Department of Hydraulics, Hydraulic Machines and Environmental Engineering with works in the field of modeling and simulation of the electro hydraulic control systems.

NICOLAE VASILIU graduated in Hydropower Engineering from University POLITEHNICA of Bucharest in 1969. He became a Ph.D. in Fluid Mechanics after a research stages in Ghent State University and Von Karman Institute from Bruxelles. He became state professor in 1994, leading the ENERGY & ENVIRONMENT RESEARCH CENTRE from the University POLITEHNICA of Bucharest. He worked always for the industry, as project manager or scientific advisor.

PETRU-CRISTINEL IRIMIA received the M.Sc. degree in Mechanical Engineering, Automotive Section, from the University Transilvania of Brasov in 1991. He developed a long carrer in R&D, having a deep education in finite element analysis with high level LMS software. He is serving currently as general manager of LMS Test Division - CAE Division - Engineering Services from Romania. He is Ph.D. candidate within the University POLITENICA of Bucharest in the field of automotive hydraulic remote steering system energy management.



ANNEX A: Computing scheme for a bent axis electro hydraulic servomotor with three-way control valve



ANNEX B: AMESIM simulation network for a bent axis electro hydraulic servomotor with three-way control valve

THERMAL MODELING OF A BATTERY PACKAGE USED ON ROAD VEHICLE POWERTRAIN

Valerian CROITORESCU[1]
Jan ANTHONIS[2]
Nicolae VASILIU[3]

University POLITEHNICA of Bucharest
[1]Automotive Engineering Department, [3]Fluid Power Laboratory
313 Splaiul Independentei st., 6th Sector,
060042, Room JC004, Bucharest, ROMANIA
[2]LMS International
68 Interleuvenlaan,
3001 Leuven, BELGIUM

E-mail: valerian.croitorescu@gmail.com, jan.anthonis@lmsintl.com, vasiliu@fluid-power.pub.ro

KEYWORDS
Hybrid vehicle, battery, thermal modeling, simulation, losses

ABSTRACT

Recent years bring new European Union regulations concerning harmful emissions for road vehicles. The hybrid electric and electric mobility is a technology revolution for the automotive industry. In response to the stringent regulations and requirements enforced, vehicle manufacturers are developing new strategies, using hybrid electric or electric powertrain solutions. In order to obtain the goal of using alternative powertrain solutions, it is important to integrate from the early beginning different devices and to optimize their behavior. However, with a minimum of one electric energy storage package represented as the battery package, a hybrid electric powertrain triggers more concerns surrounding energy flows. During operation, there are energy losses caused by heat generation. Generated heat, emphasized as losses, is having a negative effect on the battery package efficiency.
Apart from the functional behavior of the battery package, which is the deliver and to store electrical energy, due to internal losses it also has a thermal behavior, which needs to be taken into account.
Thereby this paper aims to present the thermal behavior of a battery package in order to study its' energy flows and to determine its' efficiency as part of the powertrain.
Based on thermal resistances, a thermal network was developed to compute internal losses for different parts of the battery package. Using simulation, both the battery package and its thermal model, were integrated in a mock-up virtual hybrid electric vehicle to study its behavior during a predefined cycle.

INTRODUCTION

New European standards will require auto manufacturers to reduce CO_2 tailpipe emissions down to 130 g/km by 2015. For that it is need to predict and to minimize the harmful emissions from the model developing stage, by using a complete and detailed energy balance for a given powertrain configuration.
The automotive manufacturers are developing an array of other promising technologies. Hybrid vehicles are not the only answer to reduce harmful emissions and fuel consumption, but at this moment they are a strong competitor for other solutions aiming this purpose. Hybrid vehicles are defined as having at least two different energy storage devices for vehicle propulsion, as the fuel tank – liquid chemical storage and the battery pack – electrical storage. A hybrid electric vehicle (HEV) stands out having at least one electric motor/generator, a battery pack and other components that every ordinary vehicle has.
The hybrid electric vehicle development process, within respect to energy management, contains more than one level for designing the model. The primary level may be used for developing and optimizing the control strategies, based on maps and data tables.
Another level may include only partial physical aspects, trying to find an optimal balance between maps and physical data. Using all the previous levels, it is mandatory to add a more developed level as a need for the detailed physical subsystems.
Many differences can be noticed between a conventional and a hybrid vehicle, but the most important ones are the powertrain architecture, the control and the different modes of operation.
The battery package is the secondary mandatory energy storage component of a hybrid electric vehicle, besides the fuel tank. There are many types of batteries that are used on hybrid electric vehicles. Depending of the energy management system and power electronics the batteries are designed for plug-in hybrid electric vehicles and for hybrid electric vehicles. There are many types of battery package depending on the battery structure in terms of active materials used and battery package cells arrangement. For all of them, the concerns are regarding their power, internal resistance, storage capacity, number of charging-discharging cycles, functional behavior and lifetime. The battery package's efficiency may decrease while operating due to its thermal behavior. The battery package is generating heat that may reduce the lifetime and may lead to failure while operating.
The goal of this paper is to study the thermal effects of a battery package and to emphasize the effects of the high temperature on the battery package efficiency.
To investigate the thermal effects, the basic geometric design choices act as a major constraint. Based on geometrical considerations a thermal model was developed.
The thermal model computes the heat production and temperature distribution of the battery package.
The thermal model was integrated in the design of the hybrid electric vehicle.
The battery package used as part of this powertrain architecture consists in Lithium Ion (Li-Ion) cells.

The result of this defined modeling approach is that design parameters for the battery package can be linked to multiple vehicle performance parameters.
The paper will focus on power losses estimation with respect to the battery package functional behavior based on geometrical design considerations for its component parts.

SIMULATION OF A VIRTUAL VEHICLE MODEL

The HEV (figure 1) equipped with the Li-Ion battery package will be simulated using the AMESim software tool. The construction of the model for simulation, basically consists of linking sub-models from various software`s libraries. The sub-models are shown as icons. After linking the icons (which represent each component of the vehicle) in sketch mode and choosing the proper sub-model in sub-model mode, the parameters will be set. Each icon covers a fragment of C code, written using the specific equations for the system.

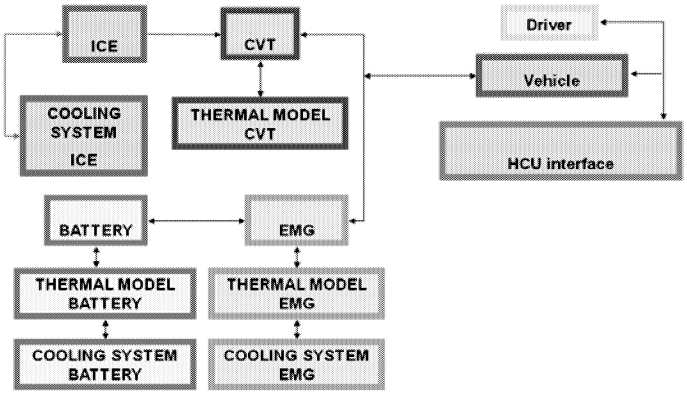


Figure 1: Model of the HEV equipped with a Li-Ion battery package

Using a pre-defined mission profile for setting the vehicle control speed with regard to the selected driving cycle, it is possible to input the ambient data (wind speed, density and ambient temperature) as real parameters. Acceleration and brake control take into account the anticipation of the speed control.
The driver used in this model controls/calculates braking, acceleration and gear shifting using the vehicle speed. The driver is not able to calculate the gearbox ratio or clutch control.
The Hybrid Control Unit (HCU) performs regulations for idle speed and maximum speed. Also the HCU controls the batteries (the state of charge), analyzing them in order to minimize the energy consumption. The HCU contains two different functional modules/levels, based on power demand and the hybrid mode selection. The HCU estimates the load demand based on current load due to road conditions, mass, speed, inertia. The hybrid modes are selected from predefined operating modes, based on power demand, driver's request and vehicle state.
The model for the electric motor can be used as an electric motor and a generator, being independent from the technology of the machine and converter. It is using the torque reference, the battery voltage and the rotary velocity for being able to compute the efficiency tables. It is controlled by the HCU to deliver the motor torque and the losses.
The battery is an internal resistance model, which characterizes the battery with a voltage source and an internal resistance. The battery output voltage is calculated taking into account the equivalent internal resistance and the input current.
The transmission includes a continuous variable transmission to get the engine torque and to deliver it to the wheels.
The vehicle models a simple vehicle load when considering no longitudinal slip between the tire and the ground.
The hybrid powertrain architecture used in the simulation is a parallel configuration equipped with continuous variable transmission.
The electric motor/generator is connected to its thermal model, exchanging information about temperature and power losses.
The parameters used in the simulation are the benefit of existing vehicle and components. The most important are shown in Table 1.

Table 1: The most important parameters used in simulation

Sub-model	Parameter	Value / Unit
Electric motor	Type	SRM
	Maximum power	30kW
	Maximum torque	200 Nm
Battery	Voltage	~300 V
	Rated capacity	6.35 kWh
	Cells in series per battery bank	96
	Battery banks in parallel/series	1/1
Transmission	Type	CVT
	Efficiency	~ 0.85 - based on efficiency tables

FUNCTIONAL BEHAVIOR OF THE BATTERY PACKAGE

One of the hybrid electric powertrain represents the battery package efficiency. The battery package is characterized using several parameters with regard to the charging capabilities, in order to store large amount of recoverable energy during short periods of time, the battery capacity, their lifetime, taking into account the number of cycles for charging and discharging.
In addition, the initial parameters often provided by the manufacturer are the electrical capacity and the external voltage. The capacity varies with the discharge current. Therefore, for different decreasing values of current consumption and low battery voltage, the capacity decreases (figure 2).

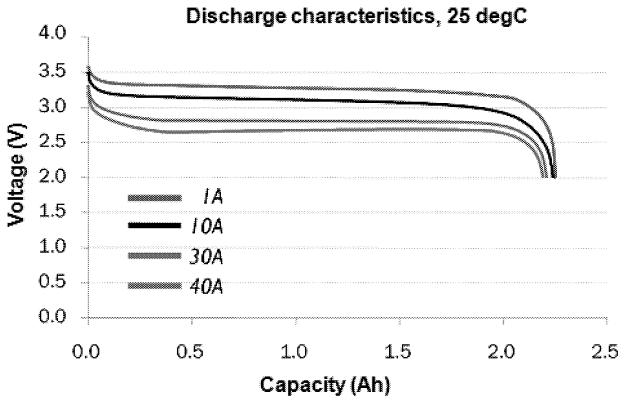


Figure 2: Discharge characteristics for different current consumption, at 25 degC

The battery package is characterized as a voltage source and internal resistance. The battery simulation model determines the losses using the temperature. The voltage and the resistance are also temperature subject of dependency.
A basic battery model can be represented by an equivalent simplified circuit. The battery is serial connected with a resistor and the circuit is considered an ideal open circuit. Thus, the output voltage is determined using the internal resistance and the current.
Each cell is considered an electrical circuit. The cell circuit contains the current and the internal resistance. The cell is having its own voltage and it is also considered an open circuit.
The battery output voltage is calculated as follows $V=V_0-I*R$, where V [V] is the output voltage, V_0 [V] is the open circuit voltage, R [Ohm] is the equivalent internal resistance, I [A] is the input current.

GEOMETRICAL MESHING AND THERMAL MODEL OF THE BATTERY PACKAGE

The geometrical parameters are very important in modeling the power losses and the thermal effects, also as input parameters for the AMESim blocks used to develop the thermal model.
It is important to mesh the battery package. Its design consists of banks in serial and parallel arrangements; each battery bank consists of cells (figure 3).

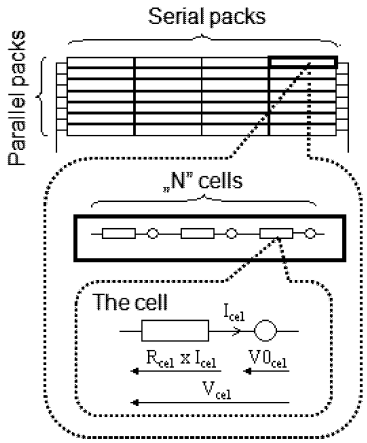


Figure 3: Battery package architecture

The battery package was divided in simple parts: the battery cells, the insulation between them and the internal air. For each of the components a thermal mass was associated.
The thermal masses were defined based on the made off materials and the shapes volumes. For the masses common determination it is need to use simple geometrical shapes. Therefore, for each of the components simple geometrical shapes were associated in order to define the heat transfer.
Taking into account geometrical considerations (partially presented in figures 2 and 3), the battery package was defined as further is presented.
The cells were considered cylinders. The insulation is wrapping the cells (figure 4), being present into entire areas between the cells.
The insulation was considered as triangular prisms, each of it corresponding to three neighboring cells (figure 5). The internal air is covering the surfaces under and above the cells. It is limited by the interior side of the battery package frame. The internal air was considered as two parallelepiped shapes, one under and the other above the cells and the insulation between the cells. The frame was considered as a hollow parallelepiped.

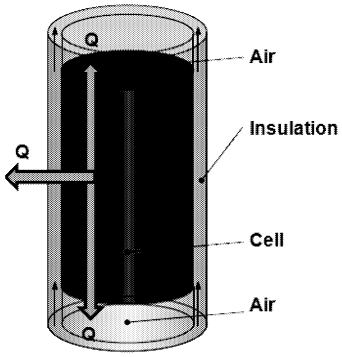


Figure 4: Battery cell 3D sectional view

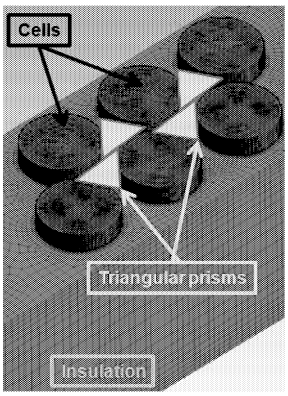


Figure 5: Insulation geometric meshing

The heat is transferred to the ambient and inside the battery package. Inside the battery package the heat exchange is possible between all its components (figure 6). The internal air has a very strong influence on cooling, being required to be taken into account.



Figure 6: Heat exchange between battery package parts

Starting from the known losses that can be met inside of a battery package and associating the thermal behavior of the most important components of a battery package (figure 7), a thermal simplified network was built (figure 8).

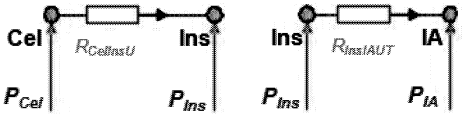


Figure 7: Heat sources and losses distribution

The thermal masses were computed based on geometrical considerations. Based on conduction and convection the detailed thermal network was used to model the thermal effects between the components.
The cells, the insulation and the internal air were considered thermal masses, separated as presented above (figure 6). Due to heat transfer, thermal resistances appear between all the thermal

masses. Each resistance corresponds to conduction and convection heat transfer.

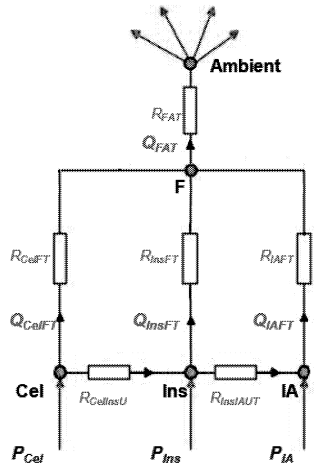


Figure 8: Li-Ion battery thermal network (simplified)

The thermal model is using all the heat flows – ambient and cooling system – in order to compute the temperature, based on power losses.

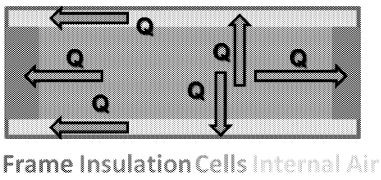


Figure 9: Battery package heat flows

The thermal model developed in AMESim contains three different parts: the losses distribution, the temperature distribution in solid parts and the temperature distribution in internal air.

Losses distribution

The heat is generated by the losses. Each part of the battery package is able to release heat to the nearest part.
Power losses for the battery package are generated while charging and discharging. They are presented as reaction power losses, polarization power losses and resistive power losses (figure 10).

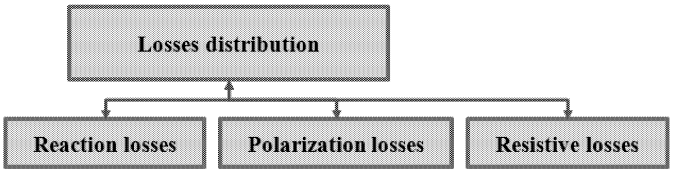


Figure 10: Losses distribution

The reaction power losses are the result of heat exchange while the batteries are charged or discharged due to the resulting reaction heat. The reaction heat is emitted when the lithium ions from the positive electrode interleaves the negative electrode.
The polarization power losses are interpreted as the heat energy loss from the emergence of polarization heat. Polarization heat occurs due to the inside battery atoms needs of diffusion and displacement, during reaction processes, while charging and discharging.

Resistive losses, known also as Joule losses, are very important for computing the total losses of the battery package. The resistive heat generates the resistive losses. The resistive heat occurs due to the interaction between electrons with atoms of the conductor.

Temperature distribution in solid parts

Being an energy transfer mode due to a temperature gradient, the conduction is taking place between all solid parts of the battery package. Convection is a major mode of heat transfer and major mode in mass transfer in fluids.
Each solid part has its own thermal properties due to the material they are made of. Besides material density, the specific heat and the thermal conductivity have a temperature dependency that cannot be neglected. Taking into account the various shapes assigned to battery package parts and the how heat exchange is done, it is necessary to model the conduction and convection for all of them (figure 11).

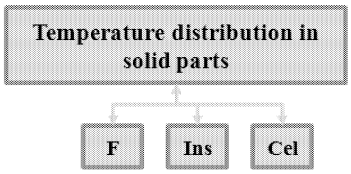


Figure 11: Temperature distribution in solid parts

As it was specified above, the cells were approximated by cylindrical shapes. Conductive heat transfer generates different ways to determine the resistance, due to the shape of the investigated part.

Temperature distribution in internal air

The internal air temperature distribution (figure 12) is possible between the interior limits of the frame, the cells and the insulation. The internal air consists in the two sides between the cells-insulation and the frame. The temperature in the internal air (IA) varies, having the insulation and cells temperatures strong influence. The gas properties are changing with temperature modification. The gas properties (GP) and turbulent air flow, if provided by some dedicated fans, are considered to be very important, being able to introduce a high difficulty level in thermal modeling.

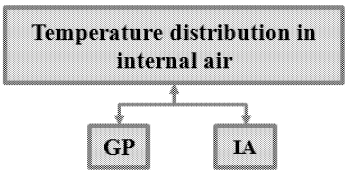


Figure 12: Temperature distributions in internal air

Therefore, the density of the internal air, the constant pressure specific heat, the perfect gas constant, the absolute viscosity and the thermal conductivity are very important to develop the thermal model corresponding to the internal air. After deep investigations, the natural and forced convection were modeled for this step, knowing the influence of Reynolds and Nusselt numbers.

Similar assumptions as for the conductive heat transfer regarding the determination method for the resistances were taken into account for the convection heat transfer.
Additionally, there were taken the following assumptions: the heat flow between insulation, cells and interior frame sides, above and under the cells, was considered turbulent, the air flow velocity was the same with dedicated fan rotary velocity. The above and under the cells internal air shapes were modeled as parallelepiped shapes.

SIMULATION RESULTS

The procedure used in this paper to estimate power losses with respect to battery package thermal behavior have been validated. In addition, the total efficiency can be estimated based on this interpretation.
The simulation was done during a defined cycle. Starting from an ambient temperature of 20 degC and taking the assumption that all battery package component parts have the same temperature, the following results were determined.
The used battery package is able to being charged or discharge using the electrical machine. Therefore, its thermal and energetically behavior during simulation cycle is changing during operation (figure 13).

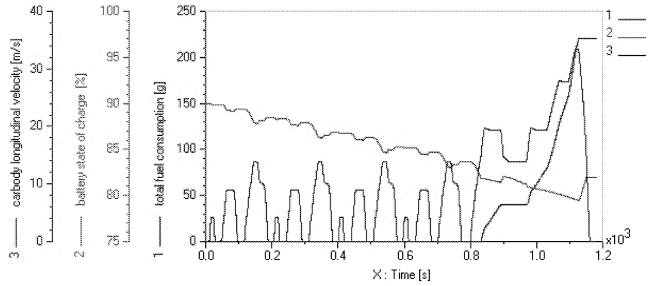


Figure 13: Vehicle velocity, fuel consumption, battery state of charge during the defined cycle

The quantity for the energy taken from or delivered to the battery package respects the HCU strategies.
The battery package operating behavior influences the temperature evolution. The end of the cycle brings the highest temperature. While braking with the electric machine, the batteries are charging and the temperature is increasing (figure 14) due to the high voltage charging demand. Immediately after the machine is providing motor torque, the batteries are delivering electrical energy, the temperature continuously increase and the power losses maintain its upward trend with temperature.

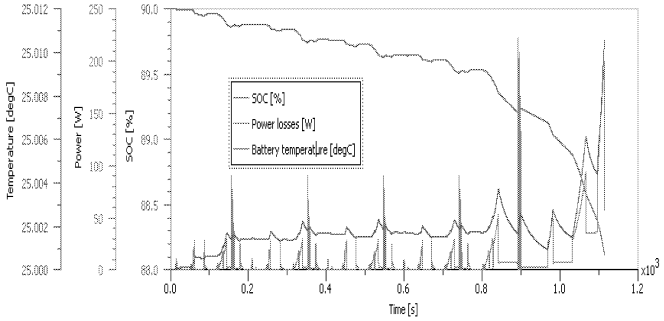


Figure 14: Battery package state of charge, power losses and temperature evolution

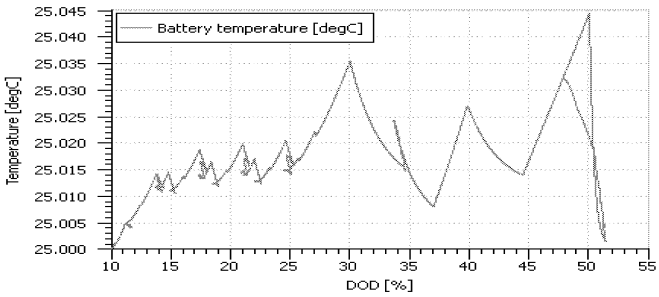


Figure 15: Battery package temperature during discharge

The cells are the most heating up components during operation. Their thermal state/condition influences the neighboring bodies/parts to heat up, influencing the global battery package temperature. As long as the internal air is having a turbulent flow, the heat transfer is possible.

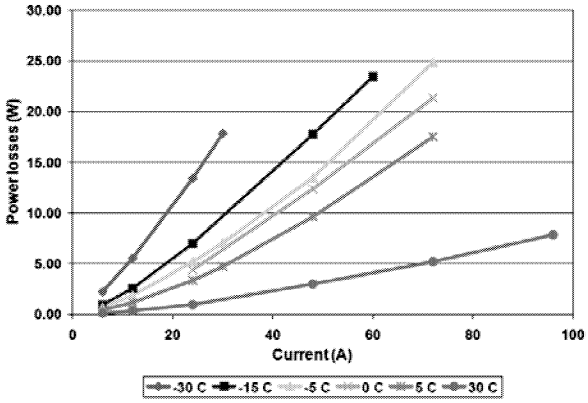


Figure 16: Battery package power losses while current increasing

Temperature distribution in internal air has an important influence in cooling the components by allowing the heat transfer between them.
During charge and discharge processes, the temperature evolution is influenced by chemical reactions inside the batteries. The ambient temperature is providing natural cooling. Adding an additional independent cooling system the batteries temperature will be characterized by rapid variations.
While the current consumption increase the power losses are proportional increasing, regardless the ambient temperature the power losses are constantly growing.
The corresponding functional model allows determining the efficiency of the battery package. Changing the ambient and functional temperature it is possible to investigate the influence of power losses regarding the thermal behavior.
The simulation results may be validated using HIL simulation test bench.

CONCLUSIONS

A detailed thermal model was created by means of integrated approach that uses geometry design for the battery package. The thermal model analysis of the battery package was based on the thermal equivalent circuit, built in respect with heat transfer fundamentals, by conduction and by convection.
An analysis of temperature rise of the battery package due to the power losses using AMESim was introduced.

The geometrical model of the battery package has been developed being able to evaluate the geometrical parameters changes influence. The thermal properties of the battery package have been evaluated in order to determine how the material proprieties affect the energy distribution. The metal solid parts of the battery package were assumed as thermal lumped masses.
Defining the battery package components as thermal masses is making possible to analyze the thermal transient effect in order to improve the thermal behavior. In addition, it allows choosing the appropriate design for further optimization processes.
In order to estimate the power losses, the battery package was investigated in respect with its geometry, material made of and functional behavior. For each part of the battery package simple shapes were assigned. Using the lumped thermal model it was highlighted the important role that the internal air is having in entirely battery package cooling.
The heating and cooling processes for the battery package during operation represent high interest for the energy losses.
Since the battery package is providing the energy for acting as hybrid or electric powertrain, it is mandatory to investigate the possibility to develop the cooling system. By connecting into a closed loop system the engine cooling system, the electrical machine cooling system, the air conditioning system and the battery package cooling system allows to reduce energy consumption. Additional cooling systems for the battery package have to be taken into account in order to monitor and control the energy management.
Hazard avoidance is strictly recommended. Therefore, the battery package temperature must be kept under 60 degC, being possible to reach this value only in exceptional cases.
Similar works on this method include the study of the cooling systems design closely related with the geometric dimensions for all the component parts and design of the electric motor/generator, battery package, transmission and engine.
In addition, based on this interpretation, the total efficiency can be estimated, allowing to determine and to avoid high energy losses and possible battery failure due to over-heating.
Finally, integrating thermal models for major hybrid powertrain components, the performance will be investigated.
Although the simulation described in this paper assumed a Li-Ion battery for HEV applications, similar assumptions and calculations are possible also for EV applications.

AKNOWLEDGEMENT

The present paper is mainly a part of first author`s PhD Thesis, being accomplished during his training period with full support of LMS International, Engineering Services.

REFERENCES

Anthonis, J., Santos, F., Croitorescu, V., Van der Auweraer, H., "Multiscale thermal and NVH models for EV/HEV Integration of an SR-based Drivetrain", Proceedings of VPC, Pune, India 2011

Buford, K., Williams, J., Simonini, M., "Determining Most Energy Efficient cooling Control Strategy of a Rechargeable Energy Storage System", SAE International, 2011

Croitorescu, V., Anthonis, J., Vasiliu, N., "Thermal Modeling of an Electric Motor Used On Road Vehicle Powertrain", Proceedings of 10th ISC, Brno, Czech Republic, 2012

Croitorescu, V., "Modern Drives Using Unconventional Energy Storage Devices – Hybrid Electric Vehicles", PhD Thesis, Bucharest, Romania, 2012

Gao, L., Liu, S., Dougal, R.A., "Dynamic Lithium-Ion Battery Model for System Simulation", IEEE Transactions on Components and Packing Technologies, vol.25, no.3, September 2002

Kennedy, B., Patterson, D., Camilleri, S., "Use of Lithium-Ion batteries in electric vehicles", Journal of Power Sources 90, 2000

Khateeb, S., Amiruddin, S., Farid, M., Selman, J.R., Hallaj-Al, S., "Thermal management of Li-Ion battery with phase change material for electric scooters: experimental validation", Journal of Power Sources 142, 2005

Pesaran, A., "Battery Thermal Models for Hybrid Vehicle Simulations", Journal of Power Sources 110, 2002

Pesaran, A., Keyser, M., "Thermal Characteristics of Selected EV and HEV Batteries", IEEE, 2001

Sakka Al, M., "Thermal Modeling of a Pouch Lithium Ion Polymer Battery", Flander's Drive Meeting LMS, Belgium, 2010

Sato, N., "Thermal behavior analysis of lithium-ion batteries for electric and hybrid vehicles", Journal of Power Sources 99, 2001

Smith, K., Wang, C-Y., "Power and thermal characterization of a lithium-ion battery pack for hybrid electric vehicles", Journal of Power Sources 160, 2006

BIOGRAPHIES

VALERIAN CROITORESCU earned his Automotive Engineer Degree in 2007, as valedictorian. He prepared the diploma project in France, inside Ecole Nationale D`Ingenieurs De Tarbes. He also attended two master of sciences programs, at University POLITEHNICA of Bucharest: `Efficiency and Security in Automotive Engineering` and `Environmental Management`. His academic records include numerous awards and certifications. In September 2007 Valerian joined the academic stuff of Automotive Engineering Department. In 2008, he founded "autojobs.ro", a specialized web portal in automotive industry having the goal to help and provide job seekers and employers the opportunity to be connected, followed by "aerojobs.ro" in 2012, aviation dedicated jobs web portal. In 2012, he earned a Ph.D. Degree at University POLITEHNICA of Bucharest, concerning Hybrid Electric Vehicles Development. A part of his Ph.D. thesis has been accomplished during a research stages at LMS International.

JAN ANTHONIS received the M.Sc. degree in mechanical engineering from the KU Leuven in 1994, and the PhD degree in 2000. From 2001 till 2007 he was post doctoral researcher at the Fund of Scientific Research in Belgium, developing researches on control and mechatronics with application in agro-machinery and agro-industry. From 2003 he is part time lecturer at the KU Leuven. In 2007 he became employee of LMS International, a spin-off company of the KU Leuven founded in 1980, specialized in noise, vibrations and mechatronics in automotive, aerospace and other mechanical industries. Here he manages a research team on control and mechatronics.

NICOLAE VASILIU is a hydropower engineer who earned the Ph.D. degree of University POLITEHNICA of Bucharest after a research stage in Ghent State University and Von KARMAN Fluid Mechanics Institute from Bruxelles. He is Professor of Fluid Control Systems, and head of Fluid Power Laboratory at University POLITEHNICA of Bucharest. He is scientific advisor for many industrial companies. Between 2006 and 2010 he managed the ROMANIAN INNOVATION FINANCING AGENCY. He is working mainly in modeling, simulation and dynamic identification of the hydraulic and electro-hydraulic control systems with LMS International from Belgium and National Instruments from U.S.A.

TURBINE SIMULATION

NUMERICAL SIMULATION AND EXPERIMENTAL VALIDATION OF THE TRANSIENTS FROM HIGH HEAD HYDROPOWER PLANTS

Radu Puhalschi, Marius Mitroi, Constantin Călinoiu and Nicolae Vasiliu
University POLITEHNICA of Bucharest
Power Faculty
Fluid Power Laboratory
313 Splaiul Independentei, 060042
Bucharest, Romania
E-mail: vasiliu@fluid-power.pub.ro

KEYWORDS
Modeling, simulation, validation, hydropower plant, by-pass valves, electro hydraulic speed and pressure control system

ABSTRACT
The paper presents the results of the theoretical researches carried out for identifying the hydraulic transients occurring in the speed and pressure electro hydraulic control systems of the high head Francis turbines set up in Raul-Mare Retezat Hydropower Plant. The authors performed the mathematical modeling and the numerical simulation of the entire control system by AMESIM and LabView languages. The experimental researches carried up in real turbine operating conditions using the power plant SCADA system have shown high intensity overpresures. The correlation between the simulated results, and the experimental ones allowed the authors to identify the real sources of the dangerous hydraulic transients. The final practical target of the paper is the proper design of a new electrohydraulic by-pass valve for obtaining a reasonable value of the waterhammer overpressure.

INTRODUCTION
Fluid transients often occur in the pipes of high head hydropower plants when the flow passing through the turbine is changed. This phenomenon is known in the literature as *waterhammer*. Pressure variation in the pipe lead to tension variation in it's walls and they depend on the variation gradient of the flow passing through the turbine. Consequently, a turbine shutdown time as long as possible is needed. At the same time, due to turbine protection needs, the shutdown of the turbine's water intake system needs to be as short as possible in case of unload. These two contradictory needs are achieved in case of Francis turbines by placing a bypass valve before the turbine, which is controlled by the turbine's governor. The actuator of this valve is always hydraulic, but the command system can be of a different kind (mechanical, hydraulic, electrical etc.). In the devoted literature (Arne Kjolle 2001; Kundur, P. 1994) several by-pass valve actuating systems are known, built by hydromchanical equipment companies with a performance level coresponding to the time period. Recent developments in the field of numerical turbine governors makes achieveing superior performance for the by-pass valve actuating system possible. Selecting and optimizing the actuating system of by-pass valves while mantaining full operating safety is a constant concern for specialists in the field. Numerical control systems, together with modern proportional hydraulic valves allow a much higher performance level for the turbine governors to be achieved, which answers the ever increasing demand for quality of electric power. A detailed study, by numerical simulation, for determining the static and dynamic performances of these systems greatly reduces both the time and the material and human resources needed for designing such systems. This paper presents the numerical symulation results for two different control systems for the by-pass valve; a hydraulic and an electro-hydraulic system. The simulations were performed for several values of wicket gate opening, to better illustrate the advantages and disadvantages of each system. The simulation evironment used was AMESIM environment.

THE CLASSICAL HYDRAULIC CONTROL SYSTEM
Figures 1 and 2 present a high head Francis turbine (Vasiliu 1976, UCMR 1976) equipped with a by-pass valve (Figure 2). The transients from the control sistem can be studied with the AMESIM network from Figure 3 (Lebrun 1997).

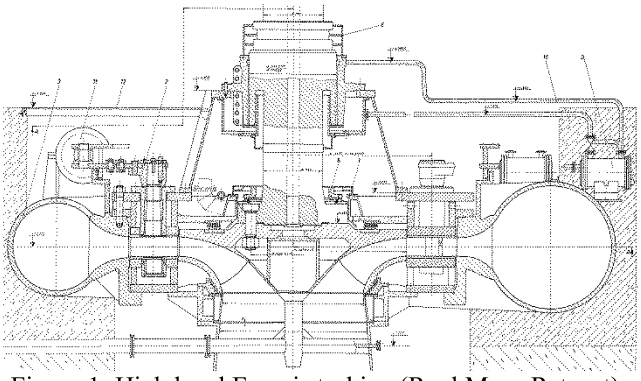


Figure 1: High head Francis turbine (Raul Mare-Retezat)

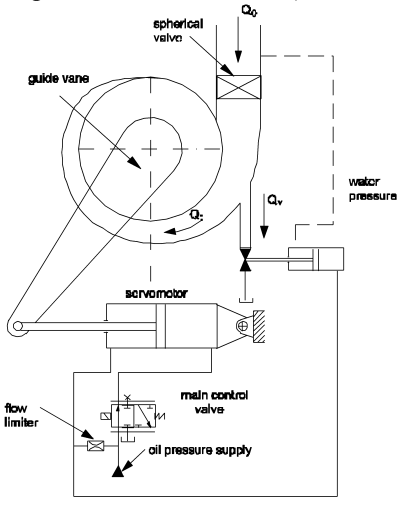


Figure 2: Hydraulic diagram of the speed governor

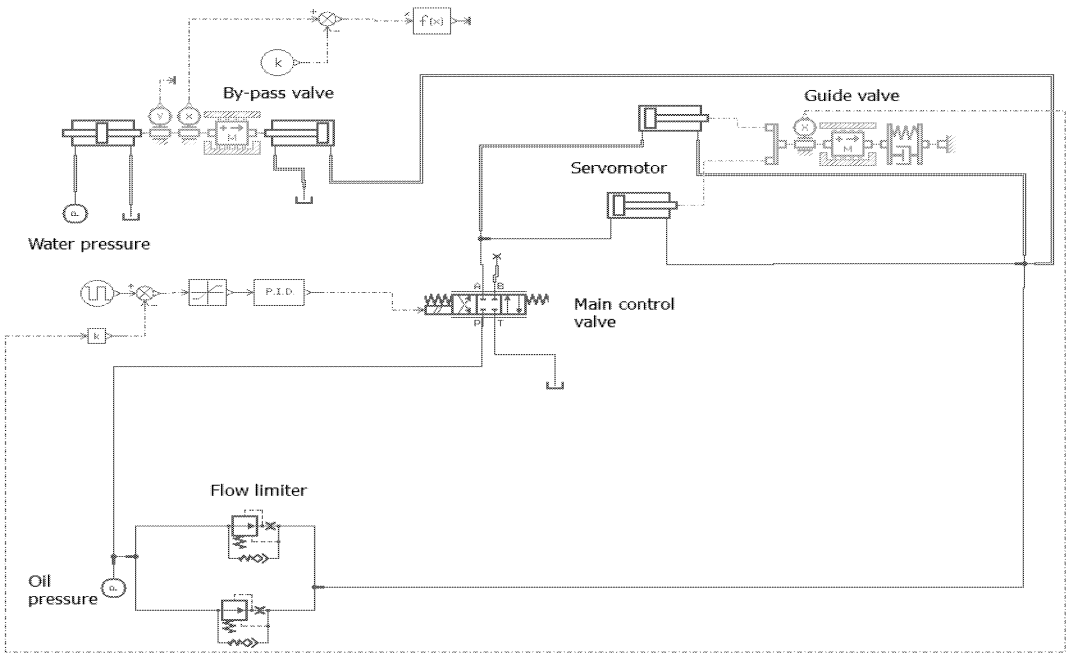


Figure 3: Numerical simulation model of the turbine governor with a by-pass valve with hydraulic control

The opening of the by-pass valve is done under the pressure forces created by the oil in the command system and the water in the pipe. The oil chamber of the command cylinder for the by-pass valve is hydraulically connected with the small chamber of the wicket gate servomotor. Therefore, on a command for the wicket gate to close, a pressure drop is created in the by-pass valve control circuit due to the flow limiter. The by-pass valve opens and a part of the water flow from the pipe is sent downstream of the turbine. After the closing of the wicket gate, the by-pass valve closes as well, with controlled speed. The variation speed of the flow passing through the pipe is reduced and consequently the pressure value that appears as the result of waterhammer is also limited [Tullis 1989, Fitzgerald 1998]. The continuity equation for the water flow is the following:

$$Q_W = Q_t + Q_v \quad (1)$$

where Q_W- dewater flow through the pipe; Q_t – water flow through the turbine; Q_v – water flow through the by-pass valve. Figure 4 presents a typical case of water flow in case of unload for a high head Frances turbine.

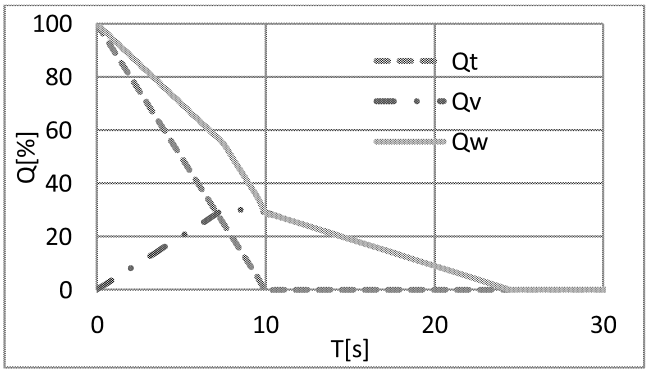


Figure 4: Water flow variation through the pipe

The numerical simulation of the turbine speed and pressure governor was performed by the realistic model presented above. When the turbine closes, under unload, an oil exchange between the by-pass valve and the wicket gate servomotor occurs. If the by-pass valve is fully opened this oil exchange stops, and the oil flow to the servomotor is therefore reduced. This phenomenon was illustrated by numerical simulation and is presented in figure 5.

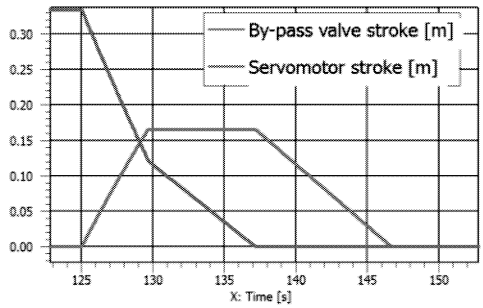


Figure 5: Servomotor piston displacement and by-pass valve opening for a wicket gate opening of 100%

By increasing the turbine shutdown time, the time during which the turbine is overloaded increases as well, which might prove a major inconvenient for certain applications. The adjustment of the parameters of the flow limiter is done manually which does not allow adapting them to the operationg parameters of the turbine from the time just before the unload. Figure 6 shows the by-pass valve opening and servomotor piston displacement for a wicket gate opening of 20%.

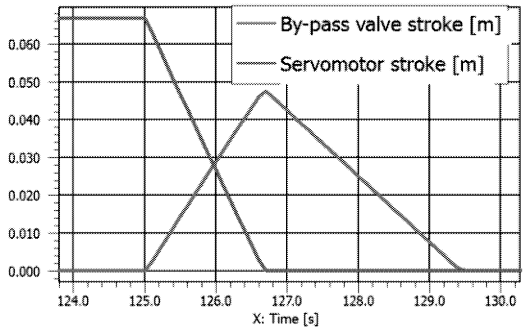


Figure 6: Servomotor piston displacement and by-pass valve opening for a wicket gate opening of 20%

The corresponding pressure evolution is presented in the Figure 7.

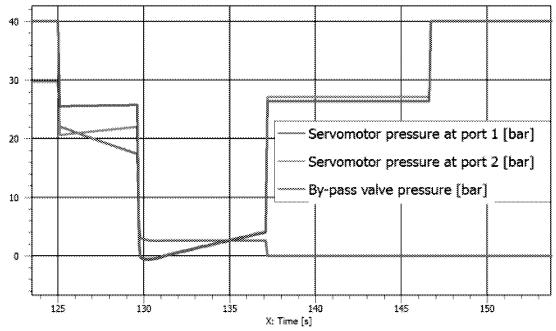


Figure 7: Pressure variation on the servomotor and by-pass valve chamber for 100% wicket gate opening

THE ELECTRO-HYDRAULIC CONTROL SYSTEM

In order to eliminate the disadvantages of the hydraulic control system presented above, the authors have developed a new system in which the hydraulic circuit that operates the by-pass valve is independent from the hydraulic circuit that operates the wicket gate servomotor. The numerical simulation model for the new control system is presented in figure 8 (Calinoiu 2011; Vasiliu 2000, 2003, 2005, 2007).

With this configuration of the control system of the by-pass valve the opening and closing speeds can be easily adjusted and adapted to the real functioning conditions of the turbine from the moment just before the overload.

Optimizing the opening and closing times of the by-pass valve can be done by numerical simulation, and the validation of the results is easy to achieve on site. Figures 9 and 10 show the variations of servomotor piston displacement and by-pass valve opening for wicket gate openings of 100% and 20%.

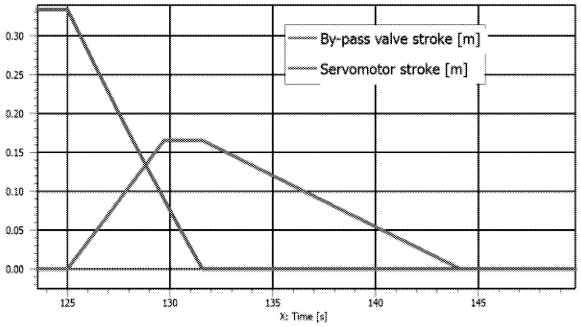


Figure 9: Servomotor piston displacement and by-pass valve opening for a wicket gate opening of 100%

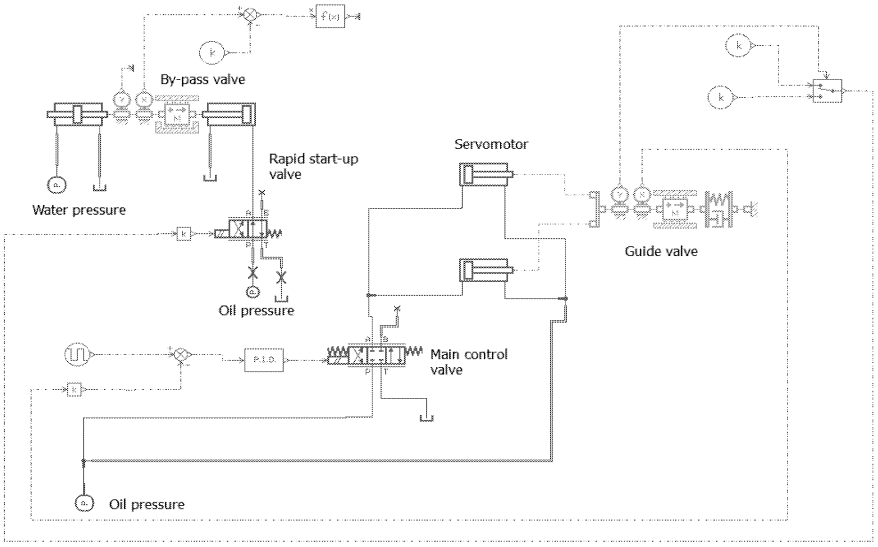


Figure 8: The numerical simulation model for the turbine governor with an electrohydraulic operated by-pass valve

The structure of this control system includes an electro-hydraulic proportional valve for the quick opening of the bypass valve. The oil flow passing through the electro-hydraulic valve is calculated with the clasical orifice relation

$$Q = K \cdot U \cdot \sqrt{\Delta p} \qquad (2)$$

where Q – oil flow; K- coefficient specific to the valve; U- command voltage; Δp- pressure drop on the valve. From equation (2) it can be seen that the flow through the valve and consequently the opening and closing speed of the by-pass valve can be adjusted by changing the value of the command voltage U or the value of the pressure drop on the valve Δp. Changing the command voltage is done with the electronic command block included in the turbine governor, and changing the pressure drop on the valve can be done with flow limiters inserted in the oil circuit.

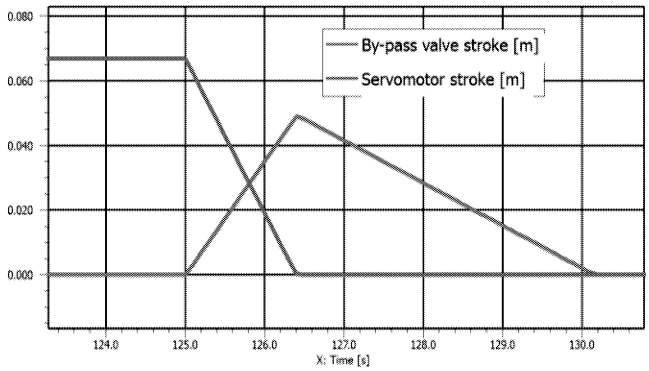


Fig. 10. Servomotor piston displacement and by-pass valve opening for a wicket gate opening of 20%

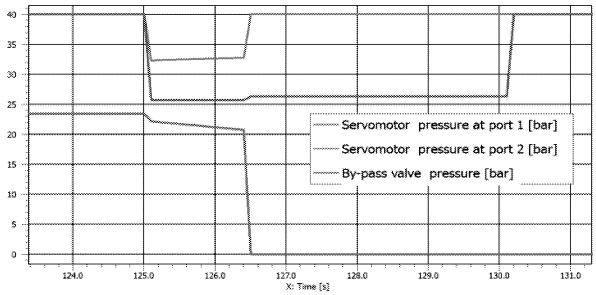


Figure 11: Pressure variation on the servomotor and by-pass valve chamber for a wicket gate opening of 20%

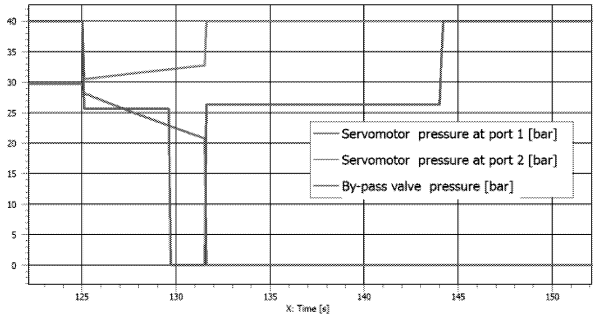


Figure 12: Pressure variation on the servomotor and by-pass valve chamber for a wicket gate opening of 100%

Figures 11 and 12 are presenting the pressure varations for the ports of the servomotor and the by-pass chamber. Tables 1 and 2 contain a summary of the servomotor closing times, the by-pass valve opening and closing times for various openings of the wicket gate in case of unload, according to the international standards promoted by IEC and ASME.

Table 1. Characteristic time values for the servomotor and by-pass valve for a classical hydraulic control system

Wicket gate		By-pass valve		
Opening [%]	Closing time [s]	Opening [%]	Opening time [s]	Closing time [s]
20	1.7	30	1.7	2.7
40	3.2	59	3.2	5.7
60	4.8	90	4.8	8.4
80	8.0	100	4.8	9.5
100	12	100	4.8	10

Table 2 . Characteristic time values for the servomotor and by-pass valve for a new electro hydraulic control system

Wicket gate		By-pass valve		
Opening [%]	Closing time [s]	Opening [%]	Opening time [s]	Closing time [s]
20	1.4	30	1.4	4.4
40	2.8	59	2.8	7.2
60	4.0	84	4.0	11.0
80	5.5	100	4.8	13.5
100	6.5	100	4.8	12.5

BY-PASS INFLUENCE ON WATERHAMMER

The phenomenon known as waterhammer represents the sudden pressure rise in a pipeline (in the form of a pressure wave) caused by a sudden change in the rate of flow in a pipeline. In the case of high head Frances turbins the pressure values that occur in the pipelines depend on the flow gradient in the pipe. If the closing time of the turbine is lower than the time needed for the pressure wave to reflect, then the pressure values are very high. This phenomenon may occur when the wicket gate opening is small, because the closing time is then very small, and the by-pass valve opens very little due to specific time constants of the equipment. The theoretical study of waterhammer is based on the two equations that fully describe it : the unsteady momentum equation applied to a section of the pipe in equation (2) and the continuity equation (3). The most common form in which these equations are found in the dedicated literature (E.B. Wiley and V.L. Streeeter 1983; Ch. Jaeger 1987) is the following :

$$gH_x + VV_x + V_t + \frac{fV|V|}{2D} = 0 \qquad (3)$$

and

$$VH_x + H_t - Vsin\alpha + \frac{a^2}{g}V_x = 0 \qquad (4)$$

where

- V- fluid velocity in the pipe *[m/s]* ;
- H- water pressure in the pipe *[m]*;
- a- celerity (speed of the pressure wave) *[m/s]* ;
- t- time variable*[s]* ;
- x- space variable *[m]* ;
- g- gravitational pull *[m/s^2]* ;
- f - Darcy-Weisbach friction factor (describes the head loss on the pipe).
- α – angle of the pipe.

As partial differential equations, no analytical solution is available for them, and therefore several methods and algorythms for numerical solutions and approximations have been developed.

The most widespread method of solving the transient flow equations is the method knows as the characteristics method. It is currently the most precise and recent developments in computational software have helped mitigate it's main drawback (sheer volume of calculations that make it, unlike many other methods, completely impractical for usage 'by hand').

The characteristics method aims to transform the general unsolvable partial differential equation into particular total differential equations which can be integrated to yeld finite difference equations which can be conveniently handled numerically.

Usually, simpler forms of equations (2) and (3), which omit the very small terms are used.

$$gH_x + V_t + \frac{F}{2D}V|V| = 0 \qquad (5)$$

and

$$H_t + \frac{a^2}{g}V_x = 0 \qquad (6)$$

We can write a linear combination of these equations using an unknown coefficient, λ, as follows :

$$\lambda\left[H_x\frac{g}{\lambda} + H_t\right] + \left[V_x\lambda\frac{a^2}{g} + V_t\right] + \frac{fV|V|}{2D} = 0 \quad (7)$$

Any two real values of λ result in a pair of equations in H and V which are completely equivalent with the system made up of equations (4) and (5) and it can be easily seen that certain values of λ can bring great simplifications to

equation (6). In general, both H and V depend on x and t. If the independent variable x is allowed to be a function of t, we get from calculus :

$$\frac{dH}{dt} = H_x \frac{dx}{dt} + H_t \qquad \frac{dV}{dt} = V_x \frac{dx}{dt} + V_t \qquad (8)$$

By examination of equations (6) and (7) we get

$$\frac{dx}{dt} = \frac{g}{\lambda} = \frac{\lambda a^2}{g} \qquad (9)$$

Equation (6) becomes the ordinary differential equation

$$\lambda \frac{dH}{dt} + \frac{dV}{dt} + \frac{fV|V|}{2D} = 0 \qquad (10)$$

The solution of equation (8) yelds the two particular values of λ :

$$\lambda = \pm \frac{g}{a} \qquad (11)$$

By substituting the values of λ in equation (8) , the particular manner in which x and t are related is given:

$$\frac{dx}{dt} = \pm a \qquad (12)$$

In a physical sense, this means that the change in position of a wave is related t the change in time by the wave propagation speed, *a*. Substituting the two values of λ in euqation (7), together with equation (11) yelds two pairs of equations, commonly known in literature as the C^+ and C^- equations:

C^+:

$$\frac{g}{a}\frac{dH}{dt} + \frac{dV}{dt} + \frac{fV|V|}{2D} = 0 \qquad (13)$$

$$\frac{dx}{dt} = +a \qquad (14)$$

C^- :

$$-\frac{g}{a}\frac{dH}{dt} + \frac{dV}{dt} + \frac{fV|V|}{2D} = 0 \qquad (15)$$

$$\frac{dx}{dt} = -a \qquad (16)$$

As can be seen, the two values of λ have been used to convert the original two partial differential equations into two total differential equations, with the restriction that each of them is valid only when the correspoding equation (13) and (15) respectively) are valid. If a visualsation of the solution is attempted in the independent variable plane(xt), each of the equations (13) and (15) plots as a straight line (Figure 13).

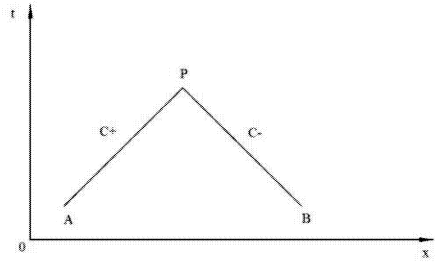


Fig. 13 Characteristics lines in the x-t plane

These are the *characteristics* lines along which equations (14) and (16) are valid. We split the pipe into N segments, each Δx in legth, like in Figure 14. The time step is calculated as $\Delta t = \frac{\Delta x}{a}$. Equation (13) holds along the positive slope AP, and equation (15) along the negative slope PB.

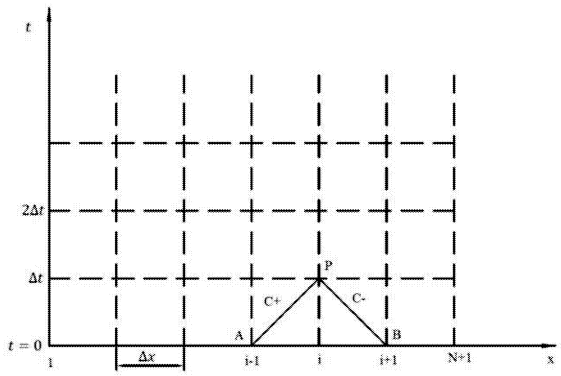


Figure 14: The characteristics grid used for simulation

If the values of V and H in point A are known, then equation (12) can be integrated between A and P, and as such written as a function of V and H in point P. Similarly equation (14) can be integrated between B and P. Therefore, in P we have a system of two equations in V and H, which can be used to determine both values.

If we multiply equation (12) by $a\frac{dt}{g} = \frac{dx}{g}$ and insert the pipe area so that we can use the flow Q instead of the velocity V, the equation becomes

$$\int_{H_A}^{H_P} dH + \frac{a}{gA}\int_{Q_A}^{Q_P} dQ + \frac{f}{2gDA^2}\int_{x_A}^{x_P} Q|Q|dx = 0 \qquad (16)$$

The variation of Q depending on x is unknown, therefore an approximation is used. First order approximation is sufficient for most transient flow problems. After integrating euqation (14) in the same way, the following system is obtained:

$$H_P - H_A + \frac{a}{gA}(Q_P - Q_A) + \frac{f\Delta x}{2gDA^2} Q_A|Q_A| = 0 \qquad (17)$$

$$H_P - H_B - \frac{a}{gA}(Q_P - Q_B) - \frac{f\Delta x}{2gDA^2} Q_B|Q_B| = 0 \qquad (18)$$

Solving for H_P, the system becomes

$$H_P = H_A - B(Q_P - Q_A) - RQ_A|Q_A| \qquad (19)$$

$$H_P = H_B + B(Q_P - Q_B) - RQ_B|Q_B| \qquad (20)$$

Here

$B = \frac{a}{gA}$ and $R = \frac{f\Delta x}{2gDA^2}$.

In general, the solution of any transient flow problem starts with assuming steady flow at the initial time, fact which allows the determination of H and Q in each pipe section at t=0. Next, based on the initial values, H and Q are calculated at the next time step (t=Δt) and so on. In the interior points of the grid, the two compatiblity equations are solved simultaneously for the variables Q_{P_i} and H_{P_i}. At the ends of the grid however, only one of the characteristic equations is available, and therefore additional boundary conditions are needed. They depend on what kind of hydraulic device is placed at each pipe end (valve, turbine, tank, another pipe etc.) In the case studied in this paper, the intake pipe of a Francis turbine has a water tank at the upstream end and the wicket gate at the downstream end. We can easily consider that the upstream head remains constant during short duration transients, and as such the upstream boundary condition can be simply expressed as $H_{P_i} = H_R$. This allows the determination of Q_{P_i} from equation (20). At the downstream end, the equation of steady flow through the wicket gate is described by the following equation:

$$Q_0 = (C_d A_G)_0 \sqrt{2gH_0} \quad (21)$$

where Q_0 – steady flow; H_0 – head loss on the wicket gate; $(C_d A_G)_0$- product of wicket gate opening surface and flow coefficient. For a different wicket gate opening, the equation becomes

$$Q_P = C_d A_G \sqrt{2g\Delta H} \quad (22)$$

where ΔH is the instantaneous head loss on the wicket gate. We can define the quantity

$$\tau = \frac{C_d A_G}{(C_d A_G)_0} \quad (23)$$

Dividing equations (22) and (21), the following is obtained:

$$Q_p = \frac{Q_0}{\sqrt{H_0}} \tau \sqrt{\Delta H} \quad (24)$$

For steady flow $\tau = 1$ and for completely closed wicket gate $\tau = 0$. Solving equations (19) and (24) both Q_p and H_p can be obtained. The practical implementation of the caracteristical method has been performed in LabVIEW with the aid of MathScript module. This has allowed the authors to combine the robustness of LabView's computational engine and it's ease of working with graphical interfaces and user-defineable parameters with the power of the MATLAB code for dealing which large amounts of equations. The block diagram of the program, consisting of the user defined inputs and the MathScript node which also generates the graphical outputs can be seen in Figure 15.

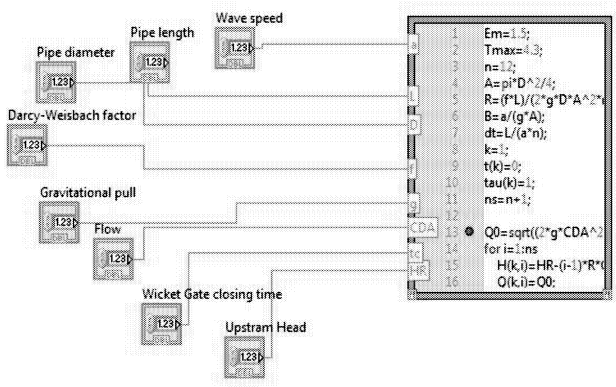


Fig. 15 LabVIEW block diagram

The interface (VI-Virtual Instrumentation as it is known in LabVIEW) has been designed by the authors to allow the user to input all the necessary parameters for a given transient flow calculation and allows the graphical display of both head (H) and flow (Q) as a function of time, as can be seen in fig. 16.

For running all the simulations, the following real set of values were used (Nicolet 2007): wave speed: 1010 m/s; pipe length: 800 m; pipe diameter: 3.6 m; Darcy-Weisbach friction factor: 0.018; upstream head: 522 m.

In the first stage, a simulation has been performed only for the pipe, disregarding the effects of the bypass valve, in order to establish a base of comparsion against which the efficiency of the by-pass valve could be measured. The results are presented in figures 17, 18, and 19.

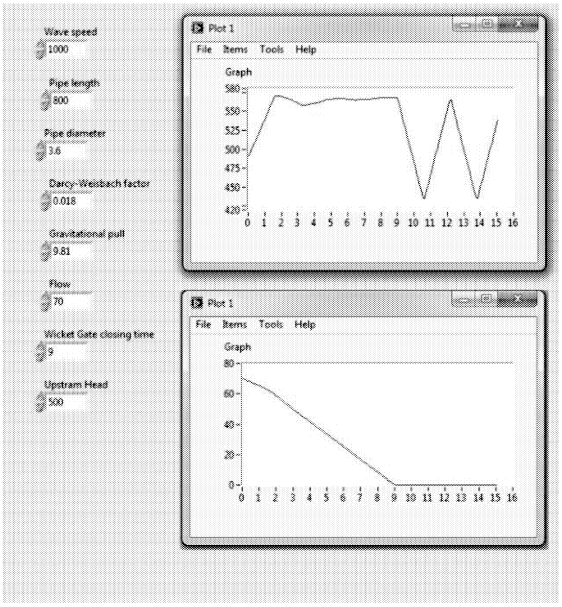


Figure 16: Screen capture of LabVIEW

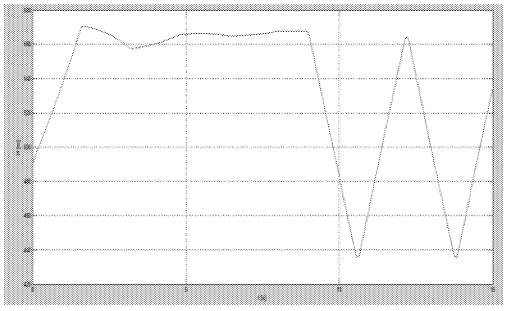


Figure 17: Pressure variation for unload without by-pass valve for 100% wicket gate opening and Ti=9s

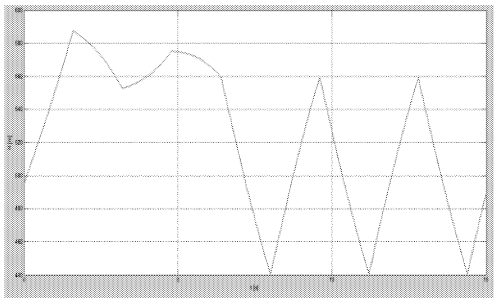


Figure 18: Pressure variation for unload without by-pass valve for 60% wicket gate opening and T_i=6.4s

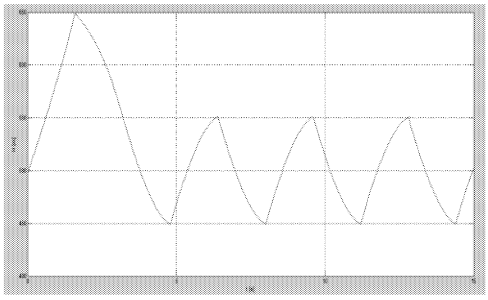


Fig.19. Pressure variation in the pipe for unload without by-pass valve for 20% wicket gate opening and T_i=2.1s

A second set of simulations also takes into account the effects of the bypass valve on the pressure variation in the pipe (Figures 20, 21 and 22).

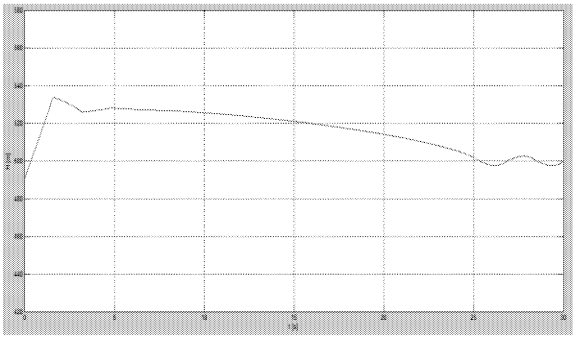


Fig.20. Pressure variation in the pipe for unload with by-pass valve for 100%, wicket gate opening and Ti=9s

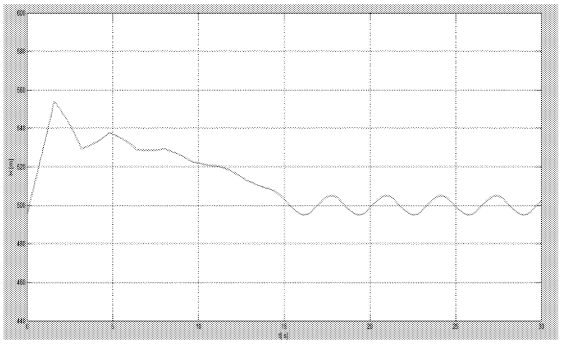


Fig.21. Pressure variation in the pipe for unload with by-pass valve for 60% wicket gate opening and T_i=6.4s

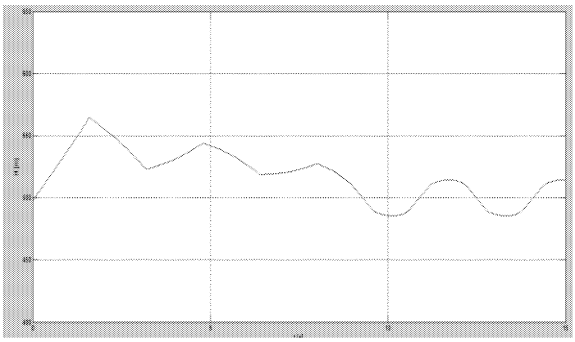


Fig.22. Pressure variation in the pipe for unload with by-pass valve for 20% wicket gate opening and T_i=2.1s

As it can be seen from comparing the two set of diagrams, the reduction in maximum head is quite significant: a) from 570 to 530 m in case of 100% wicket gate opening; b) from 590 to 550 m in case of 60% wicket gate opening; c) from 650 to 570 m in case of 20% wicket gate opening).

CONCLUSIONS

The numerical simulation of both types of speed and pressure control clearly shows the advantages of a flexible electrohydraulic solution. The digital treatement of all the signals simplifies the hydraulic diagram, and the hydromechanical connections. By the aid of the modern simulation environments (AMESIM, Matlab and LabVIEW), the authors created a complex simulation tool which allows a good prediction of the speed and pressure performances of the modern hydropower units. At the same time, the design of the classical hydromechanical by-pass valve from Figure 23 (UCMR 1976, Shevelev 1981) can be reduced to the design of a common conical seat valve, which is much simpler and cheaper.

The new design tool of Real-Time Simulation (Ion Guta et al. 2008) will be used for validating the resizing of the by-pass flow of the Raul Mare Retezat and for other hydropower stations. The modern multi stage proportional electro hydraulic servo valves with on-board electronics produced by MOOG, PARKER, BOSCH-REXROTH et al. can be used to design simple hollow-jet by-pass valves.

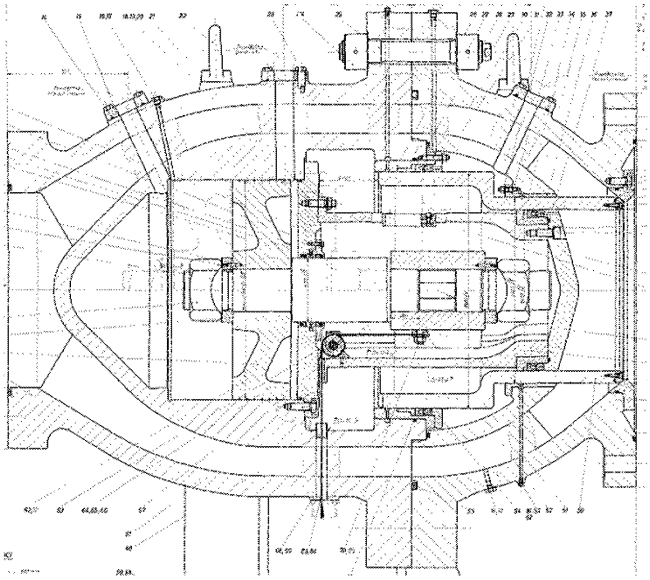


Figure 23: Classical hydromechanical by-pass valve with both water and oil piston drives

ACKNOWLEDGEMENTS

The authors are greateful for the help offered by LMS International BV from Belgium, and National Instruments Corporation from USA for their continous support through academic programs. Trying to solve complex industrial problems by the aid of high level numerical simulation environments like AMESIM and LabVIEW, the authors received always facilities and fruitful advices from the above mentioned companies.

REFERENCES

Călinoiu, C., Negoiț ă, G. C., Vasiliu, D., Vasiliu, N. 2011. *"Simulation as a Tool for tuning Hydropower Speed Governors"*. ISC'2011, Venice, Italy.

Ion Guță D., Vasiliu N., Vasiliu, D., Călinoiu C. *"Basic concepts of Real-Time Simulations (RTS)"*. U.P.B. Scientific Bulletin, Series D, Vol.70, No.4, p.291-300, ISSN 1454-2358, 2008.

Kjølle, A. 2001. Hydropower in Norway. Mechanical Equipment. Norwegian University of Science and Technology.

Kundur, P. 1994. *"Power System Stability and Control"*, McGraw-Hill.

Lebrun, M., Claude, R. 1997. "*How to create Good Models without Writing a Single Line of Code"*. Fifth Scandinavian International Conference on Fluid Power, Linköping.

Mare, J.-Ch., 1994. "*Dynamics of the Electrohydraulic Rotary Servomechanisms"*. Doctor Thesis, I.N.S.A. Toulouse.

Tullis, J.P. 1989. *"Hydraulic of pipelines: pumps, valves, cavitation, transients"*. Utah State University, USA.
Jaeger, Ch. 1987. *"Fluid Transients in Hydro-Electric Engineering Practise"*, University of Glasgow.
Shevelev D.S. and others.1981. *"Hydrolectric Power Plants" (in Russian)*, Leningrad State University, 1981.
Fitzgerald, R., Van Blaricum,V.L.1998. *"Water Hammer and Mass Oscillation"*, 3.0 User's Manual, US Army Corps of Engineers, Research Laboratories.
Nicolet, C. 2007. "*Hydroacoustic modelling and numerical simulation of unsteady operation of hydroelectric systems*". Thesis No 3751 a La Faculte Sciences et Techniques de L'ingenieur, Laboratoire de machines hydrauliques, Lausanne.
Vasiliu, N. 1976. "*Reseaches on the Hydrodinamics of the Francis Runners*". PhD Thesis, University POLITEHNICA of Bucharest.
Vasiliu, D., Calinoiu, C. and Vasiliu, N. 2000. *"Governing Hydraulic Turbines by Servo Solenoid Valves – Numerical Simulation and Test"*. 2000 ESS CONFERENCE, The Society for Computer Simulation, Hamburg.
Vasiliu, N., Calinoiu, C. 2003. *"Digital Electro Hydraulic Speed Governor for KAPLAN Turbines"*. Romanian Patent No. 120101/2003.
Vasiliu, N., Vasiliu, D. 2005. *"Fluid Power Systems"*, Vol.I., Technical Press House, Bucharest.
Vasiliu, N., Calinoiu, C., Vasiliu, D. 2007. *"Modeling, Simulation and Identification of the Electrohydraulic Speed Governors for Kaplan Turbines by AMESim"*. PTMC 2007, Bath.
Wylie, E.B., L. Streeter, L.V. 1983. "*Fluid Transients*", University of Michigan.
***ASME. 2002. *Hydraulic Turbines and Pump-Turbines"*, Performances Test Codes, ASME PTC 18.
***BOSCH-Automation Technology, 1999. *"Servo Solenoid Valves. Technical Specification 13/2"*. Stuttgart.
***IEC 61362 - *"Guide to specification of hydraulic turbine control systems"*.1998. International Electrotechnical Commission, Geneve.
***IEC 60308. 1970. *"International code for testing of speed governing systems for hydraulic turbines"*.
***LMS INTERNATIONAL, 2011."*Advanced Modeling and Simulation Environment"*, Release 10 User Manual, Leuven.
***MOOG. 1971. "*Specification Standards for Servo Valves*". Internal Standard 117, East Aurora.
***MOOG. 1971. "*Performance Estimation for Electro Hydra-ulic Control Systems*", Internal Standard 126, East Aurora.
***NATIONAL INSTRUMENTS. 2012. LabVIEW System Design Software Release 12, Austin, TX.
***Parker Hannifin. 2004. *"Servovalves series SE"*. Hydraulic Valve Division, Catalog HY14-1460, Elyria, Ohio, USA.
***UCMR. 1976. *High Head Francis Turbine with By-Pass Valve. Design Report.*

BIOGRAPHIES

Radu PUHALSCHI graduated in Applied Computer Science from University POLITEHNICA of Bucharest in 2009, and a MD in Advanced Hydraulic and Pneumatic Systems in 2011. After a stage of web designer at HP Germany, he began a PhD research work on Real-Time Simulation of hydraulic systems at the Fluid Power Laboratory from the Power Engineering Department of the University POLITEHNICA of Bucharest.

Marius MITROI, PhDc, graduated Power Engineering Faculty from University Politehnica of Bucharest in 2007. He works as Manager within Romanian R&D Financing Agency from Bucharest, Romania; he is working also in the field of modeling and real time simulation of the fluid control systems as a member of Fluid Power Laboratory of the University POLITEHNICA of Bucharest.

Constantin CALINOIU graduated in Hydropower Engineering from University POLITEHNICA of Bucharest in 1975. In 1980 he became a member of the Hydraulic Lab from the Romanian Aerospace Institute. In 1998 he became PhD and associated professor in the Fluid Power Laboratory from U.P.B. He is working mainly in modeling, simulation, and identification of the hydraulic and electro hydraulic control systems.

Nicolae VASILIU graduated in Hydropower Engineering from University POLITEHNICA of Bucharest in 1969. He became a PhD in Fluid Mechanics after a research stage in Ghent State University and Von Karman Institute from Bruxelles. He became state professor in 1994, leading the Fluid Control Laboratory from the Power Engineering Faculty of the above university. He worked always for the industry, as project manager or scientific advisor. He was also leading five years the Romanian Innovation Financing Agency.

LIFETIME EVALUATION OF THE MECHANICAL ELEMENTS FOR HYDRAULIC TURBINES BY NUMERICAL SIMULATION

Constantin DRĂGOI, Vlad Florin PÎRĂIANU
University POLITEHNICA of Bucharest
313, Splaiul Independentei, Sector 6
RO 060042 Bucharest, Romania
E-mail: {constantin.dragoi|vlad.piraianu@upb.ro}

Petru-Cristinel IRIMIA
LMS Test Division - CAE Division - Engineering Services
Str. 15A, I. Slavici Street
RO 500398 Brasov, Romania
E-mail: cristi.irimia@lmsintl.com

KEYWORDS

Hydraulic turbines, simulation by FEA, vibrations, lifetime cycle

ABSTRACT

Nowadays the capabilities of machines are pushed to the limit in order to achieve higher efficiencies, better performances, reducing the cost of production and maintenance. Suppliers of mechanical equipment invest more and more in developing modern techniques in order to study beforehand the behavior of their machines and equipment while in operation in order to decrease the cost of physical test and model test. They are looking to determine all the different challenges that equipment and equipment's components will face during transitory regimes in operation. This is one of the benefits of life cycle evaluation in which fatigue study for new equipment in designing and manufacturing. Another aspect is to determine the remaining lifetime of the equipment which is already in operation, so that it will work safely until the plan of maintenance is properly determined and then carried out. In this way, safety standards can be better respected and the cost of operation and repair is optimized. This paper is a continuation of these studies and presents improvements of the results that authors achieved in 2012. A new, more accurate and complex way to determine the remaining lifetime of components for a Kaplan hydraulic turbine is presented. The analyzed element is the turbine shaft. The main parameters taken into consideration for durability calculation are: model geometry, material properties and applied load for simulation.

INTRODUCTION

A part of the existing equipment in the operational hydroelectric power plants in Romania have between 100.000 – 200.000 hours of operation. All these devices were built into the system of unique replacement rule and therefore the maintenance and service is significantly costing. The stress of the parts are various, starting from their own weight, hydraulic forces, electromagnetic forces, vibrations and so on, both in static and dynamic regimes. The majors results published in this paper were obtained by the authors working under the frame of the research project SOP 62 557 EXCEL.

INTEGRATED SYSTEM

All the research done so far and also the current steps which are undertaken currently, have as their final goal the establishment of a functional technological model framework based on data collected and evaluated to allow development of specific tools, instruments and procedures for each equipment or its part in question. The rapid growth of computer programs together with a more profound understanding and know-how gathered regarding fatigue, advanced simulation software and finite element methods has facilitated the development of parametric design software for the targeted equipment embedded with rapid development elements (Inventor, Catia, etc). Another direction of development was towards numerical simulation and finite element analysis software, with the aim of improving and enhancing the performance of algorithms and methods used to determine the remaining lifetime based on fatigue analysis for the rotary parts of the equipments. Currently, almost all CAD programs have embedded modules of finite element analysis models which may be generated. All this has led, at present, to improve sustainability assessment through numerical simulation and testing. Such a process covers the following steps:

a) Virtual prototype. Development of the three-dimensional model (3D) using a CAD program. The resulting model (at a scale of 1:1) is a carrier information model:

- geometric information: size, mass, volume, moments of inertia, etc.
- information about the nature of the material (physical, chemical and mechanical) to the developed model is attached as a material library provided by the program; if the material is new, there is the possibility of placing it in the database together with its properties.

b) Evaluation of loads: if the stress loads are known these are converted according to parameters of interest. If there is a possibility of their acquisition by a data acquisition and processing system of the right equipment while in operation we have an ideal case.

c) FEM. The virtual prototype is transferred to a program compatible with finite element analysis. Here the model is tested at loads (stress) being put on the model (data from data acquisition system or estimated). The advantage of using finite element analysis program is to use time variable loads (Blumenfeld 1995; Bathe 1996; Zienkiewicz 2000; Constantinescu 2011).

d) Meshing model and stress application is followed by numerical simulation load software according to laboratory tests. In this stage 90% of relevant information is obtained of the actual model behavior based on a virtual prototype decision-making, which may lead to the improvement of its performance.

e) The changes to the virtual prototype requires re-start and re-evaluation of its new features and therefore the pre-diagnosis of its lifetime cycle.

f) Once the virtual prototype meets its requirement, the development of physical prototype starts, as well as the operational stress estimation and testing in the laboratory, using a suitable testing program.

Based on what was presented above, the integrated system of design/diagnosis to the availability of one equipment is shown in Fig. 1. For equipments being in operation, such a system allows to: estimate the remaining life of hydro power plant's equipment in use; determine areas and equipment which is subject to heavy fatigue; determination of limit time in operation; identify the necessary corrections in the process of rehabilitation. This paper aims to present a methodology for the life cycle approach to calculate the turbine shaft that fits a hydraulic turbine type Kaplan.

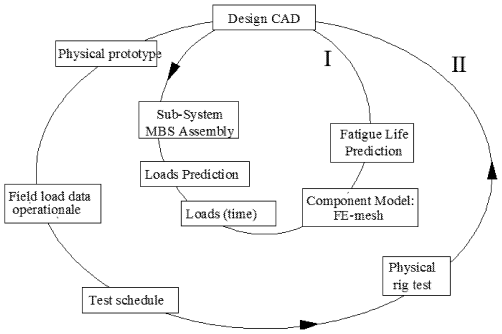


Figure 1 Integrated system

The main parameters taken into account for the calculation of sustainability are: model geometry and the application of load factor (concentrators, local tensions), material properties (mechanical properties of core and surface quality of the local tensions) and applied loads (the load, the magnitude and type of load). For analysis two methods was used: **a) tension analysis** – starting from the stress loads applied to the model, the nominal tensions are evaluated and then the information about the local elastic tensions are obtained. In the end, an estimation of the lifetime expressed in relation with elastic stress from the important area; **b) analysis of the strain** – the sustainability assessment based on distortion shall be considered as the degradation as a function of local deformed plastic. Elastic-plastic behavior is described with a model of hysteresis loop. Degradation depends on the position, size and the order of hysteresis loops.

GEOMETRICAL MODEL

Geometrical model was taken from the technical documentation as a CAD file and then transferred as a 3D Model in Mechanical Desktop. The model was completed with mechanical properties and transferred as a file .stp in ANSYS. A parallel computation was performed with LMS Virtual.Lab (www.lmsintl.com). The model was simplified in order to achieve a 3D computational model. Therefore some modification in the representation method were done to the bearing geometry and clamping flanges. The simulation surface of finite elements was obtained by using SOLID187 tetrahedral spatial element. Areas that are connecting the two flanges were imposed with a refining mesh specialized for algorithms based on valuation criteria. For the radial bearing area an orderly mesh was used. In general, uniform rectangular networks were used to create and limit the calculation errors inherent in the program. The resulting simulation model consists of more than 94360 nodes and more than 58110 elements as seen in Fig. 2.

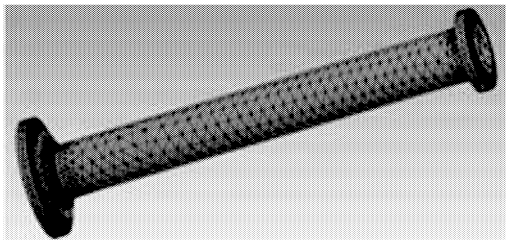


Figure 2 Meshed Model

MATERIAL PROPERTIES

In order to achieve realistic results that can be used in practice, a large number of material properties was used in the simulation. More parameters were taken into consideration for the finite elements analysis compared to classical calculation of fatigue. According to technical specification shaft material is OLC 25X (improved quality carbon steel). Its equivalent in databases AISI / SAE ASTM is Steel 1020, for which Su = 455 MPa. Its main properties are presented in Table 1. The fatigue curve for the material used was determined by empirical methods in specialized literature. A cyclic stress-strain curve dependence was used also according to existing literature.

Table 1: Material Properties

Name	Simbol	Value	UM
Elastic Module	E	206800	MPa
Ultimate Strength	S_u	455	MPa
Fatigue Strength Coefficient	s'_f	883	MPa
Fatigue Strength Exponent	b	-0,118	-
Fatigue Ductility Coefficient	ε'_f	0,16	-
Fatigue Ductility Exponent	c	-0,412	-
Cyclic Strength Coefficient	K'	1441	MPa
Cyclic Strain Hardening Exponent	n'	0,283	-

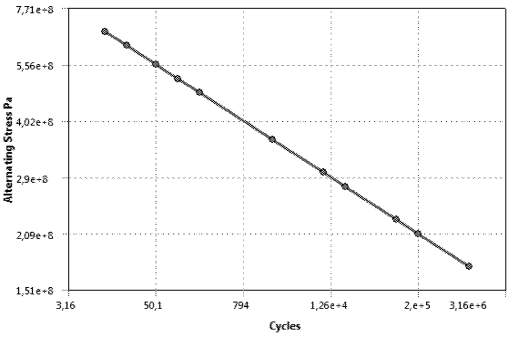


Figure 3 Fatigue curve for OLC 25X.

STRESS LOADS AND DATA ACQUISITION

The necessary parameter to define the stress analysis loads of the hydrogenerator shaft, is the torque. The torque is very difficult to measure, and therefore it has been calculated based on the delivered power from the central panel. The data was completed with information from the SCADA

system of the power plant: gross head, upper lake level, downstream level, level of trash rack debris and clogging, net head, turbine efficiency and vibrations in axial and radial bearings in order to achieve a more accurate value of the torque. The data provided was compiled, and a mathematical analysis of the torque was completed. (Fig. 4). The main parameters of the hydro power unit that were used: nominal speed (100 rot/min); mechanical efficiency of shaft (99.5%); generator efficiency (95%, considered constant to simplify the model); and overall efficiency (94.05%). Also, information from the SCADA system of the power plant was taken into consideration: as well as the upstream level in the lake, downstream level, gross head, percentage of the clogged trash rack at the intake level, used discharge and nominal discharge per each unit, turbine efficiency. Data from the SCADA system was used to obtain the gross head and to compute the net head for the turbine:

$$H_{net} = H_{gross} - \Delta h_r \quad (1)$$

where H_{gross} represents the difference between upstream level and downstream level of the water and Δh_r represent the total hydraulic headlosses: linear and local. To the hydraulic power computed with (2):

$$P_{hidr} = \rho \cdot g \cdot H_{net} \cdot Q_{used} \quad (2)$$

where Q_{used} was taken from the SCADA system and represents instant flow used in power generation measured in real time. The hydraulic power P_{hidr} was then corrected with the efficiency of the turbine η_{tb} according to manufacturer and the mechanical power P_{mec} was obtained with (3):

$$P_{mec} = P_{hidr} \cdot \eta_{tb} \quad (3)$$

The torque was then computed according to mechanical power of the turbine according to the above method and also with correlation obtained from the generator measured power P_{el} and estimated with (4):

$$M = \eta_t \times P_{el} / \omega = \eta_t \times P_{el} \times 60/2\pi n = \eta_t \times P_{el} \times 30/\pi n \quad (4)$$

Different sets of data were used when determining the remaining life time of the hydraulic turbine shaft. Of course, the values recorded of the electrical power delivered to the grid has a small variation around the reference level of 22 MW. This has to primarily with the fact that the generator is programmed to follow and to balance the grid frequency of 50 Hz. The variations around 22 MW can be also a result of grid behavior locally (strong variation of voltage and frequency) but can also be due to the actual state of the turbine and components. This refers, for example, to increased axial and radial vibrations in the axial and radial bearings, seals and other moving parts, due to age and active operating hours, for units that are scheduled to enter in full refurbishment and/or rehabilitation. A worse scenario is a unit that was in refurbishment and during installation and mounting procedures, the quality plans were not fully respected and some tolerances were accepted without consulting the manufacturer of the equipment. The data used in the simulations was as follows:

Case a) Average mechanical power at turbine shaft was determined at 22,7 MW and has a deviation of +1.637 MW and – 1.172 MW for a period of 10 minutes. The torque is $2.17 \cdot 10^6$ Nm with variations between + $0.15 \cdot 10^6$ Nm and $-0.12 \cdot 10^6$ Nm.

Case b) Average mechanical power at turbine shaft was determined at 23,75 MW and has a deviation of +0.8919 MW and –1.0718 MW for a period of 10 minutes. The torque is $2,27 \cdot 10^6$ Nm with variations between $+0.08 \cdot 10^6$ Nm and $-0,10 \cdot 10^6$ Nm.

Case c) Average mechanical power at turbine shaft was determined at 23.75 MW and has a deviation of +1.218 MW and – 1.07 MW for a period of 10 minutes. The torque is $2,27 \cdot 10^6$ Nm with variations between + $0.11 \cdot 10^6$ Nm and $-0.10 \cdot 10^6$ Nm.

The values used in simulation are shown in Figures 4 and 5.

T [s]	Q_{tur} [m³/s]	H_{net} [m]	η_{ex} [%]	P_{hidr} [MW]	P_{mec} [MW]	M [Nm]
10.1	156.233435	15.964138	0.9414	24.467435	23.033644	2.199551E+06
10.2	156.434043	15.964045	0.9414	24.498711	23.063086	2.202363E+06
10.3	158.901520	15.962902	0.9414	24.883354	23.425189	2.236941E+06
10.4	159.944681	15.962414	0.9414	25.045942	23.578250	2.251557E+06
10.5	159.262614	15.962733	0.9414	24.939636	23.478173	2.242001E+06

Case a

T [s]	Q_{tur} [m³/s]	H_{net} [m]	η_{ex} [%]	P_{hidr} [MW]	P_{mec} [MW]	M [Nm]
10.1	156.711352	15.963918	0.9414	24.541943	23.103785	2.206249E+06
10.2	156.952081	15.963807	0.9414	24.579472	23.139115	2.209623E+06
10.3	159.913054	15.962428	0.9414	25.041013	23.573609	2.251114E+06
10.4	161.164847	15.961838	0.9414	25.236099	23.757264	2.268652E+06
10.5	160.346367	15.962224	0.9414	25.108545	23.637184	2.257185E+06

Case b

T [s]	Q_{tur} [m³/s]	H_{net} [m]	η_{ex} [%]	P_{hidr} [MW]	P_{mec} [MW]	M [Nm]
10.1	157.189269	15.963697	0.9414	24.616448	23.173924	2.212947E+06
10.2	157.470120	15.963568	0.9414	24.660230	23.215140	2.216883E+06
10.3	160.924588	15.961952	0.9414	25.198657	23.722016	2.265286E+06
10.4	162.385014	15.961258	0.9414	25.426236	23.936258	2.285745E+06
10.5	161.430120	15.961712	0.9414	25.277438	23.796180	2.272368E+06

Case c

Figure 4 Calculated and measured parameters

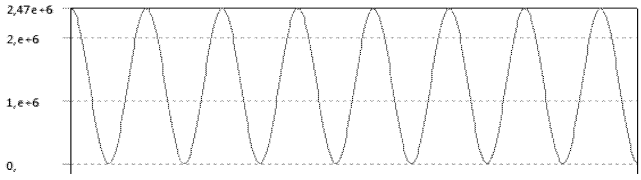


Figure 5.a. Applied stress cycles for the model. Pulsatory cycle with constant amplitude

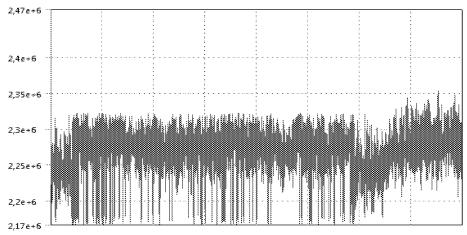


Figure 5.b. Example of real stress cycles for the model

For the nominal power of the turbine (22 MW) there is a maximum torque of M=2,47 x 10^6 Nm. The time of the data acquisition was 10 minutes with a frequency of reading of

0,1 sec. **Fixing conditions imposed:** For the radial bearing surface it an axial movement was permitted, constraining a null movement in radial direction. For mounting the flange of the hydrogenerator, null displacements were imposed in all directions. Torque is applied uniformly distributed on cylindrical surfaces of the 20 holes of the shaft's mounting flange with the turbine (Fig. 6).

MATHEMATICAL MODEL AND SIMULATION

Modal Analysis. The first preliminary analysis was the modal analysis, which determined the behavior of the shaft to six sets of frequencies. (Fig. 7).

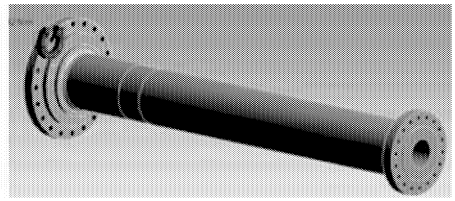
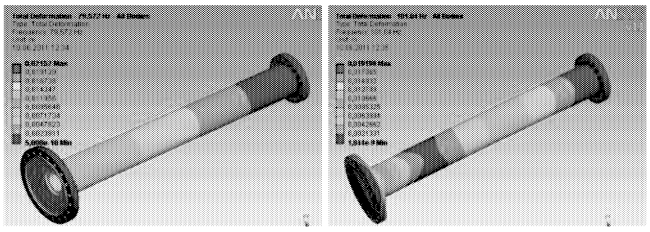
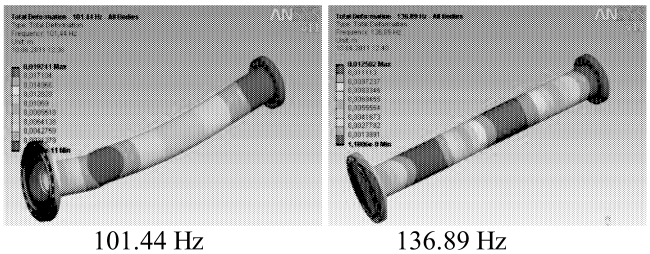


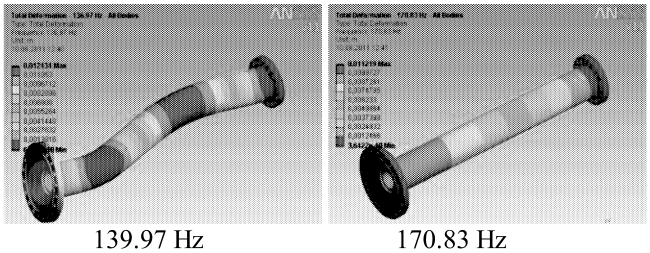
Figure 6 External load.



79.57 Hz 101.04 Hz



101.44 Hz 136.89 Hz



139.97 Hz 170.83 Hz

Figure 7 Total deformation

Pulsatory cycle with constant amplitude. A second analysis that was submitted to the model was to charge it with a constant amplitude pulse cycle (fig. 5.a) equal to the nominal value of torque corresponding to the maximum power turbines. The analysis was performed according to two theories (stress life and **foreign** life), results are presented in Figure 8. When simulations with pulsatory cycles of constant amplitude were used to compute the life time with stress life and strain life analysis methods, the results obtained were in the domain of infinite life (more than 10^6 and 10^9 cycles). Taking into consideration that the duration of studied cycle is 10 minute, and the life time of the shaft is computed between 100.000 and 200.000 operation hours, the most adequate theory to estimate the life time of a shaft is stress life theory.

Random cycle. Cycle loading with random requests (data processing) (Fig. 5.b) was performed according to theory of life stress (Fig. 9).

Table 2. Results

Study case	Lifetime	
	[cycles]	[operation hours]
Case a	1.0142×10^6	169,033
Case b	8.9445×10^5	149,075
Case c	8.9445×10^5	149,075

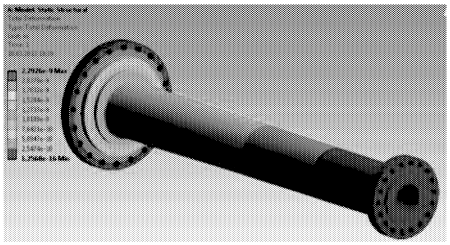


Figure 8.a. Total deformation (pulsatory cycle with constant amplitude and random cycle)

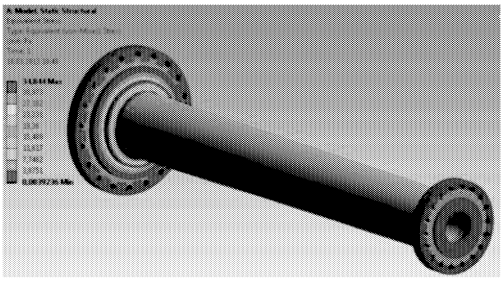
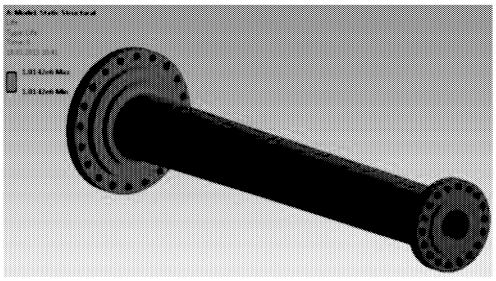
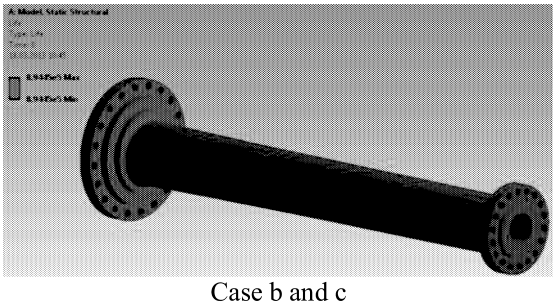


Figure 8.b. Equivalent tensions (pulsatory cycle with constant amplitude and random cycle).



Case a



Case b and c

Figure 9 Simulation results on random loading cycle.

Model sensibility analysis. Cycle loading with random requests (according to data processing) was performed according to theory of life stress (Fig. 10).

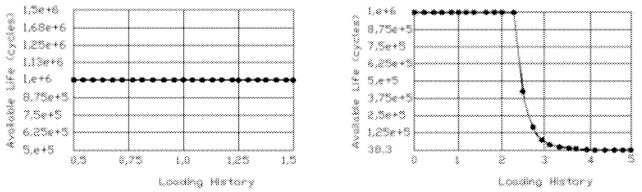


Figure 10. Model sensibility a) M_{max}=1,5 M_n b) M_{max}=5 M_n

From the analysis with a load of 150% compared to the normal one, we can observe that the remaining lifetime is not modified. In case of a load of 500% of the nominal load (accident generator), a very strong diminishing of the life time is observed, and has to be taken into consideration in order to improve the model.

CONCLUSION

From the general analysis of the 6 modules'own frequency, some conclusions were taken:
- The first six own frequencies are large enough to avoid resonance phenomena;
- The dynamic behavior of the model performed better, with no discontinuities in the obtained response;
- Model is consistent in terms of mass properties (18945 kg)

Estimation of lifetime according to theory of "stress life" is consistent with reality (100.000-200.000 equivalent operating hours) if one takes into account the simplifications of the model and the fact that the simulation was performed for normal operation load, a situation not very often found in real operation. These results were already presented in (Piraianu et al. 2000; Rosu et al. 2000; Dragoi et al. 2012).

The additional information obtained within this paper is related to the influence of pulsatory regimes upon a life time. As presented in Case a, Case b and Case c in Table 2, different variation of the torque were simulated.

The causes for the appearance of this kind are multiple: local grid fluctuations, national grid fluctuation, the requirement of the speed governor to balance the grid frequency and/or voltage (this is done by increasing or reducing the output power of the unit by adjusting the discharge), vibrations levels more than normal to the radial and axial bearings.

According to results obtained in Table 2, a decrease in life time is significant in Case b and Case c compared to Case a with about 20.000 hours of operation.

Although the same output power is generated, therefore the same income per generated efficiency, the cost of energy production is significantly higher because the unit is forced to enter in maintenance and/or repair much sooner than under normal circumstances.

The scope of this paper is to present the influence of high amplitude of the pulsatory regimes that occur during operation of power plants. The presented algorithm and method can be used as a predictive method to determine the life time of different equipment. Furthermore, based on the actual functioning hours of each shaft (or part), an estimation of dangerous regimes can be made and can be avoided to ensure a safety operation.

REFERENCES

Bathe, K.J. 1996. "Finite Element Procedures", Prentice-Hall, Inc., ISBN 0-13-301458-7.

Blumenfeld, M.R. 1995. "Introduction in Finite Element Analysis". Technical Press House, Bucharest.

Constantinescu, D. M. 2011."Durability and maintenance of metallic structures", Notes of lectures, University POLITEHNICA of Bucharest.

Dragoi, C., Piraianu, V. F., Piraianu, C. 2012. "Evaluating the Lifetime of Hydro-Mechanical Equipment for Low Head Hydropower Plants", U.P.B. Scientific Bulletin, Series D, Vol. 74, Iss.1, ISSN 1451-2358, Bucharest.

Pîrăianu, C. 2001. "Researches upon transitory phenomena on low head power plant units". Ph.D. Thesis, University POLITEHNICA of Bucharest.

Pîrăianu, C., Sorohan, Şt., Roşu, C.A., Vasiliu, N. 2000. Analysis of the Bulb Type Turbine Shafts by the Finite Elements Method. Fifth International Conference on Hydraulic Machinery and Hydrodynamics., 18-20 October 2000, Timisoara.

Pîrăianu C., Dorobanțu C., Antonescu I., Vasiliu N. 2000. Experimental Researches of the vibrations from Frunzaru Hydropower Plant. International Power Conference, U.P.B.

Zienkiewicz, O.C., 2000. "The Finite Element Method", Fifth Edition, Butterworth-Heinemann, ISBN 0-7506-5049-4.

WEB REFERENCES

www.efatigue.com
www.ansys.com
www.lmsintl.com

BIOGRAPHIES

Constantin DRAGOI was born in Romania on the 23th of Decembre, 1971. He received the M.Sc. degree in Mechanical Engineering from the University POLITEHNICA of Bucharest in 1997, and a Ph.D. degree in fluid power systems from the same university in 2001. He is currently lecturer in the Hydraulics, Hydraulic Machinery and Environmental Engineering Department, Power Engineering Faculty, University Politehnica of Bucharest.

Vlad Florin PÎRĂIANU was born in Romania on the 4 of March, 1983. He received the M.S. degree in hydropower engineering from the University Politehnica of Bucharest in 2007. He is a Ph.D. candidate in hydropower engineering within the same university, and he is working as assistant professor in the Hydraulics, Hydraulic Machinery and Environmental Engineering Department.

Petru-Cristinel IRIMIA was born in Romania on the 16th of February, 1967. He received the M.Sc. degree in Mechanical Engineering from the University Transilvania of Brasov in 1991. He developed a carrer in R&D, having a deep education in finite element analysis with high level LMS software. He is serving currently as general manager of LMS Test Division - CAE Division - Engineering Services from Romania.

MODELLING BUILDING EFFECTS ON ATMOSPHERIC DISPERSION. COMPARISON BETWEEN ADMS AND ANSYS CFD

Claudia GHEORGHE (NICOLICI)
Ilie PRISECARU
Power Plant Engineering Faculty
Politehnica University of Bucharest
313, Independentei Street, Sector 6
Bucuresti, Romania

Stefan NICOLICI
Radu M. BILEGAN
Center of Technology and Engineering
for Nuclear Projects
409, Atomistilor Street, Magurele
Ilfov County, Magurele

KEYWORDS
Atmospheric Dispersion, Gaussian models, ADMS, CFD.

ABSTRACT

This paper presents the preliminary results of a comparative study regarding the radionuclides transport around the reactor building of a Nuclear Power Plant (NPP). The aim of our work was to assess the accuracy of the RANS (Reynolds Averaged Navier-Stokes) CFD (Computational Fluid Dynamics) model developed for a continuous pollutant stack release by comparing the CFD results against the ADMS Gaussian distribution. The Ansys CFX code was set up to simulate a stable atmospheric boundary layer using the six-equation Reynolds stresses turbulence model and the Lagrange Particle Tracking (LPT) method for pollutant dispersion. The two codes were compared in terms of concentrations and plume spread at ground level for five stack-building positions relative to the mean wind direction. The main finding is that the difference between the two set results is below a factor of two depending on the building position.

INTRODUCTION

Nowadays we are witnessing an increase of awareness all over the world and of the concerns of the governments, society and industry on the risks for population and the environment determined by the location and operation of dangerous and polluting industries including the energy production units as well as by their local, regional and worldwide effects. Nuclear power plants can release radionuclides to the atmosphere under normal operating conditions or during abnormal events. As a consequence, atmospheric dispersion and radiation dose calculations for routine and accidental releases of radioactive materials are of great importance for safety and licensing requirements (IAEA 2011).

Gaussian models (widely used for regulatory purposes) behave better for smooth-plane sites and give best results for dispersion regions where there is only small variation of terrain elevation (IAEA 2002; Leroy et al. 2010). Thus, there is inherent inaccuracy if the environment around the pollutant source has a tall nearby buildings, variable surface roughness vegetation, or height variations not considered by the atmospheric dispersion model. Furthermore, presence of buildings near the source can disrupt the air streamlines. Selection of an adequate Gaussian model should consider the ratio between the height at which the effluent emission takes place and building height influencing the air movement near the point of emission.

Considering the above-mentioned aspects, in this work we have performed an investigation of the dispersion around a tall circular building employing a CFD and ADMS approaches. Some basic aspects of the ADMS model and the specific meteorological and site characteristics are presented in next Section. Further, the CFD physical models and the set up conditions are shown in the paper. The final part presents the numerical results, some discussion and the main conclusions.

DESCRIPTION OF THE ADMS MODEL

ADMS model developed by CERC (CERC 2012; Carruthers et al. 2009) represents an advanced Gaussian dispersion model that uses the boundary layer height and the Monin-Obukhov length to characterize the atmospheric stability and the vertical wind, temperature and turbulence profiles. For stable and neutral atmospheric conditions, ADMS model uses the Gaussian distribution for pollutant concentration calculation, while for unstable atmosphere a skewed distribution is employed.

The building effects are considered in ADMS by entraining a part of the plume into the downstream cavity region near the building, bringing the plume at the ground level. The same concentration profile is employed but with modified plume height and spread parameters (Robins and McHugh 2001). Further down the recirculation region, the concentrations are determined by summing up the ground level plume and the rest of the non-entrained plume. The equation used for plume concentration calculation is:

$$C = (Q/U)C_y(y, y_p, \sigma_y)C_z(z, z_p, \sigma_z, h) \quad (1)$$

where: Q is the plume strength, U is the mean wind velocity, C_y and C_z are profile functions, y_p and z_p are plume centerline coordinates, σ_y and σ_z are the spread parameters, and h is the boundary layer height.

The problem of air movement around buildings has been addressed in literature based on an initial unperturbed flow field and analyzing the perturbations of the flow properties when the streamlines meets a building. To illustrate this phenomena we assume the existence of a building located on a flat terrain and neutral stratification of the atmospheric boundary layer. Upstream of building, the vertical wind

velocity profile is logarithmic, the turbulence is homogeneous and the pressure is uniform. The presence of a building in this field will modify the wind speed and the pressure leading to an aerodynamic distorted field around the obstacle.

The size of the perturbed flow around the building, denoted B, is determined by the equivalent size of the building (height H_B, width W_B and length L_B). The building module in ADMS will take into account the effects of the building only if the source is positioned in the area B. This region is divided as shown in Figure 1. These sub-areas are (CERC 2012):

- the recirculation, R;
- wake zone, W;
- the building upstream area, U;
- the side the building, where the flow field is affected by the building;
- the lateral wake, E.

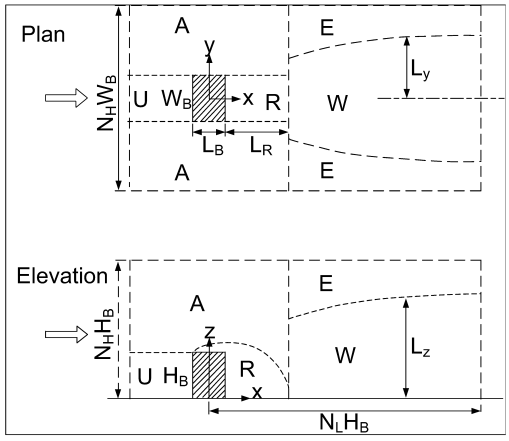


Figure 1: The Flow Region Around the Building

The upstream limit, x =-x_{min}, is determined as the minimum of the two values x_{min} respecting the two following relations:

$$\sigma_y\left(x_{\min}-\left(\frac{L_B}{2}\right)\right)=\frac{W_B}{2};\quad \sigma_y\left(x_{\min}-\left(\frac{L_B}{2}\right)\right)=\frac{H_B}{2} \tag{2}$$

There is no limit for the region downstream of the building B. The effects of the building becomes negligible at a certain distance in order to enable the attaining of the initial flow field. The transverse and vertical dimensions of area B are calculated with the following relations:

$$y_{\max}=N_H/(W_B/2);\qquad z_{\max}=N_HH_B \tag{3}$$

where:

$$N_H=1+2\min(1, W_B/H_B) \tag{4}$$

The length of the recirculation zone, L_R is determined for stable or neutral atmosphere as:

$$L_R=\frac{AW_B}{1+BW_B/H_B};\qquad A=1.8\left(\frac{L_B}{H_B}\right)^{-0.3};\ B=0.24 \tag{5}$$

The vertical limit of the area R is given as a function $z_R(x)$ =z_{Rmax}f(x) respecting the conditions: $z_R(x_R$=x=$L_B/2+L_R)$=0 and $z_R(x_{sep})$=H_B. The function parameters are as follows:

- if $L_B \geq \min(H_B, 0.5W_B)$:

$$z_{R\max}=H_B;\qquad x_{sep}=\frac{L_B}{2} \tag{6}$$

- if $L_B < \min(H_B, 0.5W_B)$:

$$z_{R\max}=H_B\left[1+0.7\left(1-\exp\left\{\frac{-(W_B-2L_B)}{H_B}\right\}\right)\right]$$
$$x_{sep}=\frac{-L_B}{2} \tag{7}$$

Downstream of the recirculation zone, in the areas W and E the perturbation theory is used to determine the flow field (Counihan et al 1974). A simplified three-dimensional model with an average velocity upstream of building, U=U_H, and a constant turbulent viscosity $D_{y,z}\infty u\cdot H_B$ is employed. The mean velocity components are calculated with the following relations:

$$u=U_H\left\{1-\hat{u}\left[\frac{W_B}{2\lambda_y}\right]\left[\frac{H_B}{2\lambda_z}\right]^2 g(\xi)h(\eta)\right\} \tag{8}$$

$$v=-U_H\hat{u}\left[\frac{W_B}{2(x-x_0)}\right]\left[\frac{H_B}{\lambda_z}\right]^2 g(\xi)h(\eta)\eta/2 \tag{9}$$

$$w=-U_H\hat{u}\left[\frac{H_B}{(x-x_0)}\right]\left[\frac{W_B}{2\lambda_y}\right]\left[\frac{H_B}{2\lambda_z}\right][g'(0)-g'(\xi)]h(\eta) \tag{10}$$

where:

$$\eta=\frac{y}{\lambda_y};\quad \xi=\frac{z}{\lambda_z} \tag{11}$$

$$g(\xi)=\left(\frac{\xi}{2}\right)\exp\left(\frac{-\xi^2}{4}\right);\quad h(\eta)=\left(\frac{1}{2\sqrt{\pi}}\right)\exp\left(\frac{-\eta^2}{4}\right) \tag{12}$$

The horizontal and vertical characteristic lengths are given by:

$$\lambda_y(x)=\left\{\frac{D_y(x-x_0)}{U_H}\right\}^{1/2};\quad \lambda_z(x)=\left\{\frac{D_z(x-x_0)}{U_H}\right\}^{1/2} \tag{13}$$

and the turbulent viscosity is:

$$D_y=\kappa u_* H_B,\quad D_z=2\kappa u_* H_B \tag{14}$$

Site meteorological characteristics dictate the rate at which the plume disperses and deposits, the degree of dispersion and the direction of plume travel. The meteorology is characterized by the prevailing wind speed, wind direction and trajectory, mixing layer heights, ambient air temperature, precipitation rate (if applicable) and atmospheric stability. For our study we have considered that the mean wind velocity is 5.4m/s, the nearest building is assumed to be 20m away from the release point. The height of the building is taken to be 45m (the height of the reactor building) and the stack height is 53m. The pollutant considered for transport in atmosphere is NO_x.

CFD MODELLING APPROACH

The aim of computational fluid dynamics is to numerically solve the partial differential equations that govern fluid flows; for our simulations we have: continuity, momentum, turbulence and particle distribution equations.

In general, the movement of particles in a continuous medium is influenced by the mass, momentum and energy transfer at the interface. In our simulations, the governing equations of the carrier phase are expressed in an Eulerian form regardless of the presence of particles. This assumption is valid as long as the volume fraction of the dispersed phase is negligible. The two conservation equations for adiabatic flows are:

- continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \vec{v}) = 0 \tag{15}$$

- Navier-Stokes equations:

$$\frac{\partial}{\partial t}(\rho \vec{v}) + \nabla(\rho \vec{v}\vec{v}) = -\nabla p + \nabla(\bar{\bar{\tau}}) + \rho \vec{g} \tag{16}$$

where $\bar{\bar{\tau}}$ is a stress tensor given by:

$$\bar{\bar{\tau}} = \mu\left[\left(\nabla \vec{v} + \vec{v}^{T}\right) - 2/3 \nabla \cdot \vec{v} I\right] \tag{17}$$

The momentum conservation equation for the dispersed phase is written:

$$m_p \frac{dv_p}{dt} = F_t \tag{18}$$

where m_p is the mass of a particle and F_t is the sum of all forces acting on a particle. The interested forces are the friction and the turbulent dispersion force. The first one is proportional to the difference between the particles and the continuous phase velocities:

$$F_D = \frac{1}{2} C_D \rho_F A_F |U_S| U_S = \frac{1}{2} C_D \rho_F A_F |U_F - U_P| (U_F - U_P) \tag{19}$$

where C_D is the friction coefficient and A_F is the cross-sectional area of the particle. The dimensionless friction coefficient expressed by Schiller Naumann model depends on the Reynolds criterion calculated using the properties of the fluid and particle diameter:

$$C_D = \frac{24}{\mathrm{Re}}\left(1 + 0.15\,\mathrm{Re}^{0.687}\right) \tag{20}$$

The turbulent dispersion force is the force acting on a particle due to variations of air velocity, carrying the particles from the areas of high concentration to areas of low concentration. This is equal to:

$$F_{DT} = \frac{C_D}{0.9} \nu_{tc} \left(\frac{\nabla r_d}{r_d} + \frac{\nabla r_c}{r_c} \right) \tag{21}$$

where C_D is the friction coefficient, ν_{tc} is the continuous phase turbulent viscosity, r_d and r_c are the volume fractions of the dispersed phase, respectively the continuous one.

In terms of pollutant dispersion phenomena, the choice of turbulence model is of great importance affecting the turbulent transport. Based on previous numerical studies (Riddle et al. 2004), in this paper we have employed the Reynolds stress model to solve the turbulent quantities. This model is based on transport equations for all components of the Reynolds stress tensor and the dissipation rate. The Reynolds stress model transport equations are solved for the individual stress components, thus considering the anisotropy of the turbulent fluctuations.

In order to assess the accuracy of the Reynolds stress turbulence model in prediction of a stable atmospheric boundary layer, the velocity and turbulent kinetic energy at inlet and outlet of the domain were compared (see Figures 2 and 3). One can observe that the CFX code can well maintain the vertical profiles of air velocity and turbulent energy.

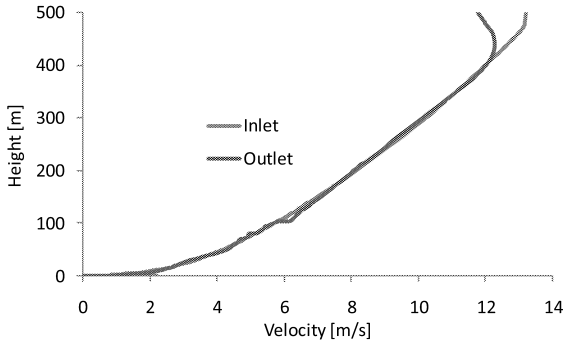


Figure 2: Vertical Velocity Profile

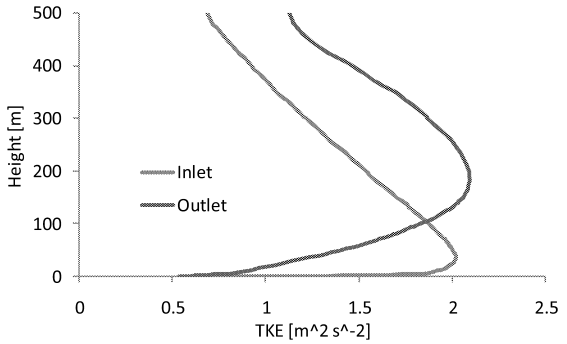


Figure 3: Vertical Turbulent Kinetic EnergyPprofile

The plume spread is modeled considering the Lagrange Particle Transport Model. For simulating the gas dispersion within the particle transport model, a large number of small size solid particles are released at the stack height and their movement is tracked through the continuum fluid. The application of Lagrangian tracking in CFX involves the integration of particle paths through the discretized domain (Ansys 2010). Individual particles are tracked from their injection point until they escape the domain or some integration limit criterion is met. The main settings used in the CFD simulations are presented in Table 1.

Table 1: Main Parameters of CFD Model

Parameter	Value
Particle source [kg/s]	1
Particle diameter [µm]	10
Stack diameter [m]	2.3
Stack height [m]	53
Building height [m]	45
Air velocity at 10m [m/s]	5.4
Rougness [m]	0.3
Friction velocity [m/s]	0.58
Boundary layer height [m]	500
Stack-building positions relative to the wind direction [degrees]	0;45;90;135;180

RESULTS AND DISCUSSION

The entrainment of plumes in building wakes being of major interest and represents a major factor in obtaining acceptable derived release limits. The most important feature of building effect upon the dispersion studies is the plume entrainment in the building wake or its rapid downwash from elevated sources, since this generates higher near-field concentrations at the ground level. The ground concentrations from CFD and ADMS models are presented in Figure 4. The plume spread at the ground level is larger for the Gaussian ADMS model, but the downwind distance at which the plume touches the ground is almost the same. Furthermore, analyzing Figure 5 we can observe that the maximum concentrations occurring in the wake of the building are higher for the ADMS model when the building is directly in front or behind the stack (0 and 180 degrees).

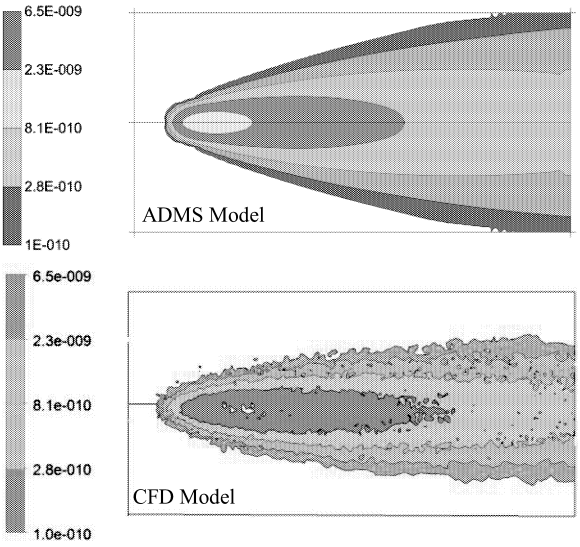


Figure 4: Ground Concentration for 0 degrees Building Position [m^3/m^3]

For all other cases (45, 90 and 135 degrees), some differences can be observed with the ground level concentrations rising more rapidly with distance for CFD model than for ADMS.

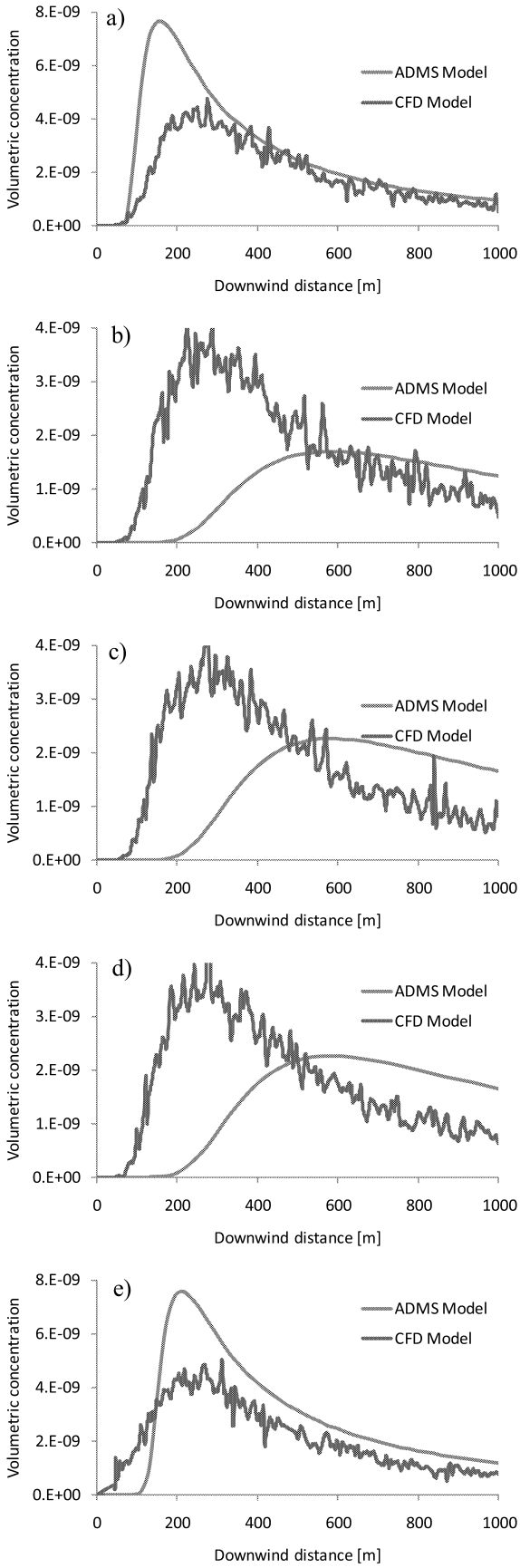


Figure 5: Ground Concentration under the Plume Centerline for Five Building Positions [m^3/m^3]
(a-0deg; b-45deg; c-90deg; d-135deg; e-180deg)

CONCLUSIONS

The standard Gaussian models used to evaluate the atmospheric dispersion of radionuclides do not model the flow directly, and their applicability to study the interaction of the air flow with buildings and complex terrain topography is limited. In this paper the authors have considered the problem of dispersion calculation around a circular building, using two approaches: the standard Gaussian distribution and the advanced CFD method.

The major conclusions are drawn as follows:

1. The results of the CFD LPT simulations have shown that the building has influenced the plume dynamics only for 180 degrees positioning, bringing the plume at ground level in the recirculation area. The ADMS model predicted that the plume is affected for both 0 and 180 degrees positions of the building.

2. The plume spread seems to be underpredicted by the LPT methodology, this aspect could be explained by the absence of the molecular diffusion of the gases or by the drag force exerted on the particles. Using the LPT approach, the pollutant is approximated by a large number of small particles on which drag and turbulent dispersion forces act. The drag force has the role to transport the particles downwind, but also it limits the lateral spreading of the particles. The only contribution to lateral dispersion is the turbulent dispersion given by the velocity fluctuations.

REFERENCES

Ansys INC.2010. "Ansys-CFX13 User Manual".

Carruthers D.J.; Weng W.S.; Hunt J.C.R.; Holroyd R.J.; McHugh C.A.; Dyster S.J. 2009. "Plume/Puff Spread and Mean Concentration Module Specifications". Tech. rept. ADMS 4.0.

CERC. 2012. "ADMS 4 Technical Specification".

Counihan J.;Hunt J.C.R.; Jackson P.S. 1974. "Wakes behind two dimensional surface obstacles in turbulent boundary-layers". Journal of Fluid Mechanics 64, 529-563

IAEA. 2011. "Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations" – IAEA Safety Standards Series, No. SSG-18, Vienna.

IAEA. 2002. "Dispersion of radioactive material in air and water and consideration of population distribution in site evaluation for Nuclear Power Plants" – IAEA Safety Standards Series, No. NS-G-3.2.

Leroy C.; Maro D.; Hébert D.; Solier L.; Rozet M.; Le Cavelier M.; Connan O. 2010. "A study of the atmospheric dispersion of a high release of krypton-85 above a complex coastal terrain, comparison with the predictions of Gaussian models (Briggs, Doury, ADMS4), Journal of Environmental Radioactivity 101, 937-944.

Riddle A.; Carruthers D.; Sharpe A.; McHugh C.; Stocker J. 2004. "Comparisons between FLUENT and ADMS for atmospheric dispersion modeling". Atmospheric Environment 38. 1029–1038.

Robins A.; McHugh C.A. 2001. "Development and evaluation of the ADMS buildings effects module". Int. J. Environ. Pollut. 16, 161-174.

ELECTRONIC NETWORKS SIMULATION

IMPROVING PERFORMANCE PREDICTION OF PARALLEL AND DISTRIBUTED DISCRETE EVENT SIMULATION: A ROUGH SETS-BASED APPROACH

László Muka and István Derka
Department of Telecommunications
Széchenyi István University
Egyetem tér 1.
H 9026 Győr
Hungary
E-mails: muka@sze.hu, steve@sze.hu

KEYWORDS
Simulation performance prediction, parallel and distributed simulation, rough sets, coupling factor approach, prediction cost.

ABSTRACT

The rapid growth of complexity of communication networks – including grids, cloud systems and services – increases the need for the use and performance prediction of parallel and distributed discrete event simulation. The paper aims at the decreasing the overall cost of simulation by improving the performance prediction of simulation particularly for the phases with uncertain and vague information. Involving rough sets-based methods, the paper introduces how the efficiency of the well known coupling factor method of the performance prediction of parallel and distributed simulation can be improved. According to the presented results, the prediction can merely be based on the maintainable lookahead feature of the designed simulation model. Using the rough set analysis, the paper also describes how the number of simulation experiments necessary for prediction can be limited without the decrease of accuracy of prediction.

INTRODUCTION

In the last decade, the need for the discrete event simulation of communication networks and services has been constantly increasing – together with the rapidly growing complexity of the communication systems – because the discrete event simulation turned to be a proficient tool for the analysis of these systems.
The simulation of large, complex systems may often require significant computing capacity which may practically be available on a computer network. Nowadays, many types of network computing resources are available for a user: for example, different clusters of computers, grids and even computing clouds. For such a situation, the appropriate simulation method for the researcher usually is the Parallel Discrete Event Simulation (PDES) method. (Fujimoto 1990). The modified parallel and distributed simulation method may be efficient in cloud environment too. (Fujimoto et al. 2010)

This research was supported by the TÁMOP-4.2.2/B-10/1-2010-0010 project and by the Széchenyi István University (15-3202-08).

The development and use of systems based on the PDES methods may be resource consuming both in time and in expertise, therefore it can be very useful to have an approximate performance-estimation method for the early stage of a simulation project (Muka and Lencse 2008, 2011).
For the *preliminary* stage of modeling and simulation, the *simulation performance prediction* provides appropriate methods that make available the examination of the influence of different software and hardware parameters such as the chosen synchronization protocol, the logical structure of the simulation model, the topology data and the hardware parameters of computing system, etc.
The two major *motivations* for the authors – working on the problems of simulation-based evaluation of communication networks and services, including cloud systems and services – are: the lack of prediction methods which can be used without the need of special tools and that support the work with imperfect, vague data.
According to motivation the following goals were set:

- First, to propose a method to improve the efficiency of simulation performance prediction which is based on easily accessible systems and tools.
- Second, to provide a new approach which can be appropriate to cope with the uncertainty and vagueness (which may reach particularly high level in case of predicting task execution process in clouds) information available in the early stages of modeling and simulation, .

The major contributions of the paper may be summarized as follows:

- A new rough coupling factor method is derived – based on rough set analysis – by simplifying the prediction evaluation using only the lookahead feature of the simulation model (the classic coupling factor method is appropriate for the performance prediction of the parallel and distributed simulation using the conservative null message synchronization algorithm).
- The paper also presents how the number of simulation experiments necessary for simulation performance prediction can be limited without decreasing the accuracy of prediction. The cost of prediction may be decreased this way.

The paper is organized as follows. The first section summarizes the related work. In the next section, first the basic concepts of the rough approach are introduced; then the rough improvement of the coupling factor approach is introduced including: the description of the basic principles of the coupling factor method, then the results of previous experiments. This section also includes the starting rough analysis of the problem the efficiency and speedup examinations and finally, the presentation of the training analysis of efficiency. The last section concludes the work.

RELATED WORK

There are two relevant groups of related work from the point of view of the analysis described in the paper:
One group is the recent simulation performance prediction approaches (for parallel and distributed simulations) and the other group is the performance prediction applications based on rough sets.

Simulation performance prediction approaches - parallel and distributed execution
Simulation execution *trace-based* methods use the event trace of simulation runs for performance prediction. According to method described in (Juhász et al. 2003), the performance prediction is based on data got from the sequential simulation execution and used in Critical Path Analysis. The difficulty with this approach is the necessity of special tools for the prediction. Another simulation trace-based method predicts the required synchronization cost (Swope and Fujimoto 1987).
Results introduced in (Varga et al. 2003), confirmed in (Lencse and Varga 2010; Lencse et al. 2013; Lencse and Derka 2013) – using results in (Bagrodia and Takai 2000) as prerequisites –, describe and verify a *practical* simulation performance predicting approach, based on theoretical considerations and on sequential simulation runs. The method predicts – only for conservative null message-based algorithm – the parallelization potential of the simulation models, and formulates requirement on how to exploit it. These methods are well suited to the requirements of preliminary modeling.
As an alternative to the conservative synchronization, the time driven version of the statistical synchronization method (SSM-T) with its loose synchronization (Lencse 1998) is a less well-known but promising PDES synchronization method that can be applied (Lencse 1999a) for the parallel simulation of certain types of systems such as communication networks. The method can achieve a good speedup that can be predicted on the basis of the frequency of its statistics exchange (Lencse 1999b), which can be used similarly to lookahead.
Analytical methods use mathematical models for prediction. More generally, for the parallel and distributed computing, (Varjasi 2006; Molnárka and Varjasi 2012) proposes a graph theoretic based potential function to take into account the influence of computer system topology. This seems to be a promising approach but it needs further theoretical and practical verification.
(Li et al. 2011) suggests an analytic upper-bound based partitioning method for null message-based parallel and discrete event simulation. The analytic upper bound-based method needs further validation too.
The hybrid approach described in (Kunz et al. 2011) uses both the simulation trace-based and the analytical approaches. In the trace-based part, the method analyses only a definite part of the sequence of events of the simulation model; while in the analytical part, for the calculation of the lower bound on the runtime of the parallel simulation, the method applies the Linear Programming approach of a simplified scheduling problem model. The application of the method requires a set of special tools which is a serious limitation.
The off-line and on-line feature of simulation may be an interesting classifier too. The paper (Ewald and Uhrmacher 2009) summarizes the requirements of automating the simulation which is the bases of online simulation.
(Karaoglanoglou and Karatza 2012)
Further promising approach: [Karaoglanoglou and Karatza 2012] introduces methods of resource discovery and categorization for large-scale grid networks. The results of the discovery process then can be used in planning and scheduling of the execution process.

Applications of rough sets for performance prediction
The paper (Pawlak 1998) introduces the well known and important for the telecommunications and other service companies churn prediction problem. The article shows how a rough sets-based prediction model of churn can easily and efficiently cope with the prediction question.
The paper (Krishnaswamy et al. 2002) presents a rough sets approach of application runtime estimation. The rough set model of runtime prediction introduced in the article based on defining the similarity (by finding the appropriate properties) derived from available past data. Using the extracted properties of similarity the runtime of applications predicted.
Parallel and distributed simulation forms complex execution environment which may make hard to define similarities and but there may also special further features (for example, the way of synchronization) which should be taken to take into account too.
The paper (Song et al. 2005) introduces how to find a solution for a remanufacturing scheduling and optimization problem in dynamic situation with uncertain data. The method proposed is a cyclic use of Linear Programming, simulation and evaluation and learning based on rough set method in an iterative process. (The problem formulation is highly similar to that of the hybrid approach described above.)

THE ROUGH SETS-BASED APPROACH FOR PREDICTION IMPROVEMENT

Introducing Concept for Rough Set Modeling

RST proposed by Pawlak (Pawlak 1982) is a new powerful mathematical framework particularly suitable for describing imprecise relations with a minimum rule set, in an information system with uncertain, vague data.
An *information system* with imbedded knowledge consists of three sets: the set of *objects* called the *universe*, set of *conditional attributes*, and the set of *decision attributes*. The objects of the universe of the case examined in the paper are computer simulation experiments. In the phase of preliminary modeling and simulation, one should face with

the problem of uncertain, imperfect data thus there is a need for an appropriate tool to handle it. The RST seems to be the required efficient method. Rough sets are defined by their lower approximation and upper approximation sets, which are precise (crisp) sets. More formally, *S= (U, A)* denotes an information system of RST (where the set U is the universe and A is the set of attributes and the sets U and A are finite, nonempty sets
$U = \{x_1, x_2, x_3, \dots x_{|U|}\}$, $A = \{a_1, a_2, a_3, \dots a_{|A|}\}$. The set of values of a called the domain of attribute a and for each attribute defined by a $f_a: U \rightarrow V_a$ function. The set $U_B(X)$ and the set $L_B(X)$ is the *B-lower* and *B-upper approximation* of the set X and defined as follows:

$$L_B(X) = \{x: [x]_{IND(B)} \subseteq X\}$$

$$U_B(X) = \{x: [x]_{IND(B)} \cap X \neq \emptyset\}, \quad (B \subseteq A)$$

The set defined by the equation

$$BN_B(X) = U_B(X) - L_B(X)$$

is the *B-boundary region* of X. If X is a crisp set then, $(X) = L_B(X)$ thus $BN_B(X) = \emptyset$ which means the boundary region is empty.
A *reduct* is the minimal subset of attributes that allows the same classification of objects of the universe as the whole set of attributes (discerning one object from all objects with a different decision). Expressing it in a formal way with indiscernibility functions:

$$IND_S(B^R) = IND_S(B), B \subseteq A, B^R \subset B$$

The set B^R is a reduct of B if B^R is independent (consisting of indispensable attributes).
Discretization is the operation of mapping the value ranges of an attribute to a *selected optimized* set of discrete values:

$$f_D: V_a \rightarrow V_{D(a)}, \; D = \{s_1, s_2, s_3, \dots, s_{|D|optimal}\}$$

The classification C $(= \{X_1, X_2, X_3, \dots, X_n\}$ $(X_i \subset U,$ $X_i \cap X_j = \emptyset, \cup X_i = U, i = 1,2,3, \dots, n)$ can be measured by of its accuracy and quality:

accuracy: $\eta_{B(C)} = \frac{|L_B(X_i)|}{|U_B(X_i)|}$

quality: $\beta_{B(C)} = \frac{|L_B(X_i)|}{|U|}$

The classification may also be described by *IF-THEN* rules. A rule can be evaluated , for example, by its generality: the coverage shows the fraction of relationship of objects of THEN and IF part of the decision class.
It can also be useful to introduce terms of *B*-positive region and *B*-negative regions which underlines the relationship with information retrieval:

$NEG_B(X) = U - U_B(X)$, *B*-negative or *B*-outside region,

$POS_B(X) = L_B(X)$, *B*-positive region.

A Rough Improvement of the Coupling Factor Approach

The Coupling Factor Approach: Principles
The coupling factor method proposed by (Varga et al. 2003) seems to be a good candidate for the rough analysis and improved formulation.

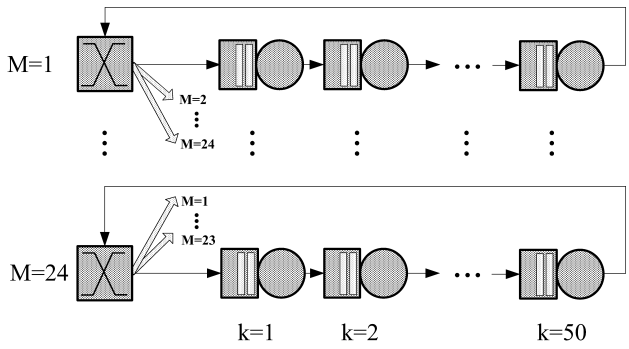


Figure 1: The Experimentation/Evaluation CQN model

The principle of the coupling factor prediction method may be formulated as an inequity:

$$L * E \gg \tau * P$$

where L is the lookahead value characterizing the model (simsec), E is the event density generated by the model (ev/simsec), τ is the latency of messages between logical process (LPs) of the model (sec), and P is the event processing computation hardware performance (ev/sec).
According to the method, the coupling factor λ is calculated according to the formula

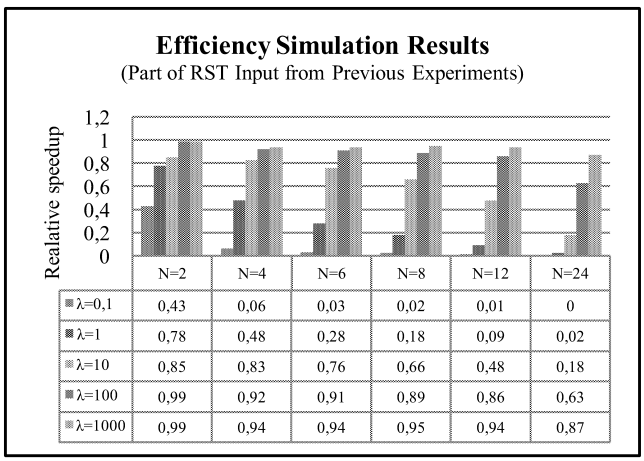
$$\lambda = \frac{L*E}{\tau*P}$$

The high value of the coupling factor λ shows the good potential of the simulation model for parallelization. The formula involves only four parameters for the performance prediction calculation. These parameters can be measured in simple sequential simulation runs.
For a separate process, the λ_N parallelization potential of a process is only a part of the whole potential:

$$\lambda_N = \frac{L*E}{\tau*P} * \frac{1}{N_{LP}}$$

where N_{LP} the number of the LPs (Lencse and Varga 2010).

Summary of the Lencse-Varga Simulation Experiments
For the present study, the results described in (Lencse and Varga 2010) are taken as starting point of the analysis.
The constraints and results of experiments (altogether 72 experiments) are summarized in the following. (Figure 2. shows the results for the relative speedup.)



	N=2	N=4	N=6	N=8	N=12	N=24
λ=0,1	0,43	0,06	0,03	0,02	0,01	0
λ=1	0,78	0,48	0,28	0,18	0,09	0,02
λ=10	0,85	0,83	0,76	0,66	0,48	0,18
λ=100	0,99	0,92	0,91	0,89	0,86	0,63
λ=1000	0,99	0,94	0,94	0,95	0,94	0,87

Figure 2: The Summary of Efficieny Simulation Results

Table 1. Efficiency and Speedup Analysis

Analysis	Generation Method					
	G1	G2	G3	G4	G5	G6
EFFICIENCY						
"H"efficiency, lookahead only rules	2	2	2	2	2	2
coverage (H=17), number	6+6	6+6	6+6	6+6	6+6	6+6
coverage (H=17), %	35+ 35	35+ 35	35+ 35	35+ 35	35+ 35	35+ 35
"L" efficiency, lookahead only rules	1	1	1	1	1	1
coverage (L =13), number	6	6	6	6	6	6
coverage (L =13), %	46	46	46	46	46	46
number of "H" rules	7	18	7	19	7	19
number of "L" rules	8	22	8	29	8	21
"H" OR "L" rules	0	0	0	8	9	0
SPEEDUP						
"a" speedup, lookahead only rules	0	0	0	0	0	0
coverage (a=15), number	0	0	0	0	0	0
coverage (a=15), percent	0	0	0	0	0	0
"n" speedup, lookahead only rules	2	2	2	2	2	2
coverage (n=15), number	6+6	6+6	6+6	6+6	6+6	6+6
coverage (n=15), %	40+40	40+40	40+40	40+40	40+40	40+40
number of "a" rules	14	34	14	40	14	34
number of "n" rules	3	8	3	5	3	8
"a" OR "n" rules	0	0	0	17	8	0

The problem class is *communication network modeling,* the selected approach is *parallel and distributed discrete event simulation*, for which the synchronization protocol is *conservative synchronization with null message algorithm.* The software environment includes: *Linux (Debian) operating system, MPI, NFS, Omnet++ network simulator*.
The hardware environment is a *homogeneous cluster* of 12 PCs.
The experimentation/evaluation model is the CQN model (Figure 1.) (delivered with OmNet++), with FCFS service discipline. For the experiments in the case, the starting number of jobs is 2 jobs/simple queue with exponential inter-arrival time and with exponential service time distributions (the expected value for arrival and service time is 10sec). The delay on links between simple queues is 1sec. The number of tandem queues is 24 (M), the number of simple queues/tandem queue is 50 (k). The switching in CQN model performed by uniform probability distribution to switch to the next tandem queue, and the delay of switching models lookahead.
The task assignment principles in the execution are the partitioning into LPs and Load Balancing Criterion. The measured value of τ is 25 μsec for the system in the case.
To define the value of λ the variables, E and P values were measured in sequential simulation runs.

Starting the Rough Analysis: Discretization, Condensing the Set of Attributes, Evaluation
The *basic sets* of the rough analysis are

$$S = (U, A), D$$

From these sets, the full set of *conditional* attributes may be defined as

$$C = \begin{Bmatrix} P, E, \tau, L, \lambda, \lambda_N, \lambda_{extrapolated}, N_{LP}, N_{cores}, \\ N_{hosts}, N_{tandemques}(M), a_{speedup}, a_{efficiency} \end{Bmatrix}$$

The full set of *decision* attributes contains two subsets:

$$D = \{d_{speedup}, d_{efficiency}\}$$

The set of the analyzed *objects* may be described as

$$U = \{o_1, \dots, o_{30}\}$$

where the object of the universe o_i is the *i*-th measurement result of the examined simulation experiments.

The discretization mapping was applied as follows:

$$V_{a_{efficiency}} = \left\{L\left(v_{a_{efficiency}} < 0.6\right), H\left(v_{a_{efficiency}} \geq 0.6\right)\right\}$$

$$V_{a_{speedup}} = \left\{a\left(v_{a_{speedup}} \geq 4\right), n\left(v_{a_{efficiency}} < 4\right)\right\}$$

$$V_{a_L} = \{l(v_{a_L} = 0.1), m(v_{a_L} = 1), h(v_{a_L} = 10), s(L = 100), e(v_{a_L} = 1000)\}.$$

Without changing the discernibility, the mapping used for V_{a_L} could be used for the set V_{a_λ} with the same discernibility.
As a constant, the τ attribute may be omitted. The P and E attributes may also be omitted because they necessary only for calculation of . The attributes N_{hosts}. N_{hosts}, N_{hosts} may be replaced by a common attribute N .

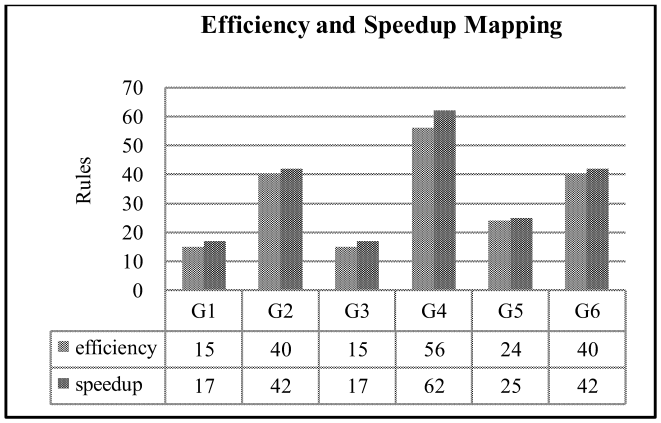


Figure 3. Efficiency and Speedup Analysis

Following the previous condensing procedure, the following sets can be used for the examination of the efficiency an the speedup:
$C_1 = \{L, \lambda, N, a_{speedup}\}$, $D_1 = \{d_{efficiency}\}$,
$C_2 = \{L, \lambda, N, a_{efficiency}\}$, $D_2 = \{d_{speedup}\}$ and
$N = \{1,2,4,8,16,24\}$.
Remark: One may be satisfied with the simplification because the evaluation itself simplifies together with it, but it should be noted that the attributes may only be omitted because they cannot change (cannot increase too) the discernibility.

Efficiency and Speedup Analysis
In this part of the RST analysis, reducts were generated for the sets S_1 and S_2 using different reduct-generation algorithms.Next, the generated set of rules were used for the re-classification of objects. (The *Rosetta* RST analysis system (Komorowski et al. 2002) was used in the examinations.) The reduct generation algorithms were as follows:
G1(Johnson), G2(Exhaustive RSES (RSES=Rough Set Exploration System)), G3(Johnson RSES), G4(Dynamic Exhaustive RSES), G5(Dynamic Johnson RSES), and the G6(Genetic).
The G1 and G2 were the most efficient generation algorithms for the examined systems: they produced 15 and 17 rules for efficiency and for speedup (Figure 3) and with the same coverage parameters as the other four algorithms with more rules (Figure 3, Table 1). The weakest is the G4 with 56 and 62 rules for the efficiency and for the speedup respectively.
G4 and G5 generate rules with double classification (8, 9 and 17, 8 for the efficiency and for the speedup respectively.
All rule sets can re-classify objects except G4 (accuracy 0,9).
In the *efficiency* analysis, rules depending only on the L attribute describe powerfully the classification according "H" and "L" efficiency values (two rules with 70% coverage for the "H" efficiency value and one rule with 46% coverage for the "L" efficiency value.
The speedup analysis shows another picture: the "n" speedup can be classified with powerful, Lookahead depending rules (2 rules with 80% coverage, depending only on Lookahead attribute). There are no efficient rules for the "a" speedup classification.

Training Analysis of Efficiency Prediction
In this part of the examinations training sets were generated by deviding the U set into training and test parts by random random splitting. The generated training sets were used for the generation of classification rules, then, the test sets were classified according to these rules.

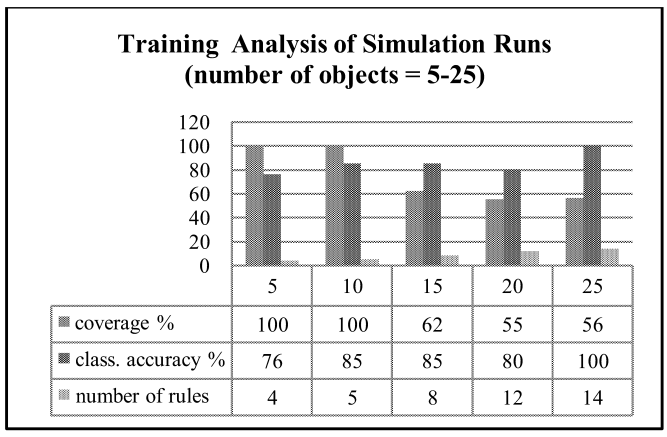
$$f_{split\left(\frac{x}{y}\right)RND}(U) \rightarrow U_{learning}, U_{test}$$

$$\left|U_{learning}\right| = 5 - 25,\ step = 5$$

The G1 algorithm was used for the reduct generation in all cases.
The result of the analysis are summarized in Figure 4.
The number of generated rules increases from 5 to 14 for the increasing size of training set.
The behavior classification accuracy shows the most remarkable feature: the accuracy reaches a good value (85%) at a small size of traing set which remains approximatly stable for the next set sizes. According to this result, *there is no need for too many simulation cycles for a satisfactory prediction*.

CONCLUSION

The development of the network computing capacities generates new challenges for the performance prediction of parallel and distributed simulation.
There have been a lot of researches made in the area but of most of the results cannot be applied without having special systems and tools.
The paper underlines and explains the increased need for the performance prediction of simulation.
Using the rough sets-based method, the paper proves that the prediction efficiency of the well known coupling factor method – the method for performance prediction of parallel and distributed simulation – can be improved: the prediction can be based only on the lookahead feature of the simulation model.
The paper concludes (using also the rough set analysis), that executing only a limited number of simulation experiments will not decrease the accuracy. This allows the decrease of the prediction cost.



Figur 4. Training Set-Size and Classification Accuracy

REFERENCES

Fujimoto, R. M. 1990. "Parallel Discrete Event Simulation", *Communications of the ACM*, vol. 33, no 10, pp. 31-53.

Swope, S. M.; Fujimoto, R. M. 1987, "Optimal performance of distributed simulation programs", In *Proceedings of the 19th conference on Winter simulation,* ACM., pp. 612-617.

Fujimoto, R. M.; Malik, A. W.; Park, A. 2010 "Parallel and distributed simulation in the cloud" *SCS M&S Magazine*, *3*, pp. 1-10.

Lencse, G. 1998. "Efficient Parallel Simulation with the Statistical Synchronization Method", *Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation (CNDS'98)* (San Diego, CA. Jan. 11-14). SCS International, pp. 3-8.

Lencse, G. 1999a. "Applicability Criteria of the Statistical Synchronization Method", *Proceedings of the Communication Networks and Distributed Systems Modeling and Simulation (CNDS'99)*, San Francisco, CA. (1999. Jan. 17-20) SCS International, pp. 159-164.

Lencse, G. 1999b. "Design Criterion for the Statistics Exchange Control Algorithm used in the Statistical Synchronization Method", *Proceedings of the Advanced Simulation Technologies Conference (ASTC 1999)* part of the 32nd Annual Simulation Symposium (San Diego, CA. April 11-15) SCS International, pp. 138-144.

Lencse, G.; Varga, A. 2010. "Performance Prediction of Conservative Parallel Discrete Event Simulation", *Proceedings of the 2010 Industrial Simulation Conference (ISC'2010)* (Budapest, Hungary, 2010. June 7-9.) EUROSIS-ETI, pp. 214-219.

Lencse, G.; Derka I.; Muka, L. 2013. "Towards the efficient simulation of telecommunication systems in heterogeneous execution environments", submitted to the *36th International Conference on Telecommunications and Signal Processing (TSP), July 2-4, 2013, Rome, Italy*.

Lencse, G. I. Derka 2013. "Testing the Speed-up of Parallel Discrete Event Simulation in Heterogeneous Execution Environments", submitted to the *2010 Industrial Simulation Conference (ISC'2010)* (Ghent, Belgium, May 22-24, 2013.)

Bagrodia, R. L. and Takai, M. 2000. "Performance evaluation of conservative algorithms in parallel simulation languages", *Parallel and Distributed Systems, IEEE Transactions on*, *11*(4), pp. 395-411.

Juhasz, Z.; Turner, S.; Kuntner, K.; Gerzson, M. 2003. "A Performance Analyser and Prediction Tool for Parallel Discrete Event Simulation", *International Journal of Simulation*, vol. 4, no. 1, May 2003. pp. 7-22.

Kunz, G.; Tenbusch, S.; Gross, J.; Wehrle, K. 2011. "Predicting Runtime Performance Bounds of Expanded Parallel Discrete Event Simulations", In *Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on* (pp. 359-368). IEEE.

Varga, A.; Sekercioglu, Y. A.; Egan, G. K. 2003. "A practical efficiency criterion for the null message algorithm", *Proceedings of the European Simulation Symposium (ESS 2003)*, (Oct. 26-29, 2003, Delft, The Netherlands) SCS International, pp. 81-92.

Varga, A.; Sekercioglu, Y. A. 2003. "Parallel Simulation Made Easy with OMNeT++", *Proceedings of the 15th European Simulation Symposium (ESS 2003)*, (Oct. 26-29, 2003, Delft, The Netherlands) SCS International, pp. 493-499.

Molnárka, Gy.; Varjasi N. 2012. "Load Balancing using Potential Functions for Hierarchical Topologies", Acta Technica Jaurinensis Vol. 4. No. 4. pp. 413-424.

Varjasi, N. 2007. "Parallel Algorithm for linear equations with different network topologies", *in Lecture Series on Computer and Computational Sciences, (2007),* Brill Academic Publishers, ISBN 978-90-04-15592-3, pp. 502–505.

Ewald, R.; Uhrmacher, A. M. 2009. "Automating the runtime performance evaluation of simulation algorithms", *In: Simulation Conference (WSC), Proceedings of the 2009 Winter*. IEEE, 2009. pp. 1079-1091.

Pawlak, Z, 1998. "Rough set theory and its applications to data analysis", *Cybernetics & Systems*, *29*(7), pp. 661-688.

Krishnaswamy, S.; Zaslavsky A.; Loke, S. W. 2002. "Predicting run times of applications using rough sets", In *Ninth International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2002)*, pp. 455-462.

Song, C.; Guan, X.; Zhao, Q.; Ho, Y. C. 2005. "Machine learning approach for determining feasible plans of a remanufacturing system", *Automation Science and Engineering, IEEE Transactions on*, *2*(3), pp. 262-275.

Li, C. H.; Park, A. J.; Schenfeld, E. 2011. "Analytical Performance Modeling for Null Message-Based Parallel Discrete Event Simulation", In *Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2011 IEEE 19th International Symposium on,* pp. 349-358., IEEE.

Muka, L.; Lencse, G., 2008. "Cooperating Modelling Methods for Performance Evaluation of Interconnected Infocommunication and Business Process Systems", *Proceedings of the 2008 European Simulation and Modelling Conference (ESM'2008),* (Le Havre, France, Oct. 27-29.) EUROSIS-ETI, pp. 404-411.

Muka, L.; Lencse, G., 2011. "Method for Improving the Efficiency of Simulation of ICT and BP Systems by Using Fast and Detailed Models", *Acta Technica Jaurinensis,* Vol.5 No. 1. 2012., pp. 31-42.

Komorowski, J.; Øhrn, A.; Skowron, A., 2002. "The ROSETTA Rough Set Software System", *In Handbook of Data Mining and Knowledge Discovery, W. Klösgen and J. Zytkow (eds.),* ch. D.2.3, Oxford University Press, ISBN 0-19-511831-6.

Pawlak, Z. 1982. "Rough sets", *International Journal of Parallel Programming*, *11*(5), pp. 341-356.

Karaoglanoglou, K.; Karatza, H. 2012. "Directing Requests and Acquiring Knowledge in a Large-Scale Grid System", *International Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS 2012,* Genoa, Italy, July 8-11, pp. 609-617.

BIOGRAPHIES

LÁSZLÓ MUKA graduated in electrical engineering at the Technical University of Lvov in 1976. He became a university doctor (dr univ) in architectures of CAD systems at the Technical University of Budapest in 1987. He got his PhD degree in Computer Science and Information Technology at the Budapest University of Technology and Economics in 2011. He has been working as an Associate Professor for the Széchenyi István University in Győr since the February of 2012. He teaches system simulation and network security. His current research interests include modeling and simulation of infocommunication systems and cloud systems and services, and improving the prediction of performance of simulation.

ISTVÁN DERKA received his Msc at the Faculty of Electrical Engineering and Informatics at the Technical University of Budapest in 1995. He worked for the Department of Informatics from 1999 to 2003 and since then he has been working for the Department of Telecommunications of the Széchenyi István University in Győr. He teaches programming of communication systems and interactive TV systems. He is an Assistant Professor. The area of his research includes multicast routing protocols and IPTV services in large scale networks.

TESTING THE SPEED-UP OF PARALLEL DISCRETE EVENT SIMULATION IN HETEROGENEOUS EXECUTION ENVIRONMENTS

Gábor Lencse and István Derka
Department of Telecommunications
Széchenyi István University
Egyetem tér 1.
H-9026 Győr, Hungary
lencse@sze.hu

KEYWORDS

parallel discrete event simulation, heterogeneous execution environments, conservative synchronisation method, relative speed-up, MPI, OMNeT++, load balancing criterion, coupling factor criterion.

ABSTRACT

This paper deals with the experimental testing and verification of the earlier proposed load balancing and coupling factor criteria for the conservative parallel discrete event simulation in heterogeneous execution environments whether they can ensure a good speed-up. The definition of the relative speed-up is extended to the heterogeneous systems in a natural way. This definition is used to measure the efficiency of the parallel simulation executed by heterogeneous systems. A closed queueing network is used as the simulation model, and it is executed on different heterogeneous test systems. Among several scenarios, it is demonstrated on the example of a heterogeneous system containing 87 CPU cores of 5 different types that a good speed-up can be achieved using the load balancing and coupling factor criteria. It is shown that the extension of the definition of the relative speed-up to the heterogeneous systems made it easy to judge the speed-up of parallel discrete event simulation in heterogeneous execution environments.

INTRODUCTION

Event-driven ***discrete event simulation*** (DES) is a powerful method for the performance analysis of information and communication technology (ICT) systems. The detailed modelling and simulation of these systems often requires a huge amount of computing power and memory. Parallelisation can be a natural solution. Kunz (Kunz 2010) points out that as the ongoing development in the hardware sector favours an increasing number of processing units over an increasing speed of a single unit thus the parallel simulation will remain an important and active field of research.

However, because of the algorithm of the event-driven DES, ***parallel discrete event simulation*** (PDES) it is not an easy task and the achievable speed-up is often limited. When doing PDES, the model of the system is divided into partitions (called Logical Processes), and the partitions are assigned to processors that are executing them. To maintain causality, the virtual times of the partitions must be synchronised. There are different methods for synchronisation (Kunz 2010). The ***conservative method*** ensures that causality is never violated. An event can be executed only if we are certain that no events with smaller timestamp exist (and also will not be generated) anywhere in the model. Unless the simulated system has a special property that the so called *lookahead* is large enough, the processors executing the partitions need to wait for each other in the majority of time, so the achievable speed-up is poor.

In the paper (Varga et al. 2003), the authors proposed a method for assessing available parallelism in a simulation model for conservative synchronization. The method requires only a small number of parameters that can be easily measured on a sequential simulation. In our paper (Lencse and Varga 2010), we checked the results of the aforementioned work for homogeneous clusters up to 24 CPU cores and also examined how the different parameters of the model influence the achievable speed-up. In our next paper (Lencse et al. 2013) we examined the criteria for a good speed-up in a heterogeneous execution environment. Our criteria were justified by several measurements in a test system. However, we could not include all the planned experiments, due to space limitations. This paper presents our further results on the examinations of different factors that influence the achievable speed-up of parallel discrete event simulation in a heterogeneous execution environment. Moreover, the definition of the relative speed-up is extended to heterogeneous systems and this extension is used in the discussion of the results of our experiments to evaluate the efficiency of parallel simulation executed by heterogeneous systems.

The remainder of this paper is organised as follows: first, a brief summary of the method for assessing the available parallelism is given. Second, our concept of heterogeneous execution environment, our criteria for a good speed up and our previous results are summarized. Third, the definition of the relative speed-up is extended to the heterogeneous systems to be able to express the efficiency of parallel simulation executed by heterogeneous systems. Fourth, our heterogeneous test environment and simulation model are described. Fifth, our further experiments and result are presented and discussed. Finally, our paper is concluded.

This topic was identified as being of importance in the parallel simulation of large systems using heterogeneous execution environments.

THE METHOD FOR ASSESSING AVAILABLE PARALLELISM

The available parallelism can be assessed using some quantities that can be measured during a sequential simulation of the model in question. The following description is taken from (Lencse and Varga 2010).

The paper (Varga et. al. 2003) uses the notations *ev* for events, *sec* for real world time in seconds and *simsec* for simulated time (model time) in seconds. The paper uses the following quantities for the assessing of available parallelism:

- ***P*** *performance* represents the number of events processed per second (*ev*/*sec*).
- ***E*** *event density* is the number of events that occur per simulated second (*ev*/*simsec*).
- ***L*** *lookahead* is measured in simulated seconds (*simsec*).
- ***τ*** *latency* (sec) is the latency of sending a message from one Logical Process (LP) to another.
- ***λ*** *coupling factor* can be calculated as the ratio of *LE* and *τP:*

$$\lambda = \frac{L \cdot E}{\tau \cdot P} \qquad (1)$$

In (Lencse and Varga 2010) we have shown that if λ is in the order of several hundreds or higher then we may expect a good speed-up. It may be nearly linear even for higher number of segments (*N*) if λ_N is also at least in the order of several hundreds, where:

$$\lambda_N = \frac{\lambda}{N} \qquad (2)$$

MODELLING AND SIMULATION IN HETEROGENEOUS EXECUTION ENVIRONMENTS

This chapter is a summary of (Lencse et al 2013).

Our Concept of Heterogeneous Execution Environments

We recommended a logical topology of two levels: a *star shaped network* of *homogeneous clusters*. This model is simple enough and can describe a typical *heterogeneous execution environment*. What is logically described as a homogeneous cluster, it can be physically, for example, a cluster of PCs with identical configuration interconnected by a switch or it can be a chassis based computer built up by several main boards, etc. The main point is that a homogeneous cluster is built up by identical configuration elements especially concerning CPU type and speed as well as memory size and speed. The homogeneous clusters are interconnected logically in a star shaped topology. The physical connection can be a switch or the topology may be different but our model considers it to be a star for simplicity.

Criteria for Achieving a Good Speed-up

We set up two criteria. The ***load balancing criterion*** requires that all the CPUs (or CPU cores) should get a *fair share* from the execution of the simulation. A fair share is proportional to the computing power of the CPU *concerning the execution of the given simulation model*. (This is very important, because, for example, using different benchmark programs for the same set of computers one can get seriously different performance results.) Thus, for the fair division of a given simulation model among the CPUs, the CPUs should be benchmarked by the same type of simulation model that is to be executed by them (but smaller in size, of course). The *lookahead* or ***coupling factor criterion*** is the same as presented and tested in (Lencse and Varga 2010) up to 24 CPU cores.

The Most Important Results

The load balancing criterion was justified by measuring the execution time of a model with different partitioning. The results of our experiments were quite close to the values computed according to the load balancing criterion. (See more details later.) The coupling factor criterion was justified by different scenarios including a simulation executed by 64 CPU cores of 4 types resulting in a good speed-up.

EFFICIENCY OF PARALLEL SIMULATION EXECUTED BY HETEROGENEOUS SYSTEMS

Relative Speed-up of Program Execution by Heterogeneous Systems

First, the definition of the relative speed-up of parallel execution of programs is extended for heterogeneous systems (in general, not only for simulation).

The conventional definition of the *speed-up* (s_n) of parallel execution is the ratio of the speed of the parallel execution by *n* CPUs and the sequential execution by 1 CPU that is equal with the ratio of the execution time of the sequential execution (T_1) and that of the parallel execution (T_n):

$$s_n = \frac{T_1}{T_n} \qquad (3)$$

The *relative speed-up* (r_n) can be calculated as the ratio of the speed-up and the number of the CPUs that produced the given speed-up:

$$r_n = \frac{s_n}{n} \qquad (4)$$

The relative speed-up measures the *efficiency* of parallel execution. A relative speed-up value of 1 means that the

speed-up is linear that is the computing power of n CPUs can be fully utilized.

When dealing with heterogeneous systems, not only the number of the CPUs but also their performance is to be taken into consideration. We were looking for a definition of the relative speed-up of heterogeneous systems that can be used to measure the efficiency of program execution by the heterogeneous systems in the same way: its value of one should mean that the computing power of all the CPUs (from different types) can be fully utilized.

Let us denote the *number of the CPU types* in a heterogeneous system by *NT*, the *number* and the *performance* of CPUs available from *type i* by N_i and P_i, respectively. The *cumulative performance* of the heterogeneous system is:

$$P_c = \sum_{j=1}^{NT} P_i \cdot N_i \qquad (5)$$

Let us denote the *execution time* of a given program by a single CPU of *type i* by T_i and let T_h denote the *execution time* of the given program by the *heterogeneous system*. The speed-up of the heterogeneous system compared to the sequential execution by one CPU of type *i* is:

$$s_{h/i} = \frac{T_i}{T_h} \qquad (6)$$

The relative speed-up of the heterogeneous system against the sequential execution by one CPU of type *i* is now defined as:

$$r_{h/i} = \frac{T_i \cdot P_i}{T_h \cdot P_c} \qquad (7)$$

We believe that this definition can be used in general for measuring the efficiency of program execution by heterogeneous systems. Thus, for simplicity, we used the word "CPU" in this section, but the expression "CPU core" could be used instead, as it is used everywhere else in this paper.

Measuring the Efficiency of Parallel Simulation Executed by Heterogeneous Systems

If the above definition of relative speed-up of program execution by heterogeneous systems is used for measuring the efficiency of parallel simulation executed by heterogeneous systems and the performance values of the CPU cores from different types are measured by benchmarking them with the same simulation model (expressing its value in events per seconds and the value of execution time in seconds) then the numerator of expression (7) gives the same values for all the values of *i* (that is its value is independent of the CPU core types): it is equal with the total number of events in the sequential simulation.

Note that the number of events in the parallel version of the simulation may be higher (e.g. due to communication and synchronization overhead) but only the events of the original sequential simulation are the essential part of the operation of the original model. Thus efficiency should consider the events of the original sequential simulation only.

Denoting the total *number of events* in the sequential simulation by N_E, definition (7) can be rewritten as:

$$r_h = \frac{N_E}{T_h \cdot P_c} \qquad (8)$$

Thus, we have shown that the value of relative speed-up calculated by (8) does not depend on which CPU core it was calculated against, it characterises the parallel simulation itself. Note that this is still true if one uses definition (7), thus one can use any of them selecting on the basis of which values are easier to measure directly in the given simulation. We will do so when calculating the relative speed-up for measuring the efficiency of simulation in the different experiments presented in this paper.

HETEROGENEOUS TEST ENVIRONMENT

Available Hardware Base

The following servers, workstations and PCs were available for our experiments at the Info-communications Laboratory of the Department of Telecommunications, Széchenyi István University. Note that this hardware base is somewhat larger than that of our previous paper (Lencse et al. 2013).

One Sun Server SunFire X4150
Two Quad Core Intel Xeon 2.83GHz CPU, 8GB DDR2 800MHz RAM, 160GB HDD, Gigabit Ethernet NICs
Altogether it means a homogeneous cluster of 8 nodes.

Three LS120 Blades form an IBM BladeCenter
Two Dual Core Opteron 280 2,4GHz CPU, 4GB DDR2 667MHz RAM, 73GB HDD, Gigabit Ethernet NIC
Altogether it means a homogeneous cluster of 12 nodes.

One Itanium Server HP Server RX2600
Two Intel Mckinley IA-64 900MHz CPU, 8GB RAM DDR 266MHz RAM, 36.4GB HDD, Gigabit Ethernet NIC
Although it has no serious computing power but it was found interesting because of its non-x86 architecture CPU.

Eleven Dell Precision 490 Workstations
Two Intel Xeon 5130 Dual Core 3 GHz CPU, 6GB DDR2 667MHz RAM, 80GB HDD, Gigabit Ethernet NICs
Altogether it means a homogeneous cluster of 44 nodes.

Eleven AMD PCs
AMD Athlon 64 X2 Dual Core 4200+ 2200MHz CPU, 2GB DDR2-667 MHz RAM, 320 GB HDD, Gb. Eth. NIC
Altogether it means a homogeneous cluster of 22 nodes.

Four Old Intel PCs (P4)
Intel Pentium 4 HT 3GHz CPU, 512 DDR-400 MHz RAM, 80 GB HDD, Fast Ethernet NIC.
Note that they use 32 bits CPUs.

Switches for Interconnection
- 3Com Baseline Switch 2948 SFP Plus (3CBLSG48)
- Cisco Intelligent Gigabit Ethernet Switch Module, 4 ports (Part Number 32R1894) in the BladeCenter
- D-Link EasySmart Switch DGS-1100-24

Software Environment

Operating Systems
Linux was used on all the computers. Sun Server and LS21 Blades: Ubuntu 12.04 LTS x86-64; Dell Precision 490 Workstations and AMD PCs: Debian Squeeze (x86_64); old Intel PC-s: Debian Squeeze (i386);

Cluster Software
OpenMPI 1.6.2 (x86_64 if not stated otherwise; i386 in some cases)

Discrete-Event Simulation Software
The widely used, open source OMNeT++ 4.2.2 discrete-event simulation environment (Varga and Hornig 2008) was chosen. It supports the conservative synchronization method (the Null Message Algorithm) since 2003 (Sekercioglu et al. 2003). We also expect that because of the modularity, extensibility and clean internal architecture of the parallel simulation subsystem, the OMNeT++ framework has the potential to become a preferred platform for PDES research.

The Simulation Model

Bagrodia and Takai proposed the Closed Queueing Network for testing the performance of conservative parallel discrete-event simulation (Bagrodia and Takai 2000). OMNeT++ has a CQN implementation among its simulation sample programs. We have found this model perfect for our purposes thus it was used in our current paper as well as in our previous ones (Lencse and Varga 2010) and (Lencse et al. 2013). The below description of the model is taken from (Lencse and Varga 2010).

This model consists of M tandem queues where each tandem consists of a switch and k single-server queues with exponential service times (Figure 1, left).

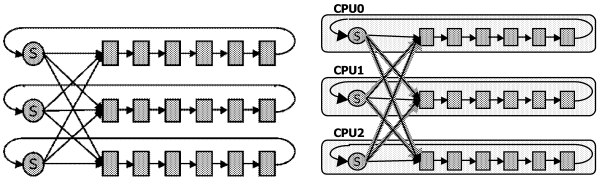


Figure 1. M=3 Tandem Queues with k=6 Single Server Queues in Each Tandem Queue – And its Partitioning

The last queues are looped back to their switches. Each switch randomly chooses the first queue of one of the tandems as destination, using uniform distribution. The queues and switches are connected with links that have nonzero propagation delays. The OMNeT++ model for CQN wraps tandems into compound modules.

To run the model in parallel, we assign tandems to different LPs (Figure 1, right). Lookahead is provided by delays on the marked links.

As for the parameters of the model, the preset values shipped with the model were used unless it is stated otherwise. Configuration B was chosen, the one that promised good speed-up. The main parameters of the CQN model were: M=24 tandem queues, k=50 queues in each tandem queue, exponential service time of the queues with expected value of 10 seconds, the delay between the tandem queues L=100 seconds and the length of the simulation was 10^6 seconds (in model time).

EXPERIMENTS AND RESULTS

Figure 2 shows the interconnection of the elements of our heterogeneous environment. Note that our experiments used different subsets of the elements.

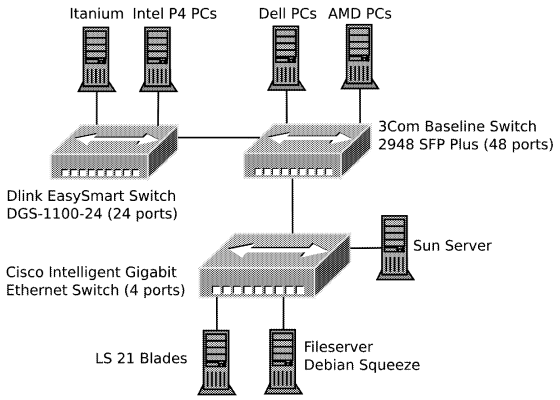


Figure 2. The elements of the Heterogeneous Execution Environment with their Interconnections

Further Validation of the Load Balancing Criterion

The load balancing criterion requires the benchmarking of the different CPU cores with the simulation model. The benchmarking was done using the CQN OMNeT++ sample model. All the experiments were performed 11 times and average and standard deviation was calculated. Unless stated otherwise, it was done so with all the following experiments, too. Table 1 shows the performance of the different CPU core types.

Table 1. The Performance of the Different 64-bit CPU Core Types (events/second)

Core Type	Sun	IBM	Dell	AMD	Itanium
Average	468 192	238 174	373 787	235 861	61 509
Std. Dev.	5 581	6 049	4 908	2 726	117

The load balancing was tested on the example of a minimal size heterogeneous system built up by one Sun core and one IBM core in (Lencse et al. 2013). Now it is validated on a similar system built up by one Dell core and one AMD core. The following series of experiments were performed: the CQN model built up by 24 tandem queues was cut into two segments: N and 24-N tandem queues were put into the segment executed by the AMD core and the Dell core, respectively, where N took its values form 1 to 23. The execution time was measured in all cases and the value of the relative speed up was calculated according to (8). Figure 3 shows the results. It can be seen that the partitioning was the best when 10 and 14 tandem queues were put into the segments executed by the AMD and the Dell computers, respectively. The exactly *performance proportional partitioning* would result in the assignment of 9.29 and 14.71 tandem queues. Thus the results of our experiments are in a good agreement with the computed "optimal partitioning".

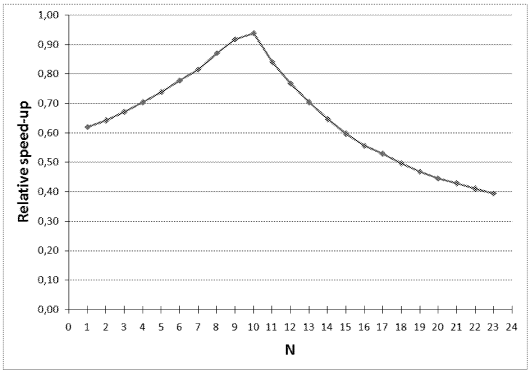


Figure 3. The Relative Speed-up of the Execution of the CQN Model in the Function of the Partitioning

A Test Including the Itanium Server

As the performance of the Itanium server is much less then that of all the others, it seems to be a reasonable question if it is worth using it at all? The following two experiments were conducted:

1. A cluster of one IBM and one AMD cores was used and 12 tandem queues were put on each of them.

2. The Itanium server was added to the above cluster. The number of the tandem queues was decreased by one both on the IBM and on the AMD cores and the two tandem queues were assigned to the Itanium server.

Note that the above partitioning in both cases was the best possible approximation of the performance proportional one.

Table 2 shows the parameters and the results of the experiments. The use of the Itanium server resulted in a 9.45% speed-up according to the conventional interpretation of speed-up. Is it a good result? To be able to judge the efficiency of the parallel simulation in the second experiment we calculated the relative speed-up according to (7) using the results of the first experiment as reference. We are really satisfied with the 0.9688 relative speed-up.

Table 2. The Parameters and the Results of the Experiments for the Itanium Server

Elements of the System	IBM+AMD	IBM+AMD+Itanium
Cumulated Performance (ev/sec)	474 035	535 544
Average Execution Time (sec)	335.97	306.96
Std. Dev. of Exec. Time	2.43	3.17
Speed-Up (Conventional)	(reference)	1.0945
Relative Speed-up	(reference)	0.9688

A Test Including the Old 32-bit Intel PC-s

Unfortunately, it is a feature of the OpenMPI, that if 32-bit libraries are used in one of the computers then they should be used on all the other ones so that the computers can communicate with each other. Thus we had to recompile everything in 32-bit mode for the 64 bit computers and benchmark them again in 32-bit mode. Table 2 shows the performance of the different CPU core types in 32-bit modes. The configuration script of OMNeT++ did not find the 32-bit MPI libraries under Debian, thus we could not test the Dell and AMD platforms. As we could not fix the problem due to lack of time, they were omitted from the 32-bit experiments.

Table 3. The Performance of the Different CPU Core Types in 32-bit mode (events/second)

Core Type	Sun	IBM	*P4*
Average	593 340	334 516	218 014
Std. Dev.	9 989	7 467	1 362

Note that the performance results of the Sun and IBM servers are significantly higher in 32-bit mode than in 64-bit mode. It is probably due to the fact that the integers and the pointers use twice as much memory space in 64-bit mode than in 32-bit mode and the longer programs can be less effectively cached.

The following two experiments were conducted:

1. The 8 cores of the Sun server and the 4 cores of one IBM Blade were used.

2. The four old Intel P4 computers were added to the above system.

The number of the tandem queues was increased in the CQN model from 24 to 240 to be able to utilize all the CPU cores. The value of the lookahead was increased from 100s to 1000s see its justification in (Lencse et al. 2013). The tandem queues were divided as close to the performance proportional partitioning as it was possible.

According to (Lencse et al. 2013), if the number of the CPU core types is denoted by NCT, the number and the performance of the CPU cores available from core type *i*

are denoted by N_i and P_i, respectively then the n_i number of the queues to be put into a segment executed by a core from type i should be:

$$n_i = \frac{240 \cdot P_i}{\sum_{j=1}^{NCT} P_i \cdot N_i} \qquad (9)$$

However, the number of the tandem queues per segments must be an integer, thus the division of the tandems could not be fully precise, some "roundings" and adjustments were done manually and there were differences made even between the load of the cores from the same core type so that the number all the tandems be exactly 240. Table 4 and 5 show the division of the tandems among the cores for the first and the second experiments.

Table 4. The Division of the 240 Tandem Queues among the Sun and IBM cores

Core type	P_i (ev/sec)	N_i	n_i	no. of cores	tandems /core	cumulated tandems
Sun	593 340	8	23.40	4 4	23 24	92 96
IBM	334 516	4	13.19	4	13	52
No. of all the cores:		12	Number of all the tandems:			240

Table 5. The Division of the 240 Tandem Queues among the Sun, IBM and P4 cores

Core type	P_i (ev/sec)	N_i	n_i	no. of cores	tandems /core	cumulated tandems
Sun	593 340	8	20.47	8	20	160
IBM	334 516	4	11.54	4	12	48
P4	218 014	4	7.52	4	8	32
No. of all the cores:		16	Number of all the tandems:			240

Table 6 shows the parameters and the results of the experiments. The use of the four old Intel P4 PCs resulted in a 9.47% speed-up according to the conventional interpretation of speed-up. Again, as with the Itanium server, we calculated the relative speed-up according to (7) using the results of the first experiment as reference to be able to judge if the efficiency of the parallel simulation in the second experiment compared to the first one. We can be satisfied with the 0.9575 relative speed-up and mention just for comparison that even the speed-up caused be the P4 PCs was a little bit higher than the speed-up caused be the Itanium server, this a is bit less good result, as the efficiency is somewhat smaller now.

Table 6. The Parameters and the Results of the Experiments for the Old Intel P4 PCs

Elements	Sun+IBM	Sun+IBM+4xP4
Cumulated Performance (ev/sec)	6 084 784	6 956 480
Average Execution Time (sec)	1 327.31	1 212.51
Std. Dev. of Exec. Time	34.91	47.15
Speed-Up (Conventional)	(reference)	1.0947
Relative Speed-up	(reference)	0.9575

Note that we calculated the *relative efficiency* in both cases because we tested if it was worth using the Itanium server or the old Intel P4 PCs. It will be done differently in the test of the largest available system.

A Test of the Largest Available System

The maximum possible number of 64-bit cores were included into the heterogeneous system. The number of the tandem queues and the lookahead were increased to 480 and 2000s, respectively. First, the partitioning was done according to the old performance values that were measured during the simulation of the system built up by 24 tandem queues and using 100s lookahead. Table 7. shows the partitioning.

Table 7. The Division of the 480 Tandem Queues Among the 64-bit Cores Using the Old P_i Values from Table 1.

Core type	P_i (ev/sec)	N_i	n_i	no. of cores	tandems /core	cumulated tandems
Sun	468 192	8	7.94	8	8	64
IBM	238 174	12	4.04	12	4	48
Dell	373 787	44	6.34	29 15	6 7	174 105
AMD	235 861	22	4.00	22	4	88
Itanium	61 509	1	1.04	1	1	1
No. of all the cores:		87	Number of all the tandems:			480

The execution time of the parallel simulation was 159.80 seconds with 2.84 standard deviation. To be able to determine its efficiency, we needed either a sequential execution time produced by any of the CPU cores or the total number of events in the sequential simulation. As the model was large we executed the sequential simulation only for a 10^5 seconds long model time interval that is only one tenth of the 10^6 model time simulated by the parallel execution. (The execution time and event number values for 10^6 seconds model time sequential simulation can be easily extrapolated the by multiplying the results with 10.) We executed the sequential simulation on all the CPU types 11 times except for Itanium, where it was executed only once. The number of all the events was approximately 170.34 million in all cases (that means 1,703.4 million events for 10^6 model time seconds). The execution time values are shown in table 8. This table also shows the speed-up values (calculated by multiplying the sequential execution time values by 10) against the sequential execution by the given CPU core types. The next line of the table shows the recalculated P_i performance values. They are much smaller than those in table 1 due to the effect called "vacationing jobs" in (Lencse and Varga 2010). Shortly summarized the effect: the long delay lines are "storing" some jobs (events) that are missing from the elementary queues of the tandem queues thus the number of events per real-time seconds is significantly decreased. The new P_i values are not even proportional with the old ones. The next line of the table shows the quotient of the new and the old P_i values for the different CPU core types: they are really different. (For this reason, we tested also another

partitioning using the new P_i values.) For computing the relative speed-up, the cumulative performance was computed from the new P_i values and we got: P_c=17 929 579 ev/sec.

The relative speed-up was calculated according to (8) as:

$$r_h = \frac{1.7034 \cdot 10^9 ev}{159.8\,sec \cdot 17929579 ev/sec} \approx 0.5945 \qquad (10)$$

This value is not at all bad for such a big and inhomogeneous cluster.

Table 8. Execution Time of the 10^5s (model time) Long Sequential Simulation and Other Derived Values

Core Type	Sun	IBM	Dell	AMD	Itanium
Avg. Ex. time (s)	593.20	960.03	778.65	972.69	6837.44
Std. Dev. of E. T.	5.38	31.40	3.87	7.49	-
Speed-up	37.1	60.1	38.7	60.9	427.9
new P_i (ev/sec)	287 154	177 432	218 763	175 122	24 907
new P_i / old P_i	0.6133	0.7450	0.5853	0.7425	0.4049

The system was repartitioned according to the new P_i values, see table 9. Note that even though the number of available cores was 87, only 86 of them were actually used as no tandem queues were put to the Itanium server.

Table 9. The Division of the 480 Tandem Queues Among the 64-bit Cores Using the New P_i Values from Table 8.

Core type	P_i (ev/sec)	N_i	n_i	no. of cores	tandems /core	cumulated tandems
Sun	287 154	8	7.69	8	8	64
IBM	177 432	12	4.75	12	5	60
Dell	218 763	44	5.85	44	6	264
AMD	175 122	22	4.68	18	4	72
				4	5	20
Itanium	24 907	1	0.67	1	0	0
No. of all the cores:		87	Number of all the tandems:			480

The execution of the simulation gave excellent results. The average execution time was 108.12 seconds with 2.34 standard deviation. The speed-up values calculated against the four actually used CPU core types are shown in table 10.

Table 10. Speed-up of the Second Heterogeneous System

Core Type	Sun	IBM	Dell	AMD
Speed-up	54.9	88.8	72.0	90.0

As the Itanium server was not used, its performance was left out from the cumulative performance: P_c=17 904 672 ev/sec. For relative speed-up, we got:

$$r_h = \frac{1.7034 \cdot 10^9 ev}{108.12\,sec \cdot 17904672 ev/sec} \approx 0.8799 \qquad (11)$$

This is excellent for such a big and inhomogeneous cluster.

CONCLUSIONS

Some of our earlier defined criteria (load balancing and coupling factor) for the possible good speed-up of parallel discrete event simulation in heterogeneous execution environments were summarized. A novel extension of the relative speed-up for heterogeneous systems was introduced.

The operation of the criteria was justified by several experiments with different size and type of heterogeneous systems.

The extension of the definition of the relative speed-up to heterogeneous systems proved to be an appropriate tool for the evaluation of the speed-up of parallel discrete event simulation in heterogeneous execution environments.

We conclude that the recommended methods are worth using and further studying.

ACKNOWLEDGEMENT

This research was supported by the TÁMOP-4.2.2/B-10/1-2010-0010 project and by the Széchenyi István University (15-3202-08).

REFERENCES

Bagrodia, R. L. and M. Takai. 2000. "Performance evaluation of conservative algorithms in parallel simulation languages" *IEEE Transactions on Parallel and Distributed Systems*, Vol. 11. No 4. 395-411.

Kunz, G. 2010. "Parallel Discrete Event Simulation" In *Modeling and Tools for Network Simulation* K. Wehrle, M. Günes and J. Gross (Eds.). Springer-Verlag, Berlin 2010. ISBN 978-3-642-12330-6

Lencse, G. and A. Varga. 2010. "Performance Prediction of Conservative Parallel Discrete Event Simulation". *Proceedings of the 2010 Industrial Simulation Conference (ISC'2010)* (Budapest, Hungary, 2010. June 7-9.) EUROSIS-ETI, 214-219.

Lencse, G., I. Derka and L. Muka. 2013. "Towards the efficient simulation of telecommunication systems in heterogeneous execution environments", accepted for the *36th International Conference on Telecommunications and Signal Processing (TSP), July 2-4, 2013, Rome, Italy.*

Sekercioglu, Y. A., Varga, A. and Egan, G. K. Egan. 2003. "Parallel Simulation Made Easy with OMNeT++". *Proceedings of the European Simulation Symposium (ESS 2003)*, Oct. 26-29, 2003, Delft, The Netherlands.

Varga, A., Y. A. Sekercioglu and G. K. Egan. 2003. "A practical efficiency criterion for the null message algorithm". *Proceedings of the European Simulation Symposium (ESS 2003)*, (Oct. 26-29, 2003, Delft, The Netherlands.) SCS International, 81-92.

Varga, A. and Hornig, R. 2008. "An overview of the OMNeT++ simulation environment", *Simutools '08: Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*. March 7, 2008, Marseille, France.

ALGORITHMS FOR ACCELERATED RARE EVENT SIMULATION WITH MARKOV CHAIN MODELLING IN COMMUNICATION NETWORKS

Izabella Lokshina
SUNY Oneonta
324B Netzer Administration Bldg.
Oneonta, NY 13820, USA
E-mail: Izabella.Lokshina@oneonta.edu

Evan Schiele
SAP America, Inc
89 Headquarters Plz,
Morristown, NJ 07960, USA
E-mail: evan_schiele@hotmail.com

KEYWORDS
Communication Networks, Queuing Systems, Rare Event Simulation, Markov Chains, Buffer Overflow Probability.

ABSTRACT

The paper recommends an effective approach to estimate the probability of buffer overflow in communication networks. The buffer overflow probability in queuing systems is defined as a rare event and can be estimated using rare event simulation with Markov chains. Two-node queuing networks are considered in this paper; and an event of buffer overflow at the second node is studied. Probability of a rare event that the content of the second buffer would exceed some high level L, starting from a certain state, is analyzed. The approach is based on the Markov additive representation of the buffer processes, leading to exponential change of measure, which is used in an importance sampling method.
The examples, considered in this paper, confirm that when the first buffer is finite, the relative error is bound independent of some high level L. However, when the first buffer is infinite, a natural extension of exponential change of measure for finite buffer case is proposed. The relative error is shown to be bound independent of L only when at the second node there is a bottleneck, i.e. buffer overflow may occur. However, when at the first node there is a bottleneck; experimental results confirm that the relative error is linearly bound to the level L.
Two efficient rare event simulation algorithms, based on the importance sampling and cross-entropy methods, are developed and applied to accelerate the overflow probability simulation with Markov chains in communication networks. Numerical examples and simulation results are provided.

INTRODUCTION

Rare events occur in communication networks when dealing with reliability and performance evaluation. Frequency of rare events is very small, e.g. 10^{-6} or less; however, rare event probability can be used to simulate, estimate and analyze many system characteristics. Queuing systems are appropriate reference models being used in different methodologies and techniques to accelerate rare event simulation in communications networks.

Estimation of rare event probability using Monte Carlo simulation requires a very long computing time and cannot easily be implemented (Bolch et al., 1998; Bucklew, 2004; Rubino & Tuffin, 2009).
Lately two basic methods of the rare event simulation were developed based on cross-entropy the approach that can be applied to a wide range of optimization tasks (Heidelberger, 1995; Kalashnikov, 1997; De Boer, 2002):

- Splitting of the sample path that for to reach definite intermediate level between the starting level and rare event (C´erou, 2005); and
- Importance Sampling (IS) generation (Bueno et al., 2000).

The probability density function is used in the IS as the rare event evaluation measure, which can be compared and changed based on the likelihood ratio of the less rare event probability density function.
One of the rare event simulation objectives is estimation of total network population overflow. Exact large deviation analysis leading to asymptotically efficient change of measure is rather difficult. Instead, a heuristic change of measure is proposed, which interchanges the arrival rate to the first queue and the slowest service rate.
A similar change of measure is suggested based on time reversal arguments. However, analysis shows that the IS estimator based on this change of measure is not necessarily asymptotically efficient. In fact, it has an infinite variance in some parameter regions (Kalashnikov, 1997).
Another rare event simulation objective is the estimation of buffer overflow that observed at individual network node. If the node of concern is a bottleneck relative to all preceding nodes, then asymptotically efficient exponential change of measure can be obtained by interchanging the arrival rate and the service rate at this particular node; and the service rates at all other nodes are kept unchanged (Kroese & Nicola, 1999; Keith & Kroese, 2002).
However, this change of measure is not asymptotically efficient when the overflow of buffer at the consequent node is considered. Effective bandwidth theory is used to derive heuristics for efficient feed-forward discrete-time queuing network simulation. This class of networks essentially resembles feed-forward fluid-flow networks.
Discrete-Time Markov Chain (DTMC) with its regular structure is a highly efficient model used for performance evaluation on the queuing system. On the one hand, the states are easily arranged as a grid in DTMC (with as

many dimensions as the number of queues). While on the other hand, any transition in DTMC corresponds to a particular elementary event at one of the queues (e.g., arrival or service completion). These events are known as transition events, and they are defined separately from the states; i.e., there is only one transition event for a service completion at a given queue, and this particular transition event corresponds to a transition out of each state in the DTMC while this particular queue is non-empty (Radev & Lokshina, 2006a; Radev & Lokshina, 2006b; Lokshina, 2011; Lokshina & Bartolacci, 2013).

However, not all transition events are enabled in every state. For example, the service completion event of the particular queue is not possible in a state where the queue is empty.

The analogous approach to continuous-time queuing networks has not yet been introduced even for Markov chains. Two-node queuing networks are considered in this paper; and the event of buffer overflow at the second node is studied.

We propose a new effective approach based on Markov Additive Continuous-Time Processes (MAP) modeling. We develop and apply the Importance sampling procedure with exponential tilting of the unbiased PDF estimation to the appropriate MAP representation, which lets us estimate the buffer overflow probability at the second node.

Unlike several heuristic changes of measure described in literature, the derived change of measure depends on the content of the first buffer. When the first buffer is finite, we confirm that the proposed simulation procedure yields the estimation with a relative error that is bound independent of the buffer overflow level. This result is much stronger than the asymptotic efficiency, which cannot be observed with other known methods.

When the first buffer is infinite, we propose a natural extension of change of measure for a finite buffer case. Applying the orthogonal polynomial model we obtain two types of simulation behavior. When the second buffer is a bottleneck, we confirm once more that simulation yields the estimation with a relative error that is bound independent of the buffer overflow level. However, when the first buffer is a bottleneck, the simulation results confirm that the relative error is asymptotic and linearly bound to the buffer overflow level.

MARKOV ADDITIVE CONTINUOUS TIME PROCESS REPRESENTATION

Markov additive continuous time process is a stochastic process (J_t, Z_t), where (J_t) is a Markov chain with the denumerable state space, and (Z_t) has stationary and independent increments during the time intervals when (J_t) is in any given state. That is, if given J_t has not changed in the interval (t_1, t_2), then for any $t_1<s_1<\ldots<s_n< t_2$, the increments $Z_{s_2}-Z_{s_1},\ldots,Z_{s_n}-Z_{s_{n-1}}$ are mutually independent, and the total increment during the interval $[t_1, t_2]$ depends on t_1 and t_2 only through the difference $t_2- t_1$.

Moreover, the transition of (J_t) from i to j has a certain probability (depending only on i and j) of triggering the transition of (Z_t) at the same time. The size of the transition in the process (Z_t) has a fixed distribution, which depends only on i and j.

The Markov additive continuous-time process (J_t, Z_t) is characterized by the family of the matrices $(M_t(s),t\geq 0)$, where (i,j)-th element of $M_t(s)$ can be defined as (1),

$$[M_t(s)]_{ij} = \mathrm{E}_i[e^{s(Z_t-Z_0)}I_{\{J_t=j\}}] \quad (1)$$

where E_i denotes the expectation operator given to the initial MAP state $J_0=i$.

Let us notice that $M_t(\cdot)$ is a generalization of the moment generating function for the ordinary random variables, as shown in (2).

$$\begin{aligned}\mathrm{E}_i[e^{s(Z_{t+h}-Z_0)}I_{\{J_{t+h}=j\}}] &= \\ =\sum_k \mathrm{E}_i[e^{s(Z_{t+h}-Z_0)}I_{\{J_t=k\}}I_{\{J_{t+h}=j\}}] &\\ =\sum_k \mathrm{E}_i[e^{s(Z_t-Z_0)}I_{\{J_t=k\}}]\mathrm{E}_i[e^{s(Z_{t+h}-Z_t)}I_{\{J_{t+h}=j\}}\,|\,J_t=k] &\\ =\sum_k [M_t(\mathrm{s})]_{ik}\,\mathrm{E}_k[e^{s(Z_h-Z_0)}I_{\{J_h=j\}}] &\end{aligned} \quad (2)$$

Consequently, if for all k and j (3) can be defined,

$$[A(s)]_{kj} := \lim_{h\downarrow 0}\frac{1}{h}\mathrm{E}_k[e^{s(Z_h-Z_0)}I_{\{J_h=j\}}-\delta_{kj}] \quad (3)$$

where δ is usual notation of Dirac, then (4) can easily be obtained,

$$\frac{d}{dt}M_t(s) = M_t(s)A(s), \qquad t\geq 0 \quad (4)$$

with $M_0(s)=I$ (the identity matrix). It is true as soon as (5) is true.

$$M_t(s) = e^{tA(s)}, \qquad t\geq 0 \quad (5)$$

The matrix $A(s)$ is known as the MAP (infinitesimal) generator.

Let us consider a simple Markov chain that consists of two queues in tandem. The calls arrive to the first queue (e.g., buffer) according to Poisson process with the rateλ. The departure time for the call in the first queue is exponentially distributed with the rate μ_1.

The calls that leave the first queue enter the second queue. The departure time in the second queue has an exponential distribution with the rate μ_2. The queuing network stability is assumed, i.e. $\lambda<\min\{\mu_1, \mu_2\}$.

The size of the first buffer may be finite or infinite; in fact, we will consider both cases. Let X_t and Y_t denote the number of the calls in the first and the second queues at the time t, respectively. Let P_i denote the probability measure under which (X_t) starts from i at the time 0 (i.e., $X_0=i$, $i\geq 0$); and let E_i denote the corresponding expectation operator.

Assuming that the second buffer is initially non-empty, say, $Y_0=1$, the probability that, starting from $(X_0,Y_0)=(i,1)$, content of the second buffer hits some high level $\mathrm{L}\in\mathbb{N}$ before hitting 0, can be estimated. This probability is denoted by γ_i and referred to it as the second buffer overflow probability, given that the initial number of calls in the first queue is i.

RARE EVENT SIMULATION

Let us consider the IS approach for the rare event simulation, where the PDF is used as the measure of rare event evaluation, which is compared and changed with the

likelihood ratio of the PDF of the less rare event (Lokshina, 2011).

First, we determine the rare event. Let $\mathbf{X} = (\mathbf{X}_1,..., \mathbf{X}_N)$ be a random vector, the values of which belong to the certain state space χ. Let $\{f(\cdot, \mathbf{v})\}$ be the family of the PDFs on χ, with respect to some base measure υ. Here $\mathbf{v}$ is a real-valued parameter (vector).

Then, for any measurable function H is obtained as (6).

$$\mathrm{E}[H(\mathbf{X})\int H(\mathbf{x})f(\mathbf{x};\mathbf{v})\upsilon(d\mathbf{x})] \quad (6)$$

In many cases f is often called the Probability Mass Function (PMF), but in this paper the generic term density, or the PDF, is used.

Let S be some real function on χ. The probability that $S(\mathbf{X})$ is greater than some real number γ, under $f(\cdot;\mathbf{u})$ can be defined. Therefore, probability can be written as (7),

$$l = \mathrm{P}_{\mathbf{u}}\left(S(\mathbf{X}) \geq \gamma\right) = \mathrm{E}_{\mathbf{u}}[I_{\{S(\mathbf{X})\geq\gamma\}}] \quad (7)$$

where $I_{\{S(\mathbf{X})\geq\gamma\}}$ is the indicator function.

If this probability is very small, like 10^{-6} or less, then $\{S(\mathbf{X}) \geq \gamma\}$ is a rare event. The simplest way to estimate l is to use the basic Monte Carlo simulation. Draw a random sample $\mathbf{X}_1,..., \mathbf{X}_N$ from $f(\cdot;\boldsymbol{u})$, and use (8) as the unbiased estimator of l.

$$\frac{1}{N}\sum_{i=1}^{N} I_{\{S(\mathbf{X}_i)\geq\gamma\}} \quad (8)$$

However, it poses serious problems when $\{S(\mathbf{X}) \geq \gamma\}$ is a rare event. In that case a large simulation effort is required in order to estimate l accurately.

An alternative is based on the IS. Take a random sample $\mathbf{X}_1,..., \mathbf{X}_N$ from the IS density g on χ, and evaluate l using the unbiased estimator, called the likelihood ratio estimator (9).

$$\hat{l} = \frac{1}{N}\sum_{i=1}^{N} I_{\{S(\mathbf{X}_i)\geq\gamma\}}\frac{f(\mathbf{X}_i;\mathbf{u})}{g(\mathbf{X}_i)} \quad (9)$$

It is well known that the optimal way to estimate l is to use the change of the measure with the density (10)

$$g^*(\mathbf{x}) = \frac{I_{\{S(\mathbf{x})\geq\gamma\}}f(\mathbf{x};\mathbf{u})}{l} \quad (10)$$

Specifically, applying this change of the measure to (9), (11) can be obtained for all i.

$$I_{\{S(\mathbf{X}_i)\geq\gamma\}}\frac{f(\mathbf{X}_i;\mathbf{u})}{g^*(\mathbf{X}_i)} = l \quad (11)$$

In other words, the estimator (9) has a zero variance, and only $N = 1$ sample needs to be produced. The obvious difficulty is, of course, that the g^* depends on the unknown parameter l. Moreover, one often wishes to choose this g from the density family $\{f(\cdot,\mathbf{v})\}$.

Now the plan is to choose the tilting parameter $\mathbf{v}$, such that the distance between the densities g^* and $f(\cdot,\mathbf{v})$ is minimal. A particular suitable measure of the distance between two densities g and f is the Kullback-Leibler distance, which is defined in (12).

$$D(g,f) = \mathrm{E}_g[\ln\frac{g(\mathbf{X})}{f(\mathbf{X})}] = \int g(\mathbf{x})\ln g(\mathbf{x})\upsilon(d\mathbf{x}) - \int g(\mathbf{x})\ln f(\mathbf{x})\upsilon(d\mathbf{x}) \quad (12)$$

Therefore, minimizing the Kullback-Leibler distance between g in (11) and $f(\cdot,\mathbf{v})$ is the same as choosing $\boldsymbol{v}$, such that $-\int g(\mathbf{x})\ln f(\mathbf{x};\mathbf{v})\upsilon(d\mathbf{x})$ is minimized, or equally, such that $\int g(\mathbf{x})\ln f(\mathbf{x};\mathbf{v})\upsilon(d\mathbf{x})$ is maximized. Formally, it can be written as (13).

$$\max_{\mathbf{v}} D(v) = \max_{\mathbf{v}}\int g(\mathbf{x})\ln f(\mathbf{x};\mathbf{v})\upsilon(d\mathbf{x}) \quad (13)$$

Applying g from (10) to (13) as the substitution, the following optimization program can be obtained as (14).

$$\max_{\mathbf{v}} D(\mathbf{v}) = \max_{\mathbf{v}}\int\frac{I_{\{S(\mathbf{x})\geq\gamma\}}f(\mathbf{x};\mathbf{u})}{l}\ln f(\mathbf{x};\mathbf{v})\upsilon(d\mathbf{x}) = \max_{\mathbf{v}}\mathrm{E}_{\mathbf{u}}[I_{\{S(\mathbf{X})\geq\gamma\}}\ln f(\mathbf{X};\mathbf{v})] \quad (14)$$

Using the IS again, with the change of the measure $f(\cdot;\boldsymbol{w})$, (14) can be re-written into (15),

$$\max_{\mathbf{v}} D(\mathbf{v}) = \max_{\mathbf{v}}\mathrm{E}_{\mathbf{w}}[I_{\{S(\mathbf{X})\geq\gamma\}}W(\mathbf{X};\mathbf{u},\mathbf{w})\ln f(\mathbf{X};\mathbf{v})] \quad (15)$$

for any tilting parameter $\mathbf{w}$, where the likelihood ratio at $\mathbf{x}$ between $f(\cdot\ ;\mathbf{u})$ and $f(\cdot;\mathbf{w})$ is W. This can be presented according to (16).

$$W(\mathbf{x};\mathbf{u},\mathbf{w}) = \frac{f(\mathbf{x};\mathbf{u})}{f(\mathbf{x};\mathbf{w})} \quad (16)$$

The basic idea of the IS method regarding the rare event is to accelerate its frequency with iterative tilting of the unbiased estimation of the PDF.

The acceleration of conditional probability of rare event X_1 with one-parameter function $f(x)$ is shown in Fig. 1. The conditional probability of rare event $P_f\{x>X_1\}$ is changing with conditional probability of $g(x) - P_g\{x>X_1\}$.

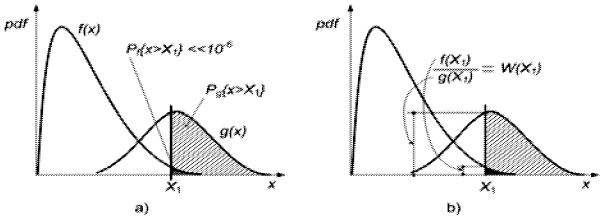


Figure 1: Conditional Probability of Rare Event $P_f\{x>X_1\}$ - (a) and Its Acceleration with Likelihood Ratio $W(X_1)$ - (b)

N independent samples, which distribution $g(x)$ can be evaluated with the likelihood ratio $W(X_1)$, are generated at each iteration of the IS simulation.

The optimal solution of (15) can be written as (17).

$$\mathbf{v}^* = \arg\max_{\mathbf{v}}\mathrm{E}_{\mathbf{w}}[I_{\{S(\mathbf{X})\geq\gamma\}}W(\mathbf{X};\mathbf{u},\mathbf{w})\ln f(\mathbf{X};\mathbf{v})] \quad (17)$$

It can be obtained by solving the following stochastic program, which can be considered as a stochastic counterpart of (15) and written according to (18),

$$\max_{\mathbf{v}}\hat{D}(\mathbf{v}) = \max_{\mathbf{v}}\frac{1}{N}\sum_{i=1}^{N} I_{\{S(\mathbf{X})\geq\gamma\}}W(\mathbf{X}_i;\mathbf{u},\mathbf{w})\ln f(\mathbf{X}_i;\mathbf{v}) \quad (18)$$

where $\mathbf{X}_1,..., \mathbf{X}_N$ is a random sample from $f(\cdot\ ;\mathbf{w})$.

The solution of (18) can be obtained by solving the following system of equations with respect to $\mathbf{v}$ in (19),

$$\frac{1}{N}\sum_{i=1}^{N} I_{\{S(\mathbf{X}_i)\geq\gamma\}}W(\mathbf{X}_i;\mathbf{u},\mathbf{w})\nabla\ln f(\mathbf{X}_i;\mathbf{v}) = \mathbf{0} \quad (19)$$

where the gradient is defined regarding $\mathbf{v}$. This confirms the expectation that the differentiation operators and the function $\hat{D}$ can be interchanged in (18) with respect to $\mathbf{v}$.

The advantage of this approach is in the fact that often the solution of (19) can be calculated analytically.
In particular, this happens when the random variable distributions belong to the natural exponential family. The cross-entropy program (19) is useful only when the probability of the target event $\{S(\mathbf{X}) \geq \gamma\}$ is not too small, say $l \geq 10^{-5}$. At that point, the above program is useful in terms of potentially more accurate estimator determination.
However, in the rare event context, (say, $l \geq 10^{-6}$), the program (19) is useless to the rarity of the events $\{S(\mathbf{X}_i) \geq \gamma\}$, because the random variables $I_{\{S(\mathbf{X}_i)\geq\gamma\}}$, i=1,…,N and the associated derivatives of $\hat{D}(\mathbf{v})$ vanish with high probability, as given in the right-hand side of (19), for reasonable sizes of N.

EXPONENTIAL CHANGE OF MEASURE

Let us first think that the first buffer has a finite capacity N. In this case the state space of the driving process (X_t) is finite in $\{0,\ldots,N\}$. Let us consider Markov additive process (X_t, Z_t).
To create the corresponding MAP generator (i.e., a matrix A(s) in (5)), the infinitesimal expectations $\mathrm{E}_i[e^{s(Z_h-Z_0)}I_{\{X_h=j\}}-\delta_{ij}]$ as $h\downarrow 0$, for all i, j in $\{0,\ldots,N\}$, have to be determined, where Z_0=1 and δ_{ij}=0 for $j\neq i$.
For instance, since the downward transition of (X_t) leads to the upward transition of (Z_t), (20) is used for i=1,…,N, as $h\downarrow 0$.

$$\begin{aligned}&\mathrm{E}_i[e^{s(Z_h-Z_0)}I_{\{X_h=i-1\}}-\delta_{i,i-1}]=\\&=\mathrm{E}_i[e^{s(Z_h-Z_0)}\mid X_h=i-1]\mathrm{P}_i(X_h=i-1)=\\&=e^s(\mu_1 h+o(h))=\mu_1 h e^s+o(h)\end{aligned} \tag{20}$$

Therefore, the (i,i-1)-th element of the matrix $A(s)$ exists and is equal to $\mu_1 e^s$. Other elements of the matrix $A(s)$ can be defined similarly. Consequently, (5) holds with $A(s)$ for Markov additive process (X_t, Z_t) with the given (N+1,N+1)-tri-diagonal matrix (21).

$$\mathbf{G}_N(s)=\begin{pmatrix}-\lambda-\mu_2+\mu_2 e^{-s} & \lambda & & \\ \mu_1 e^{s} & -\lambda-\mu_1-\mu_2+\mu_2 e^{-s} & \lambda & \\ & \ddots & \ddots & \ddots \\ & & \mu_1 e^{s} & -\mu_1-\mu_2+\mu_2 e^{-s}\end{pmatrix} \tag{21}$$

Let us note that the MAP generator $G_N(s)$ is now equal to the matrix (22).

$$\hat{\mathbf{Q}}^{(n)}(u)=\begin{pmatrix}-\lambda-\mu_2+\mu_2 u & \lambda & & \\ \mu_1/u & -\lambda-\mu_1-\mu_2+\mu_2 u & \lambda & \\ & \ddots & \ddots & \ddots \\ & & \mu_1/u & -\mu_1-\mu_2+\mu_2 e^{-s}\end{pmatrix} \tag{22}$$

Next, the change of the measure based on the family of the matrices $\mathbf{G}_N(s)$ is defined. For any $s\geq 0$, $k_N(\mathrm{s}) := \log(sp(M_t(s)))/t$ has to be defined, where $sp(M_t(s))$ denotes the spectral radius (or, the maximum eigenvalue) of $M_t(s)$.
Using (5) we identify $k_N(s)$ as the largest positive eigenvalue of $\mathbf{G}_N(s)$.

Let $\mathbf{w}(s)=\{w_k(s),0\leq k\leq N\}$ represent the correspondent right-eigenvector. When the first buffer has the infinite capacity, the process (X_t,Z_t) is still the Markov continuous-time additive process, but the state space of the Markov process (X_t) is now infinite.
Equation (5) is still used, but $A(s)$ is now given by the infinite-dimensional tri-diagonal matrix (23).

$$\mathbf{G}(s)=\begin{pmatrix}-\lambda-\mu_2+\mu_2 e^{-s} & \lambda & & \\ \mu_1 e^{s} & -\lambda-\mu_1-\mu_2+\mu_2 e^{-s} & \lambda & \\ & \ddots & \ddots & \ddots \\ & & & \end{pmatrix} \tag{23}$$

Let us note that the MAP generator $G(s)$ is now equal to the infinite tri-diagonal matrix (24).

$$\mathbf{Q}(u)=\begin{pmatrix}-\lambda-\mu_2+\mu_2 u & \lambda & & \\ \mu_1/u & -\lambda-\mu_1-\mu_2+\mu_2 u & \lambda & \\ & \ddots & \ddots & \ddots \\ & & & \end{pmatrix} \tag{24}$$

IMPORTANCE SAMPLING ALGORITHM

The IS algorithm for accelerated rare event simulation with Markov chain modeling in communication networks is proposed in this paper. This efficient algorithm consists of the following six key steps that are provided below:
Step 1. *Set t := 1 (initialization of iteration counter). Define the likelihood ratio v_1:=0 (in this case Monte Carlo simulation is appropriate).*
Step 2. *Generate a sample $\mathbf{X}_1,\ldots,\mathbf{X}_k,\ldots,\mathbf{X}_N$, from the density $f(\mathbf{X}_k;\mathbf{v}_{t-1})$ in such way that for the ρ-th part of samples (ρ=0.01) the condition of the rare event $S(\mathbf{X}_k)>M$ is executed.*
Step 3. *Determine the full paths and sort ascending in following way $S_{(1)}\leq S_{(2)}\leq\ldots\leq S_{(N)}$.*
Step 4. *Calculate conditional probability $\gamma_t = S_{[(1-\rho)N]}$.*
Step 5. *For each $S(\mathbf{X}_k)>\gamma$ define the rare event indicator $I_{\{S(\mathbf{X}_i)\geq M\}}$=1 and then determine the likelihood ratio for the next iteration $\mathbf{v}_{t+1}$ according to (15), (16) and (19).*
Step 6. *If $\gamma_t<M$ then t:=t+1 and repeat Steps 2, 3 and 4.*
The Markovian network that consists of tandem queues with a feedback is shown in Fig.2 and used as a simulation example, with the entry parameters follow: $\boldsymbol{\lambda_1}=\boldsymbol{\lambda_2}$=1, $\boldsymbol{\mu_1}=\boldsymbol{\mu_2}$=6, p=0.5. The simulation results for Poisson distribution and fixed number of calls n_1 and n_2, for four different overflow cases are shown in Table 1.
As could be seen in Table 1, the overflow probability exponentially decreases with an increase of the fixed number of calls in queues n_1 and n_2. The exponential behavior depends more on the number n_1.

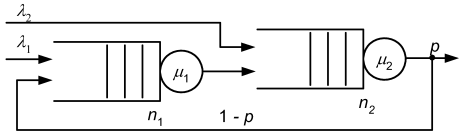


Figure 2: Two-node Queueing Network with Feedback

Table 1: Simulation Results

Overflow level		λ_1	λ_2	μ_1	μ_2	p	Overflow probability
n_1	n_2						
25	25	0.13	0.11	0.29	0.18	0.42	$2.22\text{x}10^{-8}$
50	50	0.23	0.18	2.15	2.26	0.31	$1.15\text{x}10^{-15}$
60	40	1.18	1.42	4.91	3.58	0.44	$6.29\text{x}10^{-25}$
100	40	2.02	1.67	5.82	2.94	0.37	$1.86\text{x}10^{-50}$

CROSS-ENTROPY ALGORITHM

The multi-level cross-entropy algorithm for accelerated rare event simulation with Markov chain modeling in communication networks is suggested in this paper. The idea is to introduce a sequence of the reference parameters $\{\mathbf{v}_t, t \geq 0\}$ and a sequence of the levels $\{\gamma_t, t \geq 1\}$, and iterate in both γ_t and $\mathbf{v}_t$. The initialization is done with choosing a not very smallρ, say $\rho = 10^{-2}$, and defining $\mathbf{v}_0 = \mathbf{u}$. Next, let γ_1 ($\gamma_1 < \gamma$) be such that under the original density $f(\mathbf{x}; \mathbf{u})$, the probability $l_1 = \mathrm{E}_{\mathbf{u}}[I_{\{S(\mathbf{x})\geq\gamma_1\}}]$ is at least ρ.

After, let $\mathbf{v}_1$ be the optimal cross-entropy reference parameter to estimate l_1, and repeat the last two steps iteratively with the purpose to estimate the pair $\{l, \mathbf{v}^*\}$.

In other words, the iteration of the algorithm consists of two main phases. In the first phase γ_t is updated, in the second phase $\mathbf{v}_t$ is updated. Particularly, starting with $\mathbf{v}_0 = \mathbf{u}$, the subsequent γ_t and $\mathbf{v}_t$ are obtained as described below.

Phase 1 includes adaptive update of γ_t. For a fixed $\mathbf{v}_{t-1}$, let γ_t be a (1-ρ)-quintile of $S(\mathbf{X})$ under $\mathbf{v}_{t-1}$. That is, γ_t satisfies (25) and (26).

$$\mathrm{P}_{v_{t-1}}\left(S(\mathbf{X}) \geq \gamma_t\right) = \mathrm{E}_{\mathbf{u}}[I_{\{S(\mathbf{X})\geq\gamma_t\}}] \tag{25}$$

$$\mathrm{P}_{v_{t-1}}\left(S(\mathbf{X}) \leq \gamma_t\right) \geq 1 - \rho \tag{26}$$

where $\mathrm{X} \sim f(\cdot\,; \mathbf{v}_{t-1})$. The simple estimator $\hat{\gamma}_t$ of γ_t can be obtained by drawing a random sample $\mathbf{X}_1,\ldots,\mathbf{X}_N$ from $f(\cdot\,; v_{t-1})$, calculating the performances $S(\mathbf{X}_i)$ for all i, ordering them from the smallest to the biggest: $S_{(1)} \leq \ldots \leq S_{(N)}$ and finally, evaluating the (1 - ρ) sample quintile as (27).

$$\hat{\gamma}_t = S_{[(1-\rho)N]} \tag{27}$$

Let us note that $S_{(j)}$ is called $j^{\text{-th}}$ order-statistic of the sequence $S(\mathbf{X}_1),\ldots,S(\mathbf{X}_N)$.

Let us note also that $\hat{\gamma}_t$ is chosen such that the event $\{S(\mathbf{X}) \geq \hat{\gamma}_t\}$ is not too rare (it has a probability of around ρ), and therefore, the reference parameter updated with a procedure such as (27) is not void of the meaning.

Phase 2 contains an adaptive update of v_t. For fixed γ_t and $\mathbf{v}_{t-1}$, let us derive $\mathbf{v}_t$ from the solution of the following cross-entropy program according to (28).

$$\max_{\mathbf{v}} D(\mathbf{v}) = \max_{\mathbf{v}} \mathrm{E}_{\mathbf{v}_{t-1}}[I_{\{S(\mathbf{X})\geq\gamma_t\}} W(\mathbf{X}; \mathbf{u}, \mathbf{v}_{t-1}) \ln f(\mathbf{X}; \mathbf{v})] \tag{28}$$

The stochastic counterpart of (28) is shown as follows: for fixed $\hat{\gamma}_t$ and $\hat{\mathbf{v}}_{t-1}$, derive $\hat{\mathbf{v}}_t$ from the solution of the program according to (29).

$$\max_{\mathbf{v}} \hat{D}(\mathbf{v}) = \max_{\mathbf{v}} \frac{1}{N} \sum_{i=1}^{N} I_{\{S(\mathbf{X}_i)\geq\hat{\gamma}_t\}} W(\mathbf{X}_i; \mathbf{u}, \hat{\mathbf{v}}_{t-1}) \ln f(\mathbf{X}_i; \mathbf{v}) \tag{29}$$

Therefore, at the first iteration starting with $\hat{\mathbf{v}}_0 = \mathbf{u}$, the target event is artificially made less rare with temporarily use of the level $\hat{\gamma}_1$, as being less than γ that for to get a good estimate for $\hat{\mathbf{v}}_1$.

The value $\hat{\mathbf{v}}_1$ obtained in this way will make the event $\{S(\mathbf{X}) \geq \gamma\}$ less rare in the next iteration, so the value $\hat{\gamma}_2$ can be used in the next iteration, which is closer to γ itself.

The algorithm terminates when the level is reached at some iteration t, which is at least γ and after the original value of γ can be used without getting too few samples. As mentioned before, the optimal solution of (28) and (29) can be often obtained analytically, in particular when $f(\mathbf{x}; \mathbf{v})$ belongs to the natural exponential family.

The above rationale results are placed in the multi-level cross-entropy algorithm for accelerated rare event simulation with Markov chain modeling in communication networks. This efficient algorithm consists of the following five key steps that are provided below:

Step 1. *Define* $\hat{\mathbf{v}}_0 = \mathbf{u}$. *Set t = 1. (Iteration = level counter).*

Step 2. *Generate a sample* $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N$ *with the density* $f(\cdot;\boldsymbol{v}_{t-1})$ *and compute the sample (1 - ρ) - quintile* $\hat{\gamma}_t$ *performance according to (28) with* $\hat{\gamma}_t < \gamma$. *Otherwise, set* $\hat{\gamma}_t = \gamma$.

Step 3. *Use the same sample* $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N$ *to solve the stochastic program (29). Denote the solution by* $\hat{\mathbf{v}}_t$.

Step 4. *If* $\hat{\gamma}_t < \gamma$, *then set t = t + 1 and repeat from Step 2. Else, proceed with Step 5.*

Step 5. *Estimate the rare event probability l using (30),*

$$\hat{l} = \frac{1}{N} \sum_{i=1}^{N_1} I_{\{S(\mathbf{X}_i)\geq\gamma\}} W(\mathbf{X}_i; \mathbf{u}, \hat{\mathbf{v}}_T) \tag{30}$$

where T denotes the final number of iterations, or number of the levels used.

As a simulation example, let us apply this algorithm to the tandem queue with feedback as was given in Fig.2, but at this time with the following entry parameters: λ=0.2; μ_1=0.8; μ_2=0.2; p=0.5. The simulation results for generating N=10,000 samples are shown in Table 2.

As could be seen in Table 2, overflow is obtained when we have j=6 iterations. The accuracy increases up to N=1,000,000 in the seventh iteration, and the overflow probability is obtained equal to $\hat{\ell}_{IS} = 1.342\text{e}^{-15}$.

Table 2: Simulation Results

Iteration	Repetitive trails	λ	μ_1	μ_2	p	Estimation
1	10^4	0.2	0.8	0.2	0.5	-
2	10^4	0.216	0.643	0.258	0.364	1.426e^{-15}
3	10^4	0.198	0.621	0.287	0.322	1.462e^{-15}
4	10^4	0.196	0.614	0.279	0.318	1.436e^{-15}
5	10^4	0.195	0.614	0.282	0.320	1.372e^{-15}
6	10^4	0.196	0.612	0.284	0.322	1.318e^{-15}
7	10^6	0.196	0.612	0.284	0.322	1.342e^{-15}

CONCLUSIONS

An effective approach to accelerate rare event simulation with Markov chains in communication networks was considered in this paper. Two efficient simulation algorithms were developed and applied to accelerate the overflow probability simulation that in queuing systems is identified a rare event; one based on an Importance sampling method, and another based on a Cross-entropy method.
Simulation results were shown for both algorithms. Recommended approach is appropriate and particularly efficient being used for reliability and performance evaluation in communication networks, while more reliable networks must be described by rarer events.

ACKNOWLEDGEMENTS

The authors would like to thank our colleague Cees J. M. Lanting from CSEM, who contributed his time, ideas, insights and review, and kindly shared his views on ongoing developments during the preparation of this paper.

REFERENCES

Bolch, G., Greiner, S., Meer, H, Trivedi, K. (1998). Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications, NY, John Wiley& Sons.

Bueno, D. R., Srinivasan, R., Nicola, V., van Etten, W., Tattje, H. (2000). *Adaptive Importance Sampling for Performance Evaluation and Parameter Optimization of Communication Systems.* IEEE Transactions on Communications, 48(4), 557-565.

Bucklew, J. (2004). An Introduction to Rare Event Simulation. Springer Series in Statistics, XI. Springer-Verlag, Berlin.

C´erou, F., LeGland F., Del Moral P., and Lezaud P. (2005). "Limit theorems for the multilevel splitting algorithm in the simulation of rare events". *In Proceedings of the 2005 Winter Simulation Conference*, San Diego, USA, 682–691.

De Boer, P., Kroese, D., Rubinstein, R. (2002). "Estimating buffer overflows in three stages using cross-entropy." *Proceedings of the 2002 Winter Simulation Conference*, San Diego, USA, 301-309.

Heidelberger, P. (1995). *Fast Simulation for Rare Event in Queueing and Reliability Models.* ACM Transactions of Modeling and Computer Simulation, 5 (1), 43-85.

Kalashnikov, V. (1997). Geometric Sums: Bounds for Rare Events with Applications: Risk Analysis, Reliability, Queueing. Kluwer Academic Publishers, Dordrecht.

Keith, J., Kroese, D.P. (2002). "SABRES: Sequence Alignment by Rare Event Simulation." *In Proceedings of the 2002 Winter Simulation Conference,* San Diego, USA, 320-327.

Kroese, D., Nicola, V.F. (1999). Efficient Simulation of a Tandem Jackson Network. *In Proceedings of the Second International Workshop on Rare Event Simulation RESIM'99*, 197-211.

Lokshina, I. (2011). "Algorithms to Accelerate Rare Event Simulation with Markov Chain Modeling in Wireless Telecommunications Networks." *In Proceedings of the tenth IEEE Wireless Telecommunications Symposium 2011(WTS 2011)*, NYC, NY, USA, 1-6.

Lokshina, I., Bartolacci M. (2013). *Accelerated rare event simulation with Markov chain modeling in wireless communication networks.* Int. J. Mobile Network Design and Innovation, forthcoming.

Radev, D., Lokshina, I. (2006a). *Performance Analysis of Mobile Communication Networks with Clustering and Neural Modelling*. Int. J. Mobile Network Design and Innovation, Vol. 1, Nos. 3/4, 188-196.

Radev, D., Lokshina, I. (2006b). "Rare Event Simulation with Tandem Jackson Networks." *In Proceedings of the Fourteen International Conference on Telecommunication Systems: Modeling and Analysis (ICTSM 2006),* Penn State Berks, Reading, PA, USA, 262-270.

Rubino G, Tuffin B. (2009). Rare Event Simulation using Monte Carlo Methods, UK, John Wiley& Sons.

BIOGRAPHIES

IZABELLA LOKSHINA, PhD is Professor of MIS and Chair of Management, Marketing and Information Systems Department at SUNY Oneonta, USA. Her main research interests are in area of complex system modeling (communication networks and queuing systems) and artificial intelligence (fuzzy / neuro-fuzzy systems and neural networks); Izabella.Lokshina@oneonta.edu

EVAN SCHIELE, PMP, MBA is a Program Delivery Manager with SAP America. His main research interests are in area of complex system modeling (business applications of information and communication technology) and ERP systems; evan_schiele@hotmail.com

Supervising cathodic protected oil pipe nets with CI-based methods

Matthias Reuter
Department of Modeling and Simulation
TU-Clausthal/ CUTEC Institute GmbH
Clausthal-Zellerfeld, Germany
matthias.reuter@tu-clausthal.d

KEYWORDS
Automatic Supervision of Pipe Networks, CWA, Neural Clusters with SOM Architecture, Early Detection of Faults by Neural Nets, DLS-Spectra and CI – Based Monotoring

ABSTRACT
In this paper we present a tool, which supervises automatically a cathodic protected oil pipe net by CI-based methods. To achieve this aim we use special pre-processing strategy of the DLS-Spectra and new kinds of neural clusters working with the principle of "Computing with Activities (CWA)". We will show that this strategy leads to a save detection of faults in the cathodic protection, trends of wear effects and sudden uprising of so far unknown sensor states. Furthermore we will show that metrological and urbane influences on the cathodic sensor net can be classified as noise contribution, which leads to a high reduction of the false alarm rate.

I. INTRODUCTION

Cathodic protection (CP) of pipes is a common way to slow down corrosion process significantly. The principle of CP is quite easy: the pipe to be protected is connected with another sacrificial metal, which acts as anode. Whenever a pipe is buried in the underground, the anode is also buried in a so called anode ground bed. By this way the sacrificial anode corrodes instead of the pipe. Today most pipes are routinely protected by a coating supplemented with cathodic protection. These systems consist of a DC power source, which is often an AC powered rectifier and an anode, resp. an array of anodes [1][2].

Depending on the surrounding geological and metrological parameters plus the material of the pipe, different amounts of amperes and volts are used, to build up an effective CP. In consideration of the fact that anode and cathode have to be in different places, two cables run out of the DC power box: one (positive) to the anode and the other (negative) to the cathode. For supervising the CP, preferably the second cable leads through junction boxes. In that way the behavior of voltage and amperes can be reordered over time.

Up today the interpretation of these data - and thus the supervision of the CP - is mainly done by human supervisors. Our aim was to implement a software tool which enables an automatic analysis of the CP-data in which an automatic alarming should be integrated. In this context it´s important to mention, that in urban areas thousands of data have to be supervised to guarantee a stable supply of industry every day. This includes the detection of larger damages of the pipe net in an early stage.

Another requirement to such a system is, to supervise the behavior and the status of each sensor, because frequently faults in the data communication are a daily occurrence in the sensor-control room.

II. DATA STRUCTURE AND BASIC REQUIREMENTS FOR A CP SUPERVISING SYSTEM

As explained before CP is an electric technique, so we have volt and amperes as parameters. The network status has to be evaluated by the behavior of one or both of these parameters over time. In Figure 1 a typical record of voltage from one CP-pipe sensor is shown over time. When the record lasts over several years, we can define several time depending changes in the recorded data. First we want to list some changes, which are – for our task – more or less noisy contributions; means are not in our focus for the supervision:

- Very short sharp changes in the voltage (green spikes)
- Very long lasting smoothly changes in the voltage (sinus like fluctuations over time)
- Sensor failures (marked by an "i" and followed by a yellow chart)

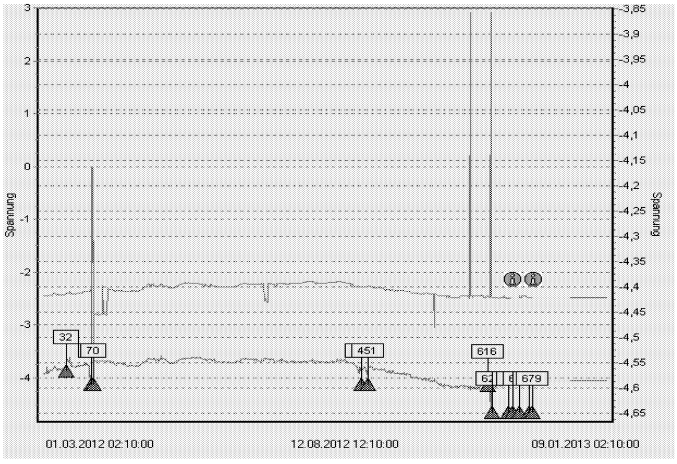


Fig 1: typical chart of voltage of a CP – sensor

The occurring sharp spike in the chart of the voltage can have several causes. Mainly in urban areas the traffic of trains, subways or trams evoke large induced voltage spikes, in which these will only be recorded if the traffic is just in time with the daily measurements. Another cause for such disturbances can be bicycles leaned against the junction boxes or e.g. thunderstorms.

The very long lasting smoothly changes in the voltage (sinus like fluctuations over time) are effected by the seasonal change in the conductivity of the ground. They are a very good indicator for the metrological influences of the underground in which the pipes are buried, but (normally) have only a lower influence on the corrosion of the pipes.

One big problem is the sensor failure of the systems (marked by an "i" followed by a yellow chart). Here we have to consider in which way the data are reported to the control room. If they are transferred by stick, a failure is nothing else than the information that no one picked up the data, if they are sent automatically by WLAN, internet or classic wire, we have to find out, if we have a male function in the broadcast or in the sensor itself. At least this task has to be done mainly by the maintenance personal stuff.

As discussed before those three items have not to be taken into account when the analysis of the records is done automatically, means in his case, a CI-based algorithm has to detect and identify them as "no relevant information".

Remark: Usually records are taken one or two times a day, so the management of the net has to decide, if a quick response like "fault acknowledgement" should be taken into account for these three items or not. But as experience shows such kinds of faults have a very short length of time.

Used methods to eliminate noisy contributions of the voltage charts

Figure 3 shows an enhanced clipping of the chart shown in Figure 1. The green chart represents the original data over the time, the "i" mark shows that a sensor blackout occurred whereby this blackout was a short one (3 days).

To eliminate the seasonal (low band) shift of the chart we used DLS-spectra [3][4], because this operation is extreme robust and adaptive. Furthermore – against the background that DLS-spectra have four parameters – the supervising system can be calibrated without problems on different geological and metrological surroundings.

The DLS analysis method implies a combination of three different spectra; whereby the combination of two of them leads to the noise detection and reduction while the third spectrum inherently describes the trend of the data over time.

In Figure 2 an example of such a noise reduction and trend analysis of a process is shown.

Following the theory of the DLS-Spectra the trend of the original spectra (black dots in Fig. 2) is smoothed by using two different average calculating methods: First the local trend - in our example in a time slot of 21 samples - of the spectrum is calculated by an average algorithm. In that way noisy contributions are diminished. The resulting spectrum is shown as red line in Fig. 2. Parallel a second more global average calculation is done to extract the global trend, shown as green line in Fig. 2. The results of the second calculation incorporate in the next calculation step as a kind of zero line; means the difference between the red and the green line is calculated. This is the yellow line in Fig. 2. The green and the yellow spectra of the DLS-calculation are the spectra of interest for the analyzer. The green line indicates the global alteration of the process over the time - the trend of the process - whereby the yellow line indicates the points of time where the process alters itself. In that way all significant alterations and the global trend over time can be detected.

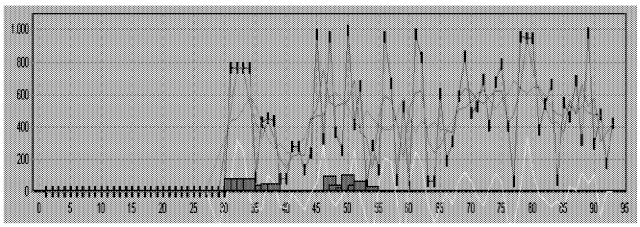


Fig. 2: Chart of a DLS-Spectrum

Furthermore the trend of the DLS-spectra is a very good and robust indicator for sensor breakdowns or sensor fail functions. In the first case we can expect three zero lines and in the second case a so far unknown increase of the DLS-spectra contributions. It is important to mention, that these outliers have to be estimated by a kind of classical expert system as DLS-spectra are a classical method of signal processing.

If we apply the DLS-algorithm on the chart of Figure 1, we got at least 4 charts, shown in Figure 3. For better understanding, in Figure 3 is shown an enlarged cutout. Beneath the original data (green, red), we see the seasonal influence on the original green data (turquoise and pink line) and on the upper hand the high frequency contributions (grey line) of the recorded data.

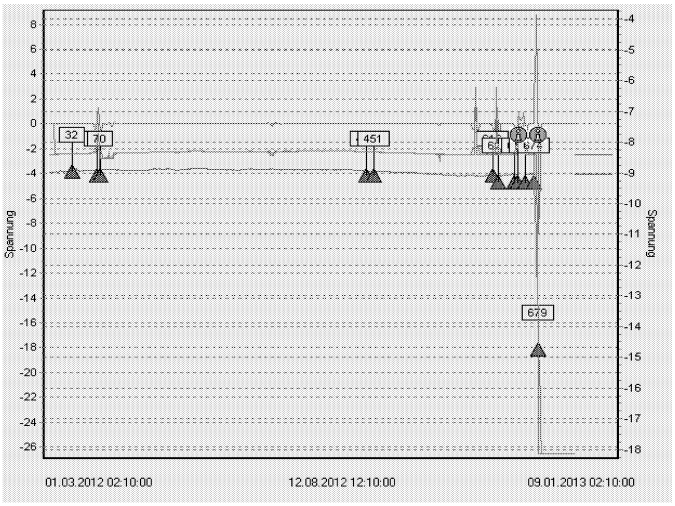


Fig. 3: DLS-operation applied on the chart in Figure 1 (enlarged cutout)

After the seasonal influence has been eliminated of the recorded data, the following calculation step has to pick out the relevant changes.

To suppress the false alarm rate we analyzed the two different canals of the sensor equipment parallel. The requirement for an occurring error under this precondition is that both canals show the same error behavior at the same time, as shown in Figure 4.

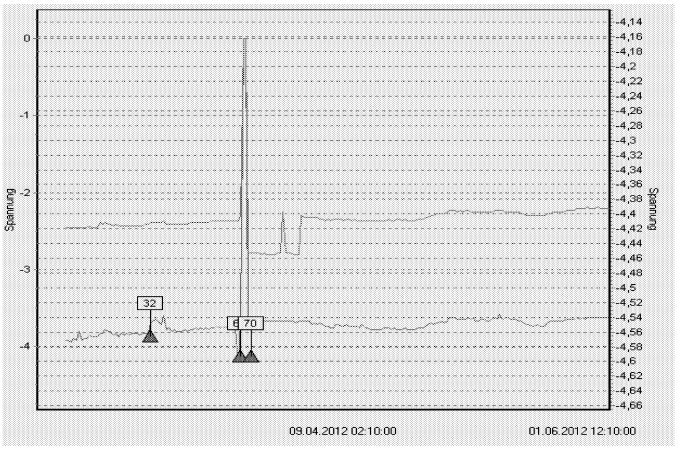


Fig. 4: Chart of an error event of a two canal analysis

It can be pointed out clearly, that at time stamps 32, 60 (covered) and 70 both canals (red and green chart) show the same trend to higher voltage. Furthermore it is easy to see, that at the time stamps 60 and 70 a large change can be detected, while at the time stamp 32 only a slightly change is observable.

A non error event is shown at Figure 5. Here only one canal (green) shows a slightly abnormal behavior.

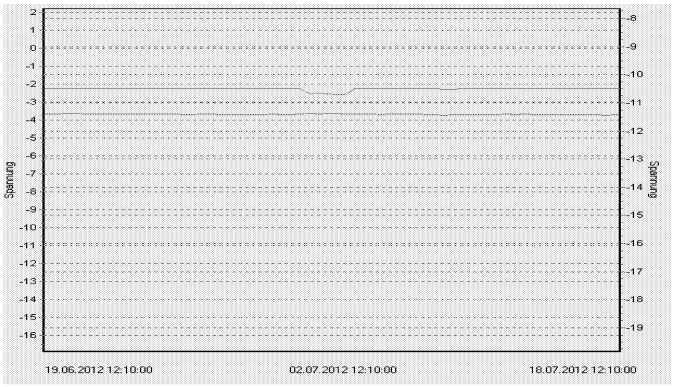


Fig. 5: Chart of an non error event of a two canal analysis

Coming back to Figure 4 we have to realize the following fact: The main problem in this chart is to distinguish between a pure noisy event in the sensor data and a real error event.

Everyone has to handle this problem in the past, because all sensors show a noisy behavior like popcorn or band noise. Finally such a problem can be solved in only one way: We need a kind of "fingerprint" of the sensor as up today every sensor has its individual behavior, surly also depending on its environmental surrounding conditions.

Solving this problem with common expert systems [7]-[13]is nearly impossible, as besides the individuality also the material fatigue has to be taken into account. For that reason we decide to take the individual "fingerprint" from every sensor via a small neural net.

III. INDIVIDUAL SENSOR FINGER PRINTS BY NEURAL NETS

Neural nets can be divided in two major classes: Supervised learning systems and self organizing neural systems. Both of them have advantages and drawbacks, but both systems fail more or less when a high flexible "fingerprint" is needed. Otherwise we have learned in the last few years that biological systems have no problem to handle such a task. The main reason for that is the fundamental difference how common simulated neural nets and biological neural nets work. In short the difference of both systems can be formulated in the following way: While the simulated neural nets code the information of a system in the activity of single neurons, biological neural nets code the information in the overall activity pattern of a neuron ensemble[14]-[18]. So – if we want to have more powerful neural nets - we have to reinterpret the way how a neural net works in that way that not a single neuron but the overall activity of a neural cluster bears the (classification) information [5][6]. This interpretation is independent of the neural architecture. If artificial neural net work architecture can not be interpreted by this paradigm, we will not find it in biological systems.

The difference between the single neuron interpretation and the overall activity based interpretation is the following: In the past the winner neuron (means – depending on the learning

rule - the neuron with the larges or lowest activity) represents the classification result. Now we use the overall activity of the neuron cluster which represents the classification result. This bears new aspects about the capability to store and handle information. First of all the neuron cluster can code a really large number of different patterns (classifications results). For example, if a technical system is able to distinguish 4 different activity states of a neuron, a net of 10 * 10 neurons will have the capacity to code 4^{100} patterns in the neuron cluster. Or for a technical application: the capacity of a 10*10 neuron cluster is to distinguish (or store) 4^{100} machine states. Furthermore - as the information is distributed over the whole neuron cluster (like in a hologram) - a change of the cluster input leads immediately to a change of the activity pattern of the neuron cluster. And – more important – a trained cluster will react to slightly input changes in a slightly way or to large changes in a large way.

It is this feature we need for a robust but still sensitive "fingerprint" of a sensor.

To create such a sensor "fingerprint", different outputs of the sensor are presented to a common self organizing map with Kohonen structure [15]. After a defined range of training circles the neural net shows a high individual activity behavior, representing sensor behavior via an implicit model of the sensor. In Figure 6 two "fingerprints" of two different sensors are shown. As we can see the global structure (activity pattern) is more or less equal, but on the second view non noisy based differences can be pointed out (green area).

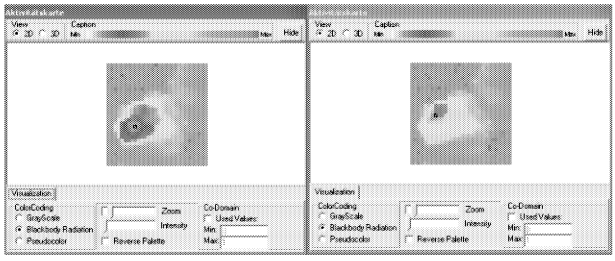


Fig. 6: Sensor "fingerprints" based on the activity pattern of Kohonen like neural nets.

By implementing this "fingerprint" we have a major advantage: We can integrate them to distinguish between “sensor-based” errors (changes in the green area in Fig 6) and real errors (changes in the yellow/red main area in Fig 6).

IV. Application Results and Discussion

After integrating the discussed methods in software architecture, we trained our supervising system with 500 data sets and evaluate the system against 700 test sets. Based on the more than satisfying results, we make a release for field application one year ago.

Since this time the supervising works accordantly and even some sensors show material fatigue the false alarm rate kept at a level round about 4 %.

Figure 7 shows a typical screen shot of our system after a daily analysis has done. As we can see, the analyzed pipe segment has not a stable (means very noisy) behavior over the time. Otherwise all this noisy contributes were not evoked by real CP-error events. Otherwise we have to realize, that the sensor equipment has fallen out since more than one week. Therefore a purple line is shown at the end of the chart and a special alarm will be given by the system.

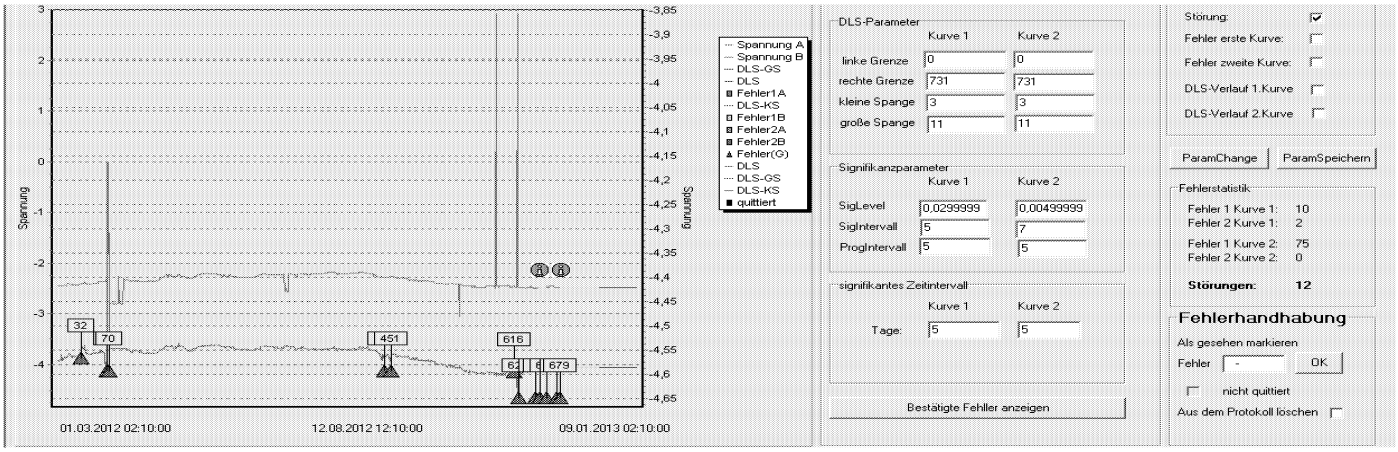


Fig. 7: User Interface of the supervising system

REFERENCES

[1] J.H. Morgan 'Cathodic Protection' National Association of Corrosion Engineers (NACE) 1987, 2nd Edition.

[2] H H Uhlig, Wiley Corrosion and corrosion control, New York, 1985, 3rd edition.

[3] M. Reuter, "Difference Spectra and Their Use in Acoustical Processing", in: Artificial Neural Networks, ed. by. T. Kohonen, Elsevier Science Publisher B.V., Amsterdam, Netherlands, 1991, pp. 1587-1595,.

[4] M. Reuter, S. Bostelmann, „Analysing trends of Quality Data by Means of New Pre-Processing Methods“, WSEAS Transactions on Business and Economics, I1, Vo. 1, 2004, pp 47-51.

[5] M. Reuter, “Of the Stability of Closed Self Organising Maps (gSOMs) for Predictive Control”, ISC’2008, Lyon, France, 2008.

[6] M. Reuter, H. Tadjine, “Computing with Activities III: Chunking and Aspect Integration of Complex Situations by a New Kind of Kohonen Map with Whu-Structures (Whu-Soms),”)”, IFSA2005, Beijing, VRC, 2005.

[7] Leite, D. F.; Costa Jr., P.; Gomide, F.; “Evolving Granular Classification Neural Networks”, International Joint Conference on Neural Networks, 2009, pp: 1736-1743.

[8] S. Ozawa, S. Pang, N.Kasabov, “Incremental Learning of Chunk Data for Online Pattern Classification Systems”, IEEE Transactions on Neural Networks, vol. 19-6, 2008, pp: 1061-1074.

[9] R. Garnett, S.J. Roberts, “Learning from Data Streams with Concept Drift”, Technical Report PARG-08-01, Dept. of Engineering Science, University of Oxford, Jun. 2008, 64p.

[10] R. Elwell, R. Polikar, “Incremental Learning in Nonstationary Environments with Controlled Forgetting”, International Joint Conference on Neural Networks, 2009, pp: 771 - 778.

[11] W. Pedrycz, “Knowledge-based clustering: from data to information granules”, Wiley, 1st edition, 2005, 336p.

[12] S. Pang, S. Ozawa, N. Kasabov, “Incremental Linear Discriminant Analysis for Classification of Data Streams”, IEEE Transactions on SMC - Part B, vol. 35-5, 2005, pp: 905-914.

[13] H. Zhao, P.C. Yuen, J.T. Kwok, “Incremental Linear Discriminant Analysis for Classification of Data Streams”, IEEE Transactions on SMC - Part B, vol. 36-4, 2006, pp: 873-886.

[14] Izhikevich EM (2007). *Dynamical systems in neuroscience*. Cambridge, Massachusetts: The MIT Press.

[15] Haken H (1996). *Principles of brain functioning*. Springer.

[16] Pikovsky A, Rosenblum M, Kurths J (2001). *Synchronization: a universal concept in nonlinear sciences*. Cambridge University Press.

[17] Singer W (1993). "Synchronization of cortical activity and its putative role in information processing and learning". *Annu Rev Physiol* **55**: 349–374

[18] Fingelkurts An.A., Fingelkurts Al.A., Kähkönen S.A. (2005) Functional connectivity in the brain – is it an elusive concept? // Neuroscience & Biobehavioral Reviews. 28(8): 827-836.

[19] V. Barnett, T. Lewis, “Outliers in Statistical Data”, Wiley Series in Probability and Mathematical Statistics, 3rd edition, 1994, 604p.

[20] Kohonen, T.: Self-Organizing Maps, Springer Berlin, 2001



MATTHIAS REUTER was born in Dresden, Germany. He studied physics and geophysics, got his doctor in engineering, and his PD in computer science. He worked as scientist at the University of the German Army in Hamburg, the University of Dortmund and the Clausthal University of Technology. Currently he is Head of the Department of Modeling and Simulation at the CUTEC Institute in Clausthal-Zellerfeld, Germany, and Professor for Computer Science at the Technical University of Clausthal, Germany. Furthermore, he is still a member of the German Air Force in the rank of a colonel.

Dr. Reuter served as steering committee chair and program chair of many international conferences. In 2007, he got the innovation price of the German Defense Industry and in 2008 the Lifetime Award of the World Automation Congress. Furthermore, he serves on Board of Associate Editors for IEEE Systems Journal. He developed a couple of new neural nets structures and learning strategies. His major investigations areas are neuronal net dynamics, robotics, neuroscience and oscillating hierarchical network.

TIME RESPONSE OF OPERATING SYSTEMS SIMULATED BY LOAD BALANCING IN GRIDSIM

Peter Kvasnica
Igor Kvasnica
Rozmarína Dubovská
Náměstí Svobody 301, 500 02 Hradec Králové,
E-mail: peter.kvasnica@tnuni.sk
E-mail: rozmarina.dubovska@uhk.cz

KEYWORDS
gridsim resource, scheduling, entity, load balancing, simulation package, gridlet.

ABSTRACT

Computer is a machine of computation resources that may be utilized different users. Maximization of the resources utilization is based on the implementation of load balancing algorithm. For this reason it is important to (i) minimize the idle time, (ii) minimize the overload, and (iii) minimize the control overhead. From this point of view, we design a dynamic and distributed load balancing scheme. The distributed character of this algorithm does not reduce the communication overhead, but also the idle time of the resources. The paradigms for generation parallel and distributed computing lie in connecting computers. The resources are usually geographically distributed in *multiple administrative domains.* The above mentionedfeatureof parallel computing can be utilized in load balancing the operating systems. We apply the load balancing approach by means of simulation capability of GridSim. The GridSim entities (user, broker, resource, information service, statistics, shutdown, and report writer) send events to other entities to signify the request for service, to deliver results, or to initiate internal actions. The toolkit supports modeling and simulation of heterogeneous Grid resources (both time- and space-shared), users and application models.

INTRODUCTION

The explosion of the computation power of personal computer, speedup managing algorithms used in operating systems and high-speed technology in computer processors gives birth of the idea of using computer resources in a more affective way. The availability of powerful computers and high-speed networks as low-cost commodity components are changing the way we do large-scale parallel and distributed computing. These computers enable sharing, selection and aggregation of suitable computational and data resources for solving large-scale data tasks in science, engineering, and commerce. The computers consist of four key layers of components: hardware, firmware (drivers), operating systems, and applications. The user essentially interacts with an application and operating systems with resources that hides the complexities of computing. The user or broker discovers resources that the user can access using information services, negotiates for access costs-effective services for fulfillment of tasks.

The Distributed Computing Environments

The distributed computing environments from our point of view comprise coupled resources - processors, that do not share memory or clock, they are interconnected by a communication network. These processors are referred to a number of names, such as *sites*, *nodes*, *computers*, or *hosts*. We use *site* to indicate the location of a machine, and *host* to refer to a specific system at a site. The design of the distributed system provides efficient and convenient sharing of resources. The four major reasons for building distributed systems are: *resource sharing, computational speedup, reliability, and communication* (Silberschatz et al. 2002).
The computation can be partioned into subcomputations that can run concurrently, then a distributed system allows us to distribute the subcomputations among the various sites, and thus provide *computation speedup* is provided. If a particular site is currently overloaded with jobs, some of them may be moved to other one, slightly loaded sites. This movement of jobs is called *load sharing*.
Nowadays two general categories of network oriented operating systems are known. *Network operating system* provides an environment in which users can access remote resources by their logging into the appropriate remote machine. In *distributed operating systems*, the users access remote resources in the same manner as they do local resources. Data and process migration from one site to another one is under control of the distributed operating systems (Abramson et al. 2000). A *data migration* is transfer the entire file form one site to another site. A logical extension of computation migration is a *process migration*. When a process is submitted for execution, it is not always executed at the site in which it is initiated. The entire system, or part of it, may be executed at different sites. This scheme may be used for several reasons:

- Load balancing
- Computation speedup
- Hardware preference
- Software preference
- Data access (Silberschatz et al. 2002).

A distributed system consists of policies for the use of the resources and the resources itself. The policies are load balancing, scheduling, fault tolerant, etc. The effects of the load balancing into today's application is that it is often too

easy to create highly available, scalable, and predictable applications. There are some differences between the Grid and traditional distributed systems. Grid Load Balancer (Arif and Arabnia 2003) attempts to distribute the computational load among the available processors.

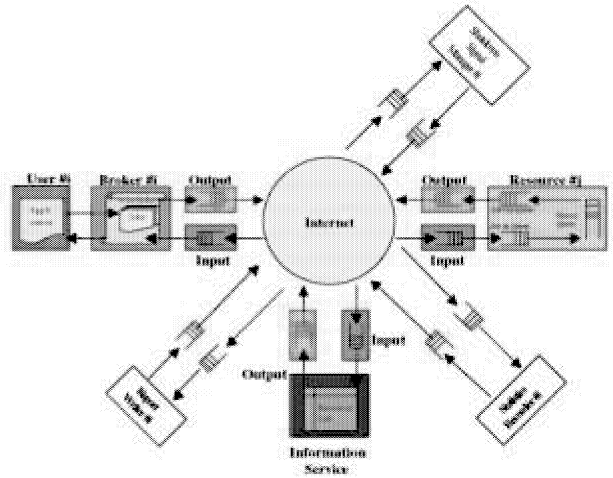
Table 1: Classification of load balancing schemes.

	statis	dynamic
Type	Require prior knowledge	No prior knowledge required
	Cannot cope with changes	Can cope with changes
	No overhead of information gathering	Information gathering overhead
Granularity	Application	System
	Minimizes the individual task's response time	Maximizes the resource utilization

Performance optimization leads to maximize the throughout, i.e., maximizing the number of jobs completion and minimizing the response time of individual job. A load balancing policy tries to maximize the resource utilization. An optional load balancing policy is to minimize the idle time, overloading, and control overheads (Qureshi and Manuel 2007). We get too lost in the realm of intelligent application routing, virtualized application services, and shared infrastructure deployments. It is important to remember to remember that none of these things are possible without a firm basis in basic load balancing technology. The Grid load balancing in Table 1 classifies the schemes on two board classes, type and granularity.
We will design and development of an optimal load balancing strategy for operating system. The use of simulation is common to analyze the effectiveness of the implemented scheme. We implement on the GridSim the operating system structure and load balancing of processes on enhanced GridSim simulation tool.

LOAD BALANCING IN GRIDSIM

Load balancing mechanisms can be broadly categorized as centralized or decentralized, dynamic or static, and periodic or non-periodic. The entire of load balancing is to create a system that virtualizes the "service" from the physical computer that actually run that service. The primary services can be summarized as *scalability*, *high availability*, and *predictability* (Varga 2001).
In a centralized algorithm, there is a central scheduler which gathers all load information from the nodes and makes appropriate decisions. However, this approach is not scalable for a vast environment like the Grid. In decentralized models, there is not usually a specific node known as a server or collector. Instead, all nodes have information about some or all other nodes. This leads to a huge overhead in communication. Furthermore, this information is not very reliable because of the drastic load variation in the Grid and the need for frequent updating.
Static algorithms are not affected by the system state, as their behavior is predetermined. On the other hand, dynamic algorithms make decisions according to the system state. The state refers to certain types of information, such as the number of jobs waiting in the ready queue, the current job arrival rate, etc. Dynamic algorithms tend to have better performance than static ones. Some dynamic load balancing algorithms are adaptive; in other words, dynamic policies are modifiable as the system state changes (Yagoubi and Meddeber 2010).
The testbed size is limited to a few resources and domains, and testing scheduling algorithms for scalability and adaptability, and evaluating scheduler performance for various applications (Buyya 1999). In a cluster or parallel computing it is hard to perform scheduler performance evaluation in a *repeatable* and *controllable* manner as resources and users are distributed across multiple computers with their own policies.
A wide range of heterogeneous resources are single or multiprocessors, shared and distributed memory machines such as PCs, workstations, SMPs, and clusters with different capabilities and configurations. They can be used for modeling and simulation of application scheduling on various classes of parallel and distributed computing systems (Howell and McNab 1998). The information services for resource discovery, and interfaces for assigning application tasks can be used to simulate resource brokers or Grid schedulers.
Each user has own private resource broker and hence it can be targeted to optimize for the requirements and objectives of its owner. This means that all users need to submit their jobs to the *central* scheduler (Buyya and Murshed 2002). A multi-layer architecture and abstraction is employed for the development of GridSim platform and its applications.



Figures 1: A generic view of the World-Wide Grid computing environment (Buyya and Murshed 2002)

Discrete Event Model

SimJava is a general purpose discrete event simulation package implemented in Java. Simulations in SimJava contain a number of entities, each of which runs in parallel in its own thread. Entities have access to a small number of simulation primitives:

- sim_schedule() sends objects to other entities via ports,

- sim_hold() holds for some simulation time,
- sim_wait() waits for an event object to arrive.

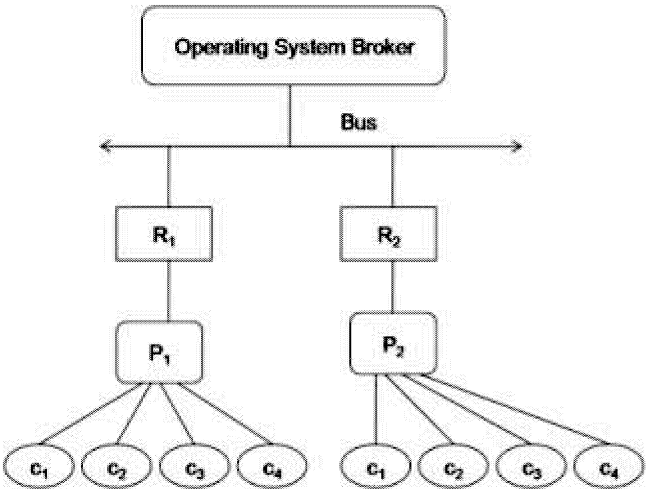
GridSim supports entities for simulation of single processor and multiprocessor, heterogeneous resources that can be configured as time- or space-shared systems. A simulation environment needs to abstract all the entities and their time-dependent interactions in the real system. It needs to support the creation of user-defined time-dependent response functions for the interacting entities (Qureshi et al. 2011). GridSim based simulations contain entities for the users as it is shown in Figure 1. Important entities are *user, broker, resource, Grid information service, input and output.*

LOAD BALANCING IN COMPUTER SYSTEMS

Load balancing is especially important for networks where it's difficult to predict the number of requests that will be issued to a server. Load balancing can also refer to the communications channels themselves. Load balancing is techniques which aim to spread tasks among the processors in a parallel processor to avoid some processors being idle while others have tasks queueing for execution. This technique may be performed either by heavily loaded processors sending tasks to other processors, by idle processors requesting work from others, by some centralised task distribution mechanism; or some combination of these.
GridSim (Hwang and Xu 1998) is a Java-based grid simulator with a clear focus on Grid-based economy and provides resource producer and resource consumer to bargain each other as in a real marketplace. These entities can have customized characteristics.

Enhanced GridSim Tool

In GridSim, each independent task may require varying processing time and input files size. Such tasks can be created and their requirements are defined through *Gridlet* objects. A Gridlet is a package that contains all the information related to the job and its execution management details.



Figures 2: Enhacing GridSim for Load Balancing strategy

Information define as job length expressed in MIPS, disk I/O operations, the size of input and output files, and the job originator (Buyya et al. 2000). These basic parameters help in determining execution time, the time required to transport input and output files between users and remote resources, and returning the processed Gridlets back to the originator along with the results. Machine entity (ME) is treated as dump entity object in GridSim 5.2 and is not able to participate in any decision making activities. We propose that the machine entity should be active and should participate in load balancing at its level. Machine is Processing Entity (PE) manager, in our solving it is processor's core. It is responsible for task scheduling and load balancing of its PEs. The name this enhancement is GridSim the Enhanced GridSim (EGridSim), see Fig.2.

Load Balancing on GridSim

Shared memory multiprocessor computers can be classified as UMA (Uniform Memory Access) or NUMA (Non Uniform Memory Access) computers. In UMA computers each processor can access any memory area with the same average cost, which simplifies the load balancing. The major drawback of UMA architectures is that the number of processors is limited by the contention on access to the shared memory bus. NUMA architectures allow a greater number of processors because processors and memory are distributed in nodes (Baker et al. 2000). Memory access times depend on the processor that a process is executing and on the accessed memory area. The load balancing on these machines is more complex, since moving a process to a node that is distant from its memory area can increase process execution time.
Load balancing for parallel environments is a problem that has been deeply studied for a long time. The load balancing (LBEGS) scheme works at three levels: *Broker*, *Resource*, and *Machine*. When a new job arrives at a machine, it submits it to a PE, which is lightly loaded. Also, after a specific interval of time, it checks the load of all PEs and normal classifies them - lightly loaded, over loaded. We shall first present our proposed scheme. Then two other schemes are presented with which we have compared the performance of our scheme (Qureshi et al. 2011).

OverloadedPEList is a list of PEs with overloaded gridlet;
UnderloadedPEList is a list of PEs with uderloaded gridlet;
NormalPEList is a list of PEs with normal tasks;
for all PEs belonging to this machine periodically do
Apply Algorithm 4 and calculate PELoad which is the work load of PE;
if (PELoad > ∂)
Add This PE to OverLoadedPEList;
elseif (PELoad < ∂)
Add This PE to UnderLoadedPEList;
else
Add This PE to NormalPEList;
endfor;
…
Make a list of gridlets which we want to transfer from overLoaded PE to underLoaded PE;
if (load of this underLoaded PE + load of this gridlet <= ∂) then
Shift this gridlet of overLoaded PE to underLoaded PE;
Move UnderLoadedPEList to next node of list;
Move OverLoadedPEList to next node of list;
else
Move OverLoadedPEList to next node of list;
end while
Calculate load of machine;
Update load of this machine;
if (OverLoadedPEListSize/Number of PE >= η) then

Trigger its Grid Resource that "machine is overloaded";
endif

Figures 3: Algorithm pseudo code for Load Balancing GridSimS scheme (Qureshi et al. 2011)

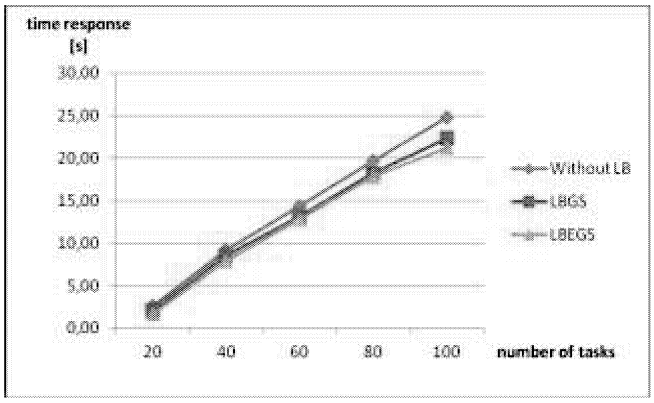
The load balancing algorithm of tasks in operating systems is presented in Fig. 3, show the PE level load calculation, machine level load calculation, and resource load calculation. The PE level load is denoted by ∂, machine level load is denoted by η, and the resource level load is denoted by ρ.
In the load balancing on GridSim LBGS, load balancing is performed only at the Broker level, while in LBEGS, the load balancing is performed at broker, resource, and machine levels. From our opinion is better LBEGS, because the algorithm is nonpreemptive scheduling policy.
We need abstract these physical entities and simulate their behavior using process oriented, discrete event 'interrupts' with a time interval as large as the time required for the completion of the smallest remaining-time job.
In this paper we also evaluate the performance of the Windows load balancing algorithm in different UMA architectures. It using the 2-level scheduling domain hierarchy built by the Windows version. To perform this evaluation we use two different approaches: benchmarking and simulation. The simulation model is developed using the JavaSim simulation tool GridSim (Congdon and Subramaniam 2008).
All algorithms are simulated in GridSim 5.2. We used operating system Windows 7 Professional with SP1 on an Intel PIV 2.6 GHz, with 1024 megabytes of RAM and 160 GB of hard disk. The simulator for these experiments is GridSim version 5.2 with JDK 1.7 updates 6 and JRE 1.7. Following parameters can be used to simulate application schedulers for single or multiple administrative domains, distributed computing systems and power resources:
∂ = 0,6 PE level load, η = 0,75 machine level load, ρ = 0,8 resource level load, n = 2 number of resources, m = 8 number of PE, BUS speed is 1066 MHz, LAN speed is 100 Mb/s.

RESULTS FROM PACKAGE GRIDSIM IN PRACTICE

The GridSim resource simulator uses internal events to simulate the execution and allocation of PEs' share to Gridlet jobs. One or more PEs can be put together to create a computer, one or more computes can be put together to create a Grid resource. A single PE type resource is typically controled by time-shared operating systems that use roundrobin scheduling policy for multitasking. When jobs arrive, time-shared systems start their execution immediately and share resources among all jobs, for preemptive scheduling policy.
The algorithms related to LBEGS are the load balancing scheme of GridSim (LBGS) are similar to Algorithm in Fig. 3. The algorithms related to WLB and LBGS are described in the work (Qureshi et al. 2011).

Simulation Time (WLB, LBGS and LBEGS) with Constant PEs

The results presented here are valid for the algorithms WLB, LBGS, and LBEGS. The number of tasks are taken by different numbers of jobs starting at 20 and ending at 100 with a step of 20, and the numbers of PEs are kept constant at 8, it means number of cores. For each of the tests performed, LBEGS obtained better performances or equal to the other algorithms considered. It is a valid reference to evaluate other techniques. The ELBG is derived directly from LBGS by adding a new computational step, called the ELBG-block, to manage the shifting of badly positioned codewords.
The response time is the total simulation time. The Fig. 4 shows that LBEGS has a lower response time. It is due to (i) gridlets that are assigned to a resource considering the current load, (ii) imbalanced loads that are mostly handled locally.



Figures 4: Simulation Time Different Algorithms Load Balancing in GridSim

WLB has a higher response time, as it uses a nonpreemptive scheduling policy. The response time of the LBGS lies between WLB and LBEGS. Although LBGS uses the preemptive scheduling policy and considers the current workload of a resource like LBEGS, it is slower than LBEGS.
The results presented here are for LBGS and LBEGS. The readings are taken by different numbers of jobs-operating system processes starting at 20 and ending at 100 with a step of 20. When seeing the graph in Fig. 4, it is evident that the gain in response time is higher in LBEGS, than the other two as compared.

Table 2: Result in [s] simulation time-shared Operating system resource Event handler.

number of tasks-gridlets	20	40	60	80	100
Algorithm WLB	2,56	9,16	14,32	19,68	24,8
Algorithm LBGS	2,08	8,48	13,16	18,24	22,4
Algorithm LBEGS	1,68	7,89	12,84	17,84	21,2

RESUME

The GridSim toolkit provides a comprehensive facility for simulation of different classes of heterogeneous resources, users, applications, resource brokers, and schedulers. It should also be noted that resource allocation within high-end processors could also be performed using the space-shared schedulers.

The broker gets job maps to resources (scheduling), stages the application and data for processing (deployment), it starts job execution, and finally it gathers results. It schedules self-events for simulating load balancing depending on the scheduling policy and the number of tasks in the queue or in execution. Load balancing can be useful in applications with redundant communications links,simulation tool or etc. The response of the dynamic changing requirements of jobs such as resources cost, availability, load, and other configuration parameters are simulated. These parameters characterize the processes running in the operating system. The GridSim resources use internal events to simulate these resources behavior and resource allocation operating system processes. Results are possible to be used in operating sytems mobile devices that have multiple picture processing, internet connections ensuring network access if one of the connections fails, and other.

REFERENCES

Abramson D, Giddy J, Kotler L. High performance parametric modeling with Nimrod/G: Killer application for the global Grid? *Proceedings International Parallel and Distributed Processing Symposium (IPDPS 2000),* Cancun, Mexico, 15 May 2000. IEEE Computer Society Press, 2000.

Arif Wani M, Arabnia H R. Parallel edge-region-based segmentation algorithm targeted at reconfigurable multi-ring network. J Supercomput 25(1):43–63. 2003.

Baker M, Buyya R, Laforenza D. The Grid: International efforts in global computing. *Proceedings of the International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet*, Rome, Italy, 31 July-6 August 2000.

Buyya R (ed.). *High Performance Cluster Computing: Architectures and Systems*, vol. 1. Prentice-Hall: Englewood Cliffs, NJ, 1999.

Buyya R, Abramson D, Giddy J, Nimrod G. An architecture for a resource management and scheduling system in a global computational Grid. *Proceedings 4th International Conference and Exhibition on High Performance Computing in Asia-Pacific Region (HPC ASIA 2000),* Beijing, China, 14-17 May 2000. IEEE Computer Society Press, 2000.

Buyya R, Murshed M. GridSim: a toolkit for the modeling and simulation of distributed resource management and scheduling for Grid computing. *Concurrency and Computation: Practice and Experience* 2002, 14:1175-1220.

Congdon P, Subramaniam S. Hewlett-Packard Company – Load Balancing Algorithms. WCB/McGraw-Hill, 2008.

Howell F, McNab R. SimJava: A discrete event simulation package for Java with applications in computer systems modelling. *Proceedings of the 1st International Conference on Web-based Modelling and Simulation*, San Diego, CA. Society for Computer Simulation, 1998.

Hwang K, Xu Z. Scalable Parallel Computing - Technology, Architecture and Programming. WCB/McGraw-Hill, 1998.

Qureshi K, Manuel P. Adaptive pre-task scheduling strategy for heterogeneous distributed raytracing system. Int J Comput Electr Eng 33:70–78. 2007.

Qureshi K, Rehman A, Manuel M. Enhanced GridSim architecture with load balancing. J Supercomput 57:265–275. 2011.

Silberschatz A, Galvin P B, Gange G. *Operating System Concepts.* 6th ed., John Wiley and Sons., Phenix, 2002.

Varga A. The OMNeT++ discrete event simulation system. *Proceedings of the European Simulation Multiconference (ESM 2001)*, Prague, Czech Republic, 6-9 June 2001.

Yagoubi B, Meddeber M. Distributed Load Balancing Model for Grid Computing, Revue ARIMA, vol. 12, pp. 43-60, 2010.

MANUFACTURING SIMULATION

Simulation of a winding production unit for the supply automation of bobbins

Brahmadeep[1]
Sébastien Thomassey[2]
[1, 2]GEMTEX - ENSAIT / Roubaix, France
[1, 2]University of Lille - North of France / Villeneuve d'Ascq, France
[1]E-mail: brahmadeep.deep@ensait.fr
[2]E-mail: sebastien.thomassey@ensait.fr

KEYWORDS
Simulation, Production model, Lean tools, Optimisation, Scheduling, Automation, Arena Simulation

ABSTRACT

This paper aims to explain the production flow and scheduling of yarn rewinding process in a yarn dyeing unit, comparing the various scenarios of production (semi-automatic and fully automatic) and optimising the production process with scheduling using the computer simulation tools. The ultimate goal of this project is to create a new intelligent model which is simulated and the processes involved are to be optimised with the consideration of all the parameters and constraints related to a fully automated production management of yarn bobbin rewinding process. Since the manual operation involves defaults, waste of time, errors and with the growing competitiveness globally, the companies in Europe need to automate as much as possible their production lines. The expected impacts are to increase the productivity and profitability, to have the possibility to customize the production, to develop production tools, implementation of the lean manufacturing tools.

INTRODUCTION

Industries in Europe are facing with pressures to reduce their production costs and to increase the efficiency of their existing staff or equipment. In the same time, the customer demand is becoming more personalised which requires industries to produce small lot size, in short lead time and in unique specifications. The only way for producer is to improve their productivity, reactivity, flexibility, and quality. This trend of mass customization and personalized production leads to fundamental changes in the material flow, plant layout and work organisation (Acaccia et al. 1999). Producers have to rely on smart automation of their production system coupled with an adequate use (qualified staff, computer aided setup), efficient and intelligent planning management, and overall supervision and decision systems. This could be very challenging for some companies such as textile companies, and more particularly dyeing and winding production units which are commonly faced with large batches, long process and set up times. This project concerns the development of automated feeding of winders in order to fight against outsourcing caused by the cheaper workforces and to fight with the lost employment in the textile sector in Europe.

The concept behind this project is based on automation, optimal flow management (internal supply chain management) and implementation of the yarn rewinding process (post-dyeing) with the support of planning and simulation. Companies with the most optimised and automated production tools will be privileged in order to obtain the best product at an acceptable cost.

Objectives

The objectives of this research are the following ones:

- To identify and to analyse the present scenario of the supply chain inside the production plan.
- To develop a standard simulation model with consideration of all parameters and constraints.
- To simulate different scenarios with the help of the developed standard model.
  i. Manual input of bobbins.
  ii. Semi-automatic input of bobbins.
  iii. Fully automatic input of bobbins by robots.
- To analyse the results obtained from the simulations.
- To optimise the efficiency of the system by the scheduling of production.
- To implement the results in the actual production.

State of the art

The simulation has been widely recognised as the best and most suitable methodology for investigation and problem-solving in real-world complex systems in order to choose correctly, understand why, explore possibilities, diagnose problems, find optimal solutions and transfer R&D results to real systems (Banks. 1998). There are various types of simulation models of which the Discrete Event Simulation Model is the one on which our simulation is based upon. Discrete event simulation involves the modelling of a system as it progresses through time and is particularly useful for modelling queuing systems (Law and Kelton 1991). There are many examples of queuing systems: manufacturing systems, banks, fast food restaurants, airports, etc. A major facet of discrete event simulation is its ability to model random events based on standard and non-standard distributions and to predict the complex interactions between these events. For instance, the effects of a machine breakdown on a production line can be modelled.

The steps that should be present in any simulation study to guide a model builder in a thorough and sound simulation

study are mentioned with the logic of a simulation project in (Banks, J et al. 1996). The reference model is the standard model with respect to the present scenario of the supply chain or manufacturing process (Biemens and Vissers. 1989) (Haren and Williams. 1990). It is the model from which other models are developed as per the various parameters and constraints. In our case, we will develop a domain-specific simulation model for manufacturing system i.e. yarn rewinding process.

Simulation Software is the heart and brain of a simulation model. It is the actual place where all the possibilities are constructed, processed, analysed, optimised and later the derived conclusions get implemented to the real situation. The analysis in (Hlupic. 1999) shows that simulation software with good visualization/animation properties are easier to use but limited in case of complex and non-standard problems. For this project we will use the Arena Simulation. It is simulation software by Rockwell Corporation (www.arenasimulation.com) and it is used in different application domains: from manufacturing to supply chain (including logistics, warehousing and distribution) from customer service and strategies to internal business processes. Arena provides the user with objects libraries for systems modelling and with a domain-specific simulation language, SIMAN (Hhuente. 1987). Simulation optimizations are carried out by using Optquest (Hhuente. 1987). Moreover Arena also provides the users animation at run time.

The use of SIMAN language of simulation can be demonstrated in this case of semiconductor manufacturing (Bodner and McGinnis. 2002). The coding is the base for Arena Simulation software. The specific system modelled is a cluster tool, which was an enclosed processing system that integrates a number of different processing chambers around a central material handling system (robot/s).

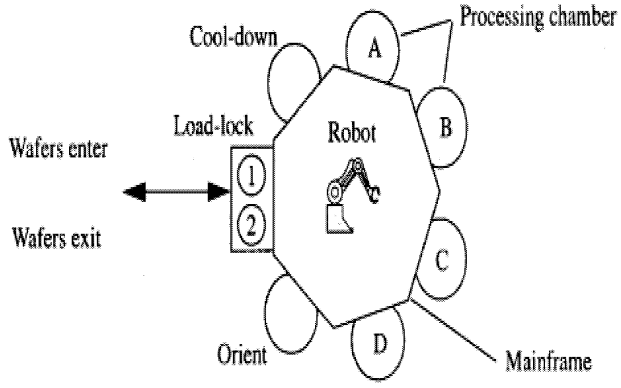


Figure 1: Cluster tool (Bodner and McGinnis. 2002)

A sample cluster tool configuration is shown in Figure 3. The simulation was able to demonstrate the concepts of TRANSPORTER, WAIT-SIGNAL and BUSY-IDLE modules. These modules play an important role while modelling an automated manufacturing process or where there is a scheduling/queuing/waiting is involved. The eventual goal is to provide simulation users with this structured approach so that they build and modify complex models.

In a case of garment manufacturing unit, the main challenges were similar to that of our project like the labour force for production processes, small quantities with few repetitions, frequently changing styles and short delivery times. The aim was to meet the challenges inherent in producing a greater range of products, but at the same time trying to keep the economy-of-scale benefits of mass production (Sepulveda and Akin. 2004) by using computer simulations. The company was trying to eliminate the effect of long lead times, high logistic costs and improve customer satisfaction. To achieve the goals of the company, various simulation model alternatives were designed in order to test and compare the models and to find the most suitable result which can be implemented later. Various situations were considered for the simulation in order to analyse and obtain the best suitable situation and later implement it. They took five situations namely A0: Initial simulation and model validation, A1: Additional workload for customised products, A2: Reducing the duplication of products, A3: Enlargement of workplaces and A4: Investment into appropriate equipment (lcha et al. 2011). Each situation were simulated and analysed for the results which are shown in figure 4.

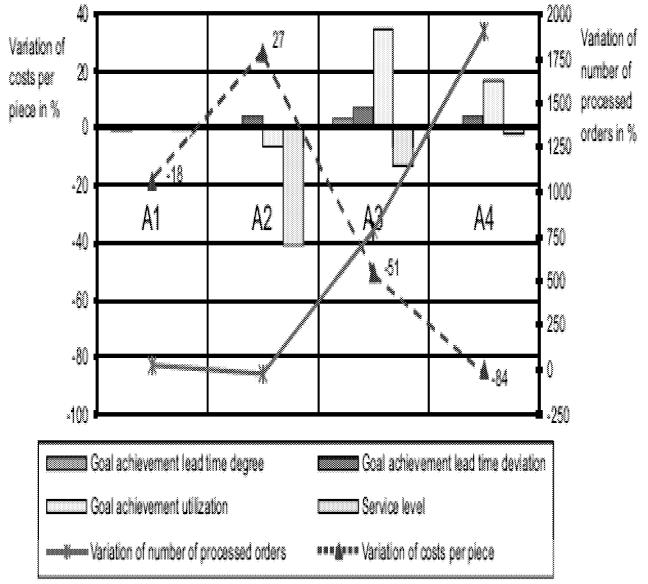


Figure 2: Result comparison analysis (lcha et al. 2011)

The results of the simulated integration of customized production into the existing production system led to many positive findings. After virtually reorganizing, as well as eliminating the potential bottlenecks, the number of processed customized orders can be increased twenty-fold. Additionally the costs per piece can decrease by 84% in comparison to the rather expensive situation in alternative A1. Some alternatives, such as Alternative A4, obtain a high service level.

The simulation results from the above case studies show that the companies achieve their goals in this competitive world by analysing different scenarios by simulations and later implementing the proposed organizational changes. This technique of simulation can be adapted to other manufacturing industries like in our case of yarn rewinding.

METHODOLOGY

The schema of the methodology is stated below in Figure 3.

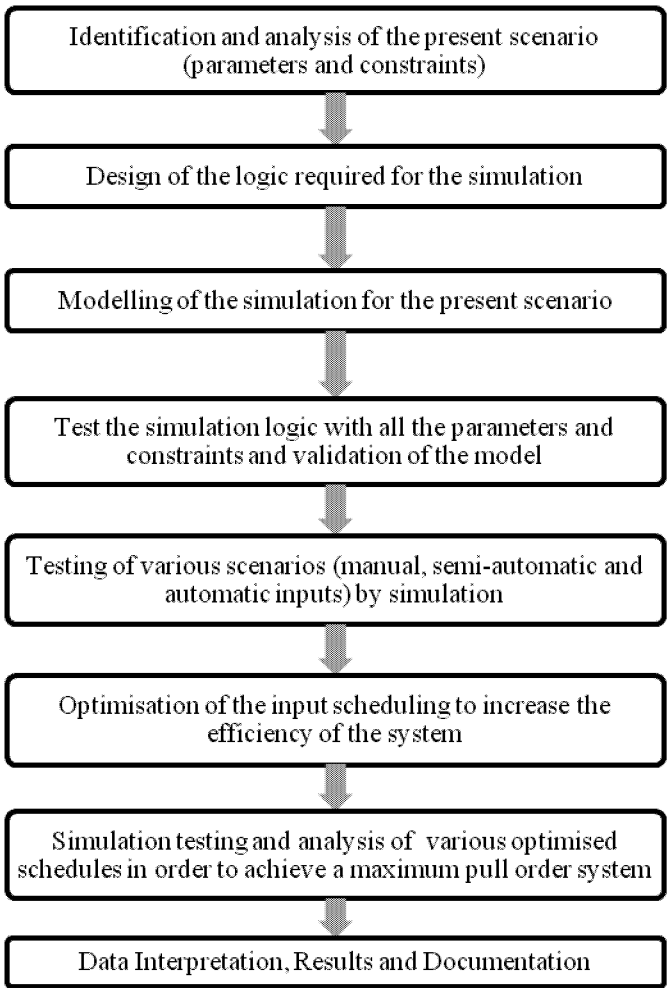


Figure 3: Methodology for the project (Refer the following sub-headings and texts for the explanation)

The following sub-sections describe its main steps with further detail.

Identification and analysis of the present scenario (parameters and constraints)

Four rewinding machines (machine number 83, 86, 87 and 85). Each machine is having six sections (Sec I, II, III, IV, V and VI) and each section has ten winders. See figure 4 for the general layout of the machines and the sections on the production floor. The machine parameters and constraints are same for the ten winders in a section i.e. each section can be identified as system of 10 winders with same machine parameters and constraints. Table 1 describes the machine/section parameters in relation to the yarn bobbins. In the case of multiply yarn winding, if they underwent the self-twisting process then they are considered to have the same machine constraints as of the single ply yarns. Only the non-self-twisted multiply yarns are considered to have the constraints of a multiply yarn on the machine.

The flow of materials in the rewinding section starts from the storage area where the dyed yarn bobbins from the dyeing section are kept waiting for the transfer onto the winding machines. The bobbins of the same lot are batched together as per the lot size in a trolley at the waiting area from where an operator manually transfers the trolley to the respective machine and to the respective section then manually

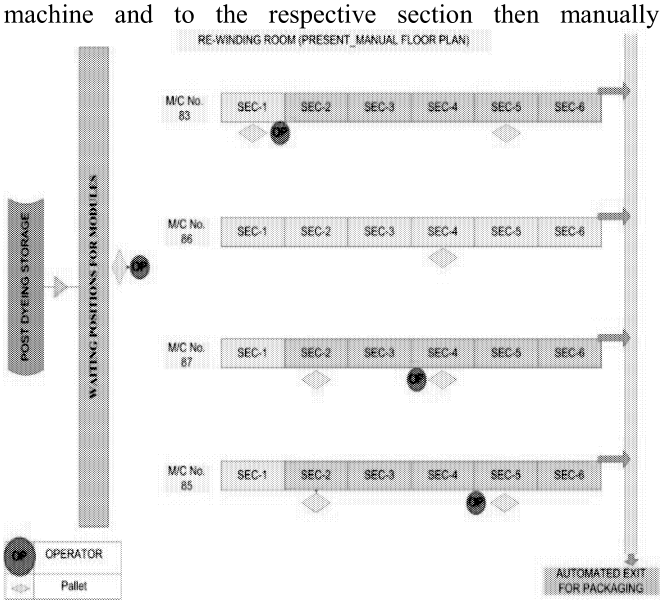


Figure 4: Layout of machines and sections

transfers the bobbins from the trolley to the machine winders. A trolley can only stay at one section only so all the bobbins on the trolley must be put on a single section only. Normally, two trolleys are allocated to each of the sections and only one operator per machine. After the winding process is over for a bobbin it is automatically transferred to the packaging area by the overhead conveyers.

Constraints	M/C 83	M/C 86	M/C 87	M/C 85
Colour Control	Yes	No	No	No
Variopack (Tension Control)	Yes	No	Yes	Yes
Thickness control	Yes	Yes	Yes	Yes
Multiply winding	Only Sec I	Yes	Only Sec I	Only Sec I

Table 1: Machine constraints

Design of the logic required for the simulation

The bobbins were categorised on the basis of machine parameters and constraints and a total of 13 varieties were obtained. Table 2 explains the various possibilities of the winding machines for each respective category of the bobbins.

Category	Winder
1	M83
2	M85, M87 or M83 if place is available
3	M83
4	M85, M87 or M83 if place is available
5	M83 (section I)
6	M85 (section I), M87 (section I) or M83 (section I) if place is available
7	M83

8	M85, M86, M87 or M83 if place is available
9	M83
10	M85, M86, M87 or M83 if place is available
11	M83 (section I)
12	M86, M85 (section I), M87 (section I) or M83 (section I)
13	Not going to the 4 machines in the study

Table 2: Possibilities of winding for each category with respect to the machines

Testing was done for the validation of the simulation model. For the validation purpose and to test the conditions and decision modules in the model, push and pull system of material flow was taken as to demonstrate on the accuracy of the model. The push flow of materials corresponds to the bobbins which arrive at the rewinding section from the dyeing section and are stored temporarily. The pull system corresponds to the flow from the storage to the machine winders by the operators which depends upon the delivery

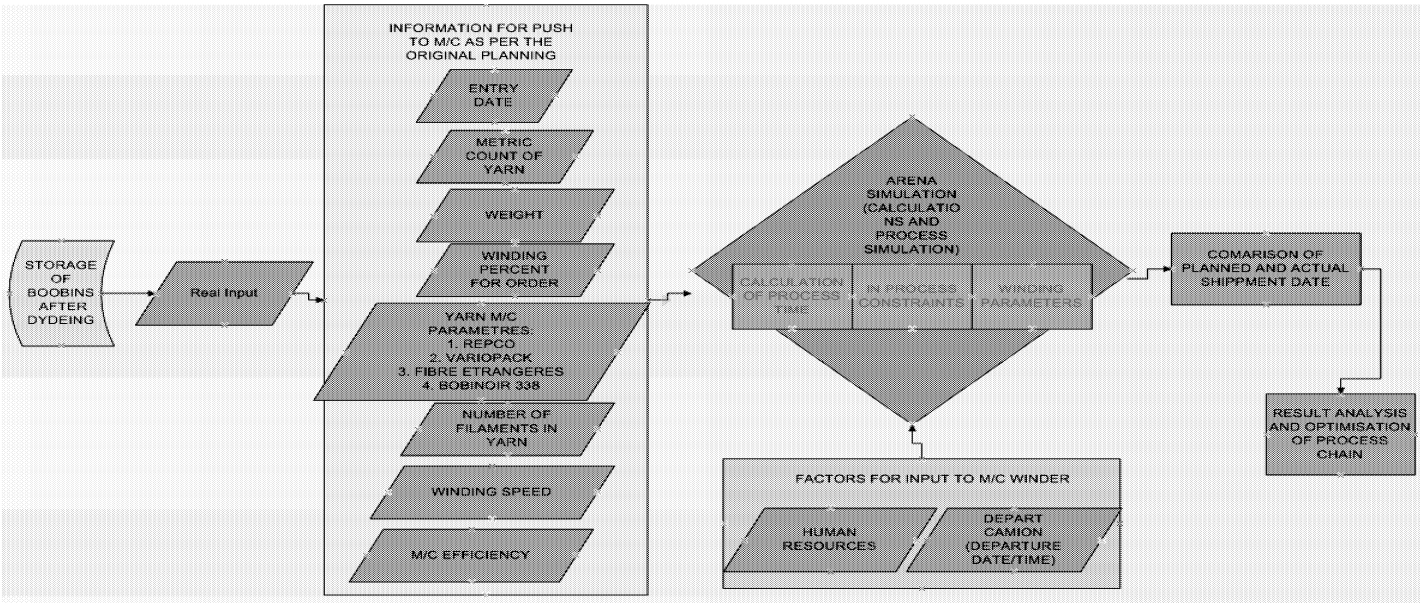


Figure 5: Simulation logic

With the consideration of all the parameters a simulation logic diagram was designed on the basis of which the actual modelling of simulation was done on the simulation software 'Arena Simulation' by 'Rockwell Automation'. The logic is demonstrated in the Figure 5.

Modelling of the simulation for the present scenario

The modelling of the model was done in the Arena Simulation with all the parameters and machine constraints. The decision making and distribution design for the flow of material was done with the consideration of 13 categories of the bobbins.

In the model the inputs for each category for the entry of the bobbins was scheduled in the excel sheets which were imported into the model. One month duration was taken for the scheduling of the inputs. Decision and Hold modules were used for the distribution of the bobbins to their respective machine sections. The distance and the speed of the operator to transfer the bobbin trolley to the machines were taken as per the time study. The operator speed was taken 28 m/s and 23 m/s with load and without load resp. The transporters in the simulations were assigned as the trolleys. The average machine speed of each winder was calculated and was taken to be 26 minutes per bobbin at 80% efficiency of the winders.

Test the simulation logic with all the parameters and constraints and validation of the model

date/time of the order, machine WIP and availability of the human resources.

The following sub-headings are the explanations for the next steps in the methodology shown in Table 1which are to be done in the near future:

Testing of various scenarios (manual, semi-automatic and automatic inputs) by simulation

The testing of the various scenarios is to be done where the model would be subjected to the changes in the parameters and constraints depending upon the type of situation for example: manual or semi-automatic or fully-automatic. Basically it is the experimentation to find the possible SWOT's (Strength, Weakness, Opportunity and Threats) in the system under different situations.

Optimisation of the input scheduling to increase the efficiency of the system

For the system to run on maximum efficiency with minimum idle time and delivery of the products on time, the scheduling of the inputs to the system must be optimised to the best possible level. The optimisation of the scheduling of the inputs for the bobbins is to be done using the simulation optimisation tools which uses various complex game theories and calculations to reach on desired schedule order(s).

Simulation testing and analysis of various optimised schedules in order to achieve a maximum pull order system

The final step after the optimised schedule order(s) is proposed from the optimisation tool is to test run of the original simulation model to for the verification of the information and to have a final conclusion on the results.

RESULTS AND DISCUSSIONS

Results

The project is in the initial stage, only the first model for the validation purpose was simulated. The simulation of the first model showed the flow and transfer of materials from the storage area to the machine and finally the dispatch to the packaging area. This simulation was a part of the manual transfer of materials. The demonstration showed that the first model is managing the decision and distribution of bobbins properly as per the machine parameters and constraints with the fixed number of human resources.

Limitations

The limitations in the initial stages of this simulation were:

- The experimentation of the model after the validation of the model.
- The time is required for the study of the system, design and to evaluate the model.

Scope and Future

The research work is in the initial stage, therefore there is a lot of research and development left to be done with the earlier mentioned methodology.

- The testing of various scenarios (manual, semi-automatic and automatic inputs) by simulation is to be done.
- The optimisation of the input scheduling is to be done.
- The key performance indicators are to be defined and integrated in the simulation model.
- The simulation testing and analysis of various optimised schedules in order find the best possible schedule which in process will achieve a maximum efficiency in the pull order system.

CONCLUSION

The simulation modelling of an internal manufacturing supply chain of yarn rewinding section in a yarn dyeing factory gives numerous possibilities and scenarios for the testing of the present and future proposed models. The optimisation of the scheduling for the inputs, the human resource requirements, finding the bottlenecks and reducing the lead times with least WIP (work in progress) can be identified. These results obtained will help to implement the proposed plans which lead from a push flow to a pull flow system, generating maximum profit and staying competitive in the global market.

REFERENCES

Acaccia G.M ; Conte M ; Maina D ; Michelini R.C ; Molfino R. 1999. *Integrated Manufacture of High Standing Dresses for Customized Satisfaction, Globalisation of Manufacturing in the Digital Communication Era*, 511–523.

Banks, J., Carson J.S., and Nelson B.L., 1996. *Discrete-Event System Simulation*. 2nd ed. Prentice Hall. Upper Saddle River. N.J.

Banks J. 1998. *Handbook of simulation*, Wiley Interscience, New York.

Biemens, F. P., and Vissers, C. A., 1989, "Reference Model for Manufacturing Planning and Control," *Journal of Manufacturing Systems*, 8(1), 35-46.

Bodner D A and McGinnis L F Proceedings of the 2002 Industrial Engineering Research Conference. A Structured Approach to Simulation Modelling of Manufacturing Systems, School of Industrial and Systems Engineering Georgia Institute of Technology Atlanta, Georgia 30332-0205, U.S.A.

Heilala and Juhani. 1999 Manufacturing operation simulation in make-to-order manufacturing. VTT Industrial Horizon, Espoo. (January)

Hlupic V. 1999. *Discrete-Event Simulation Software: What the Users Want, Simulation*, Vol. 73, No. 6, 362-370.

Hhuente D.J. 1987; "Critique of SIMAN as a programming language", *In Proceedings of the ACM Annual Computer Science Conference* 1987, 384-385.

lcha Gert Zu, H Koruca alil Ibrahim, Borkircher Mikko, 2011. *Computers in Industry*. 568–577

Law A.M and Kelton W.D. 1991. *Simulation Modeling and Analysis*, 2nd ed. McGraw-Hill, New York.

Sepulveda J.A., Akin H.M. 2004, Modelling a garment manufacturer's cash flow using object-oriented simulation, in: R.G. Ingalls, M.D. Rossetti, J.S. Smith, B.A. Peters (Eds.), Proceedings of the Winter Simulation, vol. 2, Association for Computing Machinery et al., New York, NY, et al., 2004, pp. 121–128.

Van Haren, C. R., and Williams, T. J., 1990, "*A Reference Model for Computer Integrated Manufacturing from the Viewpoint of Industrial Automation*," in Proceedings of CIMCON '90, Gaithersberg, MD, 42-62.

WEB REFERENCES

Arena by Rockwell Corporation, www.arenasimulation.com.

AUTHORS

SEBASTIEN THOMASSEY
ENSAIT / GEMTEX
2 allée Louise et Victor Champier
59100, Roubaix, France
phone: +33 (0)3 20 25 64 64
e-mail: sebastien.thomassey@ensait.fr

BRAHMADEEP
ENSAIT / GEMTEX
2 allée Louise et Victor Champier
59100, Roubaix, France
phone: +33 (0)3 20 25 64 64
e-mail: brahmadeep.deep@ensait.fr

IMPACT OF INDUSTRIAL SORTING ON VARIOUS PISTACHIO CATEGORIES IN LINK WITH AFLATOXIN CONTAMINATION THANKS TO A MONTE CARLO SIMULATION

Nathalie Wesolek, Alejandra Ramirez-Martinez, Alain-Claude Roudot
LERCCO (Laboratoire d'Evaluation du Risque Chimique pour le COnsommateur)
UBO (Université de Bretagne Occidentale)
6 av Victor Le Gorgeu
29238 Brest CS 93837, France
E-mail: nathalie.wesolek@univ-brest.fr

KEYWORDS
AFB1, early split, mouldy, cracked, industrial process.

ABSTRACT

Aflatoxin B1 (AFB1) is the main aflatoxin found in pistachios, and has genotoxic carcinogenic effects with no threshold. It is crucial to predict the impact of sorting on various pistachio categories, which are early split, cracked and mouldy pistachios, and thus on AFB1 contamination level. This work assesses the distribution and incidence of aflatoxin in pistachio nuts after forced air drying, as well as in final packages, thanks to a Monte Carlo simulation. A 10 kg sample size of Iranian pistachios for export to Europe is considered. Differences between good and bad lots are taken into account with lots split into 3 classes for mean contamination levels in export packages. Simulation results are in good accordance with observed contamination levels in Iran. Our assessment shows that sorting enables to divide the incidence of early split and mouldy pistachios by three, for any incidence level prior to sorting.

INTRODUCTION

Aflatoxins are mycotoxins produced by *Aspergillus* moulds. They contaminate pistachios in the orchard, when *Aspergillus* spores, carried by insects or wind get into contact mainly with pistachios having their hull ruptured. Indeed, the hull protects in-shell pistachios when it is not split. After harvest, contamination does not increase in case of proper storage condition, after industrial processing. Due to AFB1 threat for human health, the European Union (EU) has set a maximum level at 8 ppb (or µg/kg) for ready to eat pistachios (Regulation 165/2010) in 10 kg samples (Regulation 178/2010). Depending on harvest year, aflatoxin can cause a high rejection rate of Iranian pistachios at the point of entry into the EU. Rejections are reported through the RASFF (Rapid Alert System for Food and Feed) and cause great financial losses. A ban on Iranian pistachio imports due to high levels of aflatoxin occurred in September 1997. The embargo was lifted three months later following assurances that the quality of Iranian pistachio would be improved. Aflatoxin level in pistachio consignments is difficult to assess, knowing that this contaminant is heterogeneously distributed in pistachios. Per example, it has been estimated that only one pistachio out of 10,000 to 1,000,000 might be highly contaminated (Schatzki and Pan 1996), or one out of 25,000 nuts (Sommer *et al.* 1986). For this reason, a new statistical method must be found to assess the contamination incidence for individual pistachios, as well as the contamination level and their impact on the contamination level distribution of pistachio samples. The proposed method uses the Monte Carlo technique, as this technique enables to take into account the whole range of variability of the variables. As regards the industrial process, Iranian pistachios are dehulled, sorted by water floatation, dried by forced air, manually sorted and put into export packages. So, the statistical method is applied here to pistachios just after they are dried by forced air, as well as after manual sorting, it says at the final step before export. The objective of this study is to assess the impact of manual sorting at industrial level on specific categories of pistachios having visual defects, and thereby on the aflatoxin contamination level of export packages.

DATA USED AND METHODS

Contamination Incidence and Levels

Contamination incidence and levels in individual pistachios were calculated from results reported in a publication from Hadavi (2005). In this study, pistachios were collected from six commercial orchards in Iran and aggregated into samples, based upon hull appearance. It is reported that, by this sampling procedure, the heterogeneous distribution of contaminated nuts was well covered, as practiced earlier in literature. Sound, early split, cracked (growth split) pistachio samples for each orchard were separated according to fluorescence or non-fluorescence, as well as kernel aspect. Indeed, the bright greenish yellow colour under an ultraviolet light, referred to as the 'BGY' fluorescence (BGYF) is partially linked to aflatoxin contamination. In addition, a brown coloured or mouldy kernel is a good indicator of mould growth. The data taken into account for the simulation were sample contamination levels for early split and cracked pistachio samples having sound kernels, as well as pistachios having brown spotted or mouldy kernels, or with the presence of fungal colony. More precisely,

pistachios were categorized according to visual defects into three categories, namely early split, cracked and mouldy pistachios. As the sample size was small enough (from 1.1 g to 168 g of kernels), we considered that a maximum of one kernel per sample was contaminated, taking into account a kernel weight of 0.52 g. Indeed, aflatoxin incidence is a rare event in pistachios, even in mouldy kernels, due to the fact that *Aspergillus* is not the sole mould capable of infesting the nut. In addition, when *Aspergillus* is present, it only synthetizes the toxin in specific conditions. Then, the calculation of the incidence of contaminated pistachios per category and individual contamination levels was achieved. Individual contamination level distributions were fitted to theoretical distributions for each pistachio category. Simulated lots were split into three different quality levels for the mean contamination level, numbered by ascending contamination level: 1st class, 2nd class, 3rd class. We considered a constant incidence level of pistachio categories within lots of the same quality levels. The incidence of pistachio categories used for pistachios after processing was deduced from a document from the UNECE (United Nations Economic Commission for Europe), defining the maximum tolerances for « extra pistachios » (UNECE, 2001): 2% of early split pistachios and 1% of cracked pistachios. So, for 2nd and 3rd class lots in final packages, the incidence levels of early split and cracked pistachios were set at 2% and 1% respectively. However, for 1st class lots, these categories incidences were divided by 2. As regards the incidence of the mouldy pistachio category, we kept it as an adjustment variable. The final calculation of contamination incidence was achieved thanks to the Bayes' theorem in Equation (1):

$$P(\text{contaminated} \cap \text{category}) = P(\text{contaminated} / \text{category}) \times P(\text{category}) \quad (1)$$

with P(contaminated), the probability that the individual is contaminated and P(category), the probability that the individual is in a given category.

Monte Carlo Technique using @Risk

The @Risk software is a Monte Carlo software for risk analysis and risk assessment using Monte Carlo random simulations. It is used to determine how often specified events may occur thanks to the generation of Probability Density Functions (PDFs). This probability risk analysis approach is opposed to the deterministic risk analysis calculation. Indeed, when achieving a deterministic risk evaluation, values are assigned for discrete scenarios to see what the outcome might be in each. On the contrary, in a probability risk evaluation, values are given to all the possible scenarios, taking into account their probability of occurrence. It is used to estimate the effect of variability on the outputs of models which use a variety of input parameters. Monte Carlo simulation is the replication of a procedure with input data drawn randomly from a parametric distribution. However, Monte Carlo simulation can be used with two different sampling types under @Risk: the Monte Carlo (MC) and the Latin Hypercube (LH) sampling types. The LH sampling is a particular MC sampling: more precisely it is a constrained MC (i.e. random) sampling scheme, involving a stratification of the input distribution. The constraint refers to the way each variable is sampled: the statistical cumulative distribution is split in n equally probable intervals on the cumulative probability scale, and then a random value is selected within each interval. The number of n equally probable interval is equal to the number of iterations of the Monte Carlo simulation. A simulation with 500 iterations, per example, would split the probability into 500 segments, each representing 0.2% of the total distribution. For the first segment, a number would be chosen between 0.0% and 0.2%. For the second segment, a number would be chosen between 0.2% and 0.4%. This number would be used to calculate the actual variable value based upon its distribution. Once each variable has been sampled using this method, a random grouping of variables is selected for each Monte Carlo calculation. Independent uniform selection is done on each of the variable's generated values. Each value must only be used once. LH sampling would create a more representative sampling of the distribution, ensuring that tail samples of the input distributions are drawn (e.g. for 1000 iterations, exactly one value above the P99.9 would be drawn, whereas for MC sampling either none, one or several samples may be drawn). So, LH sampling can be used to force the sampling of tails of distributions, along with a great number of iterations. This is why, in this study, thanks to the @Risk 5.7 software (Palisade Corp.), PDFs were modeled, using Monte Carlo simulations with 10,000 iterations under the LH sampling type.

Modeling of Pistachio Categories and Contamination

Calculations were achieved for each pistachio category separately, at first. The contamination distribution for 10 kg samples for a given number of contaminated individuals per sample is given in Equation (2):

$$Cs = \frac{Ci}{Ni} \times Nic \quad (2)$$

with *Cs*, 10 kg samples concentration distribution; *Ci*, individual pistachios concentration distribution; *Ni*, the number of individual inshell pistachios in a 10 kg sample; *Nic*, the number of contaminated individual inshell pistachios in a 10 kg sample. The probability of a given number of contaminated individuals (0, 1, 2, 3, etc.) was calculated thanks to a binomial distribution, knowing the final calculation result of the contamination incidence. These probabilities were then grouped together, using a discrete distribution. A conditional function was used to match these probabilities with concentration levels distributions in 10 kg samples. The result was a PDF for the contamination level in 10 kg samples, considering only the contamination stemming from one pistachio category. In a second step, PDFs obtained for each pistachio category were added up in order to obtain the PDFs of final concentration in 10 kg samples for each of the 3 classes.

Simulation Adjustment

Two observed distributions of AFB1 contamination in 10 kg samples were used. These distributions were generated by

the Green Corridor Project 2004 (INC website: International Nut and Dried Fruit Council; personal communication from the INC for the sampling plan used). The goal of the Project was to monitor conditions and practices at critical control points through the delivery chain starting from pre-harvest to utilization in Europe for aflatoxin contamination, but only for Iranian pistachios out of 2004 crop. Both the distributions after the drying step and in final packages were used to adjust simulation results to real data. In the Project, the samples that we considered were taken from Iranian processing terminals, with 40 samples of pistachios after drying and 41 samples of export packages. For the simulation, the incidence levels of mouldy pistachios in final packages for each of the three lot classes had to be set in order to ensure that all ranges of observed concentration in the samples could be generated. Moreover, a binomial distribution was used to combine the 2nd and 3rd lot classes together and another binomial distribution enabled to mix this combination with the 1st lot class, thus obtaining the PDF for final concentrations in 10 kg samples in export packages. As regards the PDF for these concentrations in 10 kg samples after the drying step, we considered that the proportions of the different pistachio lot classes were the same as in export packages. Indeed, the proportions of the different qualities of lots remained the same, as we considered pistachios from the same crop year before and after manual soting. We took into account that the decrease in contamination level after sorting was due to the removal of a certain proportion of early split and mouldy pistachios. This assumption was based on literature data, as per example Bonjar (2004) which stated that the elimination of early split nuts during sorting will reduce substantially moulds and aflatoxin in pistachio nuts. Moreover, in our study, we have separated early split nuts with sound kernels from mouldy pistachios, but most of mouldy kernels stem from early split appearance inshell pistachios in general. Now, these early split appearance inshell pistachios are the pistachios main visual defects that are removed during manual sorting, as they are dark spotted so, easily recognizable from sound pistachios. Moreover, Shakerardekani *et al.* (2012) reported that the presence of *Aspergillus* can cause a dark-greyish appearance of the shell, so we could deduce that such pistachios were likely to be removed by manual sorting too. For this reason, we have set that the decrease in early split pistachio and mouly pistachio categories would be the same in proportion.

RESULTS AND INTERPRETATION

Contamination Distribution Fitting and Interpretation

Goodness of fit tests were achieved for calculated individual contamination levels for each pistachio category, using @Risk. Test statistics for the Kolmogorov-Smirnov (K-S) test for lognormal distribution, triggered by the software were compared to critical values for this test at a 5% risk level, as shown in Table 1. So, a lognormal distribution can not be rejected as the true parent distribution at a 5% risk level. Cumulative distribution functions for contamination levels, along with their regression curves, were plotted using Excel in Figure 1, 2, 3 for early split, cracked and mouldy pistachio categories, respectively.

Table 1: K-S Test for Contaminated Pistachios from Each Category for LogNormal Distribution Fitting

Pistachio category	Test statistic	Critical value for $\alpha = 0.05$
Early split	0.1175	0.301425
Cracked	0.1960	0.563275
Mouldy	0.1671	0.309360

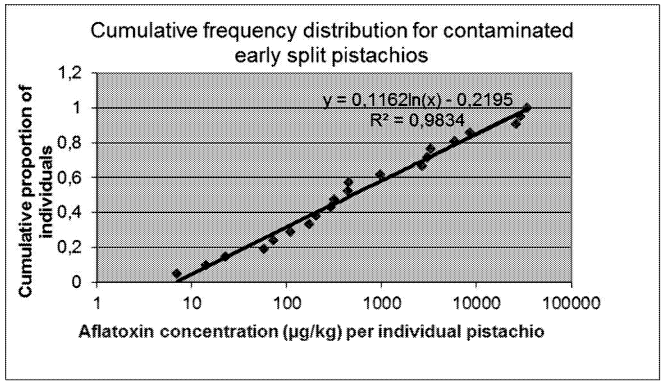


Figure 1: Regression for the Early Split Pistachio Category

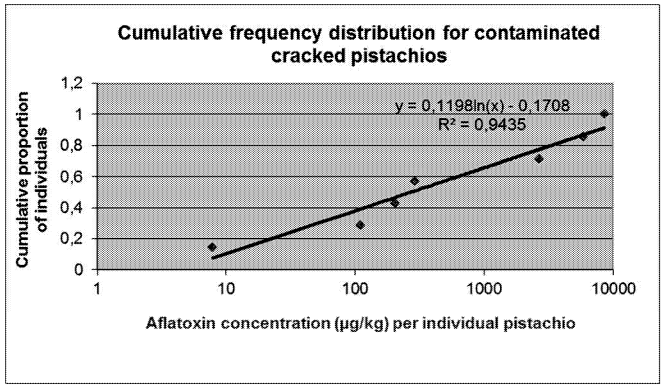


Figure 2: Regression for the Cracked Pistachio Category

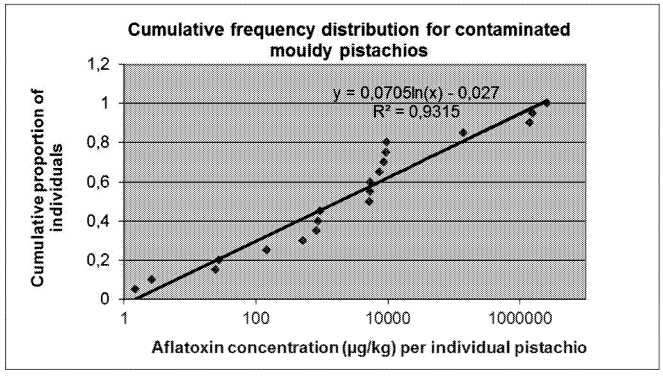


Figure 3 : Regression for the Mouldy Pistachio Category

As can be observed in Table 2, the mean individual contamination level was higher in early split than in cracked pistachio category, the highest mean contamination being in mouldy pistachio category.

Table 2: Mean Contamination Level and SD

Pistachio category	Mean (ppb)	Standard deviation (SD)
Early split	5,542	10,525
Cracked	2,527	3,412
Mouldy	287,646	706,790

Indeed, Doster and Michailides (1994) reported that most early split nuts had the hull split before any pistachio fruits had cracked hull, which may explain why cracked fruits had less kernel mould than early splits. The fact that early split and cracked pistachio categories (having apparently sound kernels) were contaminated can be explained by the fact that the mould was not present anymore, but had synthetized aflatoxins while it had been growing.

Simulation After the Last Sorting Step

The incidence of mouldy pistachio categories was set at 0%, 0.01% and 0.25% for the 1st, 2nd and 3rd lot classes, respectively. Final calculations of contamination incidence for each pistachio categories and each lot class were achieved (Table 3). As expected, final concentration PDFs in 10 kg samples obtained for each of the 3 classes with @Risk, covered well the range of concentration and variability that can be observed for the Green Corridor Project. Mean and P95 concentrations are reported in Table 4.

Table 3: Final Calculation of Contamination Incidence

Pistachio category	1st class	2nd class	3rd class
Early split	0.000075456	0.000150912	0.000150912
Cracks	0.0000305666	0.0000611332	0.0000611332
Mouldy	0	0.000008	0.000169074

Table 4: Parameters Obtained for the Simulated PDFs

PDF parameters	1st class	2nd class	3rd class
Mean (ppb)	0.502	3.49	48.25
P95 (ppb)	2.08	10	100

Table 5: Comparison of the Probabilities Obtained

Obtained probabilities	< 2 ppb	2-5 ppb	5-10 ppb	10-25 ppb	> 25 ppb
Green Corridor Project	82.9%	4.9%	4.9%	2.4%	4.9%
Simulation	83.8%	6.0%	2.8%	2.8%	4.6%

Then, a binomial distribution was used to combine the 2nd and 3rd lot classes together with a probability at 0.6 for the 2nd lot class. This combination was then mixed with the 1st lot class, this latter class having a probability set at 0.7. Simulated and observed (from Green Corridor Project) PDFs for final concentrations in 10 kg samples in export packages were compared. Obtained probabilities for each concentration range are reported in Table 5. So, the simulation was validated, as the probabilities were very similar.

Simulation Before the Last Sorting Step

As stated earlier, before sorting, the proportions of the different pistachio lot classes were the same as they were in export packages. However, the proportion of early split and mouldy pistachio categories were not the same. Incidence levels for these categories were multiplied by 3 and thus, the final calculation result of the contamination incidence was as reported in Table 6. Such values for incidence levels were set, for the simulation PDF for contamination levels in 10 kg samples to correspond to observed PDF values from the Green Corridor Project. A comparison of simulated and observed probabilities is reported in Table 7.

Table 6: Final Calculation of Contamination Incidence

Pistachio category	1st class	2nd class	3rd class
Early split	0.00022637	0.00050723	0.00050723
Cracks	0.0000305666	0.0000611332	0.0000611332
Mouldy	0	0.000024	0.00045274

Table 7: Comparison of the Probabilities Obtained

Obtained probabilities	< 2 ppb	2-15 ppb	15-700 ppb
Green Corridor Project	68%	13%	19%
Simulation	69%	19%	12%

DISCUSSION

The good accordance between incidences of early split and mouldy pistachio categories used for the simulation before sorting and literature data was tested for further validation. Simulation results were compared to contamination levels reported by Shakerardekani *et al.* (2012) before sizing, grading, separation of non split pistachios and manual sorting. In this latter publication, 3 kg of Ahmad-Aghaei and 4 kg of Kale-Ghoochi pistachio samples before the grading stage, were split into categories, among which yellow-brown stain and dark-grey stain pistachios. We calculated the total proportion of pistachios having stain defects (after sizing and grading). Aflatoxin concentrations in the two samples were computed, considering that the contamination only came from these stain defects pistachios, and that tiny, damaged and non-split pistachios were removed by automatic sorters during processing. On order to put these data in parallel with our simulation data, we calculated global incidences of early split and mouldy pistachio categories for each of the lot classes, before sorting. Table 8 enables to observe that incidence levels for Ahmad-Aghaei pistachios are very close to 3rd class lot simulation data. In addition, incidence levels for Kale-

Ghoochi pistachios are very similar to 1st class lot data. For this reason, the concentration result for Kale-Ghoochi pistachios was compared to the simulated concentration distribution for the 1st class lots in Table 9.

Table 8: Total incidence of stained and mouldy-like pistachios

From Shakerardekani *et al.* (2012)		Our simulation	
Ahmad-Aghaei	Kale-Ghoochi	1st class lot	3rd class lot
7.08%	2.35%	3.00%	6.75%

Table 9: Aflatoxin concentration results (ppb)

From Shakerardekani *et al.* (2012)		Our simulation	
Ahmad-Aghaei	Kale-Ghoochi	1st class lot	3rd class lot
28.73	1.44	Mean: 1.35 90% CI: [0; 5]	Mean: 151 90% CI: [6; 572]

The same process was achieved between concentrations for Ahmad-Aghaei pistachios and the distribution for 3rd class lots. We could deduce that experimental data were in the range of the 90% Confidence Interval (CI) for the simulation. The model implied that manual sorting enables to divide both the incidences of early split and of mouldy pistachio categories by 3. These categories consist mostly of dark stained pistachios, easily recognizable from unstained pistachios. For this reason, they are readily removed by workers on the processing line. We found that the proportion of stained pistachios removed would be the same for any quality level for the mean concentration value of the lot (1st class, 2nd class, 3rd class). As simultated results fitted observations well, we propose that stained pistachios that are not removed consist of the proportion of pistachios that are not properly oriented on the line, it says not in a way for workers to see them. More precisely, the stain on the shell is placed on the lower layer or is hidden by other pistachios, etc. Our suggestion in order to enhance the quality level of contaminated lots is to assess the proportion of dark stained pistachios removed during sorting for a lot and then, in case it is above 2%, to improve the quality of the lot by achieving a second sorting step. Knowing that 70% of lots have less than 2% of dark stained pistachios removed, this would only imply further processing on 30% of lots, thus not being too costly.

CONCLUSION

Our results validate the use of the Monte Carlo technique upon aflatoxin contamination data for various pistachio categories, as well as for pistachio samples before and after final sorting. We found that early split and mouldy pistachios have the higher impact on final contamination levels. Moreover, final sorting enables to divide these categories incidences by 3. For this reason, we suggest to make a second final sorting on the 30% of lots having more than 2% of dark stained pistachios removed during the (first) sorting step. However, the incidence of the three pistachio categories as function of the lot mean concentration is going to be further validated in a near future.

REFERENCES

Bonjar, GHS. (2004). Incidence of aflatoxin producing fungi in early split pistachio nuts of Kerman, Iran. *Journal of Biological Sciences* 4 (2): 199-202.

Doster, MA., Michailides, TJ. 1994. Aspergillus molds and aflatoxins in pistachio nuts in California. *Postharvest Pathology and Mycotoxins* 84(6): 583-590.

Hadavi, E. 2005. Several physical properties of aflatoxin-contaminated pistachio nuts: Application of BGY fluorescence for separation of aflatoxin-contaminated nuts. *Food Additives and Contaminants* 22(11): 1144–1153.

Regulation (EU) No 165/2010 of 26 February 2010 amending Regulation (EC) No 1881/2006 setting maximum levels for certain contaminants in foodstuffs as regards aflatoxins. *Official Journal of the European Union.*

Regulation (EU) No 178/2010 of 2 March 2010 amending Regulation (EC) No 401/2006 as regards groundnuts (peanuts), other oilseeds, tree nuts, apricot kernels, liquorice and vegetable oil. *Official Journal of the European Union.*

Schatzki, TF., Pan, JL. 1996. Distribution of aflatoxin in pistachios. 3. Distribution in pistachio process streams. *Journal of Agriculture and Food Chemistry* 44:1076–1084.

Shakerardekani, A., Karim, R., Mirdamadiha, F. 2012. The effect of sorting on aflatoxin reduction of pistachio nuts. *Journal of Food, Agriculture & Environment* 10 (1): 459-461.

Sommer, NF., Buchanan, JR., Fortlage, RJ. 1986. Relation of early splitting and tattering of pistachio nuts to aflatoxin in orchards. *Phytopathology* 76:692–694.

UNECE. 2001. *Revision of UNECE standards inshell pistachio nuts, Transmitted by Turkey*, TRADE/WP.7/GE.2/2001/6.

WEB REFERENCES

INC, International Nut and Dried Fruit Council. Green Corridor Project. Available on line: http://www.nutfruit.org/en/green-corridor-iranian-pistachio-project_36887

ACKNOWLEDGEMENTS

Authors would like to thank to mexican CONACyT for the Financial support given through the post doctoral fellowship of A. Ramirez-Martinez.

AUTHOR BIOGRAPHY

NATHALIE WESOLEK was born in Strasbourg, France and after a technical degree in Food engineering and Biology at a College of Advanced Technology, she moved to Brest in order to graduate at a Microbiology and Food Safety Engineering School: ESMISAB. Since the end of 2009, she has been recruited at the Université de Bretagne Occidentale (UBO) in Brest, on sampling schemes performance evaluation for chemical contaminants in food and exposure assessment. She is now achieving a PhD on the application of Bayesian methods as a mean of evaluating mycotoxins sampling schemes in tree nuts.

PRODUCTION OPTIMIZATION

USING SHEWHART CONTROL CHARTS FOR CONTINUOUS MODEL OPTIMIZATION

Rodrigo Rezende Amaral
Peter Ottoy
Veronique Limère
Department of Industrial Management
Ghent University
Technologiepark 903, 9052 Zwijnaarde - Belgium

KEYWORDS
Continuous Improvement, Statistical Process Control (SPC), Shewhart Control Charts

ABSTRACT

This paper intends to provide a view on different possibilities of using Shewhart Control Charts in order to detect sources of imprecision or inaccuracy in models for industrial processes. In the classical approach, control charts are commonly used to detect abnormal variations of a single process variable, which can be the error of a process model. The techniques explored in this work help linking this error to other variables from the model, allowing a fast detection of incorrect parameterization.

INTRODUCTION

The current operational standards of most industrial processes do not allow companies to predict or drive process variables exactly to predefined values. Some of the reasons for this are:

- Insufficient or insufficiently implemented knowledge of the processes
- Inadequate, inaccurate or imprecise measurement of process variables
- Unexpected machine behavior (lack of maintenance, improper use, inappropriate design etc.)
- Non-conforming raw materials
- Incorrect execution of planned actions (human error)
- Influence of external environment

In order to achieve satisfactory productivity levels, companies must apply methodologies to improve and sustain the levels of accuracy and precision in their processes. Statistical Process Control (SPC) is one of these methodologies. Among the tools deployed for the implementation of SPC are the Control Charts.

Control Charts

Originally developed by Walter A. Shewhart and also known as Shewhart Charts, the Control Charts are used to determine whether a process is in statistical control. In short, a control chart consists of points plotted in chronological sequence representing the individual measurements of a process variable or a statistic calculated by grouping those measurements into so-called rational subgroups. The statistic can be, for instance, the mean, the range or the standard deviation of the subgroups.

Besides the points, three horizontal lines are also calculated and plotted. They are called *centerline* (CL), *upper control limit* (UCL) and *lower control limit* (LCL). The centerline is set equal to the expected value of the points. The control limits are usually set equal to the centerline plus and minus three theoretical standard deviations of the plotted values.

The plotted statistic determines the control chart type. The most used chart types are summarized in Figure 1. Not all of them were included in the original set proposed by Shewhart. The *Cumulative Sum Chart* (CUSUM), for instance, was developed by E. S. Page, and the *Exponentially-weighted Moving Average Chart* (EWMA) by S. W. Roberts.

The choice of the chart type and the control limits depends on the type of data, the expected behavior of the process, the kind of process variation which it is supposed to detect, among others. These design issues are discussed into more detail in Benneyan (1998). NIST/SEMATECH (2012) explains how the points and limits should be calculated for each chart type.

Chart type	Plotted statistic
X-chart (or I-chart)	individual measurements
mR-chart	moving range of individual measurements
mA-chart	moving average of individuai measurements
Xbar	subgroup means
R-chart	subgroup ranges
s-chart	subgroup standard deviations
EWMA-chart	exponentially weighted moving average
CUSUM-chart	cumulative sum of ind. measurements or averages
c-chart	count of defects in subgroup
u-chart	defects per unit in subgroup
np-chart	number of detective units in subgroup
p-chart	proportion of defective units in subgroup

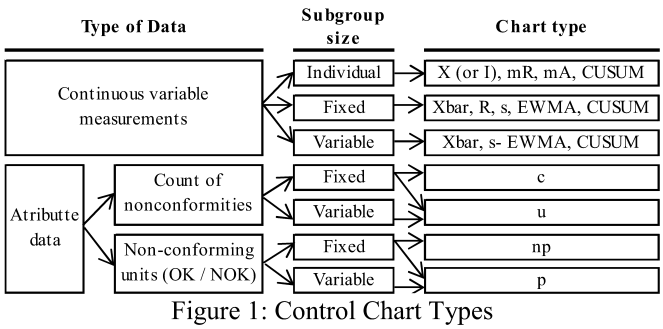


Figure 1: Control Chart Types

An example of an X-chart can be observed in Figure 2. The plotted variable is the difference between the actual and the expected duration of a certain operation in a batch process. Each point represents one batch. A value above zero means that the operation took longer than expected and a value below zero means that it took shorter. The distance of a point to the centerline indicates the time difference in seconds.

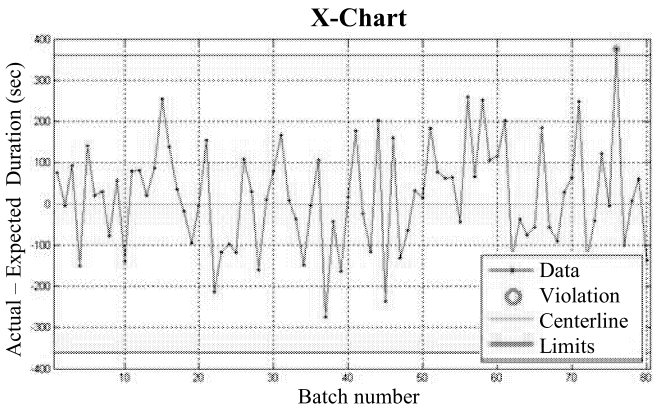


Figure 2: Example of an X-Chart

Control Rules

The general rule used to evaluate control charts states that a point out of the control limits is an indication that the process might be out-of-control. The reason for this is that, assuming the points follow a normal distribution, only about 0.27% of them are expected to lie outside the limits ($\mu \pm 3\sigma$) when the process is under statistical control. Thus, whenever it happens, the existence of special causes for variations should be investigated. For an in-control process, the last mentioned percentage will be false alarms.

Some authors and organizations have created extra rules to detect trends or systematic patterns on the charts. In Western Electric Co. (1956) the quality staff of the Western Electric Company defined a set of rules known nowadays as Western Electric Rules. Instead of having only one upper and one lower control limit, they first divided the chart into three zones, as described in Table 1.

Table 1: Chart zones for Western Electric Rules

Zone A	Within $\pm 2\sigma$ of the centerline and $\pm 3\sigma$
Zone B	Within $\pm 1\sigma$ and $\pm 2\sigma$ of the centerline
Zone C	Within $\pm 1\sigma$ of the centerline

Next they considered that an out-of-control alarm should be given whenever one of these situations occurs:

1. Any single point falls out of zones A, B or C (out of the control limits);
2. Two out of three consecutive points fall in or beyond zone A on the same side of the centerline;
3. Four out of five consecutive points fall in or beyond zone B on the same side of the centerline;
4. Eight consecutive point fall on the same side of the centerline.

Figure 3 shows how these situations could appear on a chart.

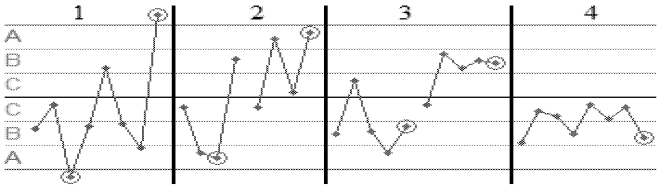


Figure 3: Western Electric Rules

Western Electric has also suggested other rules which depend on the type of data and chart, on the probability distribution and on the kind of variation that should be detected.

The introduction of control rules helps identifying abnormal behavior, but also increases the rate of false alarms, since the described situations can occur even when the process is under statistical control. When designing control charts, the costs of checking and acting upon false alarms and the costs of not detecting an out-of-control situation should be considered. Therefore an important design parameter is the *in-control average run length* (ARL), which is the number of points that are plotted on a control chart before a false out-of-control condition is indicated.

Control Charts, Modeling and Simulation

In the context of modeling and simulation, control charts can be used to monitor the correspondence between model calculations and the real process. For instance, the expected duration of an operation can be the output of a simulation exercise. As another example, the final temperature of a batch in a chemical process can be estimated using a model which considers the amount and the properties of each raw material used, the machine configuration, the duration of the process, etc. In both cases, it can be very useful to use a control chart to follow up the difference between the actual and the estimated value, i.e. the model error.

This approach can work in different ways. The estimated values can be seen as a reference for the process variables. A control chart would detect when the actual values begin to deviate from this reference, which indicates that an unexpected change has occurred and should be investigated. One possibility is that there is an operational problem which should be eliminated. Another possibility is that there were changes on variables whose effects were not correctly estimated by the model. In this case, the information can be used to improve the model and increase the predictability of the process.

However, as it will be demonstrated further in this paper, following up only the model error can be insufficient to detect variations arisen from the imprecision or inaccuracy of the model. After a discussion on this subject, solutions addressing this issue are presented.

PROBLEM DESCRIPTION

A control chart displays the evolution of a single variable, which is normally the output of a process, or, as stated in the end of last section, the difference between the actual value of a variable and the value estimated by a model. Nevertheless, the output of a process (and of a model) usually depends on

many input variables. These variables can be categorical (e.g. machine type, product class, raw material supplier etc.) or continuous (e.g. amount of raw material added to the process, duration of an operation, external temperature etc.). The effect of each variable on the output of the model is described by a model parameter.

Although having models for the processes provides many advantages for an enterprise (predictability, stability, etc.), building and keeping them updated can be very costly. Collecting and analyzing data or re-evaluating the models every time there is a change in the process are examples of activities which incur non-negligible costs.

The balance between the costs and benefits of model accuracy is then a goal to be achieved. It means that, even in an optimal solution from the business point of view, some model parameters can be incorrectly estimated. For instance, collecting the adequate amount of data to determine all parameters with good precision can be too expensive. The economic solution is to allow some imprecision to the model.

The wrong estimation of a single parameter, however, is not always easily detected by following up the model error alone. In the next subsections, the reasons for it are discussed. For a better understanding, the distinction between parameters of categorical and continuous variables is made.

When the effect of a categorical variable is incorrectly estimated

The chart presented in Figure 2 shows the difference between the actual and the expected duration of a certain operation for each batch in a process. Consider that the value of the expected duration is the output of a model which takes into account the product type, the machine where the product was processed, among others. Product types and machines can be enumerated and therefore constitute categorical variables. Each existing product type or machine is a category entry.

Now suppose that the model slightly underestimates the duration of the operation for one of the product types. The average value of the plotted variable for this category entry will be higher than the others. Figure 4 illustrates such a situation. The marked points represent the batches from the product type in question.

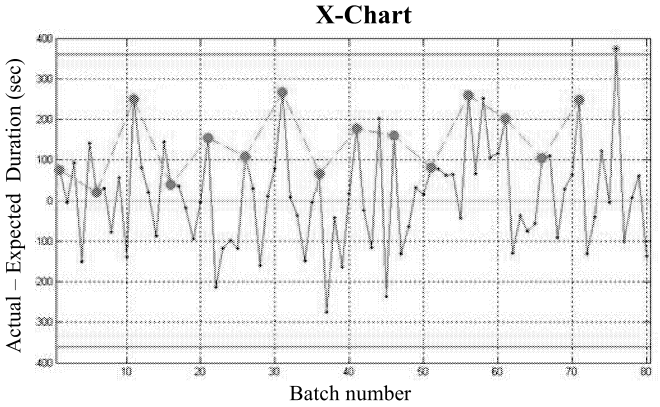


Figure 4: Model underestimates the output for a product type

As batches from different types can be mixed on the chart, the control rules suggested by Western Electric would not indicate an out-of-control situation, since they are defined based on combinations which can occur with consecutive points. The incorrect estimation of the parameter linked to the product type is thus not detected.

This problem is even more difficult to detect in charts with grouped data (e.g. Xbar-chart, p-chart) or with auto-correlated points (e.g. mA-chart, CUSUM-chart). In the first case, the extra difficulty arises because each subgroup may contain data from different category entries and in the second case because the value of each point depends on the values of previous points. In both cases, it is not possible to link an individual point to a unique category entry.

When the effect of a continuous variable is incorrectly estimated

The model considered in the last example would probably use continuous variables as input as well. For instance, the amount of each raw material added in the process would certainly have an impact on the duration of the operation and should therefore be included in the model.

Suppose that the effect of the quantity of one raw material on the duration of the operation is underestimated by the model. In this case, the model error would tend to be higher when more of this raw material is added.

However, as the amount of each raw material added in a batch can vary, the incorrect estimation may not be detected by the chart. This can be explained considering Figure 4 again. The marked points would now represent batches in which a higher quantity of the raw material in question was added. The values of the marked points have indeed a higher average than the other, but, again, no alarms are given when the control rules are applied.

Detecting the problem with other statistical tests

At this point, a careful reader may suggest that statistical techniques such as ANOVA could be used to detect the incorrect estimation of the effect of a categorical variable, or that multiple regressions could be used with the same objective for continuous variables. ANOVA would detect whether the average model error by different groups are statistically different from each other, while multiple regressions would show if the prediction error is dependent on one or more continuous variables.

In practice, these approaches may fail because the deviations caused by the incorrect estimation of a model parameter may not always be present but start at some point in time. For instance, the specifications of a product type can be changed but the consequences of that for the duration of the process not immediately perceived by the process engineers. In such a situation, the model would not be modified and from that point the deviations would start to appear. However, the process engineers would hardly think of it when applying ANOVA or multiple regression techniques.

PROPOSED SOLUTION

The solution proposed by the authors consists in the generation and analysis of charts for each model parameter separately. For complex models, this would generate a large number of charts, and in most cases it would become unfeasible or unprofitable to assess each one of them manually. Therefore, an algorithm based on the CUSUM-chart proposed by E. S. Page was developed in order to make an automated analysis of the generated charts. The output of the whole procedure is only a list of parameters which should be reviewed.

The intention is not to replace the existing monitoring systems, but to run in parallel with them. The suggested automated analysis can be implemented as an add-on feature to the computer programs already used for SPC in an organization. It is important, however, that the calculations are continuously made, in the same way as for the common charts. This allows the constant evaluation of the influence of the model parameters on the error.

In this section, the generation of the charts and the algorithm used to automatically assess them are described.

Generating charts for categorical variables

In order to generate charts for parameters related to categorical variables, the original data should be iteratively filtered by the entries of each categorical variable, and for each of those entries a control chart should be made. In the previous example, one X-chart would be generated for each product type, each machine, etc. Table 2 is an example of how data would be filtered following this procedure.

Table 2: Example of how data is filtered by category entries

Original data			
Batch order	**Model error (sec)**	**Product type**	**Machine**
1	91	A	1
2	-2	B	1
3	99	B	2
4	-149	C	1
5	122	A	2
6	-82	A	2
7	62	B	1
8	-133	B	1
9	221	B	1
10	-92	C	2
11	28	C	1
12	-89	B	2
13	125	A	2
14	-17	A	2
15	-156	A	1

Data to be plotted on the new charts (1 column = 1 chart)				
Product A	Product B	Product C	Machine 1	Machine 2
91	-2	-149	91	99
122	99	-92	-2	122
-82	62	28	-149	-82
125	-133		62	-92
-17	221		-133	-89
-156	-89		221	125
			28	-17
			-156	

Generating charts for continuous variables

When a parameter related to a continuous variable is correctly estimated, there should be no correlation between the model error and the variable. As a consequence, the expected value for the linear regression coefficient between these variables is zero. If successive regressions of the model error on the continuous variable are calculated every n batches using the Ordinary Least Squares method (OLS), it is expected that the linear regression coefficients will be a random variable following a normal distribution $N(0, \sigma_\varepsilon / \text{sqrt}(n \text{ Var}(X)))$, where X is a vector containing the values of the model error and σ_ε is the standard deviation of the regression error.

Thus, in order to check if the parameter is – and stays – correctly estimated, the authors suggest the follow up of these regression coefficients on a chart.

Table 3 shows an example of how this could be done. The regression coefficients calculated every 50 batches should be the plotted variable.

Table 3: Example of Calculations for Linear Regression Coefficient Follow Up

Batch order	**Model error (sec)**	**Raw material 1**		**Raw material 2**	
		Individual value (kg)	**Linear Regression coefficient (sec/kg)**	**Individual value (kg)**	**Linear Regression coefficient (sec/kg)**
1	91	101		31	
2	-2	121		25	
3	99	115	-0.006	29	0.008
...	...	...		...	
50	-137	98		33	
51	52	99		31	
52	-4	71		27	
53	-80	111	0.011	28	-0.012
...	...	...		...	
100	211	97		35	
101	-35	87		24	
102	138	79		29	
103	-47	123	-0.002	33	-0.007
...	...	...		...	
150	-248	115		27	
151	76	105		25	
152	33	99		31	
153	-15	88	0.007	26	0.001
...	...	...		...	
200	-110	116		17	

The Adapted CUSUM Algorithm

The CUSUM-chart proposed by E. S. Page is used to detect small but persistent shifts of a variable. To achieve this objective, the variables C^+ and C^- are created according to the following equations:

$$C^+_i = \max[\, 0 \,,\, x_i - (T + K^+) + C^+_{i-1} \,]$$

$$C^-_i = \max[\, 0 \,,\, (T - K^-) - x_i + C^-_{i-1} \,]$$

$$C^+_0 = C^-_0 = 0$$

X is the vector containing the values of the process variable. In the present discussion, it would be the values of the model error filtered by categorical variables or the calculated regression coefficients. T is the target mean for the variable, which should be zero in both cases. K^+ and K^- are called *threshold* parameters and are commonly set as $0.5\sigma_X$.

For a given *i*, C^+ will increase if the difference between T and x_i is greater than K^+, and decreases otherwise, but not below zero. The slope of increasing or decreasing is $x_i - (T+K^+)$. Analogously, C^- will increase if the difference between T and x_i is smaller than $-K^-$, and decreases otherwise, but not below zero. The slope, in this case, is $(T - K^-) - x_i$.

If the process is stable, both C^+ and C^- tend to become zero after eventual increases. However, if the process mean is shifted, one of them will indefinitely increase (C^+ in case of a positive shift and C^- in case of a negative shift). Based on this fact, two control limits are defined: h^+ and h^-. If C^+ is greater than h^+ or C^- is greater than h^-, an out-of-control alarm is given. The limits h^+ and h^- are commonly set as $5\sigma_X$.

Although the CUSUM-chart accomplishes the objective of detecting small shifts of a variable, it presents some limitations which restrain its real applicability. The first limitation is that the algorithm does not provide formal means to detect when a temporary out-of-control process comes back to normal. If a process is positively shifted for a long period, for instance, the variable C^+ will steadily increase, becoming much higher than h^+. When the shift is over, the value of C^+ tends to decrease, but it can take long until it is under h^+. In this period, the algorithm neither informs if the process is in control nor is able to detect a new shift.

The second limitation lies in its lack of robustness against outliers. A single outlier can increase the variables C^+ or C^- to values greater than their respective limits and sets an unnecessary alarm. Albeit it also happens when using other charts, it is an especially negative factor for the CUSUM-charts, since they are mainly designed to detect small and persistent shifts.

To illustrate these limitations, the following data series was generated:

- 30 points following a normal distribution N(0,1)
- 70 points following a normal distribution N(1,1)
- 15 points following a normal distribution N(0,1)
- 1 outlier with value -6
- 30 points following a normal distribution N(0,1)

Considering this data series corresponds to a process variable which has a normal distribution N(0,1) when in-control, an X-Chart and a CUSUM-chart were created and are presented in Figure 5. The shift starting on the 31st point is detected, but the values of C^+ remain above h^+ after the process has returned to control, what happens around the 101st point. Furthermore, the single outlier on -6 (116th point) raises C^- to a value above h^- and sets the alarm for negative shifts.

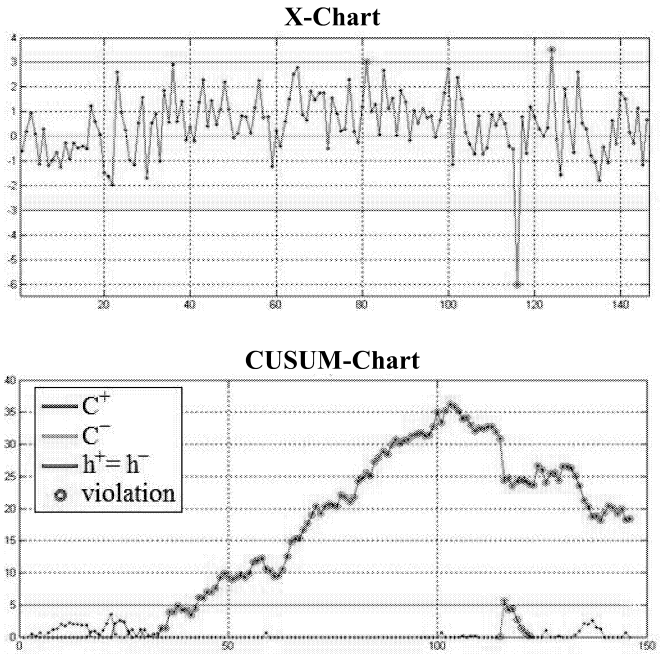


Figure 5: X-chart and CUSUM-chart

Two major changes on this algorithm were proposed by the authors addressing those limitations. The objective of the first major change was to allow the algorithm to detect when a temporary out-of-control process comes back to control. In order to accomplish this objective, the calculation of C^+ and C^- was slightly modified by the introduction of two new variables and two new parameters:

$$C^+_i = \max[\, 0 \,,\, x_i - (T + K^+) + D^+_i \,]$$

$$C^-_i = \max[\, 0 \,,\, (T - K^-) - x_i + D^-_i \,]$$

$$D_i^+ = \begin{cases} C^+_{i-1} & \text{if } C^+_{i-1} < h^+ \\ R^+ * h^+ & \text{otherwise} \end{cases}$$

$$D_i^- = \begin{cases} C^-_{i-1} & \text{if } C^-_{i-1} < h^- \\ R^- * h^- & \text{otherwise} \end{cases}$$

$$C^+_0 = C^-_0 = D^+_0 = D^-_0 = 0$$

The parameters R^+ and R^- could be called *reset factors* and should be in the interval (0,1]. The authors suggest $R^+ = R^- = 0.75$. This does not allow the values of C^+ and C^- to reach extreme values even when the shift on the process persists for a long period. However, since these variables are reset to values smaller than h^+ and h^-, it is not possible to define if the process is out-of-control only by comparing the value of a point with the limits. The workaround for this new problem is to change the alarming strategy. The alarm is set on when C^+ (or C^-) becomes higher than h^+ (or h^-) and only set off only when C^+ (or C^-) becomes zero.

The objective of the second major change was to increase the robustness of the algorithm against outliers. This is done by limiting the increase of C^+ and C^-. The complete algorithm is described by the following set of equations:

$$C^+_i = \max[\,0\,,\, G^+_i + D^+_i\,]$$

$$C^-_i = \max[\,0\,,\, G^-_i + D^-_i\,]$$

$$D_i^+ = \begin{cases} C^+_{i-1} & \text{if } C^+_{i-1} < h^+ \\ R^+ * h^+ & \text{otherwise} \end{cases}$$

$$D_i^- = \begin{cases} C^-_{i-1} & \text{if } C^-_{i-1} < h^- \\ R^- * h^- & \text{otherwise} \end{cases}$$

$$G^+_i = \min[\,x_i - (T + K^+)\,,\, L^+ * h^+\,]$$

$$G^-_i = \min[\,(T - K^-) - x_i\,,\, L^- * h^-\,]$$

$$C^+_0 = C^-_0 = D^+_0 = D^-_0 = 0$$

The parameters L^+ and L^- could be called *maximum increase factors* and should be in the interval [0,1). The authors suggest $L^+ = L^- = 0.5$. As a consequence of this modification, the maximum values by which C^+ and C^- can be increased are respectively $(L^+ * h^+)$ and $(L^- * h^-)$. This significantly reduces the probability of an alarm being set by a single outlier.

Figure 6 shows a chart created using the adapted algorithm on last example's data. As previously, the positive shift is detected, but now the alarm is set off a few points after the process returns to its normal behavior. Besides that, the single outlier does not raise any alarm.

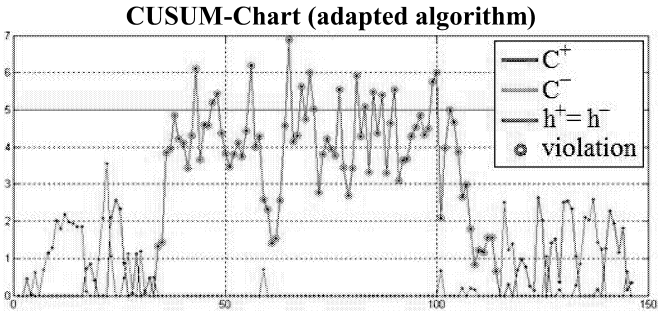


Figure 6: CUSUM-chart created using the adapted algorithm

Using the Adapted CUSUM Algorithm to assess the generated charts

After generating charts for categorical or continuous variables (in the first case by filtering the data accordingly to the categories and in the second case by calculating the regression coefficients), the adapted CUSUM algorithm should be applied on the data of each of them. If the CUSUM alarm is active for the last point of a chart, the model parameter related to it should be reviewed.

There is an exception for the case of categorical variables. If a CUSUM alarm from the same type (positive or negative) is active in all charts created for a category (e.g. charts of all product types), the model parameters linked to this category do not have to be reviewed. In such a situation, the process is indeed out-of-control (shifted), but this is independent of the categorical variable.

Another remark when dealing with categorical variables is that category entries for which not many points are available should not be considered. The reason for this is that the Adapted CUSUM algorithm requires a reasonable sample size to raise an alarm. The authors recommend that only category entries with more than 20 data points should be included in the analysis.

It is important to emphasize that the whole procedure, i.e. the generation of the charts, the execution of the Adapted CUSUM algorithm and the assessment of the outcome, can be continuously and transparently performed by computer systems. The final output is only a list of parameters which should be reviewed.

REAL APPLICATION OF THE SOLUTION

The techniques described in the previous section were applied in the steelshop of a large steelmaker in Belgium, whose production was around 4.7 million tons of steel in 2012. Details about the implementation of the computational tools and examples of processes which were improved can be found in Amaral (2012). In this section, an example extracted from this reference is presented.

In general, an integrated steel plant combines several production installations in one site. The steelshop is the department provided with equipment to transform the hot metal produced by the blast furnaces into steel. One of the steps required for this is to blow oxygen into a mixture of hot metal and metal scrap in order to decrease its carbon level. This is a typical batch process: the mixture is placed into a vessel with capacity around 300ton and oxygen is blown during around 17 minutes. The mixture is then cast into a ladle for further processing and the vessel is filled in again.

This exothermic process strongly increases the temperature of the mixture. At the company in question, the average temperature of the hot metal is around 1250ºC and at the end of the oxygen blowing it reaches around 1650ºC. The exact achieved temperature of each batch is very important for the logistics and costs of the steel plant. It has to be in an acceptable range for the further processing of the steel. If it is too low, the process has to be repeated, causing production delays and extra costs. If it is too high, it is necessary to reduce it by adding cooling material, what causes non-negligible energy and material losses. In extreme situations, very high temperatures can break the equipment, which can lead not only to huge economic losses but also to risks to the safety of employees and installations.

In spite of the importance of achieving the desired temperature with good precision, the current standards of the process do not allow its online measurement and control. Only the initial temperature is measured and a model is used to determine how the process parameters (amount of oxygen to be blow, duration of the blowing process, amount of cooling material etc.) should be arranged in order to achieve the aimed temperature.

The effect of the cooling material on the temperature is one of the factors included in the model. For each type of material, it is described by a cooling coefficient, usually expressed in °C/(kg of added material/ton of steel). For instance, if 1200kg of determined material whose cooling coefficient is 2°C/(kg/ton) is added to a batch of 300ton, the temperature is expected to decrease approximately 8°C (=2·1200/300).

In the past, the correctness of the estimation of the cooling coefficients was periodically checked by the process engineers. They gathered data from recent operations and made the regression between the model error and the amount of each type of cooling material added to the batches. If the linear coefficient of this regression was significantly different than zero, the cooling coefficient should be reviewed.

Depending on the workload of the department, this periodic review could take weeks or months, during which the incorrectness of the estimation could cause considerable losses. The graphs presented in Figure 7 are examples of how it actually happened. The first graph, made using data from November 2011, shows that the model error was independent of the amount of cooling material from the certain type (*slibbri*). This indicates that the cooling coefficient for this material was correctly estimated. However, the second graph, made using data from April 2012, shows a different situation: the model error was higher (in absolute terms) when more *slibbri* was added.

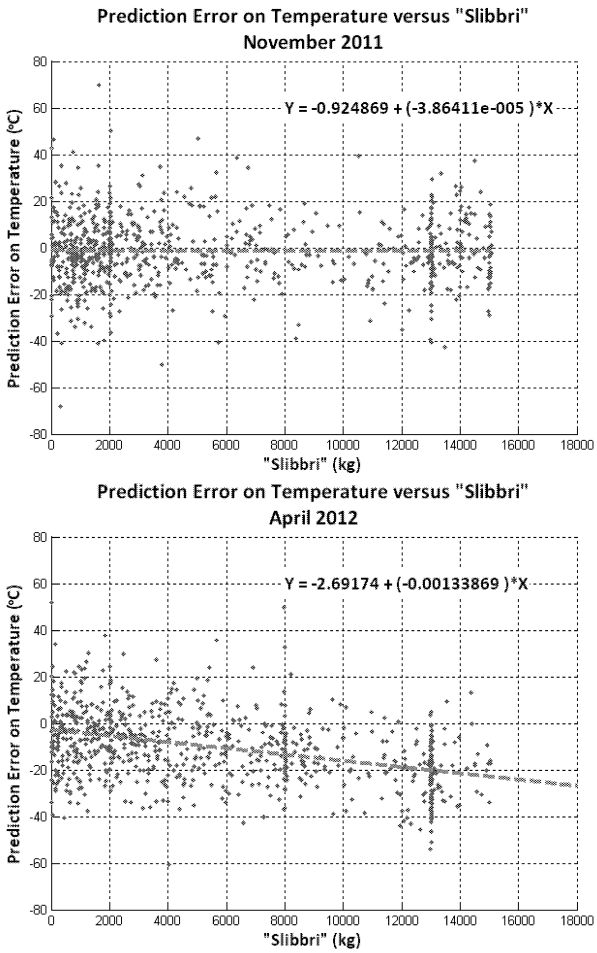


Figure 7: Model error vs. Amount of Cooling Material

The reason for this is that the constitution of the material had been changed by the supplier but the effect of this change was unnoticed by the process engineers. In other words, the actual cooling coefficient had changed, but the model was not adapted. As a consequence, the achieved temperatures began to be lower than expected, especially when higher quantities of *slibbri* were added, resulting in higher inaccuracy and imprecision (i.e. economic losses). Since the amount of *slibbri* added to each batch varies and in many cases it is relatively small, the problem could not be detected by charts which followed the model error only. Although the change happened in January 2012, it was only perceived during a model review made in May 2012.

The follow-up of regression coefficients was the solution implemented in the company to avoid this problem. If this approach had been applied before, the problem described above would have been detected promptly, as shown in Figure 8. Each point on the chart represents a linear regression coefficient of the model error on the amount of cooling material. The regressions were made each 100 batches. The Adapted CUSUM Algorithm would have given an alarm after the second point below the LCL.

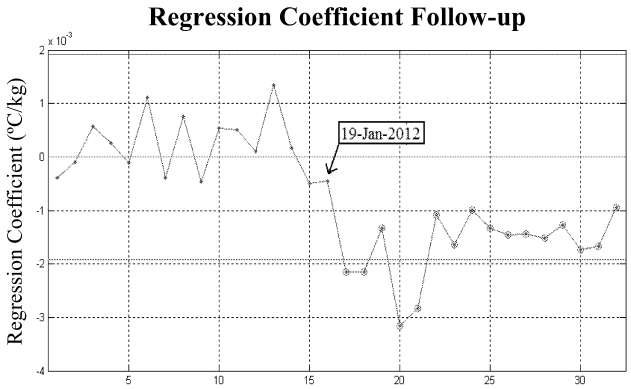


Figure 8: Regression Coefficient Follow-up

CONCLUSION

The techniques presented in this work have a great potential to enhance Statistical Process Control in industrial activities. The proposed solution combines well-known statistical methods and although a few modifications were introduced, incorporating the presented techniques to the current control systems in enterprises where the SPC concepts are maturely understood and applied would not be costly.

The proposed methods were recently applied in a considerably large industrial plant with very satisfactory results. In most cases, like in the example given in the last section, the implementation of the solution was relatively simple and inexpensive. It should be highlighted, however, that it was only possible because the company was already provided with good control and measurement systems, computer programs for statistical calculations, mathematical models and competent personnel.

The automatic correction of model parameters is a point that could be studied in further developments. This could be, for instance, triggered by the alarms given when using the

proposed methods. The choice of the parameters for the Adapted CUSUM Algorithm and their impact on the *in-control average run length* (ARL) constitute another issue which could be further studied.

REFERENCES

Amaral, R. R. 2012. "A new SPC tool in the steelshop at ArcelorMittal Gent designed to increase productivity" Master Dissertation. Faculty of Engineering and Architecture, Department of Industrial Management, Ghent University

Benneyan, J. 1998. "Technical Aids". *International Journal for Quality in Health Care*, 1998. Volume 10, Number I: pp. 69-73.

Western Electric Co., Inc., 1956. "Statistical Quality Control handbook" 2nd ed. Easton: Mack Printing Company, 1956.

WEB REFERENCES

NIST/SEMATECH, 2012. "e-Handbook of Statistical Methods", *http://www.itl.nist.gov/div898/handbook/*, accessed on 31st January 2013

DATABASE-DRIVEN 3D SIMULATION
A METHOD SPECIFICATION USING THE UML METAMODEL

Martin Hoppen
Michael Schluse
Juergen Rossmann
Institute for Man-Machine Interaction
RWTH Aachen University
D 52074 Aachen
Germany
E-mail: hoppen,schluse,rossmann@mmi.rwth-aachen.de

KEYWORDS
Meta, Object-oriented, Virtual Reality, Software engineering, Database management systems

ABSTRACT

Many benefits derive from combining database technology with 3D simulation systems and similar applications. Nevertheless, only few systems follow up this approach so far. Compared to them, our approach for runtime database synchronization uniquely provides a tight integration of real-time 3D simulation and object-oriented database technology including persistence, collaboration, arbitrary data schema, and versioning. To enable its adoption, we previously stated the necessary metamodel-based requirements. These suggest to use the UML as a pivot metamodel with model transformations to and from a simulation database and an external database. This guarantees the underlying UML modeling concepts (e.g. class, property, association, instance specification) to implicitly exist in both databases. It also allows to indirectly refer to the databases' metamodels by using the corresponding UML metaclasses. This is the basis for our new generalized method specification presented in this paper, which is independent from actual database implementations.

INTRODUCTION

For our research and development, database technology already offers many benefits for 3D simulation in different fields of application like industrial automation applications (Schlette and Hoppen 2012), Virtual Testbeds for space missions (Rossmann et al. 2011b), or geoinformation systems for city and forestry scenarios (Rossmann et al. 2011a), see Figure 1. As shown in (Hoppen et al. 2012), an efficient integration can be achieved by coupling a 3D simulation system's internal runtime database (*SimDB*) with an external database (*ExtDB*). ExtDB provides persistence and versioning, a common application schema, and an integrated communication hub for collaboration or distributed simulation. SimDB however provides a real-time capable cache and drives the 3D simulation itself.



Figure 1: Exemplary applications realized using the presented approach (city simulation, planetary navigation, forestry simulation, simulated planetary landing mockup).

The basic idea for this flexible yet efficient database coupling is to synchronize the two systems during runtime. For a pair of SimDB and ExtDB, synchronization is performed on schema, data and functional level. On system startup the schema description of ExtDB is synchronized to SimDB so both "speak the same language". During runtime, data is replicated from ExtDB to SimDB and kept in sync using data synchronization. Local changes within SimDB are made persistent within ExtDB where a versioning mechanism may preserve previous values. In case of a multi-client setup, this implies local changes being distributed via the central database to all connected clients. Furthermore, application schemata like SEDRIS or CityGML may en-

code functionalities necessary for the simulation. Functional synchronization is a means for making this simulation logic (e.g., geometries, transformations, dynamic parameters) locally available within SimDB by translating it to structures with semantics known to the simulation system.

A core concept of this architecture is decoupling the simulation system from ExtDB, hiding it from all its internal components. ExtDB's schema as well as its instance data can only be accessed by means of replicas within SimDB. Besides a general separation of concerns, this transparency makes SimDB work as an "intelligent" cache for ExtDB. It provides real-time data access for simulation as well as visualization and speeds up repeating access patterns, which would otherwise lead to repetitive queries to ExtDB.

Synchronization requires both systems to be compatible to a certain extent (Hoppen et al. 2013). A metamodel can describe the abstract syntax of a database's modeling language, i.e. how schema constructs and their instances can be modeled. The compatibility of two metamodels can be determined by a transformation in between using a model transformation language. This can also be expressed by a pivotal metamodel with transformations to and from SimDB's and ExtDB's metamodel. Requiring a pivot can be used to enforce its modeling standards on both databases. In the context of 3D simulation, object oriented modeling is well suited for the typically hierarchically structured data. Thus, the Unified Modeling Language (UML) (OMG 2013a) is a reasonable pivot. Besides the metamodels, certain requirements also apply to the databases' runtime systems, like a reflection interface or a change notification service.

Furthermore, a pivotal metamodel can be used to make indirect statements about the databases' metamodels. With UML as a pivot, UML metaclasses like Class, Association and Property can be used to address the corresponding metaclasses within SimDB and ExtDB. In this paper we now use this approach to specify our synchronization methods in a generalized form, based only upon the aforementioned set of requirements and described in terms of the UML metamodel. This identification of universal methods decoupled from specific choices of SimDB and ExtDB allows for their broad usage within different systems.

The rest of this paper is organized as follows: In the next section we summarize the requirements for SimDB and ExtDB. Section METHOD SPECIFICATION establishes the basis for the new method specifications for schema (Section SCHEMA SYNCHRONIZATION) and data (Section DATA SYNCHRONIZATION) synchronization. Section RELATED WORK introduces work related to our approach. Finally, Section FUTURE WORK AND CONCLUSION delineates next steps and gives a conclusion.

REQUIREMENTS

As stated in the introduction, there are certain requirements that have to be fulfilled by SimDB and ExtDB for our database synchronization approach to be applied. A more detailed insight is given in (Hoppen et al. 2013). Most of these requirements apply to the databases metamodels. According to (Seidewitz 2003), a metamodel is a model that specifies a modeling language. A well known example is the UML's metamodel. Likewise, the metamodel of a database specifies how schemata are modeled. When coupling SimDB and ExtDB, correspondences between schema components have to be identified or created. Using metamodels, schema components are modeled as metaclasses. The Meta Object Facility's (MOF) (OMG 2013b) meta-metamodel is used to define a meta-modeling language for metamodels. Rules for correspondences between metaclasses can be formally described using a model transformation language like QVT (OMG 2013c). This implies the requirement for a metamodel of SimDB and of ExtDB as well as a model transformation in between to be given.
In data synchronization, data conforming to the initially synchronized schemata is synchronized between the two databases. This requires knowledge about their respective instantiation concepts and how they relate. I.e., how data conforming to a schema is instantiated and how instance data in SimDB corresponds to instance data in ExtDB. Thus, a concept for ontological instantiation (Atkinson and Kuhne 2003) has to be part of each metamodel and has to be taken into account within the transformations in between.
Adding an additional pivotal metamodel with transformations from and to SimDB's and ExtDB's metamodel has two main advantages. Firstly, it can guarantee certain modeling capabilities within SimDB and ExtDB by choosing an appropriate pivot. Due to the existence of the transformations, the pivot's modeling concepts have to exist in both databases, at least implicitly. Secondly, the previous requirement definitions postulated on SimDB and ExtDB do not give any insight on the individual metaclasses and their semantics. For example, it is unknown, which metaclasses describe schema structures and which describe instances thereof. However, to use the requirements for a specification of general methods for database synchronization, there has to be some knowledge about such semantics. A known pivotal metamodel with known semantics, i.e. a known theory (Seidewitz 2003), can resolve this lack of information. Using UML (i.e. UML Superstructure's language unit Classes::Kernel) as a pivotal metamodel provides object oriented modeling concepts and metaclasses with well-known semantics.
Some requirements also apply to the databases' runtime systems. For synchronization, both must provide access to the schema description. Such access methods make the system reflective (Maes 1987). Schema information

has to be determinable from instantiated data, e.g., to determine an object's class. Furthermore, new data has to be instantiated using schema structures, e.g., given a class, instantiate a new object. Another aspect is to query data, e.g. to retrieve all objects of a class. Thus, a reflection interface is required by SimDB and ExtDB. Other requirements are instance data identification and change notification.

METHOD SPECIFICATION

Based on the aforementioned requirements, the core of this paper specifies the methods for database synchronization in a new generalized form. Although these methods refer to the runtime and not to the metamodel level they will be stated in terms of metaclasses and instances thereof. Instances of metaclasses constitute a model, which itself represents (or specifies) a system under study (Seidewitz 2003). In this context, this is a database system during runtime. This rationale legitimates to use the same notions for the system as for the model. By fulfilling the given requirements, a UML metaclass can be used as a surrogate to indirectly address the otherwise unknown metaclasses within SimDB or ExtDB that it corresponds to by model transformation. To simplify nomenclature, the notion of a *pseudo metaclass* is introduced. For example, SimDB-Class is a pseudo metaclass acting in place of those one or more metaclasses within SimDB that are transformed into UML-Class and vice versa. By use of this notion, the semantics of a UML metaclass can easily be transferred to the context of SimDB and ExtDB. Reflection shall be provided by access to reified versions of (instances of) metaclasses. An overview of the approach is given in Figure 2. The excerpt of the UML Classes::Kernel package (abbreviations in parentheses) is used for schema and data synchronization.

SCHEMA SYNCHRONIZATION

A schema is synchronized from ExtDB to SimDB to provide a common data schema for the whole distributed system. Once during system startup, ExtDB's schema is analyzed using its reflection interface giving access to the reified instances of metaclasses describing the schema. The schema sync process then yields a schema mapping M_S between the two databases. There are two approaches to schema synchronization.

Schema Synchronization by Transfer

Schema synchronization by transfer dynamically creates a schema within SimDB. This transfer is entirely performed at runtime. Provided write access, it uses reflective methods to instantiate metaclasses of SimDB's metamodel, similar to QVT *enforce domain* rules. Algorithm 1 gives an overview. The result of this process

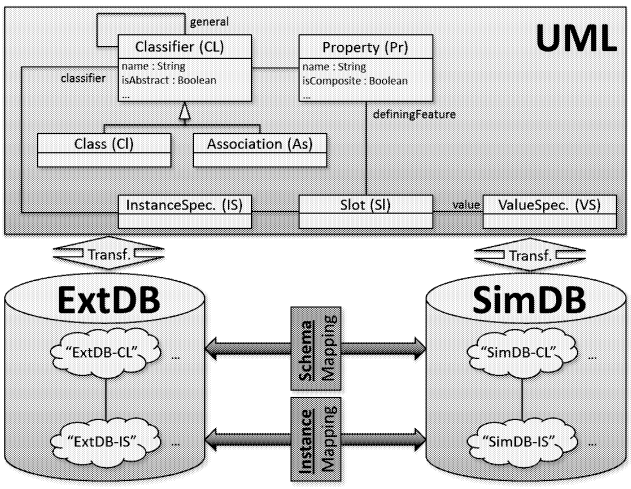


Figure 2: Transformations guarantee UML concepts – runtime synchronization is performed between the DBs.

is a set of newly created runtime schema components within SimDB and a schema mapping containing the correspondences to ExtDB. Use cases for this approach are schemata previously unknown to the simulation system or frequently changing schemata.

Algorithm 1 Schema transfer

for all classifiers CL_{Ext} with type T **do**
  $CL_{Sim} \leftarrow$ new SimDB-T
  Copy all metaattribute values of CL_{Ext} to CL_{Sim}
  $M_S(CL_{Ext}) \leftarrow CL_{Sim}$
  for all owned attributes/ends Pr_{Ext} of CL_{Ext} **do**
    $Pr_{Sim} \leftarrow$ new SimDB-Property
    Copy all metaattribute values of Pr_{Ext} to Pr_{Sim}
    Add Pr_{Sim} to owned attributes/ends of CL_{Sim}
    $M_S(Pr_{Ext}) \leftarrow Pr_{Sim}$
  end for
end for
for all classifiers CL_{Ext} with type T **do**
  if T is Association **then**
    for all member ends Pr_{Ext} of CL_{Ext} **do**
      Add $M_S(Pr_{Ext})$ to members ends of $M_S(CL_{Ext})$
    end for
  end if
  for all direct generalizations CL^{gen}_{Ext} of CL_{Ext} **do**
    Add generalization $M_S(CL^{gen}_{Ext})$ to $M_S(CL_{Ext})$
  end for
end for

Schema Synchronization by Matching

In contrast, in schema synchronization by matching, SimDB already contains a predefined version of the schema, which only has to be matched against ExtDB's schema. Note that matching does not imply any translations between generally different schemata. E.g., it

does not match a name attribute in ExtDB to separate first and last name attributes in SimDB. It only applies to identical schemata represented using the metamodels of SimDB and ExtDB. The basic procedure for schema matching is very similar to the approach given in Algorithm 1. Instead of creating new classifiers or properties, equivalents are searched within SimDB and the postprocessing only verifies association ends as well as generalizations. The only result is the schema mapping M_S (or failure). The procedure does not alter SimDB itself, resembling QVT *checkonly domain* rules. Schema matching is less flexible than schema transfer. But when used with a static schema predefined at compile time of the simulation system (where applicable), it can have the advantage of static type safety. In contrast, dynamically created schemata can only be accessed using reflection, implying runtime type checking. Thus, use cases for schema sync by matching are stable (rarely changed) schemata, statically predefined in SimDB, a SimDB implementation that only provides read-only reflection, or schemata dynamically defined in SimDB by another process (e.g. XMI import). The two principles can also be mixed. A use case could be a static base schema synced by match with changing application schemata synced by transfer.

DATA SYNCHRONIZATION

One of the core features of the presented approach is making ExtDB's data transparently accessible to components of the simulation system by means of SimDB. Using data synchronization, requested data is loaded (i.e. replicated) from ExtDB to SimDB, changes are tracked and resynchronized, and data can be unloaded from SimDB. Due to restricted space in the context of this paper, we will focus on data loading and unloading. For each pair (original and replica) of data instances an entry in the instance mapping M_I is stored. The metaclass InstanceSpecification represents an (ontologically) instantiated Classifier. For classes, an instance is called an object, for associations a link. Furthermore, UML metaclass Slot, representing an instantiated Property, and ValueSpecification as the abstract base metaclass for miscellaneous types of values are of particular interest here.

Loading

Methods for loading data from ExtDB to SimDB can be divided into two categories: While basic operations are used internally, interface methods provide data access to other components of the simulation system. To maintain transparency, the latter must not expose ExtDB's nature, its pseudo metaclasses, or its runtime interface. A basic operation in data loading is to load a single object. Here, an instance specification IS_{Ext} (with a class-type classifier) is replicated from ExtDB to SimDB with all its values. The correspondence between the original object IS_{Ext} and its replicate IS_{Sim} is stored as an entry in the central instance mapping M_I. Algorithm 2 illustrates the approach. After checking whether the object is already loaded it (ontologically) instantiates the corresponding SimDB-Class, replicates all values and stores a new mapping entry.

Algorithm 2 Load a single object.

```
function loadObject(IS_Ext)
  if IS_Ext ∈ M_I then
    return M_I(IS_Ext)
  end if
  Cl_Ext ← class of IS_Ext
  Cl_Sim ← M_S(Cl_Ext)
  IS_Sim ← instantiate Cl_Sim
  for all Slots Sl_Ext of IS_Ext do
    Pr_Ext ← defining feature of Sl_Ext
    Sl_Sim ← create SimDB-Slot for IS_Sim with M_S(Pr_Ext)
    for all VS_Ext ∈ Sl_Ext do
      add replicateValue(VS_Ext) to Sl_Sim
    end for
  end for
  M_I(IS_Ext) ← IS_Sim
  return IS_Sim
end function
```

For structured 3D data, a superimposed *hierarchical* loading function is used to load an object and all of its ancestors. Here, an ancestor is recursively defined as another object aggregating this object or one of its other ancestors using an association with composite aggregation. A direct ancestor is called a parent, its counterpart is called a child. An object without a parent is a root object. Following the UML, a part object may only be included in at most one composite at a time. Thus, an object may have at most one parent at a time. Furthermore, associations with composite aggregation must be binary associations. Therefore, only $0..1 : n$ associations are considered here.
To reflect parent-child-relations and consistency within SimDB, it is a reasonable approach to load a part object only with its composite object. Therefore, Algorithm 3 can be used to recursively load an object including all its ancestor objects from ExtDB into SimDB.

Algorithm 3 Recursively load an object.

```
function loadObjectWithAncestors(IS_Ext^child)
  IS_Sim^child ← loadObject(IS_Ext^child)
  IS_Ext^link ← findParentLink(IS_Ext^child)
  if IS_Ext^link exists then
    IS_Ext^parent ← other member end of IS_Ext^link
    loadObjectWithAncestors(IS_Ext^parent)
    loadLink(IS_Ext^link)
  end if
  return IS_Sim^child
end function
```

Algorithm 4 is used to load a link. Links are only identi-

fied by their current values. For each member end of an association, a link holds a slot with a single instance of the corresponding target type as a value. It is defined and identified only by this tuple. In case of a binary association connecting classes, this is a pair of objects. However, links "belong" (not in the sense of ownership) to multiple instances. So, a management using M_I is advantageous. For the sake of simplicity, links are assumed to be unique and non-ordered. A prerequisite for link loading is that all participating member objects have to be already loaded into SimDB. Therefore, the tuple t_{Ext} of linked objects is retrieved and checked for instance mapping entries. When all member objects are loaded, a new SimDB link is instantiated, registered in the instance mapping, and slots are created for each member end with instance values from the tuple.

Algorithm 4 Load a single link.

function LOADLINK(IS_{Ext}^{link})
  $IS_{Sim}^{link} \leftarrow M_I(IS_{Ext}^{link})$
  if IS_{Sim}^{link} exists **then**
    return IS_{Sim}^{link}
  end if
  tuple $t_{Ext} \leftarrow$ member objects of IS_{Ext}^{link}
  if $\exists\, IS_{Ext} \in t_{Ext} : M_I(IS_{Ext}) =$ undefined **then**
    return null
  end if
  $As_{Sim} \leftarrow M_I($ classifier of $IS_{Ext}^{link})$
  $IS_{Sim}^{link} \leftarrow$ instantiate As_{Sim}
  $M_I(IS_{Ext}^{link}) \leftarrow IS_{Sim}^{link}$
  for all $IS_{Sim} \in M_I(t_{Ext})$ **do**
    $Pr_{Sim} \leftarrow$ member end for IS_{Sim}
    add new Sl_{Sim} for Pr_{Sim} with IS_{Sim} to IS_{Sim}^{link}
  end for
  return IS_{Sim}^{link}
end function

The previously introduced methods allow to load given objects from ExtDB. Yet, components of the simulation system have only indirect access to ExtDB's contents via the replicated data in SimDB. Thus, appropriate query mechanisms have to be provided that must be transparently accessible from the simulation system. Using reflection is a way to provide data access without exposing ExtDB's characteristics. Internal data requests of the simulation system can transparently be formulated in terms of (reified linguistic instances of) SimDB's pseudo metaclasses. By applying the schema mapping from the previous level of synchronization, such requests can be translated into corresponding requests in terms of ExtDB's pseudo metaclasses.

One example is to load all child objects of a given (parent) object IS_{Sim} as described Algorithm 5. Here, *findAllChildLinks* collects links of all associations with composite aggregation. A more selective variation is to filter for child links of specific associations instead.

Recursively combining the previous approach, a subtree of objects can be loaded. Here, for a root object given

Algorithm 5 Load children of a parent object.

function LOADALLCHILDOBJECTS(IS_{Sim})
  $IS_{Ext} \leftarrow M_I^{-1}(IS_{Sim})$
  for all $IS_{Ext}^{link} \in$ FINDALLCHILDLINKS(IS_{Ext}) **do**
    $IS_{Ext}^{child} \leftarrow$ object at other end of IS_{Ext}^{link}
    $IS_{Sim}^{child} \leftarrow$ LOADOBJECT(IS_{Ext}^{child})
    LOADLINK(IS_{Ext}^{link})
  end for
  return all loaded IS_{Sim}^{child}
end function

in SimDB an object subtree defined by composite aggregation is loaded from ExtDB into SimDB including all its descendants. Consistent with previous definitions, a descendant of an object is defined recursively as another object composed by this object itself or one of its descendants. Accordingly, a child is a direct descendant of an object. To load an object subtree, Algorithm 5 needs to be recursively applied to load all child objects of a given root object.

An example for a subtree of hierarchical structured 3D city data is given in Figure 3. Building *b1* is its root object, the parent of Room *r1* and *r2* and an ancestor of all objects within the subtree (except itself). Door object *d* is a child of *r1* and a descendant of *b1* (and also *r1*). Calling *loadObjectWithAncestors* on *d* loads *d* as well as *r1*, *b1*, and *c* (when not already loaded).

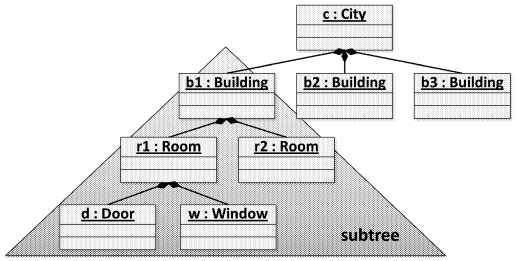


Figure 3: Example: Hierarchical 3D city data.

Besides associations with composite aggregation, general, *n*-ary links between objects may also be of interest. An interface function loading all objects connected with links of an association As_{Sim} to the given reference object IS_{Sim} is given in Algorithm 6.

Algorithm 6 Load linked objects.

function LOADLINKEDOBJECTS(IS_{Sim},As_{Sim})
  $IS_{Ext} \leftarrow M_I^{-1}(IS_{Sim})$
  $As_{Ext} \leftarrow M_S^{-1}(As_{Sim})$
  for all IS_{Ext}^{link} of As_{Ext} with member IS_{Ext} **do**
    for all $IS_{Ext}^{member} \in$ member objects of IS_{Ext}^{link} **do**
      LOADOBJECTWITHANCESTORS(IS_{Ext}^{member})
    end for
    LOADLINK(IS_{Ext}^{link})
  end for
end function

Data Unloading

Besides loading instance data it can also be unloaded. Unloading an instance means its controlled removal from SimDB to reduce the amount of space occupied during runtime. For objects, this removal includes all descendant objects and the corresponding links. Unloading a parent link is identical to unloading the child object (and its descendants). Unloading any other link just unloads the link itself.

A prerequisite for unloading an instance is that all local changes within SimDB are resynchronized to ExtDB. If an instance is not *clean* in this way, the corresponding changes would be lost. Changes within ExtDB however can be ignored as these global changes are only of interest if the related data is (still) loaded. All corresponding descendant objects within SimDB must be clean as well. They cannot be unloaded themselves and without their ancestor their connection to their subtree of composite aggregations is broken. For example in Figure 3, if Room *r1* was to be unloaded, Door *d*, Window *w* and *r1* itself must be clean, too. The same applies to dependent links connected to the object or its descendants. Likewise, they can only be safely removed from SimDB when they are clean.

Algorithm 7 illustrates a depth-first search to check a subtree rooted at object IS_{Sim} with parent link IS_{Sim}^{plink} for the prerequisites mentioned above. IS_{Sim} is considered *recursively clean* when all its child objects are recursively clean and its parent link, all its non-composite links, and itself are clean. While checking this, the algorithm collects all recursively clean objects and their respective parent and non-composite links in the set S_{clean}.

Algorithm 8 implements two a strategies to unload a subtree. Without pruning, only a completely clean subtree is unloaded. Otherwise, all recursively clean subsubtrees are unloaded (pruned) to free as much space as currently possible.

Algorithm 7 Can the subtree be unloaded?

```
function canUnload(IS_Sim, IS_Sim^plink, S_clean)
  ok ← true
  for all IS_Sim^link ≠ IS_Sim^plink with member IS_Sim do
    if IS_Sim^link is child link then
      IS_Sim^child ← object at other end of IS_Sim^link
      if not canUnload(IS_Sim^child, IS_Sim^link, S_clean) then
        ok ← false
      end if
    else
      if IS_Sim^link clean then
        ok ← false
      else
        add IS_Sim^link to S_clean^link
      end if
    end if
  end for
  if ok and (IS_Sim and IS_Sim^plink clean) then
    add IS_Sim, IS_Sim^plink to S_clean
    S_clean ← S_clean ⋃ S_clean^link
    return true
  else
    return false
  end if
end function
```

Algorithm 8 Unload a subtree.

```
function unloadSubtree(IS_Sim, pruning)
  IS_Sim^plink ← parent link of IS_Sim
  allClean ← canUnload(IS_Sim, IS_Sim^plink, S_clean)
  if allClean then
    remove IS_Sim
  else if pruning then
    for all IS_Sim^clean ∈ S_clean do
      remove IS_Sim^clean
    end for
  end if
end function
```

Streaming

Based upon loading and unloading, a streaming strategy can be realized. Here, streaming means an on demand loading and unloading of instance sets. It allows the simulation system to access huge data sets stored within ExtDB piece by piece, i.e. of which only parts are needed at a time within SimDB. These instance sets are usually determined by applying a filter on a property and varying its parameters. Three types of streaming are presented: spatial streaming, hierarchical streaming and streaming for data processing.

To move through a huge virtual environmental model like a city (or a forest) model, a spatial filter can be used to select only those buildings around a user's current position. If he moves, previously loaded sets of buildings can be unloaded and new sets can be loaded, based on the new position. Thus, this is a *spatial streaming* approach. A simple approach to discretize data management in such scenarios with movement over a 2D surface is to use a filter based on bounding rectangles. The area is uniformly subdivided into tiles (squares). For each tile the user approaches all objects are loaded and logically associated with the tile. Objects may be associated to several of these tiles, depending on their geometric features. Nevertheless, they will only be loaded once into SimDB. When a tile is left behind by the user, objects are disassociated from it. Objects without remaining tile associations can be unloaded.

Another use case for streaming is to navigate through hierarchical structured data e.g. in a user interface for SimDB. E.g., in an industrial context one may want to explore a factory model's structure by navigating from a factory root object via the corresponding links to the associated factory floors, from one of the factory floors to its associated working cells, and so forth. In user in-

terfaces, this can be presented using a tree view. On opening a branch in such a view, all associated objects can be loaded, on closing the branch they can be unloaded implementing *hierarchical streaming*.
A third type is *streaming for data processing* purposes. This approach is used when large amounts of data in ExtDB shall be processed in smaller subsets within SimDB. The idea is similar to an iterator or a database cursor. Each subset is loaded, processed in some way, resynchronized if necessary, and unloaded. Streaming for data processing can use different iteration strategies. It may apply spatial streaming to process data spread over an area, hierarchical streaming to process a subtree of objects, or other specialized approaches. For example in an annual process, tree instances from a forestal inventory database are loaded in tiles by a spatial filter, processed by simulating their growth, resynchronized, unloaded, and so forth.

Retain-And-Release Management

So far, data loading and unloading is considered to be used by one mechanism of the simulation system at a time, e.g. only by spatial or hierarchical streaming. When multiple mechanisms are used in combination, repeated attempts to load the same instance always yield the same single replicate within SimDB. However, unloading an instance can only be done once. Thus, a conflict of interest may arise when the same instance is loaded by several mechanisms and at some point unloaded by one of them. For example, one component of the simulation system initially loads a building instance using spatial streaming. Another component also loads it using hierarchical streaming yielding the identical replicate. When the first component unloads the building it is no longer available to the other, too.
To avoid this problem, it is necessary to *retain* an instance from being unloaded as long as it is required by some component of the simulation system. Likewise, the instance has to be *released* when it is no longer required.
A strategy for this retain-and-release management is reference counting. Here, instance mapping entries maintain a reference counter. Each loading request for an instance increments its counter, each unloading request decrements it. Unloading itself is only pursued for a zero reference counter. Counting may also be optional per load or unload request.
Note that the approach for spatial streaming in the previous subsection already employs a retain-and-release management by associating loaded objects with corresponding tiles. However, using the integrated strategy presented above, loading objects for a tile retains them once and unloading releases them once. Again, only completely released objects are really unloaded.
Instead of unloading immediately, it can be deferred by scheduling a kind of garbage collection for completely released instances. This way, unload tasks are accumulated and processed later. This strategy can be used when memory in SimDB is not essential. It may also be advantageous when instances are required shortly after being released. Retaining an instance once again will preserve it from garbage collection. Furthermore, the point in time for unloading may be chosen depending on the current work load to optimize overall system performance.

RELATED WORK

Regarding our model-driven approach to specify methods and their requirements, Generic model management uses a similar approach (Bernstein and Melnik 2007). Our approach can be seen as an implementation of the *ModelGen* operator that automatically translates a schema from one metamodel into another, including mapping creation. For the *Match* operator the emphasis is more on matching schemata different in themselves whereas our approach only matches equal schemata expressed using different metamodels. However, other implementations (Smith and McBrien 2008, Atzeni et al. 2009) do not focus on simulation and runtime data management or do not use industry standards like MOF or QVT.
In general, databases have been integrated into 3D systems in many different ways, e.g. (Julier et al. 2000, Watanabe and Masunaga 2002, Schmalstieg et al. 2007, Dassault Systèmes 2008, Walczak 2012). But no approach provides our tight integration of real-time 3D simulation and object-oriented database technology including persistence, collaboration, arbitrary data schema, and versioning.

FUTURE WORK AND CONCLUSION

As a next step, we will supply a more generalized specification of our methods for change management and resynchronization in the context of data synchronization in another publication. We also want to apply the requirements definition to our current prototypical implementation. This prototype uses the 3D simulation system VEROSIM and the geo database SupportGIS Java. Furthermore, we plan to transfer the generalized methods presented in this paper to another implementation of ExtDB, e.g. PostgreSQL.
To conclude this paper, we presented a new generalization of our methods for database driven 3D simulation as informally presented in (Hoppen et al. 2012). This specification is based on the generalized requirements as defined in (Hoppen et al. 2013). It gives a more detailed insight in these methods. In particular, we use concepts from the UML metamodel to give a general method specification. Therefore, the notion of pseudo metaclasses is introduced to indirectly refer to details of SimDB's and ExtDB's metamodel. Thus, method descriptions can be decoupled from the actual

databases. Based upon pseudo metaclasses, the approach for schema synchronization and approaches for loading, unloading, streaming, and retain-and-release strategies in data synchronization are specified. The universality of these specifications allows them to be applied to different implementations of SimDB and ExtDB more easily - provided they fulfill the stated requirements. However, besides these theoretical basics, a prototypical implementation of the approach has already proven its practical suitability in many applications, as seen in (Schlette and Hoppen 2012, Rossmann et al. 2011b;a) and Figure 1.

REFERENCES

Atkinson C. and Kuhne T., 2003. *Model-Driven Development: A Metamodeling Foundation. Software, IEEE*, 20, no. 5, 36–41.

Atzeni P.; Bellomarini L.; Bugiotti F.; and Gianforme G., 2009. *A runtime approach to model-independent schema and data translation.* In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology.* ACM, New York, NY, USA, EDBT '09, 275–286.

Bernstein P.A. and Melnik S., 2007. *Model management 2.0: manipulating richer mappings.* In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data - SIGMOD '07.* ACM Press, New York, New York, USA, 1–12.

Dassault Systèmes, 2008. *ENOVIA V6 Architecture Performance Capability Scalability a Product Lifecycle Management Whitepaper Prepared by ENOVIA , a Dassault Systèmes Brand Executive Summary.*

Hoppen M.; Schluse M.; and Rossmann J., 2013. *A Metamodel-Based Approach For Generalizing Requirements In Database-Driven 3D Simulation.* In *Proceedings of the 3rd International Workshop on Model-driven Approaches for Simulation Engineering.* To be published.

Hoppen M.; Schluse M.; Rossmann J.; and Weitzig B., 2012. *Database-Driven Distributed 3D Simulation.* In *Proceedings of the 2012 Winter Simulation Conference.*

Julier S.; Baillot Y.; Lanzagorta M.; Brown D.; and Rosenblum L., 2000. *Bars: Battlefield augmented reality system.* In *NATO Symposium on Information Processing Techniques for Military Systems.* 9–11.

Maes P., 1987. *Concepts and experiments in computational reflection. ACM Sigplan Notices*, 22, no. 12, 147–155.

Rossmann J.; Schluse M.; Waspe R.; and Moshammer R., 2011a. *Simulation in the Woods: From Remote Sensing based Data Acquisition and Processing to Various Simulation Applications.* In *Proceedings of the 2011 Winter Simulation Conference.* 984 – 996.

Rossmann J.; Wantia N.; Springer M.; Stern O.; Müller H.; and Ellsiepen M., 2011b. *Rapid Generation of 3d Navigation Maps for Extraterrestrial Landing and Exploration Missions : the Virtual Testbed Approach.* In *Advanced Space Technologies for Robotics and Automation.* 1–7.

Schlette C. and Hoppen M., 2012. *Modular Control System Architecture.* In S. Haag, M., Härer (Ed.), *SCALAB Scalable Automation for Emerging Lab Production : Final Report of the MNT-ERA.net Research Project 'SCALAB'.* 29–40.

Schmalstieg D.; Schall G.; Wagner D.; Barakonyi I.; Reitmayr G.; Newman J.; and Ledermann F., 2007. *Managing complex augmented reality models. IEEE Computer Graphics and Applications*, 27, no. 4, 48–57.

Seidewitz E., 2003. *What models mean. IEEE Software*, 20, no. 5, 26–32.

Smith A. and McBrien P., 2008. *A Generic Data Level Implementation of ModelGen.* In A. Gray; K. Jeffery; and J. Shao (Eds.), *Sharing Data, Information and Knowledge*, Springer Berlin / Heidelberg, *Lecture Notes in Computer Science*, vol. 5071. 63–74.

Walczak K., 2012. *Dynamic Database Modeling of 3D Multimedia Content.* In W. Cellary and K. Walczak (Eds.), *Interactive 3D Multimedia Content*, Springer London, London, chap. 4. 55–102.

Watanabe C. and Masunaga Y., 2002. *VWDB2: A Network Virtual Reality System with a Database Function for a Shared Work Environment.* In K. Tanaka (Ed.), *Information Systems and Databases.* Tokyo, Japan, 190–196.

WEB REFERENCES

OMG, 2013a. Unified Modeling Language (UML). http://www.uml.org/ accessed 02/25/2013.

OMG, 2013b. Meta Object Facility (MOF). http://www.omg.org/mof/ accessed 02/25/2013.

OMG, 2013c. MOF Query/View/Transformation (QVT). http://www.omg.org/spec/QVT/ accessed 02/25/2013.

SIMULATION SYSTEM FOR LINEAR CUTTING PLANS

Ovidiu-Dorin Alupei-Cojocariu, Nicolae Vasiliu, Szilard Feher

University POLITEHNICA of Bucharest
313 Splaiul Independentei, Bucharest
Romania, 060042
E-mail: {ovidiu.alupei|vasiliu1946|feher.szilard}@gmail.com

KEYWORDS

Product Data Management (PDM), simulation of linear cutting plans, cutting stock, steel profiles, linear programming

ABSTRACT

Linear cutting plans, also known as nesting plans, are often found in a variety of industries that use one-dimensional raw work pieces. Each part has a certain length, an identification code and can appear once or more than once in the work plane. Mass production can be achieved having similar small series or even unique products that are considered as one whole task and are manufactured up on an accurate cutting plan that the system can supply. In order to simulate the efficiency of the cutting process, one must combine the power cutting algorithms with a sound result representation, offering an analysis tool. An application that gathers all the similar parts from several similar products forms a stock that must be trimmed and manufactured together. This is a problem of linear mathematics that, if solved, could lead to a substantial decrease in material consumption (Kravitz 2012; Garey and Johnson 1979). The application is developed in AutoLISP under AutoCAD and offers the advantages of a solid programming tool, coupled with a precise drawing editor where the results of cutting plans are estimated (Alupei-Cojocariu 2004).

INTRODUCTION

Linear parts are often found in structures that have minimal mass and maximal overall strength of the product. A variety of materials like wood, steel, plastic are used, these having a uniform cross section and a certain length. The products can involve several parts that have the same cross section, with repetitive parts that have the same length; they also have a suite of parts that are non-repetitive. Mass production is not a reality nowadays. Products can be demanded in limited series and that raises problems in economical material use. Repeatable parts form clusters of pieces must be processed one after another or intercalated with other parts.

FIRST STEP: BUILDING A DATABASE

Modular products with linear parts are often conceived using a CAD environment that can generate a report as BOM (Bill of Materials) presenting all the parts, their names (codes), length, mass, profile type, number of identical items. The parts are gathered into assemblies that must be processed in order to obtain the prototype or to replicate into series as the project demands. One example is offered by the application TowerCAD that is able to assist the CAD implementation of sustain towers used into electrical power lines. A tower is mainly conceived from long bars (angles) L steel profiles with equal wings, assembled by bolts and gussets or welded into the structure. A large variety of parts are used in each construction due to the equal resistance structure. Even if the parts have the same cross section, their length is variable from base to top of the tower. The parts are identical only on symmetrical sides and are fit for a group technology. When a whole line is manufactured, a certain number of towers are identical, forming a large or small series; others form a single construction without repeatability. A plant that manufactures the towers must have the skills and means to manage the whole product data volume in order to minimize the material consumption and maximize the profit. Less left over material leads to a larger profit or a larger negotiation limit for contracts. Having an overall image about the manufacturing plan through a consistent database of parts and assemblies, a chronology of contracts and the real lasting of manufacturing machines can bring competitive advantages for this market.

Figure 1 shows a pack of product data used in order to launch the command YAZ. The information is based on the bill of material from TowerCAD.

Type	Command	Launch	Assembly	Code	Material	Length	Repeat	Mass
long bar	YAZ	12.10.2011	A	A-51	L160x12-OL52	3300	10	98.01
long bar	YAZ	12.10.2011	A	A-52	L160x12-OL52	3300	10	98.01
long bar	YAZ	12.10.2011	B	B-68	L120x10-OL52	1088	1	19.8016
long bar	YAZ	12.10.2011	B	B-69	L120x10-OL52	1088	1	19.8016
long bar	YAZ	12.10.2011	C	C-56	L90x8-OL52	2851	4	31.0759
long bar	YAZ	12.10.2011	C	C-57	L120x10-OL52	4125	4	75.075

Figure 1: Database of angle parts

The table gathers a subset of more comprehensible table of 275 parts used on demonstration.

SECOND STEP: ANALYZING THE MACHINES

The modern manufacturing of steel towers nowadays involves the use of NC Machines that punch the holes, mark the parts and trim the part from the long L workpiece.

Each NC Machine has a buffer of programs that stores the succession of punch operations for punch tools and coordinates of holes, the code of the part for marking purposes and the material type. The programs are automatically obtained from TowerCAD and transferred online or offline to the NC Machine memory. The succession of parts must be stored as workbench data, according to the cutting plans. Each machine has a mark unit with 2-4-8 mark places filled with marking tools. This information is important because when the cutting plan has a single extra code over the marking capacity of the machine, the process must be interrupted to fill the mark unit with new codes and is need to recalibrate the workpiece with some technological cuts that affect the cutting scheme. Knowing the NC Machines model and mark unit capacity can offer an accurate cutting plan involving all material losses. The marking tools are assembled from mark letters offering the codes of processed parts. For example, we can use NC Machines from FICEP (http://www.ficepcorp.com/anglelines.cfm) that offers the whole manufacturing capacity for angles.

THIRD STEP: DEFINING AN APPLICATION

The product data must be transformed into a document which is used in the management of the manufacturing process. The Product Data Management provides the Database that must be processed using a cutting algorithm, flexible and adapted for the NC Machine model. The application must be fitted on the data processing chain, having minimum interference with other applications. Due to the excellent features of TowerCAD, that offers an integrated CAD environment for steel structures, AutoCAD was the work platform that was asked for the analyse of the cutting plans as part of the Computer Integrated Manufacturing platform (AutoDesk, 2012).

The language chosen for the proof of concept and final application programming environment was AutoLISP for integration with TowerCAD.
The main specifications are as follows:

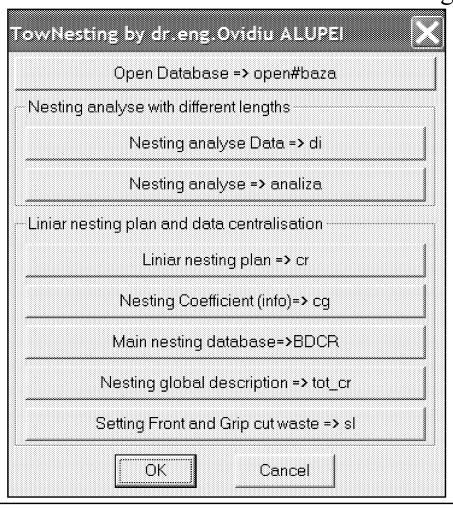
- Complete integration with AutoCAD;
- Import of database from .csv files generated with TowerCAD and other PDM activities;
- Integrated launch platform for multiple database sources in order to input the number of repetitive manufacturing of identical assemblies;
- Integrated database for NC Machines features;
- Flexible selection tool from the database;
- Complete automatic document generation for PDM ;
- Visual representation of the cutting results for simulation and validation purposes;
- Tool for variant analyze in order to optimize the selection of workpiece's lengths;

FOURTH STEP: BUILDING THE APPLICATION

The application was developed following the specifications and was named TowNesting. It was developed in AutoCAD using Visual LISP (Stein 2011) and is available on all AutoCAD platforms since R14. It is full integrated with TowerCAD and cannot be used separately for software security reasons (Alupei-Cojocariu 2004).

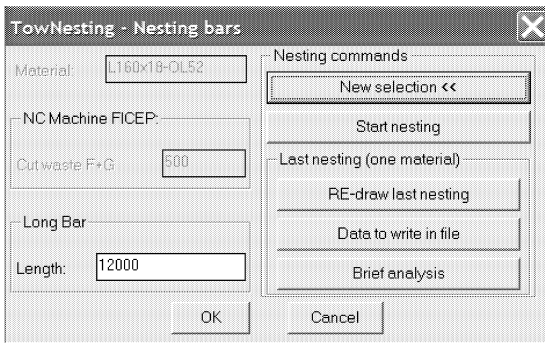
Table 1: TowNesting – Commands and results

Command TC=>Main Menu of TOW-Nesting 1D

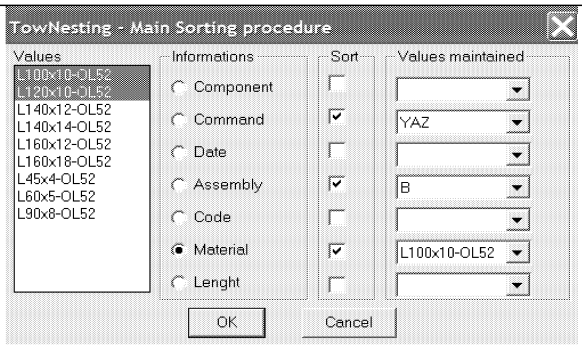


Main nesting database=>BDCR
Provide a general analysis of nesting process
Select all desired nesting results
You will generate these written results:
I have write into file D:\FICEP \DEMO\Bars_cutt.csv
I have write into file D:\FICEP \DEMO\Receittes_cutt.csv
I have write into file D:\FICEP\DEMO\Remains_cutt.csv

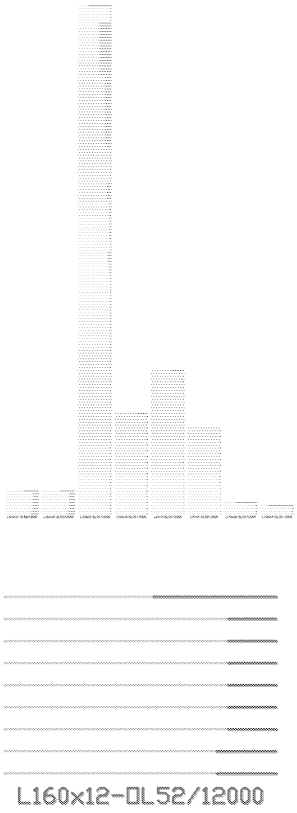
Linear nesting plan => CR
Main Nesting command
Start nesting performs the nesting for selected information
The length of Raw bar can be edited by user
The Material is set following New selection…
NC Machine FICEP is pre-selected outside the application as technological cuts list and loaded through Setting Front and grip cut waste =>sl (from TC Menu)



New selection => SORT#
starts the command Filtering the Data
You can use any filter from middle column, validating the selection by checking boxes
Multiple item selection allowed



Start nesting
Starts the nesting suite
Draws the results automatically
Writes the results automatically into separate files
The results in files are not overwritten but appended having access to a complete history of tries.
For each written file, a written message is displayed like
I wrote the cutting plan on file D:\FICEP \DEMO\L60x5-OL52.ret



Each column has the material name and length at the bottom
Cyan lines are effective nested parts.
Magenta lines are the calculated waste material.
Last bar shows an incomplete use of the raw material due the finish of listed parts.

The results of nesting:

Tower Company
Zincate Metallic towers
NC Machine FICEP 3
Cutting plan for material: L90x8-OL52
Overall length:11000 Cutting length:10620
Total grip+front cut waste: 380

1...1.) 2x(D-68) 8x(D-78) =>1 bar
2...2.) 1x(D-68) 1x(D-69) =>1 bar
3...4.) 2x(C-68) =>2 bars
5...5.) 3x(D-55) =>1 bar
6...6.) 1x(D-55) 2x(D-56) =>1 bar
7...7.) 2x(D-56) 1x(C-55) =>1 bar
8...8.) 3x(C-55) =>1 bar
9...9.) 3x(C-56) =>1 bar
10...10.) 1x(C-56) =>1 bar
Total:10 bars
==========Continues…

Receittes_cutt.csv

BC	Material	Length (mm)	N B	Mass (kg)	Type
YAZ	L45x4-OL52	12000	31	1019.9	bars
YAZ	L45x4-OL52	-	-	0	remain
YAZ	L45x4-OL52	-	-	24.24	waste
YAZ	L140x14-OL52	-	-	60.02	waste

A file provide a table with condensed results of the whole selected nesting results

Remains_cutt.csv

BC	Material	NC	Length (mm)	Mass (kg)
YAZ	L120x10-OL52	3	2283	41.54
YAZ	L120x10-OL52	3	2283	41.54

This file provides information about the waste of material that is more than 1000 mm long. The list can be used for a database containing shorter bars that can be used in other products, aiming to assure a complete use of the original raw material.

The results of the command **Nesting global description => tot_cr**

Tower Company / Zincate metallic towers
Batch no.: YAZ / NC Machine3
Cutting plan for material: L45x4-OL52
Total length[mm]:12000 Base length[mm]:11820 Cutting sets Front/Grip: 40/140
Global coeficients n/n-1[%]:97.18/97.21
---
1) Nr. bars:10; Ccr[%]:98.46; Waste[mm]:185
2x(A-131) 1x(A-132)
---
11) Nr. bars:1; Ccr[%]:96.61; Waste[mm]:407
10x(A-132) 1x(A-133) 1x(A-97)
---
12) Nr. bars:1; Ccr[%]:96.61; Waste[mm]:407

11x(A-133) 1x(A-97)

And follows…

Nesting analyze

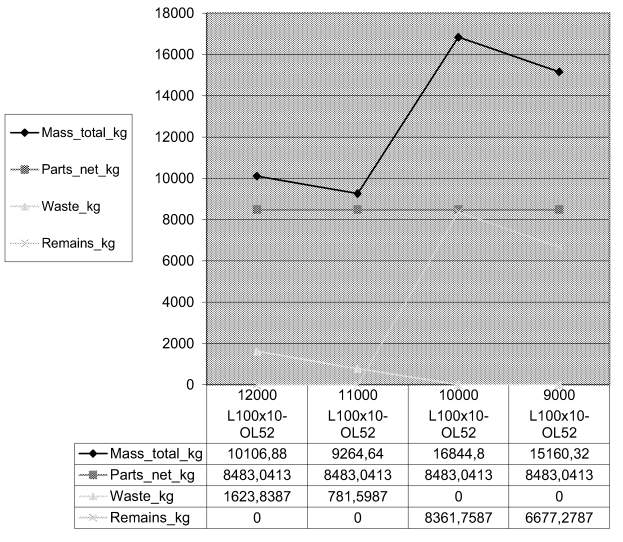
The initial data needed for analyze is the following file: settings_analyze.csv

NAn	Profile-Material	Length (mm)
1.	L100x10-OL52	12000
2.	L100x10-OL52	11000
3.	L100x10-OL52	10000
4.	L100x10-OL52	9000
5.	L120x10-OL52	12000
6.	L120x10-OL52	11000
7.	L120x10-OL52	10000
8.	L120x10-OL52	9000

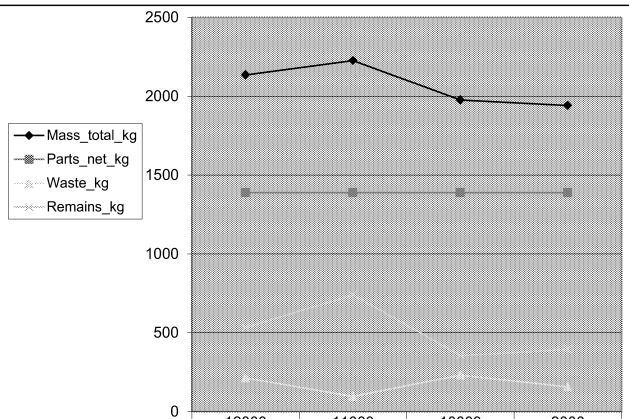
The first column shows the analyzed material name, the second content the length of the raw material.
Using several raw lengths is possible to reach an optimal use of the material with less waste.

NA Nr	Bars Nr	Mass_T kg	Parts_N kg	Waste kg	Remains kg
1.	56	10106,88	8483,041	1623,839	0
2.	56	9264,64	8483,041	781,5987	0
3.	112	16844,8	8483,041	0	8361,759
4.	112	15160,32	8483,041	0	6677,279
5.	8	2135,891	1389,32	213,3574	533,2131
6.	9	2226,311	1389,32	96,3312	740,6598
7.	10	1977,038	1389,32	232,383	355,335
8.	11	1941,529	1389,32	155,9648	396,2438

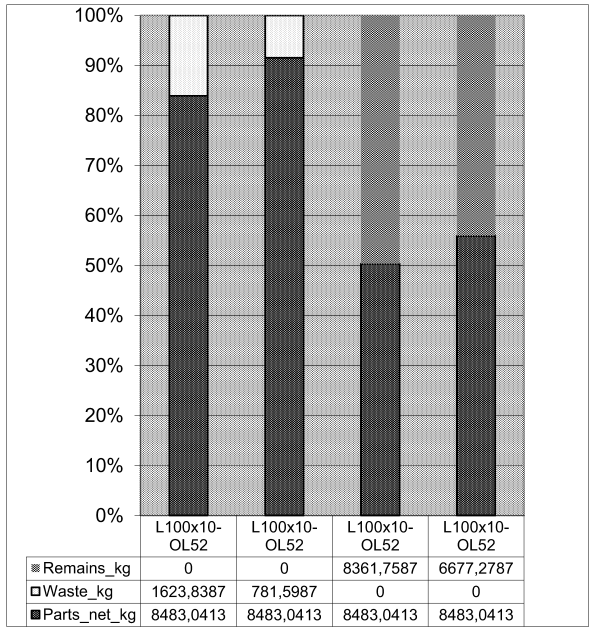
A special designed file can offer a graphic image of the nesting analyze: results_analyze_graphics.xls



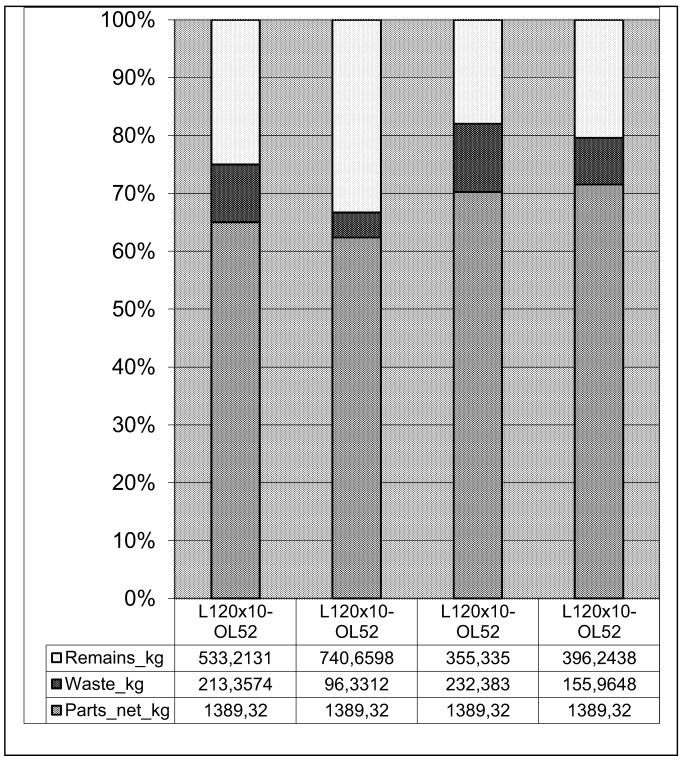
	12000 L100x10-OL52	11000 L100x10-OL52	10000 L100x10-OL52	9000 L100x10-OL52
Mass_total_kg	10106,88	9264,64	16844,8	15160,32
Parts_net_kg	8483,0413	8483,0413	8483,0413	8483,0413
Waste_kg	1623,8387	781,5987	0	0
Remains_kg	0	0	8361,7587	6677,2787



	12000 L120x10-OL52	11000 L120x10-OL52	10000 L120x10-OL52	9000 L120x10-OL52
Mass_total_kg	2135,8906	2226,311	1977,0381	1941,5286
Parts_net_kg	1389,32	1389,32	1389,32	1389,32
Waste_kg	213,3574	96,3312	232,383	155,9648
Remains_kg	533,2131	740,6598	355,335	396,2438



	L100x10-OL52	L100x10-OL52	L100x10-OL52	L100x10-OL52
Remains_kg	0	0	8361,7587	6677,2787
Waste_kg	1623,8387	781,5987	0	0
Parts_net_kg	8483,0413	8483,0413	8483,0413	8483,0413



CONCLUSIONS

The PDM problem of linear nesting is solved using TowNesting application; Simulation of different variants of nesting plans are available in graphic environment of AutoCAD and in Excel worksheet; Transition to different material cross sections is easy due to open .csv PDM database; AutoCAD and AutoLISP offers an attractive PDM environment for tower plants. The overall manufacturing costs decrease through a better use of the material and a faster document release.

REFERENCES

Alupei-Cojocariu O.D.2001."*AutoLISP Programming handbook".* BREN Publishing Press House, Bucharest. ISBN 973-8143-37-3 *(in Romanian).*

Alupei-Cojocariu, O.D. 2004-2010. "*TowerCAD User reference*", Bucharest.

Alupei-Cojocariu, O.D. 2010-2012. "*TowNesting User reference*". Bucharest.

Alupei-Cojocariu, O.D. 2002. „*Integrierte Produktion*", Frankfurter Hochsculzeitung. Dezember 2002. 83. Pag.17-19.

Kavitz, Hanan. 2012. "*Knapsack problem*". www.mathworks.com.

Garey, M.R., Johnson D.S., 1979. "*Computers and Intractability: A Guide to NP-completeness*". Freeman & Co. San Francisco. U.S.A.

AutoDesk. 2012. „*AutoLISP Developer's Guide". AutoCAD 2013.* January 2012.

Stein, David M. 2011. "*The Visual LSP Developer's Bible".* 2011 Edition.

BIOGRAPHIES

OVIDIU-DORIN ALUPEI-COJOCARIU graduated in Manufacturing Technology Department from University POLITEHNICA of Bucharest in 1986, and a PhD in Mechanical Engineering in 1999. He become an AutoDesk Authorized Trainer in 1996 and teaches courses about CAD/CAM, Databases for Engineering and Management, Management of RDI Projects and Product Data Management as Associated Professor in the University POLITEHNICA of Bucharest. He has developed many AutoLISP first class applications during 1996-2012, mainly for the industry.

NICOLAE VASILIU graduated in Hydropower Engineering from University POLITEHNICA of Bucharest in 1969. He became a Ph.D. in Fluid Mechanics after a research stages in Ghent State University and Von Karman Institute from Bruxelles. He became state professor in 1994, leading the ENERGY & ENVIRONMENT RESEARCH CENTRE from the University POLITEHNICA of Bucharest. He worked always for the industry, as project manager or scientific advisor.

SZILARD FEHER graduated in Mechanical Engineering from University POLITEHNICA of Bucharest in 2009, and a MD in Advanced Hydraulic and Pneumatic Systems in 2011. The same year he began a PhD research work on mobile electro hydraulic servo systems at the Fluid Power Laboratory from the Power Engineering Department of the University POLITEHNICA of Bucharest. He serves currently as technical advisor for the company HIDRAULICA SRL from Brasov, a PARKER industrial applications developer.

Automotive Production Location Cost Optimization with Productivity and Sustainability Constraints

Matthias Klumpp, Viktor Kuchshaus
Institute for Logistics and Service Management, FOM University of Applied Sciences
Leimkugelstraße 6, D-45141 Essen, Germany
E-mail: matthias.klumpp@fom-ild.de, www.fom-ild.de

KEYWORDS
Automotive Industry, GAMS, Production Location Cost Optimization, Carbon Taxation, Productivity Adjustment

ABSTRACT
Especially the automotive industry is one of the most globally oriented in purchasing, production and distribution – therefore optimization and cost analysis instruments for the strategic production location question are in high demand within this industry. This contribution outlines how a standard cost minimization approach modelled in an GAMS LP setting can be *enhanced* by a productivity adjustment for less developed countries as well as a carbon tax regime simulation in order to understand the magnitude of potential political decisions regarding e.g. a carbon tax regime on automotive supply chains and location decisions.

INTRODUCTION
Location problems and their connected cost optimiziation e.g. for transport costs are a long-standing research question in operations research (e.g. Miehle 1958, Domschke and Drexl 1996, Isard 2003, Schätzl 2003, Mainz 2005, Staudacher 2005, Eschlbeck 2006, Hübner 2007, Balcik and Beamon 2008). Especially in times of globalization and global competition this question has increasing importance for companies world-wide. The automotive industry is a good example for this development. The following figure shows an example for the expected top 15 car sales markets which will influence production and supply chain set-ups in the coming years.

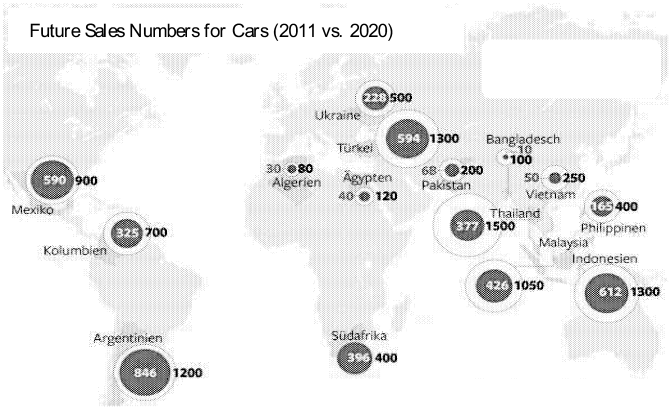


Figure 1: Future Car Markets (CAR UDE 2012)

The basic and most important factor for many location and also outsourcing decisions in this industry has been the *production cost perspective*. Therefore a basic GAMS LP optimization model has been established for 43 interesting production location countries world-wide in order to provide a basic setting for further constraint implementations.

BASIC GAMS LP COST MODEL
For a global comparison regarding possible automotive (parts) production locations altogether 43 countries were analyzed and included in the GAMS model. These were taken from different industry reports and OR as well as business prognosis publications in order to provide a broad picture of possible future developments and also selection options on the globe – e.g. all continents besides Australia are represented in the sample.

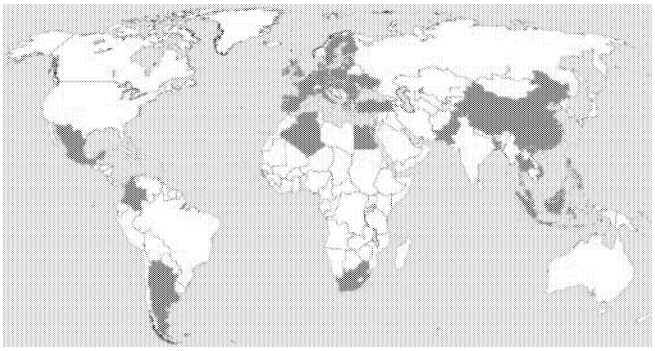


Figure 2: Global Distribution of Analyzed Countries (Own Created Map)

The LP cost minimizing model is integrating production (labor, energy) and logistics transport costs, the exemplary data from the following table was used within the model. For each country a exemplary business locations was identified in proximity to a seaport in order to calculate transportation and later also CO_2 levels as realistically as possible, e.g. Rosario in Argentina. The modeled scenario assumes an OEM final assembly plant in the US and the question is to choose an optimal production location for small parts (electrical connections) for this value chain including production (wage) and transportation costs. A specific number of production units and material parts were assumed to be delivered (produced, shipped).

	i	A	kM(i)	kEn(i)
No	Country	Location	Wage [€/h]	Electricity [€/kWh]
1	Aegypten	Nasr City	5.57	0.027
2	Algeria	Batna Oran	0.85	0.040
3	Argentina	Rosario	12.07	0.046
4	Bangladesh	Rangpur	0.17	0.054
5	Belgium	Opglabbeek	39.30	0.110
6	Bulgaria	Yambol	3.50	0.065
7	China	Hubei	1.03	0.074
8	Denmark	Vanløse	38.90	0.099
9	Deutschland	Essen	30.10	0.125
10	Estland	Tallinn	8.10	0.072
11	Finland	Uusikaupunki	29.70	0.076
12	France	Issy-les-Moulineaux	34.20	0.085
13	Greece	Volos	16.40	0.101
14	Indonesia	Bandung	0.53	0.055
15	Ireland	Belfast	27.40	0.116
16	Italy	Turin	26.70	0.153

17	Kolumbia	Bogotá	0.79	0.046
18	Lettland	Daugavpils	5.90	0.098
19	Litauen	Klaipéda	5.50	0.105
20	Luxembourg	Luxemburg	33.70	0.100
21	Malaysia	Kuching	0.45	0.082
22	Malta	Valletta	11.90	0.180
23	Mexico	Cuautitlan	4.92	0.027
24	Netherlands	Eindhoven	31.10	0.103
25	Oesterreich	Graz	29.20	0.098
26	Pakistan	Punjab	0.34	0.062
27	Philippines	Tarlac City	1.52	0.137
28	Poland	Poznań	7.10	0.101
29	Portugal	Ovar	12.10	0.099
30	Rumaenia	Ploiesti	4.50	0.080
31	Sweden	Malmö	39.10	0.089
32	Slowakia	Prievidza	8.40	0.128
33	Slowenia	Ljubljana	14.40	0.099
34	Spain	Barcelona	20.60	0.114
35	South Africa	Uitenhage	0.56	0.046
36	Thailand	Pluak Daeng	0.58	0.066
37	Tschech R.	Plzen	10.50	0.111
38	Turkey	Bursa	1.92	0.048
39	Ukraine	Uzhgorod	1.48	0.023
40	Ungarn	Kecskemét	7.60	0.095
41	UK	Manchester	20.10	0.098
42	Vietnam	Đà Nẵng	0.28	0.050
43	Zypern	Nicosia	16.50	0.167

Table 1: Country Locations and Cost Factors
(Own Calculations based on Cotton Bangladesh 2010, Bureau of Democracy, Human Rights and Labor Vietnam 2009, Economic Populist 2012, Statistisches Bundesamt 2012, U.S. Bureau of Labor Statistics 2012)

Rank	Country	CO_2 [t]	Total Cost [€]
1	Vietnam	281.610	182.044.920
2	Malaysia	292.070	185.631.500
3	Algerien	241.050	186.153.780
4	Indonesia	300.660	187.749.890
5	Philippines	278.640	198.715.000
6	Thailand	311.670	203.158.800
7	South Africa	317.960	205.102.140
8	Kolumbien	222.110	212.125.340
9	Litauen	258.130	218.604.400
10	Turkey	298.260	225.208.730
11	Mexico	166.890	227.791.850
12	Estland	263.250	234.477.600
13	Lettland	279.160	243.941.500
14	Aegypten	299.940	245.227.850
15	Bulgaria	307.080	246.164.600
16	Bangladesh	379.660	250.380.500
17	Malta	260.040	257.715.000
18	Argentinia	282.670	261.980.740
19	China	317.090	265.802.800
20	Portugal	251.680	278.406.400
21	Poland	291.720	282.090.400
22	Slowenia	284.630	285.148.400
23	Greece	276.330	286.652.400
24	Rumaenia	336.140	289.929.600
25	Ukraine	328.450	290.604.070
26	Zypern	289.060	297.638.100
27	Ungarn	241.700	302.178.500
28	UK	348.040	303.345.500
29	Spanien	249.310	304.767.400
30	Slowakei	334.960	315.841.300
31	Tschech Rep.	310.050	332.506.400
32	Pakistan	466.950	355.949.300
33	Ireland	248.760	357.531.400
34	Italy	269.490	359.590.100
35	Finland	247.720	361.941.600
36	Netherlands	249.770	374.540.400
37	Deutschland	258.600	380.514.300
38	Oesterreich	300.200	391.755.500
39	France	253.860	400.594.500
40	Denmark	252.100	412.082.400
41	Sweden	251.920	413.022.500
42	Luxemburg	269.510	415.269.400
43	Belgium	255.950	430.559.400

Table 2: Country Total Cost Results
(Own Calculations)

PRODUCTIVITY ADJUSTMENT CONSTRAINT

As a further *constraint extension* the different work productivity situation in these countries were modeled by a productivity factor change in relation to the average GDP per head in the countries. The following adjustment factors (table 3) were used. The results are depicted in the following table 4 for this productivity adjustment.

No	Threshold Value GDP (Relation to Average x)	Production Cost Increase y
1	x > 125%	0%
2	125% > x >= 100%	10%
3	100% > x >= 75%	20%
4	75% > x >= 50%	30%
5	50% > x >= 25%	40%
6	25% > x >= 0%	50%

Table 3: Adjustments for Country Productivity (Own Source)

Ranking (new)	Rang (alt)	Country	Adjustm.	Pers. Cost old [€]	Pers. Cost new [€]
1	1	Bangladesch	50%	992.800	1.489.200
2	2	Vietnam	50%	1.635.200	2.452.800
3	3	Pakistan	50%	1.985.600	2.978.400
4	4	Malaysia	30%	2.628.000	3.416.400
5	6	Suedafrika	40%	3.270.400	4.578.600
6	5	Indonesien	50%	3.095.200	4.642.800
7	7	Thailand	40%	3.387.200	4.742.100
8	8	Kolumbien	40%	4.613.600	6.459.000
9	9	Algerien	40%	4.964.000	6.949.600
10	10	China	40%	6.015.200	8.421.300
11	11	Ukraine	40%	8.643.200	12.100.000
12	12	Philippinen	50%	8.876.800	13.315.000
13	13	Tuerkei	30%	11.213.000	14.577.000
14	14	Bulgarien	30%	20.440.000	26.572.000
15	15	Rumaenien	30%	26.280.000	34.164.000
16	16	Mexiko	30%	28.733.000	37.353.000
17	17	Litauen	20%	32.120.000	38.544.000
18	19	Lettland	30%	34.456.000	44.793.000
19	18	Aegypten	40%	32.529.000	45.540.000
20	20	Polen	20%	41.464.000	49.757.000
21	21	Ungarn	20%	44.384.000	53.261.000
22	23	Slowakei	10%	49.056.000	53.962.000
23	22	Estland	20%	47.304.000	56.765.000
24	24	Tschechien	10%	61.320.000	67.452.000
25	25	Malta	10%	69.496.000	76.446.000
26	27	Portugal	10%	70.664.000	77.730.000
27	28	Slowenien	0%	84.096.000	84.096.000
28	26	Argentinien	20%	70.489.000	84.587.000
29	29	Griechenland	10%	95.776.000	105.350.000
30	30	Zypern	10%	96.360.000	106.000.000
31	31	UK	0%	117.380.000	117.380.000
32	32	Spanien	0%	120.300.000	120.300.000
33	33	Italien	0%	155.930.000	155.930.000
34	34	Irland	0%	160.020.000	160.020.000
35	35	Oesterreich	0%	170.530.000	170.530.000
36	36	Finnland	0%	173.450.000	173.450.000
37	37	Deutschland	0%	175.780.000	175.780.000
38	38	Niederlande	0%	181.620.000	181.620.000
39	39	Luxemburg	0%	196.810.000	196.810.000
40	40	Frankreich	0%	199.730.000	199.730.000
41	41	Daenemark	0%	227.180.000	227.180.000
42	42	Schweden	0%	228.340.000	228.340.000
43	43	Belgien	0%	229.510.000	229.510.000

Table 4: New Total Cost Situation (Own Calculation)

CARBON TAX REGIME CONSTRAINT

As a second extension of the basic model a theoretical carbon tax regime constraint was modeled. In this case the basic model was used in order to determine the tax threshold vaule where a change in the basic model optimal location (Vietnam) would occur due to rising transport costs with a carbon tax. The GAMS model for this extension reads as follows (with some *data extensions* labeled as "Weitere" for "further data" in order to save publication space here):

```
scalar
x Steuersatz /3/;

PARAMETER Kosten(i), EMGJahr(i),zaehler;
 EMGJahr('Vietnam')= 281610;
 EMGJahr('Malaysia')= 292070;
 EMGJahr('Algerien')= 241050;
 EMGJahr('Indonesien')= 300660;

   ... 39 Weitere

 Kosten('Aegypten')      = 258240000;
 Kosten('Algerien')      = 188140000;
 Kosten('Argentinien')   = 276000000;
 Kosten('Bangladesch')   = 250880000;

   ... 39 Weitere

zaehler = 0;

while(Kosten('Vietnam')=smin((i),Kosten(i)),
**smin bedeutet, dass Vietnam ab der ausgegeben Zahl, z.B. 1001, nicht mehr günstigster Anbieter ist
        Kosten('Aegypten')      = Kosten('Aegypten') + (EMGJahr('Aegypten') * x);
        Kosten('Algerien')      = Kosten('Algerien') + (EMGJahr('Algerien') * x);
        Kosten('Argentinien')   = Kosten('Argentinien') + (EMGJahr('Argentinien') * x);
        Kosten('Bangladesch')   = Kosten('Bangladesch') + (EMGJahr('Bangladesch') * x);

         ... 39 Weitere

        zaehler = Zaehler + 1;
   DISPLAY  kosten,zaehler;
);
```

Figure 3: GAMS Model Carbon Tax Constraint (Own Source)

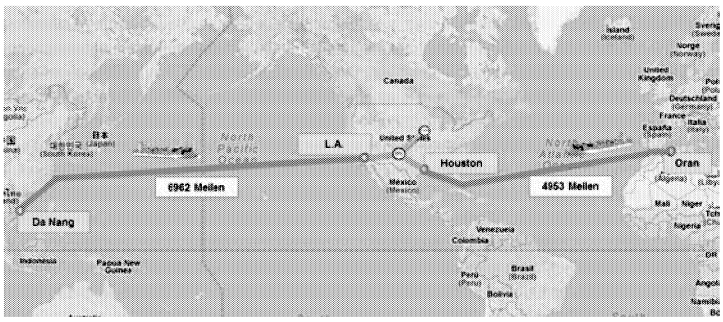


Figure 4: Supply Chain Change by Carbon Tax (Own Map Descirption)

In effect, the tax *threshold value of 130,18 € per ton CO_2* has the first "game-changing" power, i.e. with this tax level the optimal production location would no longer be Vietnam but Algeria due to the shorter transport distance to the OEM facility in the USA.

CONCLUSIONS

The implemented GAMS model and the important extensions in the direction of productivity assumptions in different countries as well as for an international carbon tax regime have shown the simulation value of such location models in terms of assumptions for future developments. These are especially important, as such simulations are run in OEM corporations for locations decisions as standards but not yet in a broader range of supplier corporations in this industry. Therefore such models can be important for decisions taken especially in smaller supplier firms within the automotive supply chain, as they also change their production locations more often than the OEMs for the final assembly point.

REFERENCES

Balcik, B., Beamonm, M.B. 2008. Facility location in humanitarian relief. In: International Journal of Logistics: Research and Applications, Vol. 11, No.2, pp. 101-121.

Bureau of Democracy, Human Rights and Labor Vientam 2009. 2008 Human Rights Reports: Vietnam, http://www.state.gov/j/drl/rls/hrrpt/2008/eap/119063.htm, Download: 17.01.2013.

Cotton Bangladesh 2010. State of Cotton & Textile Industry of Bangladesh, http://cottonbangladesh.com/October2007/StateOfCotton&TextileIndustry.htm, Download: 17.01.2013.

Domschke, W., Drexl, A. 1996. Logistik: Standorte. München: Oldenbourg.

Economic Populist 2012. China and India Really Are Cheap Labor in Manufacturing, http://www.economicpopulist.org/content/china-and-india-really-are-cheap-labor-manufacturing, Download: 17.01.2013.

Eschlbeck, D. 2006. Internationale Wirtschaft: Rahmenbedingungen, Akteure, räumliche Prozesse, München: Oldenbourg.

Hübner, R. 2007. Strategic Supply Chain Management in Process Industries: An Application to Specially Chemicals Production Network Design, Berlin, Heidelberg, New York: Springer.

Isard, W. 2003. History of Regional Science and the Regional Science Association International: The Beginnings and Early History, Berlin, Heidelberg, New York: Springer.

Mainz, M. 2005. Ökonomische Bewertung der Siedlungsentwicklung: Ansatze einer effizienten Siedlungspolitik, Göttingen: V&R.

Miehle, W. 1958. Link-Length Minimization in Networks. In Operations Research, Vol. 6, No. 2, pp. 232-243.

Schätzl, L. 2003. Wirtschaftsgeographie: Theorie, Paderborn: Schöningh.

Statistisches Bundesamt 2012. EU-Vergleich der Arbeits-kosten und Lohnne-ben-kosten für das Jahr 2011, https://www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2012/04/PD12_144_624.html;jsessionid=802F362BB88DDBD018FF7B32048B4D62.cae1, Download: 17.01.2013.

Staudacher, C. 2005. Wirtschaftsgeographie regionaler Systeme, Wien: Facultas.

U.S. Bureau of Labor Statistics 2012. Country at a Glance Mexico, http://www.bls.gov/fls/country/mexico.htm, Download: 17.01.2013.

U.S. Bureau of Labor Statistics 2012. Country at a Glance Philippines, http://www.bls.gov/fls/country/philippines.htm, Download: 17.01.2013.

BIOGRAPHIES

MATTHIAS KLUMPP studied economics and business administration after a vocational degree in logistics (Speditionskaufmann IHK, 1996) at University Leipzig and the IECS Strasbourg from 1995 to 1998 (Diplom-Kaufmann, Diplom-Volkswirt). After his PhD at University Leipzig in 2007 he started as a professor at FOM University of Applied Sciences. He is member of the scientific committee for the German national excellence cluster "LogistikRuhr" and has done research or teaching stays in Australia, China, Finland, France, Netherlands, Norway, Spain, Switzerland Turkey and the USA.

VICTOR KUCHSHAUS is logistics information systems manager in an automotive company and student in the master program logistics management at FOM University of Applied Sciences Essen.

AGV SIMULATION

COUPLING OF A DISCRETE EVENT MULTI AGENT SIMULATION WITH A REAL TIME MULTI AGENT SIMULATION

Christoph Schwarz[1], Jurij Schachmanow[2], Jürgen Sauer[3], Ludger Overmeyer[4] and Georg Ullmann[2]
[1]OFFIS - Institute for Information Technology
[2]IPH - Institut für Integrierte Produktion Hannover
[3] Carl von Ossiettzky University Oldenburg
[4]Leibnitz University Hannover
email: Christoph.Schwarz@offis.de
email: Schachmanow@iph-hannover.de
email: juergen.sauer@uni-oldenburg.de
email: ludger.overmeyer@ita.uni-hannover.de
email: ullmann@iph-hannover.de

KEYWORDS
AGVs, Simulation, Multi Agent Systems, JADE, MASON

ABSTRACT

In this paper we present a coupled multi agent simulation of an automated guided vehicle system. We use two different simulation concepts one based on JADE and one based on MASON and a coupling between these two systems. With this approach we can use specific advantages (e. g. realistic messaging behaviour of JADE) of both systems.

INTRODUCTION

Simulations of Material Flow Systems are usually discrete event simulations (DES, see Wenzel (2006)). In DES, the events are usually represented as flows of goods at fixed times. With these simulations bottlenecks within the material flow system and feasible clock rates for the facilities may be identified. In DES the system analyst models the concrete entities which are involved in the logistics or the material flow such as cranes, floor conveyors, automated guided vehicles (AGVs), etc.. In recent years multi-agent based simulations (MABS) are increasingly in the focus of research (see Davidsson (2001)). It could be shown that MABS are a capable tool for the simulation of intralogistic processes (see for example Becker et al. (2006) or Gehrke et al. (2006)). It also can be used for the analyses of cognitive logistic systems (Ommen et al. (2010), Ommen et al. (2010), Beth et al. (2009), Overmeyer and Heiserich (2009)) and to study the impact of individual decisions of forklift drivers (Schwarz and Hahn (2012)).
Usually it is necessary to make a compromise between simulation speed and realism in MABS. Especially the communication between agents on different platforms is often a bottleneck for the simulation speed. For our current research in decentralised control of AGV systems we would like to have both attributes. In order to be able to have a fast simulation to make many simulation runs for parameter tuning without the loss of a realistic communication behaviour we developed a coupled simulation. The simulation consists of a MASON part which is able to run many simulation runs in a short amount of time and a JADE part which is able to simulate the communicational behaviour of the agents in a realistic manner. In this paper we present our approach along with an evaluation on the simulation speed.

JADE MULTI AGENT SIMULATION

The first attempts to implement AGV systems by an agent system were already made (see for example Radosavac et al. (2011) or Heiserich and Overmeyer (2010). Existing decentralized solutions, however, are not consistently decentralized - they use centralized algorithms or are just logically but not physically distributed. E.g. a centralized dispatching algorithm decides which one of the transport orders should be processed next and by which AGV. For this reason we are trying to implement a completely decentralized, agent-based system, which uses only decentralized algorithms and is distributed among the (hardware) entities in the AGV system.
For our purpose we chose a well known framework called JADE (Java Agent Development Framework[1]), that was developed by Tilab (Telecom Italia Lab). It simplifies the implementation of multi-agent systems through a middle-ware that complies with the FIPA specifications [2] and through a set of graphical tools that supports the debugging and deployment phases. JADE is widely used in science and is being continuously further developed. The latest version is 4.2.0 from 26/06/2012. One of the reasons for choosing JADE is the platform independence, because it is fully implemented in Java. An-

[1]http://jade.tilab.com/
[2]http://www.fipa.org/

other important reason is the compatibility with FIPA standards. This makes our system also compatible with other agent-based system that follows the FIPA conventions. Following we briefly introduce the architecture of JADE.

The core of each agent system are the agents. They can communicate with each other and perform tasks. Agents need an execution environment. In JADE they are executed in the so called containers, which themselves existing in an agent platform. The agent platform provides the agents with basic services like communication and supervise them during their life cycle: initializing, monitoring and killing (see figure 1). Each platform has at least one container – the main container. It is always started first and already contains two special agents: AMS (Agent Management System) and DF (Directory Facility). These agents are mandatory for the whole system and helps in supervising another agents.

Also in the figure 1 you may see, that there are two ways in JADE to create a distributed system. The first way is to distribute containers of the same platform on different hosts. In a logical view you then have only one platform and one main container. That leads to only inner-platform communication and a simple way to address agents with their local names. Another way to create a distributed system is to install an agent platform on each of the hosts involved in the system. A disadvantage of this way is the necessity of inter-platform communication and therefore the necessity of fully qualified agent names like nickname@hostname:port. But the other way round your agent platform and hosts are completely independent from each other. There is even no need to install strictly JADE-platforms - the only restriction to the platforms is being FIPA compatible. This is a big advantage.

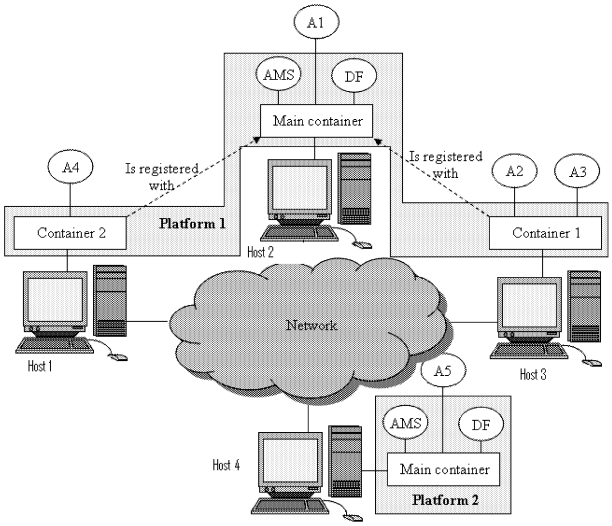


Figure 1: The JADE Architecture. Grimshaw (2013)

JADE-based AGV System

Before we introduce the whole system let us do some considerations. An AGV system mainly consists of minimum one automated guided vehicle, of a guidance control system and some infrastructure. So the main characters in the system are the AGVs: they do the transports, communicate among each other and with the stations, and pay attention for another inner-traffic. That is why we could say, that each AGV is a small agent system itself, with agents doing each his own task like communicate, drive, doing load pickups and so on. From these considerations we identified for an AGV system the agents shown in the table 1.

Table 1: Overview of the Agents in an AGV

Name	Description
Manager	Control and coordination of other agents
StatusManager	Status report and monitoring
ResourceManager	Managing of a resource
TransportOrderManager	Managing the transport orders
TravelOrderManager	Managing the travel orders
SafetyManager	Interface to the safety system
DriveManager	Interface to the driving system
LoadManager	Interface to the load pickup system
PowerManager	Interface to the power system

The nine identified agents are the core of the system. Following we present the detailed description of each of them: The "Manager"-agent coordinates the work and communication between other agents. He delegates tasks and has an overview of the systems status. So in case of conflicts or failures he can initiate the necessary actions. The "Manager" has an interface to other entities in the system. So two "AGV-Managers" can communicate with each other directly. Other agents must first contact their own "Manager" in order to communicate with external agents.

The "StatusManager" agent is a monitoring interface between the entity (AGV or station) and the outer world. All status messages from another agents are collected and prepared here, so the system user has access to important statistics. Typical statistics are fault statistics, order statistics, throughput, utilization rate of the system/vehicles, downtimes, utilization rate of transport paths, utilization rate of buffers, turnaround time, availability and energy balance as de-

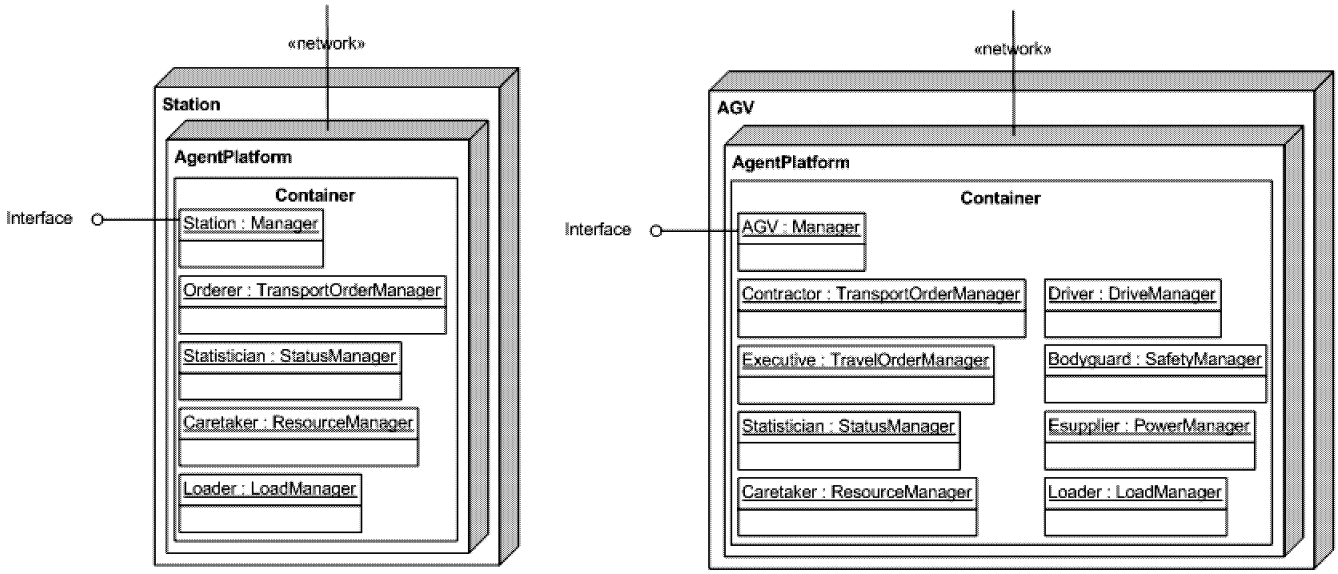


Figure 2: The Deployment of the Agents

fined in VDI guideline 4451-7(see Verband Deutscher Ingenieure (VDI) (2005)). All these statistics are stored on the entity but also send to a monitoring station. This station is the only one central component in our system and has no control but only monitoring functions.

An AGV system consists of resources of any kind: each section of the course must be reserved by the AGVs, each station must be reserved by an AGV in order pickup something there. For the purpose of managing a resource we defined the "ResourceManager" agent. The two roles of this agent are owner and reserver. In the example of an AGV reserving a station, the owner lives on the station and the reserver lives on the AGV.

The most important function of an AGV system is the transport order processing. For this purpose we defined the "TransportOrderManager" agent. He lives in an AGV and in a station. In a station the only tasks of "TransportOrderManager" are to check a list of transport orders for its feasibility and, in the affirmative, start an auction (see [3]) for allocation a transport order to an AGV. In an AGV the tasks of "TransportOrderManager" are varied: participating in auctions, completion of incomplete transport orders, decomposition to travel orders and monitoring the processing of them and so on. The "TravelOrderManager" agent receives the travel orders from "TransportOrderManager" agent. It then decomposes them to the basic commands like start/stop driving or start/stop the load suspension devices (LSD) and forwards them to the "Manager" for execution. After a command is executed the "TravelOrderManager" checks a list for new travel orders and if so, starts a new decomposition.

The "RouteManager" agent lives in an AGV and is responsible for finding a route from the transport source to the transport target. Therefore the "RouteManager" possess a virtual map. The course layout on the map is built of a graph, so finding a route is actually the same as finding a shortest path in the graph. The only restriction is that a found path can also be used by another AGVs. Thus, in addition to finding a path, the "RouteManager" must be able to reserve needed segments of the course and is therefore communicating with the "ResourceManager".

The four following agents are similar to the modules defined in VDI guideline 4451-4 (see Verband Deutscher Ingenieure (VDI) (1998)). Their main purpose is to define an interface to the hardware of an AGV.

The "SafetyManager" agent has control over the safety system of an AGV and performs the following functions: evaluating of the safety sensors like frontal and side impact protection, activating the safety actuators like the emergency stop circuit. The "SafetyManager" communicates with the "Manager" only. The "DriveManager" agent has control over the driving system of an AGV and performs the following functions: track guidance, position sensing and location determination. The "DriveManager" communicates with the "Manager" only. The "PowerManager" agent has control over the the battery of an AGV and performs the following functions: monitoring the charge status and the condition of the battery. The "PowerManager" communicates with the "Manager" only. The "LoadManager" agent has control over the LSD of an AGV and performs the following functions: evaluation of the load safety sensors and coordinating the LSD. The "LoadManager" communicates with the "Manager" only.

With these nine agents we reproduced all the functionalities of a real centralized AGV system. After imple-

[3]FIPA Contract Net Interaction Protocol Specification: http://www.fipa.org/specs/fipa00029/index.html

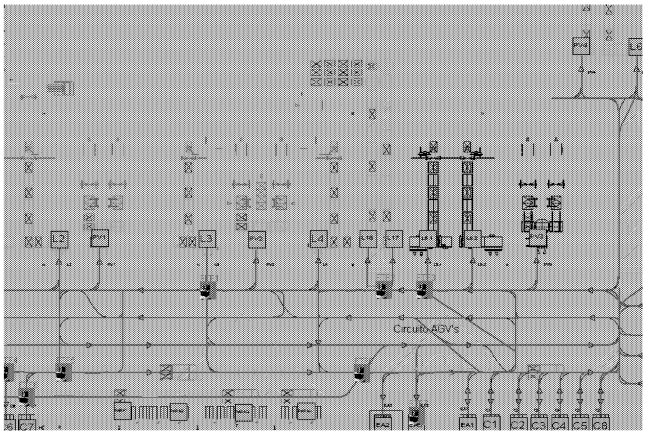


Figure 3: The MASON Simulation

menting the agents and the algorithms we planned the deployment of the agents on the system entities. In figure 2 you can see an AGV and a station each one with an agent platform and a container. Inside the container the identified agents are placed. Which of the agents are placed depends on the needed functionality.

MASON MULTI AGENT SIMULATION

The second part of out coupled simulation is based on the java library MASON (see Luke et al. (2004)). MASON was developed specifically for the construction of multi-agent simulations in which the position of the agents matters. The MASON library is modular. The core of the model library provides the basic tools of MASON. The most essential part is a scheduler for temporal control of agents. The scheduler implements a step-based simulation. Methods of an agent can either be run once in a given simulation step in the future or in a sequence at specific time intervals. To implement an agent that is to be controlled by the MASON scheduler, it must implement the method step from the interface Steppable. Then this class can be marked for execution. The agents then have to calculate their actions in their step method.
Our MASON based simulation essentially consists of the following agents: AGV-agents, station-agents, sensor-agents and agents for the dispatching of orders. A screenshot of the simulation is shown in figure 3. The AGV-agents are responsible for the movement. They also represent the different vehicles with their different properties like speed or width. The station-agents represent the loading and unloading stations. They are responsible for the connection between vehicle and station. The sensor-agent checks for collisions. It is in our case a virtual sensor which works without errors. The order dispatching agents are responsible for the dispatching of the transport orders to the vehicles if the MASON simulation works in the stand-alone mode (see section COUPLING). The Layout in the MASON simulation is modelled as an annotated graph in which every station and every crossing is modelled as node and the ways between crossings and station as edges. The edges have an attribute width which is used to determine if two or more vehicles may use the same edge at the same time.

COUPLING

Since MASON and JADE are both java frameworks we developed the important algorithms for the decentralised control of AGV systems like the route planing, order allocation, conflict resolution and so one (see Schwarz and Sauer (2012)) in a manner that is compatible with both simulations. By this we are able to use the MASON simulation and the JADE simulation as stand-alone simulations for all AGV system specific tasks. In this stand-alone modes both simulations simplify certain aspects of the real processes. The stand-alone MASON simulation simplifies the communication between the agents and the stand-alone JADE simulation simplifies the hardware and the movement of the vehicles in the plant or warehouse. In these stand-alone modes the simulations are strapped down to the needs of these subtasks and therefore work quit efficient (see section exemplary evaluation). For a realistic simulation of a given system the simulations work in a coupled mode.
In this mode the MASON simulation simulates only the vehicles and the environment. It executes the driving of the vehicles makes collision checks with the virtual sensor takes care of the battery unloading and loading (if this is relevant in the scenario) and simulates the loading and unloading of the transportation goods.
The JADE simulation is responsible for the control of the system. Since the simulation presented in this paper was designed for the evaluation of decentralised control mechanisms in AGV systems with autonomous AGVs this control is heavily based on communication. The agents communicate in an auction process to get transport orders they communicate for way segment locks which they need to drive without conflicts (again see Schwarz and Sauer (2012)) and for conflict resolutions. In the JADE simulation these agents have to use communication as in a real system (in fact the JADE agents can be deployed on different hardware platforms especially they can be deployed on the target hardware (usually small industry pcs mounted on the AGVs)).
To couple the MASON system which relies on a fixed stepping (in the example in the next section we use a stepping of one second per step) and the JADE system which relies on discrete events which have no fixed timing we use a message / freeze based coupling. The coupling is shown in figure 4. First the MASON simulation simulates one step. In this step the movement of the vehicles the progresses of loading and unloading as well as battery loading and unloading are determined. Also sensor warnings for possible collisions are calcu-

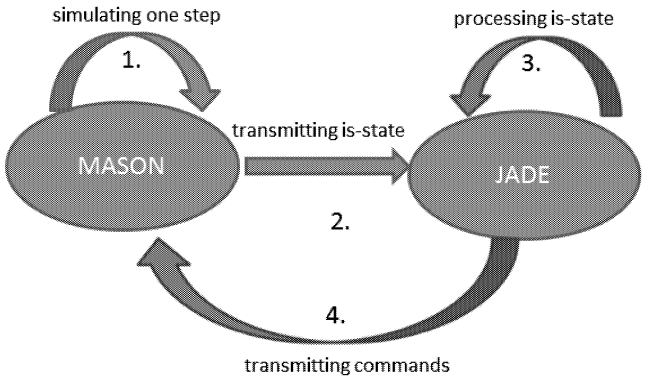


Figure 4: The coupling of the simulations

lated. The relevant part of the current is-state are than sends to the JADE simulation and the MASON simulation is frozen. In the third step the JADE simulation processes the is-state and than sends new command orders to the MASON simulation. Command orders are move or start loading / unloading orders for the vehicles or start docking / undocking orders for the station. The MASON simulation wakes up after it got these commands. These cycle repeats until the end of the simulation. If needed the MASON simulation can also be waked up after a fixed period of time (usually the time which is simulated in one step). This is needed if very time consuming algorithms are used in the JADE simulation which may take longer than this period of time. The simulation then takes another step without new commands from the JADE simulation and the JADE commands may become obsolete. This can lead to an undesired system behaviour. If this is the case simpler algorithms or faster PCs on board of the AGVs have to be used.
Technically we use predefined messages which we send over sockets. In these messages every line represents an event or a command. So only two messages have to be send in each simulation step. One from the MASON simulation to the JADE simulation and one in the opposite direction. For example the MASON simulation sends the events of the following types:

- AGV ID reachedNode ID,
- AGV ID finishedLoading ID,
- AGV ID finishedUnloading ID,
- AGV ID hasSensorWarning ID distance,
- AGV ID hasBattery percent,
- STATION ID hasCoupled ID,
- STATION ID hasUnCoupled ID .

For example the JADE simulation sends the commands of the following types:

- AGV ID driveTO ID,
- AGV ID startLoading ID,
- AGV ID startUnloading ID,
- AGV ID doEmergencyBreak,
- AGV ID enterIdleMode,
- AGV ID startBatteryLoading.

EXEMPLARY EVALUATION

We used the simulation approach described in the previous sections for the simulation of a beverage bottling plant (shown in figure 3). The area is rectangular with a dimension of 79m x 125m. The plant has 33 stations at which loading and unloading of transport goods takes place and 4 parking positions. In average 46 transport orders must be performed per hour. In total 8 AGVs are in use in order to fulfil these transport orders.
In order to measure the impact of the coupling to the simulation performance we compared the running speeds of the MASON stand-alone solution with the coupled solution. Since we developed the algorithms in a way that they are usable by the MASON simulation and the JADE simulation the running time of the algorithms for calculating routes or auction bids is the same in both.
All the simulation runs were performed on a notebook using only a single core of an intel i3 processor. In this way network delays which could occur in the message sending process could be bypassed.
First we made 1.000 runs with the MASON stand-alone simulation. In average these simulations were $\approx$ 6,500 times faster than realtime (meaning the simulation of a hour real time takes $\approx$ 550 milliseconds). Afterwards we made 1.000 runs with the coupled version. Since we have used the same algorithms with the same parameters the behaviour of the vehicles and stations and thus the simulation results were identical. In these runs the simulation was $\approx$ 400 times faster than realtime. This means that the simulation is $\approx$ 16 times slower in the coupled mode than in the uncoupled mode.

CONCLUSIONS

We have developed a coupled simulation based on MASON and JADE and a set of algorithms implemented in JAVA which are usable by both simulations and can now use the benefits of both simulation frameworks. If many simulation runs are needed, for example to use heuristics like tabu search (see Glover and Laguna (1997)) or ant colony optimization (see Dorigo and Birattari (2010)), the speed of the MASON simulation can be used. If on the other hand realistic communication should be taken into consideration the coupled version can be used. In this way you can first optimize the algorithms with the

MASON simulation and then test these algorithms with the real JADE agents which can then be deployed on real hardware. As shown the coupling itself slows down the simulation only by a factor of ≈ 16 which is acceptable for simulating realistic system behaviour without the need of optimization (which can be done before in the stand-alone mode).

ACKNOWLEDGEMENTS

The research project "Decentralized, agent based self-control of Automated Guided Vehicle Systems" (17237 N) of the Federal Unified Association of Logistics (BVL) was funded via the AiF within the framework of cooperative industrial research (IGF) by the Federal Ministry of Economics and Technology (BMWi) due to a resolution of the German Bundestag.

REFERENCES

Becker M.; Wenning B.L.; Görg C.; Gehrke J.; Lorenz M.; and Herzog O., 2006. *Agent-based and discrete event simulation of autonomous logistic processes.* In *Proceedings of the 20th European Conference on Modelling and Simulation.*

Beth C.; Kamenik J.; Ommen D.; and Hahn A., 2009. *Design Aspects of Cognitive Logistic Systems.* In *2nd International Conference on Dynamics in Logistics (LDIC).*

Davidsson P., 2001. *Multi Agent Based Simulation: Beyond Social Simulation.*

Dorigo M. and Birattari M., 2010. *Ant Colony Optimization.* In *Encyclopedia of Machine Learning.* 36–39.

Gehrke J.D.; Lorenz M.; Wenning B.L.; and Becker M., 2006. *Integration of two approaches for simulation of autonomous logistic processes.* In *Tagungsband zur 12. ASIM Fachtagung.*

Glover F. and Laguna M., 1997. *Tabu Search.* Kluwer Academic Publishers, Norwell, MA, USA.

Grimshaw D., 2013. *JADE Administration Tutorial.* URL `http://jade.tilab.com/doc/tutorials/JADEAdmin/index.html`.

Heiserich G. and Overmeyer L., 2010. *Context-based Learning of Adaptive Vehicle Dispatching Strategies.* In *Proceedings Artificial Intelligence and Logistics (AILog-2010) - Workshop at 19th European Conference on Artificial Intelligence (ECAI).*

Luke S.; Cioffi-revilla C.; Panait L.; and Sullivan K., 2004. *MASON: A new multi-agent simulation toolkit.* In *Proceedings of the 2004 SwarmFest Workshop.*

Ommen D.; Kamenik J.; Beth C.; Busch J.C.; Kulas A.; Cramer E.; and Hahn A., 2010. *An Adaptable Architecture for Intelligent Conveyors.* In *ICINCO (1).* 249–254.

Overmeyer L. and Heiserich G., 2009. *Cognitive Transportation Modules for Adaptive Material Flow Systems.* In *Proceedings of the Distributed Intelligent Systems and Technologies Workshop (DIST).*

Radosavac M.; Heiserich G.; and Overmeyer L., 2011. *A Reinforcement Learning Method for Dispatching Strategies in Intralogistics.* In *Proceedings of the Distributed Intelligent Systems and Technologies Workshop.*

Schwarz C. and Hahn A., 2012. *Modeling Individual Decision Makers In A Multi Agent Warehouse Simulation.* In *Proceedings of the 10th International Industrial Simulation Conference.*

Schwarz C. and Sauer J., 2012. *Towards Decentralised AGV Control With Negotiations.* In *Proceedings of the Sixth Starting AI Researchers' Symposium.*

Verband Deutscher Ingenieure (VDI), 1998. *VDI 4451-4: Compatibility of Automated Guided Vehicle Systems (AGVS). Open control system for automated guided vehicles (AGV).*

Verband Deutscher Ingenieure (VDI), 2005. *VDI 4451-7: Compatibility of Automated Guided Vehicle Systems (AGVS). AGVS guidance control system.magic.*

Wenzel S., 2006. *Planung und Simulation in logistischen Anwendungen Einleitendes Statement zum Stand der ereignisdiskreten Simulation in der Logistik*, 379–380.

DECISION-MAKING IN CONCEPTUAL AGV SYSTEMS DESIGN USING SIMULATION-BASED OPTIMIZATION

Leif Pehrsson[1, 2]
Amos H.C. Ng[2]
[1]Volvo Car Corporation, SE-405 31, Gothenburg, Sweden
[2]Virtual Systems Research Centre, The University of Skövde, SE-541 28, Skövde, Sweden
E-mails: leif.pehrsson@his.se, amos.ng@his.se

KEYWORDS

AGV Simulation, Decision-making, Multi-Objective Optimization, Innovization, Visualization.

ABSTRACT

In order to shorten the development lead-time and enable the evaluation of many alternative conceptual scenarios, the use of aggregated modeling techniques is required. On the other hand, it is believed that well-informed decision making in conceptual production systems development can be elevated to a new level through the application of multi-objective optimization together with techniques for post-optimality analysis and Visualization of high-dimensional data. An industrial case study regarding the conceptual design of an AGV system has been carried out to evaluate the usefulness of such a combined methodology. The case study, based on a real-world industrial decision-making problem, demonstrates how knowledge can be extracted from optimization data in order to support decision-making which cannot be easily done with other industrial practices.

INTRODUCTION

While analytical methods like Petri nets (Castillo et al. 2001) and queuing models (Sang and Kwon 1994) can be effectively applied to AGV systems design, simulation (in particular discrete-event simulation) is still the most practical choice for industrial practitioners (Tempelmeier 2003)(Vis 2006). This is not a surprise when considering that any complex products mix and their routings can be easily modeled using almost any simulation software. On the other hand, simulation is not an optimization tool when used standalone. The connection of meta-heuristic search methods to simulation models, referred to as Simulation-based Optimization (SBO), can be used for finding the optimal setting of many decision variables in a design optimization problem (Law and McComas, 2002). Despite the potential of finding optimal or near optimal solutions opened by SBO, the task of finding essential parameters, design principles and relations within data from SBO can be a very time-consuming and daunting task. The concept of innovization through optimization referred to as *innovization*, as introduced by Deb and Srinivasan (2009), incorporates post-optimality analysis on the solution sets generated from multi-objective optimization in order to reveal essential design rules, hidden in the Pareto-based optimal solutions. By integrating advanced visualization techniques of high-dimensional data, for example, Mosaic plots, Parallel Coordinate (PC) plots, and other glyph-based plots (Liebscher et al., 2009), into the "innovization" process, it is believed that decision-making can be supported by an elevated knowledge about the system.

The type of aggregated conceptual production systems modeling and optimization approach applied in this paper can be used for the integration of several production lines, transport solutions and logistic systems into one model. FACTS Analyzer (Ng et al., 2007), or FACTS hereafter, as a conceptual factory modeling and simulation tool, which has built-in support for multi-objective optimization for production systems design problems has been used throughout the study. Previously, we have reported the use of FACTS in a comprehensive industrial cost optimization study (Pehrsson and Ng, 2011). The purpose of this paper is to demonstrate how the post-optimality analysis and visualization techniques supported by FACTS can support the innovization process to gain insight to support decision-making through a case study of analyzing an Automated Guided Vehicles (AGV) based transport system which connects several new production cells. An essential part of the problem was to make a fast analysis of different alternative options in order to decide if the concept with an AGV system should be further developed or if alternative solutions should be considered.

AN INDUSTRIAL CASE STUDY

In contrast to other recent simulation studies of AGV system design, e.g. (Duinkerken et al. 2006)(Klaas et al. 2011), which are more focused on routing strategies, the main task involved in this case study was to evaluate and optimize the required number of AGVs and inter-station storage capacities. These are the most important parameters to be selected during the decision making of the conceptual phase of the production system design for further development. An AGV system can be fairly costly and the number of AGVs has to be

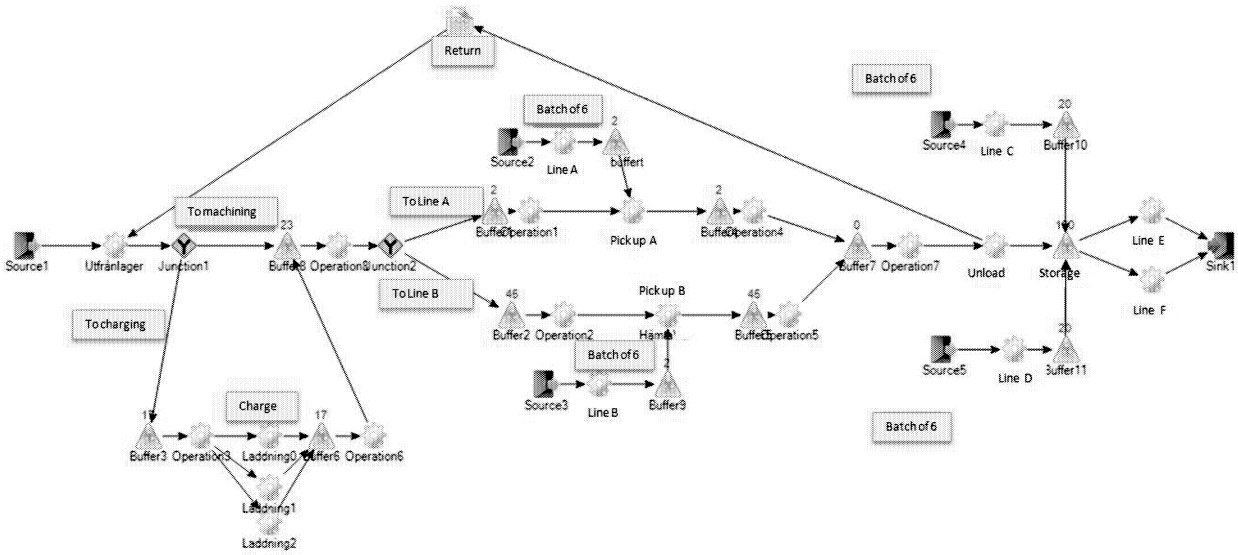


Figure 1: Conceptual flow model of the AGV system and the main production system.

kept to a minimum in order to reduce the total investment cost. Design of Experiments (DoE) is a common approach to address this level of optimization. Similar to the innovization approach, DoE facilitates the optimization and analysis of the number of AGVs and its relationship with other design variables (or factors using the term in DoE); see (Gebennini et al. 2008). Nevertheless, as will be revealed in this paper, innovization can offer a powerful, yet more interactive and intuitive approach to gain insights into solving the design problem, than DoE.

System Data

As input for the study, a conceptual layout of the production system, including all distances, was used together with information on AGV performance such as speed, battery capacity, charging time. The capacity and throughput figures of each connected line was aggregated and described through cycle time, availability, mean time to repair (MTTR) based on separate simulation and optimization of the individual lines. The availability and MTTR of the AGVs was also considered. The storage was defined by capacity and transport time. The actual parameter values can be found in Table 1.

Table 1. Parameter values in the simulation model.

Object	Cycle time	Availability	MTTR
Line A	120	75	25
Line B	80	75	25
Line C	90	75	25
Line D	229	84	25
Line E	55	85	5
Line F	55	85	5
Pick up A	66	99	5
Pick up B	66	99	5
Unload	66	99	5

Modeling

As a conceptual model, the FACTS simulation model was built following the logical flow of the AGVs, rather than the actual movements of the AGVs. Standard FACTS objects, such as operations and buffers, were used in combination with flow logic to represent the behavior of the AGV system. The carriers were modeled using the standard pallet objects; see (Ng et al. 2011) for a list of the standard FACTS objects. The completed FACTS model is shown in Figure 1.

The average speed of an AGV and the distances from the layout were recalculated into transport times and included into buffer objects. The capacities of these buffers were set to mirror the maximum number of AGVs possible to "store" along each transport according to distances taken from the layout. Junctions were used to emulate the flow logic and the routing was aggregated into percentages of the AGVs following a part of the flow. A maximum of three battery charging stations were planned and 5% of the AGV flow would follow the charging route and stay at a charging station for one hour at a time. The share of the AGV flow to pick up products from line A was 40% and the share to pick up products from line B was 60%.

Stage 1 Optimization

An optimization of the system was performed using the widely-used NSGA-II algorithm (Deb et al. 2002) with 1000 simulation evaluations (each 5 replications). The input parameters were the number of carriers, n, with range [1, 10], and the storage capacity, s, with range [0, 1200]. Main objectives were to simultaneously minimize the number of carriers, minimize the storage capacity and maximize throughput. In other words, the optimization problem was formulated as follows:

{Max($TH(n, s)$), Min(n) and Min(s)};

where TH is the system throughput, which is a function of n and s, number of AGVs and storage capacity, respectively.

Stage 1 Analysis

The throughput was above various storage capacity levels, dependent on the number of AGVs released into the system. With a storage capacity larger than 150 units, the system was found to be stable from a throughput perspective. Filtering the data for solutions with storage capacity larger than 150 units revealed nine clusters, each with a number of solutions with the commonality of having the same throughput. By observing the decision-space parameters for each cluster, it was found that each specific number of AGVs in the system was connected to a corresponding cluster. Three and four AGVs ended up in the same cluster limiting the total number of clusters to nine. Hence the number of carriers was identified as the main factor affecting the system throughput. The highest throughput was achieved by investing (or releasing) 6 carriers into the system.

Visualization

With stabilized throughput secured by a correctly dimensioned storage, the number of AGVs clearly constrained the system throughput to certain levels. This can be seen by the lines in the scatter plot showing the optimization results in Figure 2. The values on the right-hand side of Figure 2 show the number of AGVs with the corresponding throughput.

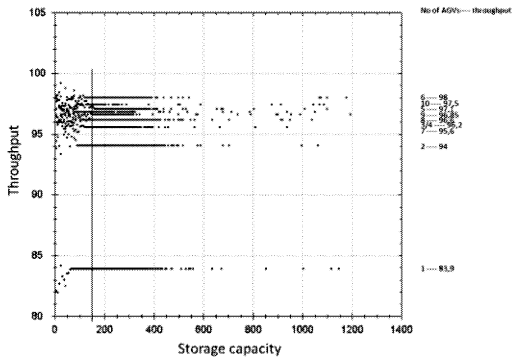


Figure 2: Initial optimization results.

Six AGVs would be required in order to reach the maximum throughput achieved through the optimization of the system. With just one AGV the throughput only became 83.5 jobs per hour (JPH) and despite that 94 JPH could be achieved by using two AGVs, the decision was to investigate alternative solutions.

The actual throughput objective was to reach a total of 100 JPH from line A, B, C, D, E and F. In order to do that other aspects of the system had to be considered for re-design. The result was somewhat unexpected and required deeper analysis.

Mathematical Modeling: A Comparison

A mathematical method for dimensioning of AGV systems (Groover, 2008) with even further conceptual analysis nature was applied. The required number of AGVs (n_c) can be estimated by the calculation of the average AGV work load (WL) and the available time (AT). The estimated n_c is then equal to WL/AT. In this case the estimation resulted was n_c=0.97. In other words, only one AGV would be required. The capacity would however be somewhat limited with only one AGV considering that a 60-minutes charging time is required every 20 hours. The combined theoretical hourly capacity from line A and B would be reduced from 56.25 to 53.44. The theoretical capacity from line A and B was also confirmed to be 56.25 by a deterministic simulation (with the availability included as tact time) of the corresponding part of the system with one AGV and no charging. Based on this the recommendation from the mathematical study would be to invest in two AGVs. There were however other parameters in the system that could be considered for optimization. The capacity of the buffers on the output conveyors from line A and line B were such parameters. The stochastic behavior of the Line A and B with 75 percent aggregated availability and 25 minutes MTTR might actually require another dimensioning of these buffers. Increased buffers would also be more forgiving for eventual effects from the simplified conceptual control logic when splitting the AGV flow to line A and B.

Stage 2 Optimization

The input parameters to the optimization were updated to include the capacities of Buffer A and Buffer B with a range of 1 to 10 transport loads with 6 jobs each. A new optimization, using NSGA-II was run with 1000 iterations, again each replicated for five times. The sum of the capacities in Buffer A and Buffer B was introduced as an optimization objective for minimization.

Stage 2 Analysis

The data from the second optimization run revealed some new opportunities. The maximum throughput from the complete system with one AGV increased

to 90.3 JPH, with two AGVs to 98.6 JPH and with four AGVs it would be possible to reach 100 JPH as shown in Figure 3.

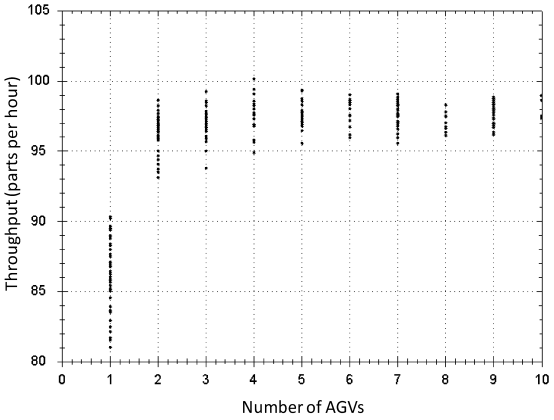


Figure 3: Number of AGVs and possible throughput from optimization run 2.

The interesting question was then what configuration would be required to run the system at a throughput of 98.6 JPH with only two AGVs.

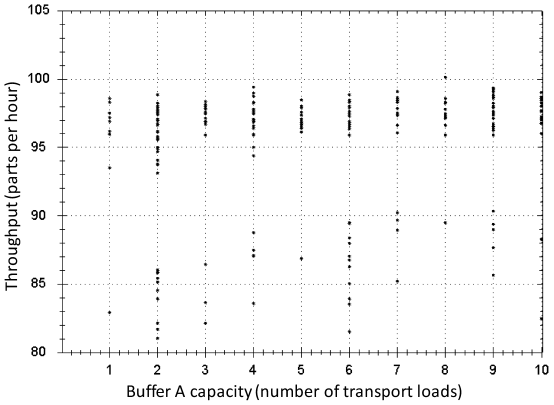


Figure 4: Buffer A capacity against throughput.

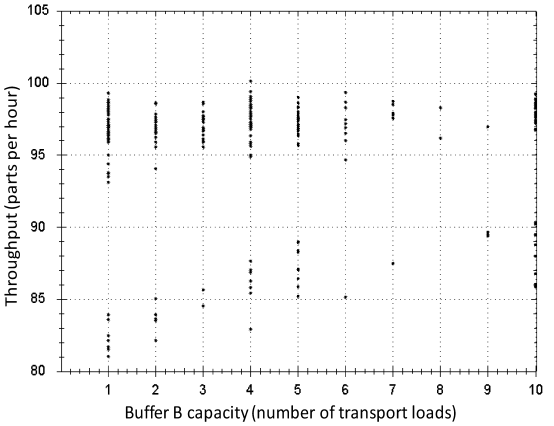


Figure 5: Buffer B capacity against Throughput.

In order to reveal more information, the new buffer parameters were plotted against throughput according to Figure 4 and Figure 5. According to these plots it is possible to reach a throughput of 98.6 JPH with a transport load capacity of just one. This is shown to be valid for both the buffers.

Apparently more information is required in order to make an informed decision. By plotting the optimization objectives in a 4D graph (3D plus color), shown in Figure 6, some insight about this industrial problem can be discovered. By observing the two clusters in Figure 6, there is apparently some interaction effect between certain parameters which can leverage the system to reach the desired performance trade-off. It will be possible to find solutions with 2 AGVs, an acceptable throughput level and one of the Buffers A or B at one transport load as long as the sum of buffers, A+B, is large enough. Even more information about the system can be revealed by using a PC plot, as shown in Figure 7, which connects objective space parameters with decision space parameters.

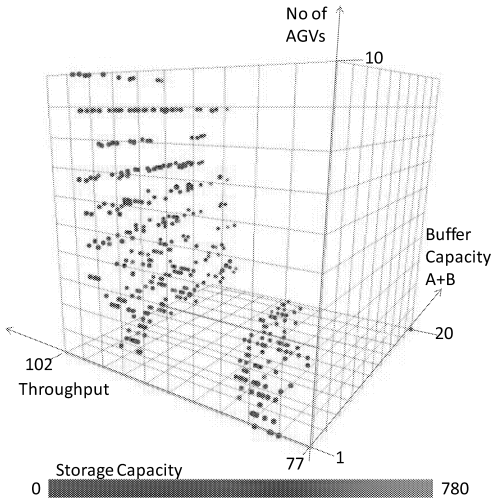


Figure 6: 4D plot of the objective space.

By filtering out solutions with throughput above 98.5 JPH and storage capacity smaller than 50 it is becoming obvious that the number of AGVs required in the system will be 6 or 7, the Buffer A capacity must be 9 or 10 and Buffer A+B capacity should be 12 or 13. The major problem with these results is the amount of AGVs required.

If the data is filtered for solutions with throughput larger than 98.5 JPH and the number of AGVs smaller than 3 another trade-off is to be found. This is illustrated through the parallel coordinate plot in Figure 8. By increasing the storage capacity to some ware between 340 and 535 while Buffer A capacity is kept at 10 the throughput and AGV-amount objectives can be met.

Decision Making

The number of AGVs required to operate the system at sufficient throughput with an acceptable dimensioning of the storage capacity was

considered to be high to motivate the investment in AGVs. The relatively large Buffer A shown to be important to fulfill the other objectives was not really a feasible solution when considering layout issues. In the end an alternative layout was found to be much more favorable and the investment in an AGV system could be omitted.

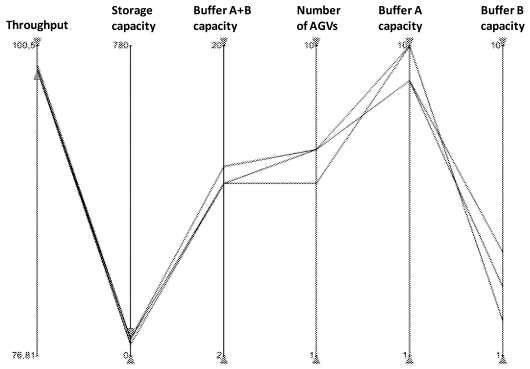


Fig 7: Parallel coordinate plot focusing on high throughput and a small storage.

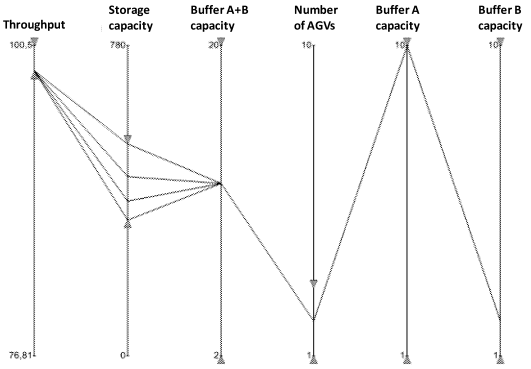


Figure 8: Parallel coordinate plot focusing on high throughput and a small number of AGVs.

CONCLUSIONS

In this paper, the use of a combination of post-optimality analysis and advanced visualization techniques, like 4D and PC plots, has shown to be very useful for the extraction of essential parameter settings from optimization data and gain knowledge about the problem to support decision-making. This is demonstrated through applying such an innovization process to a conceptual, aggregated, logical flow model of an industrial AGV system. The target was to optimize the number of AGV required, together with the simultaneous optimization of the inter-station storage capacities.

An imperative finding was the importance of including enough parameters to truly optimize the system. The application of a mathematical model for the same problem has also been performed in this study and compared with the results from the first stage of optimization. In has been found that the results obtained from the mathematical model could also be used to improve the setup of the optimization problem by including only the essential parameters in the subsequent optimization stage in order to find more favorable trade-off solutions.

ACKNOWLEDGEMENT

This study is partially funded by the Knowledge Foundation (KKS) of Sweden through the BlixtSim project. The authors gratefully acknowledge their provision of research funding.

REFERENCES

Castillo, I., Reyes, S.A., Peters, B.A. (2001). Modeling and analysis of tandem AGV systems using generalized stochastic Petri nets, *Journal of Manufacturing Systems*, Vol. 20, No.4, 236–249.

Deb, K., Agrawal, S., Pratap, A. and Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, Vol. 6, No.2, 182–197.

Deb, K. and Srinivasan, A., (2006). Innovization: Innovating design principles through optimization. In *Proceedings of the Genetic and evolutionary Computation Conference (GECCO-2006)*, The Association of Computing Machinery (ACM), New York, 1629-1636.

Duinkerken, M.B., Lodewijks, G. and Ottjes, J.A. (2006). Comparison of Routing Strategies for AGV Systems Using Simulation. In *Proceedings of the 2006 Winter Simulation Conference*, L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds. 1524-1530.

Groover, M.P. (2008). *Automation, Production Systems and Computer-Integrated Manufacturing*, Pearson Education Inc. Upper Saddle River, New Jersey 07458.

Gebennini, E., Dallari, S., Grassi, A., Perrica, G., Fantuzzi, C. and Gamberini, R. (2008). A Simulation Based Approach for Supporting Automated Guided Vehicles (AGVs) Systems Design. In *Proceedings of the 2008 Winter Simulation Conference.* S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler eds. 2156-2163.

Klaas, A., Laroque, C., Fischer, M. and Dangelmaier, W. (2011). Simulation aided, knowledge based routing for AGVs in a distribution warehouse. In *Proceedings of the 2011 Winter Simulation Conference*, S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, eds. 1673-1684.

Law, A.M. and McComas, M.G. (2002). Simulation Based Optimization. *Proceedings of the 2002 Winter Simulation Conference,* WSC2002, San Diego, California, USA, December 8-11, 2002, 41-44.

Liebscher, M., Witowski, K., Goel, T., 2009. Decision-making in Multi-Objective Optimization for Industrial Applications – Data Mining and Visualization of Pareto Data. *8th World Congress on Structural amnd Multidisciplinary Optimization*, June 1-5, 2009, Lisabon, Portugal.

Ng, A.H.C., Urenda, M., Svensson J., Skoogh, A. and Johansson, B. (2007). FACTS Analyser: An innovative tool for factory conceptual design using

simulation. In *Proceedings of Swedish Production Symposium*, Gothenburg 28-30 August 2007.

Ng, A.H.C., Bernedixen, J, Urenda, M. and Jägstam, M. (2011). Factory Flow Design and Analysis Using Internet-enabled Simulation-based Optimization and Automatic Model Generation. In *Proceedings of the 2011 Winter Simulation Conference*, S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, eds. 2181-2193.

Pehrsson, L. and Ng, A.H.C. (2011). An applied framework for simulation-based multi-objective production system optimization within production system development. In *Proceedings of the Industrial Simulation Conference*, ISC'2011, June 6-8, 2011, Venice, Italy.

Sung, C.S. and Kwon, S.T. (1994). Performance modeling of an FMS with finite input and output buffers. *International Journal of Production Economics*, Volume 37, Issues 2–3, December 1994, Pages 161-175.

Tempelmeier, H. (2003). Practical considerations in the optimization of flow production systems. *International Journal of Production Research*, 41, No. 1, 149-170.

Vis, I.F.A. (2006). Survey of research in the design and control of automated guided vehicle systems. *European Journal of Operational Research*, Vol. 170, Issue 3, 677-709.

AUTHOR BIOGRAPHIES

LEIF PEHRSSON holds a M.Sc. degree in Automation Engineering and a B.Sc. degree in Mechanical Engineering, both from the University of Skövde. He has been working for more than 15 years in various positions, including production engineering manager, within automotive industry. In year 2000, he received the Henry Ford Technology Award for his work on production system development. Started from 2008, he is employed as a Volvo Industrial PhD Program (VIPP) candidate and as a Senior Advisor at Volvo Cars Engine, Skövde, simultaneously.

AMOS H.C. NG is an Associate Professor at the University of Skövde. He holds a B.Eng. degree and a M.Phil. degree, both in Manufacturing Engineering from the City University of Hong Kong and a Ph.D. degree in Computing Sciences and Engineering from De Montfort University, Leicester, U.K. He is a member of the IET and a Chartered Engineer in the U.K. His research interests include modeling, simulation and multi-objective optimization for production systems design and analysis.

A SIMULATION MODEL USING AMR SENSORS APPLIED TO VEHICLES CLASSIFICATION AND IDENTIFICATION

Ramón V. Cirilo Gimeno, Marta Pla Castells, Antonio García Celda and Javier Martínez Plumé
IRTIC (Robotics and Information and Communication Technologies Institute)
University of Valencia
E 46980, Paterna,
Spain
E-mail: {Ramón.V.Cirilo; Marta.Pla; Javier.Martinez-Plume}@uv.es, Antonio.Garcia@irtic.uv.es

KEYWORDS
Transportation, Signal processors, Validation, Classification, ITS

ABSTRACT

This paper presents a methodology to identify and classify vehicles using Anisotropic Magneto-Resistive (AMR) sensors through the magnetic signature that these vehicles generate when passing over the sensors located on the ground. The work is based on simulating different vehicles ferromagnetic characteristics with a combination of magnets and positions passing over an AMR sensor, and on analyzing the characteristics of the obtained signals. A similarity measure is defined in order to compare different signals and a validation study is done using several reference configurations. The simulations and the posterior analysis show that the methodology works in a controlled scenario and the obtained results predict very positive results when these experiments are placed in a real environment.

INTRODUCTION

Road traffic and transport administrations are every day more interested in the advances that researches on Intelligent Transport Systems (ITS) can offer to the improvement of management tasks. A good traffic management strategy usually implies a reduction in all the well-known negative impacts that affect the people and the environment: safety (fatalities and injured people), environment (pollution, recycling), and sustainability (economy, energetic resources consumption) (Black and Schreffler 2010; Dignum and Pavon 2012; Lam and Head 2012; Kepaptsoglou et al. 2012).

It is widely assumed that road sensing is an important information source for traffic operators, especially when data about vehicles (characteristics, speed, position, category) is gathered, processed and offered in an intelligent and automatic aggregated view to decision makers (Biem et al. 2010; Horvitz and Srinivasan 2011). New ITS technologies can play an important role in this process, both considering the development of new sensing devices and the algorithms and methods to deal with the data that the new sensors can provide. The administrations and government agencies are conscious of this phenomenon and try to promote the research on ITS (European Union 2010), which has given good results for the time being.

There is a wide range of road sensing methods to allow the traffic operator be aware of traffic status in real time. The most widely used, mainly due to the maturity of the technology and the costs of implementation, are the magnetic loop sensors and the traffic cameras (Robert 2010; Bottero et al. 2013). However, the capacity of these sensors to automate surveillance tasks and trigger alerts for incident detections is still far from being perfect. Traffic cameras are used to let the operator observe certain critical areas of the road network and observe incidents in detail when they occur. Certainly, a lot of research has been done in trying to obtain information from cameras to provide the traffic operators with objective data. In particular, license plate recognition (Wen et al. 2012; Lee et al.2013), speed estimation (Becker et al. 2011), vehicle counting (Sanchez et al. 2011; Butch et al. 2011), and automatic incident detection (Chang et al. 2012; Kahaki and Nordin 2011) have attracted the main interest of researchers.

On the other hand, magnetic loops provide information about vehicle counts, classification and speeds, and gets to the Traffic Control Center in form of aggregated values. Loops are currently the most extended mean to obtain this type of information from the roads, but have problems that can hardly been solved because they are inherent to the technology they are based on. As in the case of the cameras, many researchers have made efforts to use the signal that the vehicles generate when passing over the loops (Vanajakshi et al. 2013; Soriguera and Robusté 2011). The major focus of the research tasks have been set on trying to get a more detailed classification of vehicles (Thakoor and Banu 2012), and even identifying each vehicle through the analysis of the characteristics of the signals (Osipov and Malinkin 2013). Some advances have been made on the classification by extending the types of vehicles that can be recognized when processing the signal. However, no major improvements have been obtained when trying to identify vehicles from the signals due to the lack of differentiating characteristics when comparing signals of different vehicles.

The magnetic loops work by detecting the variation of the magnetic inductance of the loop when a vehicle passes over it. They are located under the pavement of the roads, and send the sensed data to external gathering equipments accompanying the sensor.

The idea behind measuring the disturbances of magnetic fields when an object with significant amounts of ferrous metals, like vehicles, moves in the surroundings of a particular point is not new. Two approaches have been considered: a) vehicles emit magnetic fields that sensors can measure, and b) some devices can measure variations on existing magnetic fields. As very few vehicles emit magnetic fields, the second option is the only viable solution to our problem. It has been shown that measuring inductance by magnetic loops doesn't offer a signal with enough quality (Belda 2008), but a new range of sensors have been recently developed and integrated in some kind of industrial form that can measure high and low magnetic fields variations. Additionally, the earth magnetic field is present in all over the world, and the creation and evolution of AMR sensors can be applied in order to detect vehicles presence thanks to measurement of the magnetic field perturbations (Cheung et al. 2006; Taghvaeeyan and Rajamani 2011; Li et al. 2012). The data quality improves the currently offered by the magnetic loops, basically providing a more detailed signal.

The work that is presented here aims at investigating the suitability of AMR sensors to identify vehicles through the magnetic signature that they generate when passing over the sensors located on the ground. Some studies have been made in order to apply this technology in controlled scenarios (Ndoye et al.2011), but signal characteristics and processing methods have not been clearly exposed. This work is based on simulating the ferromagnetic signature of different vehicles by means of a combination of magnets and positions passing over an AMR sensor. A similarity measure is defined in order to be able to compare and identify these signatures from the obtained data, testing several configurations. The main contribution of this paper is, on the one hand, to prove that a software system can be developed and set up to distinguish all vehicle configurations with different levels of accuracy and, on the other hand, to provide a method to identify a blind configuration when its signature is compared with a set of reference signals stored on a library of the different configuration signals in the experiment.

OBJECTIVES AND METHODOLOGY

As stated before, the objective of this work is to obtain an approximation to the algorithms, techniques, processes and software development that are needed to obtain the maximum information from an AMR sensor when used to detect vehicles.

A simulation methodology is designed and proposed with this objective. It consists of a six steps process that leads to the validation of the designed algorithms and procedures applied to magnetic signatures that the AMR sensor captures. This validation uses a combination of well-known and new digital signal processing methods and matching procedures. The steps are:

1. Definition and construction of the object model that generates different configurations of magnetic fields to be measured by the AMR sensor
2. Definition and construction of the simulation platform to make the model object pass over the AMR sensor.
3. Design and construction of the software to extract and process magnetic signatures from the object at each simulated iteration.
4. Design and implementation of the algorithms and software to process the magnetic signatures signals and the similarity measure used to compare signals.
5. Experiment development, organized in phases:
   a. Phase I. Calculation of similarity measures for the same model object configuration in several passes over the sensor.
   b. Phase II. Calculation of similarity measures for different model object configurations in several passes over the sensor.
   c. Phase III. Blind model object identification based on the comparison of its signature against a library of well-known configuration signals.
6. Analysis of experimental results and conclusions.

Object Model Definition and Construction

One of the objectives of using a simulation based approach is to be able to work in small scale conditions. For this purpose, a set of experiments have been conducted in a small laboratory without the need of using real vehicles. The AMR sensor selected for the experiment was the Honeywell HMR2300 (Honeywell HMR2300) It is a three-axis smart digital magnetometer that is able to detect the strength and direction of an incident magnetic field, based on three magneto-resistive sensors in orthogonal directions (X, Y an Z). It works in a range of ±2 gauss with a sampling frequency of up to 154 samples per second. The sensor sensibility is good enough to detect observable variations of the earth magnetic field when a small ferrous object moves over it in a direction at less than 10 cm., and also for a vehicle moving along a road with the sensor set under the road surface in the middle of a lane, for example, with a separation of more than 10 cm.

Different small ferrous metallic objects produce, however, very similar variations of the field, mainly because they aren't complex structures with different distributions of mass, which is the case of real vehicles. So, it was necessary to find an object with the capacity of generating artificial variations of the earth magnetic field. We have used a set of magnets that can be linked together using steel balls at their poles. Using different combinations of magnets and balls, different controlled geometrical configurations can be built, in order to make up objects that generate different magnetic field variations depending on its orientation when moving over the sensor. An important part of the experiment was to build a tetrahedron made of magnets with different colors on the edges so that it could be easy to identify its orientation, and associate it with the signal gathered by the sensor when detecting it.

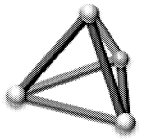


Figure 1: Tetrahedron Made of Magnets

In order to conduct a controlled experiment, a range of possible orientations was defined so that each one could represent a different vehicle. The orientations considered were based on the position and configuration (edge colors) of the base, and from all the possible combinations, a small set were selected for the experiment.

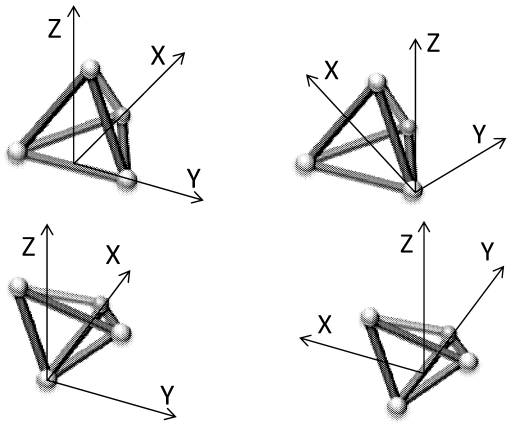


Figure 2: Reduced Tetrahedron Base Orientation Classes

Placing the tetrahedron with a triangular face on the base, for each of the 4 bases, the orientation could follow one of the two first represented in Figure 2, where the X axis indicates the direction of the movement of the object. It was also considered that rotating each figure 120º on the Z axis generates 3 more different orientations per base. This makes a total of 24 orientations. Additionally, inverting the figure on the Z axis (locating the triangular face on the top), another 24 orientations are possible. Then, an edge was selected to be the base of the figure on the direction of movement. With six edges, and a rotation of 180º (represented in Figure 2), there are 12 different orientations. It was also considered a rotation of 90º of the object, so that the base edge stood transversal to the direction of movement (X axis), so generating 12 more orientations. In sum, a total of 48 different orientations are possible, from which only 12 were selected.

The Simulation Platform

In a second stage, a system to move the object over the sensor was designed. The authors decided to use a simple, easy and widely accessible way to get a track where to fix the movement of the object. They recurred to a very popular children cars and tracks toy: the Scalextric. A small oval track was constructed and fixed on a laboratory table, and the AMR sensor was placed and fixed at a particular point under the track and the table, so that the distance between the sensor and the moving object was inferior to 10 cm.

Afterwards, a container was designed to host the object in any of the particular possible orientations defined. The container was attached to the rear part of a regular Scalextric vehicle. Being aware of the necessity of obtaining a wide range of signals, and that the car would generate its own magnetic signature, the container was designed to allow the situation of the object in 4 different places, ranging from close to the vehicle to far from the vehicle. That permitted to have 4 different signals for each of the defined object orientations, which made up a total of 192 positions/signals, from where 48 were used on the experiment (the same 12 orientations on the 4 positions).

Signal Extracting and Processing

The typical signal that is obtained from the AMR sensor when the vehicle and magnetic object is detected is not very complex, in the sense that it has not many peaks and the gradient doesn't provide very high values. Figure 3 shows an example of one of the signals obtained during the experiment.

The signal contains a noise that is clearly distinguishable and is associated to high frequency interference (possibly inherent to the sensor). Its length depends on the velocity of the vehicle passing over the sensor, and the amplitude of the signal varies in function of the configuration of the magnetic object. Taking this into account, the signal needs to be normalized in order to be comparable to other signals obtained from other experiment instances. The method for processing the signal consists in four steps: a) removing the noise by applying a low-pass filter, b) determining the beginning and ending of the signal, c) re-sampling the signal to get a normalized length (1000 samples), and d) moving the signals to the relative zero origin and re-scaling the amplitude of the X, Y and Z signals to be in the [-1, 1] interval.

In order to determine the beginning and the end of the signal, and working with a signal that has been smoothed with the low-pass filter at the first step, an energy threshold method was selected and applied to each individual signal. So, the beginning of the signal was decided to be the first point at which any of the X-Y-Z signals had energy over the threshold, and the end was calculated to be at the point where the three signals, at the same time, had energy below the threshold. The result of the application of the whole process can be seen in Figure 4, where a processed signal is represented (and can be compared to Figure 3).

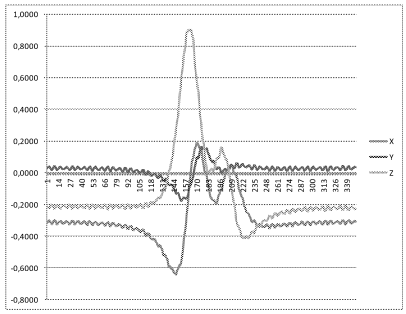


Figure 3: Example of AMR Sensor Signal

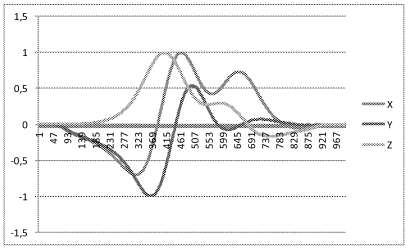


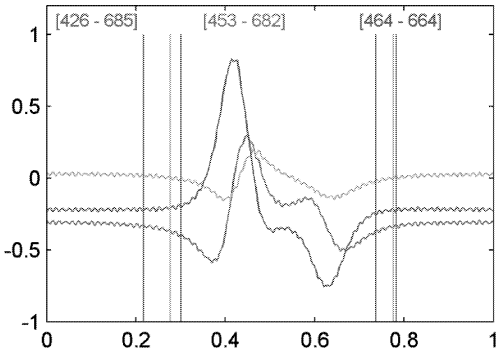
Figure 4: Example of Processed AMR Signal

Similarity Measures and Calculation

The signals obtained from the AMR sensors have three components, one per each axis of detection. The process of defining a similarity measure for signals of distinct iterations has to consider a method to compare signals and a method to merge the individual similarity values of each axis into a scalar value. There exist several methods to obtain a similarity degree between two signals. Given the characteristics of the normalized signals obtained from the previous step, it was decided to use a cosine similarity measure (Nguyen and Bai 2011), because a) the signals are sampled to 1000 units, and b) the resulting similarity value will be normalized in the interval [-1,1]. The cosine similarity method is also easy and rapid to calculate for the conditions of the signals obtained.

In order to merge the similarity values of X-Y-Z signals, two approximations were taken into account: a) the mean of the similarity values for the X-Y-Z signals (called SimM from now on), and b) the product of the similarity values for the X-Y-Z signals (called SimD from now on). The main difference between them is that SimM provides higher similarity values than SimD, and SimD is more sensible to deviations from one only composing signal to its corresponding pair (in the matching process). Finally, SimM was selected as it was considered that variations of one only component shouldn't affect the global similarity measure, providing this way a higher degree of tolerance to re-identification of signals.

At this stage, an initial similarity calculation was performed among various signals from the same vehicle and magnetic object configuration, which should have provided values close to 1. In some cases, the similarity values go far from what it was expected. An analysis of the conflictive signals showed that the cause of this to happen was that the process of establishing the beginning and end of the signal had a dependence on the vehicle speed, and this error had a considerable influence in the similarity calculation. This can be shown in Figure 5.



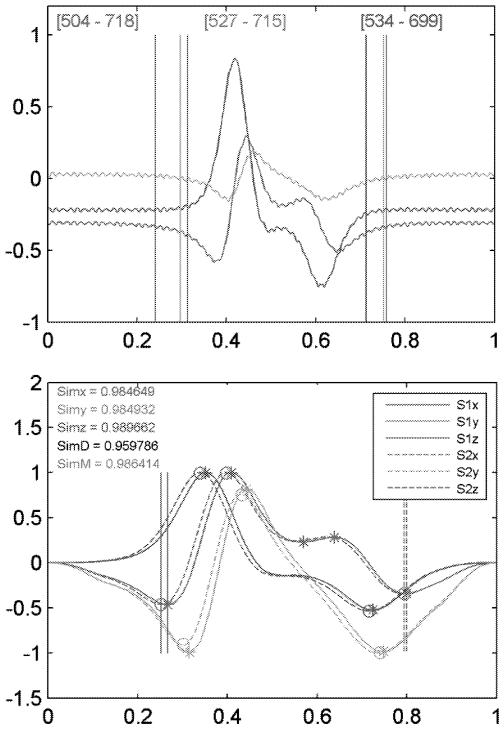


Figure 5: Compared Signals of the Same Configuration at Different Iterations with Different Delimitations.

This additional problem was solved by designing a mechanism to resample one of the X-Y-Z signals in order to best match the other one. A peak alignment procedure was designed, so that one of the X-Y-Z signals was applied a linear transform to match the selected peaks of both iterations. The selection of peaks was done by choosing the absolute maximum and minimum of each component signal, and selecting the one that had a higher distance between the peaks. This was done for both signals, but for the same component. Then, a two set approach to determine the scale factor was applied. First, it was fixed the right side of the signal and a re-sample to the left was applied (1), using x_1 as the scale factor and being K the number of samples of the signal (1000 in this case). In second place, it was fixed the left side of the signal and a re-sample to the end was applied (2) using x_2 as the scale factor. The combination of both transforms was the final transform to apply to each component of one of the original signals.

$$y_1 = K\frac{y-x_1}{K-x_1} \qquad (1)$$

$$y_2 = K\frac{y_1}{K-x_2} = \frac{K^2(y-x_1)}{(K-x_1)(K-x_2)} \qquad (2)$$

Now, x_1 and x_2 must be calculated for each particular case, and so the peaks information can be used as boundary conditions to get a system of two equations with two unknown variables. M1 and Z1 are the peaks of the reference signal 1, and M2 and Z2 are the peaks of the reference signal 2. The objective is to match M1 to M2 and Z1 to Z2. So, applying the transform (2), and using M1, M2, Z1 and Z2 as boundary conditions, the solution to the equation system is:

$$x_1 = \frac{M1Z2 - Z1M2}{M1 - Z2} \quad (3)$$

$$x_2 = K - \frac{K^2(M2 - Z2)}{M1(K - Z2) + Z1(M2 - K)} \quad (4)$$

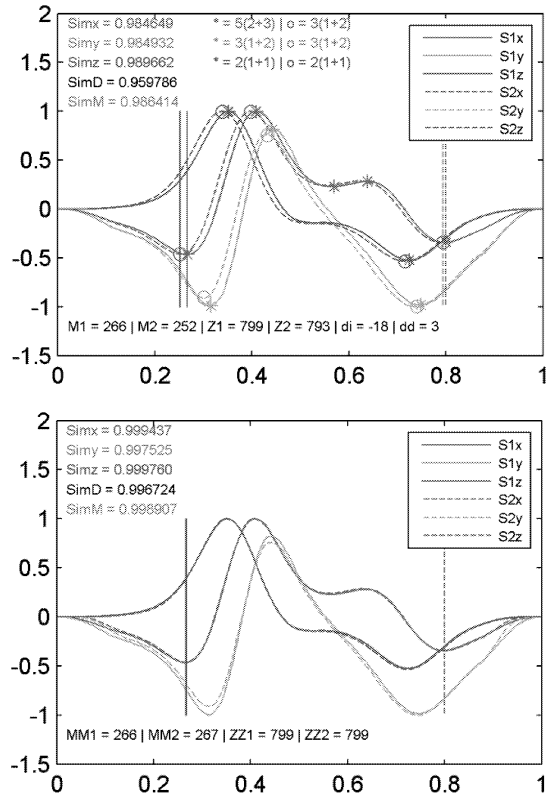


Figure 6: Compared Signals of the Same Configuration at Different Iterations with Different Delimitations

When the peak-alignment method was applied prior to the calculation of the similarity, the signals belonging to the same vehicle and object configuration produced values close to 1 as expected (Figure 6). Notwithstanding, an objective prove of the improvement was obtained by comparing the similarity values of the 48 iterations between each other, with and without applying the peak alignment method. The comparison of the dispersion of the SimM and SimD values for both approaches is shown in Figure 7 and Figure 8.

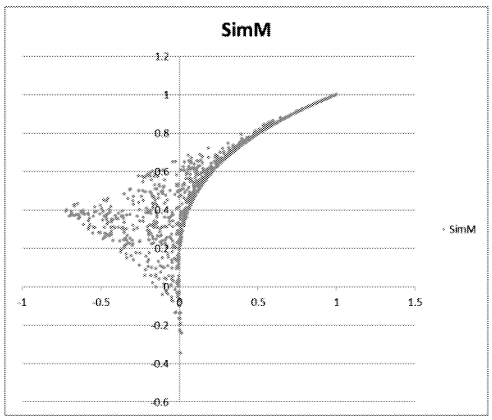


Figure 7: Scattering Graphs for SimM and SimD without Peak Alignment

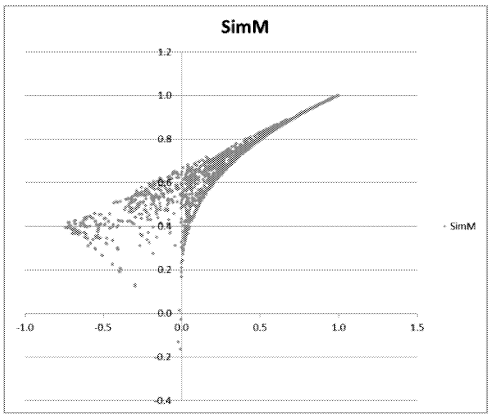


Figure 8: Scattering Graphs for SimM and SimD with Peak Alignment

When comparing both figures, it is obvious that, when using the peak alignment technique within the matching algorithm, the scattering is inferior as to the opposite case. This can be objectively asserted by comparing the average of similarity values for both SimM and SimD measures in both cases (See Table 1).

Table 1: Similarity measures average for SimM and SimD

	W/O P-Align.	W P-Align.
SimD avg	0.1637	0.2894
SimM avg	0.5409	0.6945

Therefore, peak alignment provides a higher rate of success when comparing two signals because the similarity values are also higher. As a consequence, the threshold that is used to make the decision on whether or not two signals can be considered to belong to the same object should also be increased. This is not a real problem, because analyzing the similarity values in a cross table that compares all signals to each other permits to find the frontier between same and different objects.

Experiment Layout

An experiment was designed for obtaining a set of signals corresponding to several vehicle and object configurations. As stated before, 48 different configurations were selected, and a total of 5 signals per configuration were obtained and stored in a preliminary library. All of them were obtained by passing the vehicle and the attached container with the magnetic object through the track and over the AMR sensor.

All obtained signals were processed following the procedures above explained, and a final library of 48 signals was constructed by obtaining the average of the 5 signals corresponding to each configuration. Both the preliminary and the final library were used to conduct the experiment in three phases:

1. Phase I. Calculation of similarity measures for the same model object configuration in several passes over the sensor.

2. Phase II. Calculation of similarity measures for different model object configurations in several passes over the sensor.
3. Phase III. Blind model object identification based on the comparison of its signature against a library of well-known configuration signals.

The phase I aimed at obtaining the range of similarity values for signals of the same object. This procedure allowed determining the minimum value (threshold) to be used when comparing two arbitrary signals and deciding whether or not the signals could belong to the same vehicle and object configuration (when the similarity values was over the threshold).

The phase II aimed at checking the threshold obtained in phase I with the complete set of signals by matching all of them with each other. The similarity values obtained in the 48x48 table defined a set of configurations that were equivalent, as the values were over the threshold. A set of classes was then derived from this phase of the experiment.

The phase III aimed at identifying a blind configuration by checking it against the library of 48 signals constructed with the selected configurations. 48 blind configurations equivalent to any of those stored in the library were selected for the experiment, and the similarity value of each blind configuration was ranked in reference to the library signals. The outcome of this phase allowed confirming the initial hypothesis that a configuration could be identified as belonging to one of the known configuration classes.

EXPERIMENTAL RESULTS

Phase I

The results of the Phase I of the experiment are a set of similarity measures that have been represented in a histogram based on percentages that can be seen in Figure 9.

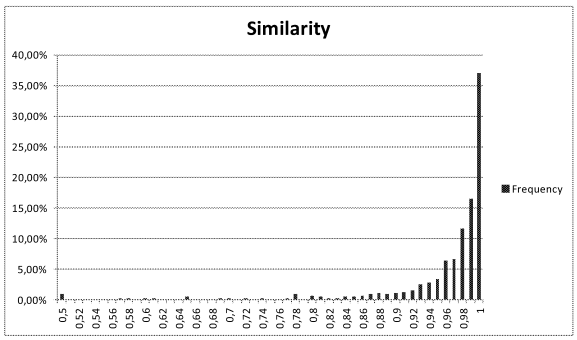


Figure 9: Histogram of Similarity for Identical Configurations

The percentage of correct matches in function of the similarity measures threshold can be seen in Figure 10.

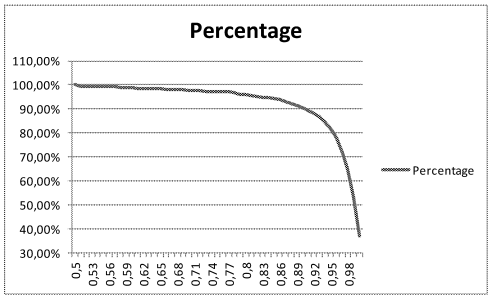


Figure 10: Aggregated Percentage of Matching Exit in Function of the Similarity Threshold.

Therefore, there is a trade-off between the percentage of correct matches and the similarity threshold to be used. The authors have considered that a 90% of exit rate in the identification of the signal is a very good one, which permits to establish a similarity threshold value for SimD in 0,90.

Phase II

The results of Phase II show a higher dispersion in the similarity values of SimD, as expected. However, values of similarity very close to 1 have been found, indicating that some configurations could belong to the same class. The distribution of similarity measures in a histogram are shown in Figure 11.

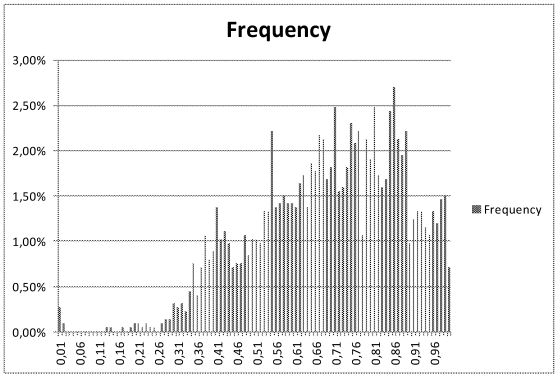


Figure 11: Histogram of Similarity for Different Configurations

Figure 11 shows an increasing trend in the frequency for similarity values as they also increase. However, there is a point where the trend changes to lower values of frequency; the value of the similarity for this decreasing point is SimD = 0,90, which is in accordance to the outcome of the Phase I

Phase III

The test with the 48 blind configurations compared to the library of identified configurations showed that 46 of them were correctly matched in the first position. That represents a 95.83% of success with an average of 99.82% of similarity.

The 2 cases that were not correctly matched in the first place, the second matching option was the correct one, with an average of 98.89% of similarity.

It's been also observed that the similarity obtained for different configuration matches was over 90% in average for the first 6 matching candidates. It is probably a consequence of the similarity of the magnetic fields that generate the 48 configurations that, as shown before, could be grouped into classes.

CONCLUSIONS

In this paper we have developed a methodology that can be applied to the use of AMR sensors to simulate the detection and identification of vehicles circulating on a road. The obtained results are promising, as the success rate of identification is very high. The signal processing methods and similarity measures proposed allow a simple but powerful application to a real case, based on the simulation experiment results.

The application of the methodology to a real case study should be considered as a next step, because some parameters that have been used in this experiment apply only to its particular conditions. This is also an initial work with the objective of obtaining a new type of sensor/software that can be used to improve the quality of the data that traffic operators use when monitoring roads and traffic. The next step will be the application of the methodology in a real environment.

FUTURE WORK

The application of the methodology has been developed under laboratory conditions, and has been proved to be effective with experimental signals that do not correspond to real vehicle signals generated by AMR sensors. The simulation aims at setting the principles to apply algorithms and matching mechanisms for signal processing in real cases, with real vehicles and in real traffic conditions.

With that objective in mind, the next steps will be a) to set a test site for gathering and analyzing real vehicle signals and check the applicability of the model b) setting the algorithms parameters and matching procedures to the real case study, and c) set up an experiment or field experiment where classification and re-identification of vehicles could be tested based on the hypothesis presented in this work.

The sensor behavior is expected to be variable (to provide different signals) depending on the relative position of the sensor to the vehicle during the detection phase, so it is possible that a combination of sensors is necessary to get good re-identification rates by using a combination of signals. In that case, new algorithms should be developed to apply the matching procedure and the similarity measure should be revised.

REFERENCES

Becker, Florian, et al. 2011. "Variational recursive joint estimation of dense scene structure and camera motion from monocular high speed traffic sequences", in Computer Vision (ICCV), 2011 IEEE International Conference on. 1692-1699.

Belda, Enrique. 2008. "Data Quality for Traffic Management. The Spanish Approach". Data Quality Workshop. Vienna.

Biem, Alain, et al. 2010. "Real-time traffic information management using stream computing". IEEE Data Engineering Bulletin, vol. 33, no 2, p. 64-68.

Black, C. S.; Schreffler, E. N. 2010. "Understanding Transport Demand Management and Its Role in Delivery of Sustainable Urban Transport". Transportation Research Record: Journal of the Transportation Research Board. Volume 2163 / 2010. 81-88.

Bottero, M.; Dalla Chiara, B.; Deflorio, F. P. 2013. "Wireless sensor networks for traffic monitoring in a logistic centre". Transportation Research Part C: Emerging Technologies, vol. 26, 99-124.

Buch, Norbert; Velastin, Sergio A.; Orwell, James. 2011. "A review of computer vision techniques for the analysis of urban traffic". Intelligent Transportation Systems, IEEE Transactions on, vol. 12, no 3, 920-939.

Chang, Yuntao; Su, Zhongming; Qian-Yu, Liang. 2012. "A New Traffic Incident Detection Method Under Low-Volume Condition Based On Automatic Vehicle Identification", in Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on. IEEE, 2012. 2853-2859.

Cheung, S.Y.; Coleri, S.; et al. 2006. "Traffic Measurement and Vehicle Classification with Single Magnetic Sensor". Transportation Research Record: Journal of the Transportation Research Board, vol.1917/2005, 173-181.

Dignum, F; Pavón, J;. 2012. "Special issue: Multi-Agent Systems applications in transport". Progress in Artificial Intelligence, Volume 1, Issue 2 , 137-138.

European Union. 2010. "DIRECTIVE 2010/40/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 7 July 2010 on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport". Official Journal of the European Union.

Horvitz, E. J.; Srinivasan, S. 2011. "Predictive models of road reliability for traffic sensor configuration and routing". U.S. Patent No 7,948,400.

Kahaki, S. M. M.; Nordin, M. J. 2011. "Vision-based automatic incident detection system using image sequences for intersections", in Pattern Analysis and Intelligent Robotics (ICPAIR), 2011 International Conference on. IEEE, 2011. 3-7.

Kepaptsoglou, Konstantinos, et al. 2012. "Quality Management in Mobility Management: A Scheme for Supporting Sustainable Transportation in Cities". International Journal of Sustainable Transportation, vol. 6, no 4, 238-256.

Lam, Debra; Head, Peter. 2012. "Sustainable Urban Mobility", in book Energy, Transport, & the Environment: Addressing the Sustainable Mobility Paradigm. 359-372. ISBN: 978-1-4471-2716-1.

Lee, Rong Choi; Hung, King Chu; Wang, Huan Sheng. 2013. "Real-Time Vehicle License Plate Recognition Based on Scanning and 2D Haar Discrete Wavelet Transform". Applied Mechanics and Materials, vol. 284, 2402-2406.

Li, X.S. et al. 2012. "High Accuracy AMR Magnetometer and its Application to Vehicle Detection". Advanced Materials Research, vol150, 443-444.

Ndoye, M; Totten, V.F.; Krogmeier, J.V. 2011. "Sensing and Signal Processing for Vehicle Reidentification and Travel Time Estimation". IEEE Transactions on Intelligent Transport Systems. Vol 12. No 1. 119-131.

Nguyen, H. V.; Bai, L. 2011. "Cosine Similarity Metric Learning for Face Verification". Lecture Notes in Computer Science. Volume 6493, 709-720.

Osipov, S. K.; Malinkin, A. Y. 2013. "Method and Device for Determining the Speed of Travel and Coordinates of Vehicles and Subsequently Identifying Same and Automatically Recording Road Traffic Offences". U.S. Patent No 20,130,038,681.

Robert, K. 2010. "Bringing Richer Information with Reliability to Automated Traffic Monitoring from the Fusion of Multiple Cameras, Inductive Loops and Road Maps", in Advanced Video and Signal Based Surveillance (AVSS), Seventh IEEE International Conference on. IEEE, 2010, 9-16.

Sanchez, Angel, et al. 2011. "Video-Based Distance Traffic Analysis: Application to Vehicle Tracking and Counting". Computing in Science & Engineering, vol. 13, no 3, 38-45.

Soriguera, F.; Robusté, F. 2011. "Estimation of traffic stream space mean speed from time aggregations of double loop detector data". Transportation Research Part C: Emerging Technologies, vol. 19, no 1, 115-129.

Taghvaeeyan, S.; Rajamani, R.; 2011. "Use of vehicle magnetic signatures for position estimation". Applied Physics Letters , vol.99, no.13, 134101-134101-3.

Thakoor, Ninad; Bhanu, Bir. 2012. "Structural signatures for passenger vehicle classification in video", in Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012. 926-929.

Vanajakshi, Lelitha, et al. 2013. "Development of a Vehicle Counting and Classification System for Heterogeneous and Lane less Traffic". International Conference on Intelligent Infrastructure.

Wen, Ying, et al. 2011. "An algorithm for license plate recognition applied to intelligent transportation system". Intelligent Transportation Systems, IEEE Transactions on, vol. 12, no 3, 830-845.

WEB REFERENCES

Honeywell HMR2300
http://www51.honeywell.com/aero/common/documents/myaerospacecatalog-documents/Missiles-Munitions/HMR2300.pdf. Last visited 01-03-2013.

AUTHOR BIOGRAPHIES

RAMÓN VICENTE CIRILO GIMENO received the M.Sc. in Physics from Universidad de Valencia, Valencia, Spain, in 1993. He is currently an Associate Professor at the School of Engineering of Universidad de Valencia (2002), and member of the LISITT research group at the Institute of Robotics and Information & Communication Technologies of the UVEG (IRTIC) since 1992. His main research interests are focused in the fields of Intelligent Transport Systems (ITS), Information Systems and Databases. He has been member of the e-Safety Observers group (European consultant board on e-Safety matters) from 2005 to 2010.

DR. MARTA PLA-CASTELLS received her BsC in mathematics in 2000, her MsC in Computer Science in 2004, and her PhD by Univ. de Valencia in 2009. He has worked in physically-based simulation and its application to computer graphics for training and serious games. Currently she works in Intelligent Transport Systems and the application of IT to traffic management.

ANTONIO GARCIA CELDA is a computer and telematics engineer by the University of Valencia in the years 2010 and 2012, respectively, and a member of the research group LISITT (Integrated laboratory of Intelligent Systems and Traffic Information Technologies) since the end of 2008. Currently, he has specialized in the field of telematics systems applied to traffic and transport, and he is finishing his MSc in "Systems and Services in the Information Society" too.

JAVIER MARTINEZ PLUMÉ is Assistant professor at the University of Valencia in Computing Science. He has been working in ITS applications since 1991, participating in several EU projects as INVAID and VERA supported by the EU Commission. In the last 5 years he has worked as ITS consultant in Bizkaia in the definition and deployment of the telematics implementations project in the metropolitan area of Bilbao. Currently he is working in the application of wireless sensors to traffic management.

SUPPLY CHAIN SIMULATION

SECTIONED WALKING WORKER LINES WITH LOOP BALANCING

Leif Pehrsson[1, 2]
Amos H.C. Ng[2]
Jacob Bernedixen[2]
David J. Stockton[3]
[1]Volvo Car Corporation, SE-405 31, Gothenburg, Sweden
[2]Virtual Systems Research Centre, The University of Skövde, SE-541 28, Skövde, Sweden
[3] Faculty of Technology, De Montfort University, The Gateway, Leicester, LE1 9BH, UK
E-mails: leif.pehrsson@his.se, amos.ng@his.se

KEYWORDS

Walking worker, assembly, simulation, optimisation.

ABSTRACT

The competition in the automotive industry moves forward and alternative production concepts emerge. Within assembly systems development the walking worker concept has been noticed as an interesting approach to level capacity to customer demand. One of the main disadvantages has been that each worker has to learn a large work content covering all stations in the line. This paper investigates an approach with a sectioned line divided into worker loops (streets) in order to overcome some of this disadvantage while maintaining the effectiveness and ability to align capacity to customer demand.

INTRODUCTION

The automotive industry is constantly striving for improved production performance in a very competitive business situation. In line with the last years significant and frequent fluctuations in customer demand alternative production concepts are emerging. Within assembly operations the traditional setup with one or more workers dedicated at each station, illustrated in Figure 1, is challenged by other concepts such as walking workers following the assembly object through the line (Wang et al., 2005) as shown in Figure 2. A setup alternative to having the workers following the assembly object is to use fixed positions without moving the work items (Huang et al., 2007), e.g. for fragile, bulky or heavy products.

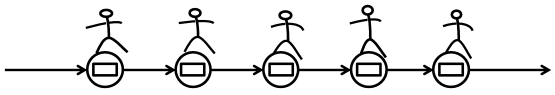


Figure 1. Traditional assembly line.

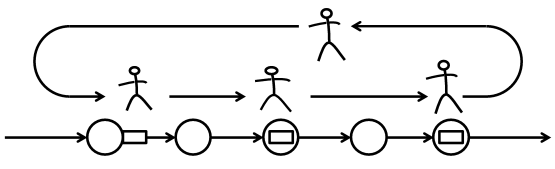


Figure 2. Walking-worker assembly line.

Another variant is to have multiple multi-function workers serving parts of a u-shaped line (Nakade and Ohno, 2003). In this paper the focus will be on workers moving along with the assembly object. The walking worker concept has many advantages including cross trained workers, less WIP with fewer buffers, improved worker accountability and robustness against station time variations as shown by several studies (Wang et al, 2009, Wang et al., 2005, Mileham et al., 2000), but also some disadvantages such as pronounced worker training requirements and higher potential for blocking due to lower performing workers. The major advantage with the walking worker line concept is probably that it enables the opportunity to align the assembly capacity to the customer demand (Wang et al., 2005) by varying the number of workers on the line without the need of rebalancing and physical changes in the setup of the line and the material facade. The capacity can be adjusted from just having one operator on the line up until the number of operators matching or surpassing the total number of stations. The concept is considered as a concept within reconfigurable assembly systems (Bi et al., 2007). It has been shown (Wang et al., 2005) that walking worker lines are quite robust against the effect of disturbances if the number of workers is kept somewhat lower than the number of stations enabling higher throughput with the same number of workers compared to traditional setups. However, as the number of workers is approaching the requirements for maximum capacity the negative effect from variation sources become more evident in limiting throughput. It has been shown that simulation modelling and combination of optimisation techniques can be an effective tool for designing walking worker assembly lines (Alzuheri et al., 2010). The expected capacity from a specific number of workers can be studied by using simulation (Wang et al., 2007) and repeated simulations for each possible number of workers can provide a complete mapping of all the available capacity steps. Balancing the system with respect to the required amount of training and the ability to align capacity towards customer demand is a serious challenge for enterprises when designing, dimensioning and implementing walking worker lines. During the past years Simulation-based multi-

objective optimisation (SMO) has emerged as a very useful tool for decision-support within production systems design and improvement (Ng et al., 2012, Pehrsson et al., 2011 and Pehrsson et al., 2011). The opportunity to use this method to solve a task concerning the optimised manning of a sectioned walking worker line is explored in this paper.

CASE STUDY

During the improvement of a mixed-model (Sarker and Pan, 1998) walking worker assembly line within an automotive manufacturer the project team was challenged to reduce the worker training requirements while keeping the opportunities to align capacity towards customer demand by varying the number of workers on the line. Behind this challenge was an expectation from management to improve quality and to increase the capacity as a result of less worker related abnormalities, errors and station time deviations. The idea to segment the line into two or more walking worker areas then raised. The segments would enable a majority of workers to be trained only to perform a limited number of operations. The flexibility to follow the customer demand and align capacity accordingly would still be covered even though some limitations might occur. Such limitations could be compensated for considering some of the workers still could be trained to master all operations on the line. Since no major hardware changes were required to try this method a test was executed rapidly after the idea emerged. Splitting the line into areas was done by dividing it into to halves with a walking worker loop area for each half as seen in the before and after Figure 3.

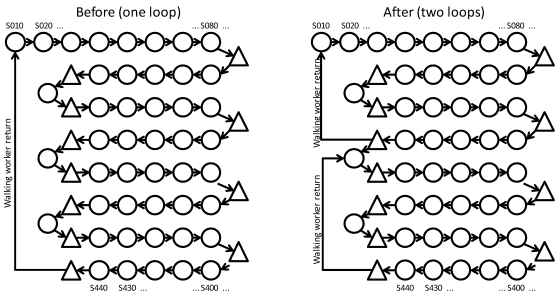


Figure 3. Before and after sectioning the line into two loops

The workers were divided equally into two teams belonging to the new areas and production started. The throughput from the line was not increased, but was actually declining. Finding out more about the causes of this system behaviour was considered a problem suitable to analyse with the help of SMO and a case study was initiated. The management wanted fast response from the improvement team and the SMO analysis and a discrete event simulation model of the line with reduced complexity and aggregated data was crated with FACTS Analyser (Ng et. al., 2007) in order to be able to create valuable decision support within an acceptable timeframe. The simplifications made were to consider all workers on the line to perform with the same pace and with the same type of variation pattern at all stations. By doing this the line could be modelled with standard operations, requiring only cycle time, availability and mean downtime data together with a pallet system carrying the products. The pallets could then mimic the worker moving with the object since an assembly object always was released into the line together with a worker. Variation in pace for a worker or between workers (Sakamoto, 2010) in combination with the workstation effectiveness can be integrated in the simulation model though station cycle times with variation. The cycle time variations in this case study were aggregated and modelled using a triangular distribution with min mode max parameters for each assembly station as shown in Table 1.

Availability was set to 99% and mean down time to 30 seconds for all stations based on data from the company. When examining the data in table 1 it can also be observed that there was some imbalance between station cycle times. The case study was also limited to include the most common variant on the line, produced 60 to 70 % of the time.

The before and after scenarios were both simulated with a total of 26 workers and five replications and the results were strikingly equal to the real line behaviour with 34.2 jobs per hour in throughput for the original line and 31.9 for the segmented line to be compared with the real-life result 33 vs. 31 jobs per hour stated by the production manager.

However, there was still no answer to the question why the throughput and the capacity decreased when segmenting the line by introducing two worker loops. Theories about blocking and starving between the two segments were discussed but could not be verified through the simulation models. An SMO study was prepared with the numbers of workers per loop defined as two input parameters connected to the simulation model. They were allowed to vary within the range [1, 22]. The widely-used evolutionary multi-objective optimisation algorithm, NSGA-II (Deb, 2001), was used for running 500 iterations, each with 5 replications.

Table 1. Cycle times for all stations with min, mode and max related to the triangular distribution used in the case study

STN	min	mode	max
S010	60,3	67	73,7
S020	65,88	73,2	80,52
S030	55,98	62,2	68,42
S040	55,26	61,4	67,54
S050	62,46	69,4	76,34
S060	0	0	0
S070	0	0	0
S080	0	0	0
S090	0	0	0
S100	71,37	79,3	87,23
S110	56,79	63,1	69,41
S120	60,21	66,9	73,59
S130	61,02	67,8	74,58
S140	48,15	53,5	58,85
S150	67,05	74,5	81,95
S160	62,37	69,3	76,23
S170	67,59	75,1	82,61
S180	45,81	50,9	55,99
S190	48,69	54,1	59,51
S200	66,15	73,5	80,85
S210	37,35	41,5	45,65
S220	43,74	48,6	53,46
S230	21,06	23,4	25,74
S240	59,31	65,9	72,49
S250	36,18	40,2	44,22
S260	31,95	35,5	39,05
S270	52,02	57,8	63,58
S280	32,31	35,9	39,49
S290	0	0	0
S300	37,89	42,1	46,31
S310	16,74	18,6	20,46
S320	59,13	65,7	72,27
S330	71,19	79,1	87,01
S340	63,09	70,1	77,11
S350	53,55	59,5	65,45
S360	43,2	48	52,8
S370	32,58	36,2	39,82
S380	36,9	41	45,1
S390	41,4	46	50,6
S400	60,66	67,4	74,14
S410	47,61	52,9	58,19
S420	40,68	45,2	49,72
S430	29,79	33,1	36,41
S440	47,34	52,6	57,86
Sum (s)	1950,75	2167,5	2384,25
(minutes)	32,5125	36,125	39,7375

When applying SMO with subject to minimise the number of workers per loop while maximising throughput some interesting knowledge about the system could be discovered as shown in Figure 4. Solutions with equal throughput performance for the one loop system and the two loop system with the same total number of workers could be found. There were actually some solutions surpassing the original solution in respect of throughput.

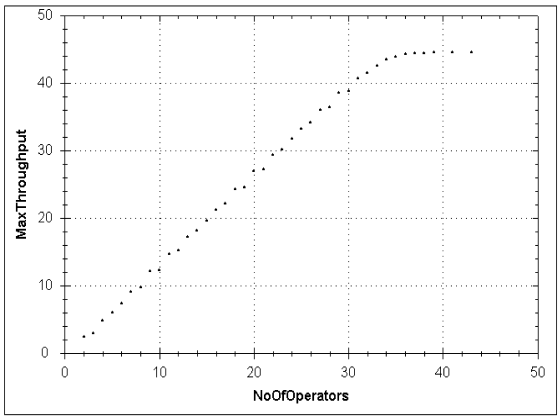


Figure 4. Results from the SMO-study with optimised number of worker per loop.

Hence not only the stations must be balanced when moving from a one loop system to a two loop system, the loops must also be balanced e.g. by releasing the right amount of workers into each loop. When splitting the line into two segments the sum of station times for each segment was not considered and due to imbalance in station times between the two segments the throughput was throttled due to a larger amount of work put on the workers in one of the loops causing idle time in the other loop. One way to address this problem is to release the optimum number of workers per section; another way might be to level out the station times between the loops through re-balancing the work content. In his case the loop imbalance could be used as an advantage when introducing to segments as some steps in the number of workers actually will result in higher throughput. The required throughput in this case could be achieved with fewer workers after dividing the line into two walking worker loops. A recipe for correct manning (Table 2) could be created in order to smoothly be able to align capacity towards customer demand.

Two benchmark simulations with completely even balanced station times were run in order to quantify the amount of losses caused by deviations in station times and in order to compare the existing situation with the performance of the system under ideal or close to ideal conditions.

The results from the benchmark simulations are shown together with the simulations of the original setup, the sectioned line setup and the sectioned setup with balanced number of workers in the two loops are illustrated in Figure 5.

Table 2. Recipe for aligned manning based on the SMO study.

No of operators	Loop 1	Loop 2	Through-put
2	1	1	2,48
3	2	1	3,06
4	2	2	4,96
5	3	2	6,11
6	3	3	7,43
7	4	3	9,17
8	4	4	9,90
9	5	4	12,22
10	5	5	12,37
11	6	5	14,83
12	7	5	15,26
13	7	6	17,30
14	8	6	18,31
15	8	7	19,75
16	9	7	21,32
17	9	8	22,19
18	10	8	24,33
19	10	9	24,64
20	11	9	27,05
21	12	9	27,35
22	12	10	29,46
23	13	10	30,30
24	13	11	31,85
25	14	11	33,27
26	14	12	34,22
27	15	12	36,11
28	15	13	36,52
29	16	13	38,67
30	17	13	38,91
31	17	14	40,80
32	18	14	41,54
33	18	15	42,69
34	19	15	43,63
35	19	16	43,95
36	20	16	44,40

A comparison between the scenarios show that the effectiveness of the system for a specific number of workers can be kept at considerably high levels with un-even balancing as long as the line is operated at throughput levels below the total limitation of the balance. Close to the maximum throughput the system utilisation drops. These findings correspond with findings in other studies (Wang et al., 2009 and Wang et al. 2005).

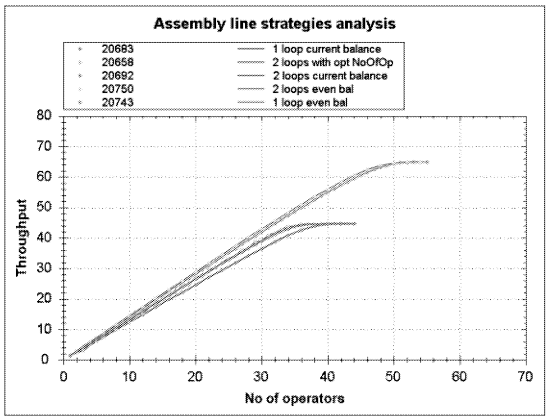


Figure 5. Results from simulations and optimisation of the number of operators per loop.

Balancing the number of operators between the loops enables the performance of the two loop system to be on par with the one loop system. It also seems like the correctly balanced two loop setup might perform somewhat better than the one loop setup. The differences are very small indeed and it is probably other factors such as worker training, efforts and skill that will make the real difference. The next step in exploring the line configuration opportunities will be to run case studies with fewer stations in smaller loops. Future studies will have to show whether the sectioning of the line and the reduced training requirements will enable assembly operations with less disturbances due to improved worker performance.

CONCLUSIONS

Walking worker lines have the ability to maintain effectiveness at reduced throughput levels with the opportunity to align the number of workers and the capacity to customer demand. Some of the disadvantages with extensive training requirements can be offset by sectioning the workers into working in loops covering parts of the line while maintaining or slightly improving the overall performance of the system. The robustness towards disturbances in a walking-worker assembly system seems to be very useful when operating lines with un-even balancing and fluctuating customer demand.

REFERENCES

Alzuheri, A., Luong, L. and Xing, K., 2010. An Integrated Design Support Methodology for Walking Worker Assembly Lines in *Proceedings of the InternationalMultiConference of Engineers and Computer Scientists,* Vol. III, IMECS 2012, March 17-19, 2010, Hong Kong.

Bi, Z.M., Wang, L. and Lang, S.Y.T. 2007. 'Current status of reconfigurable assembly systems', *Int. J. Manufacturing Research*, Vol. 2, No. 3, pp.303–328.

Deb, K., (2001). *Multi-objective optimization using evolutionary algorithms,* 3'rd ed., Wiley, Wiltshire, UK.

Huang, G.Q., Zhang, Y.F., Jiang, P.Y., 2007. RFID-based wireless manufacturing for walking-worker assembly islands with fixed-position layouts in *Robotics and Computer-integrated Manufacturing* 23, 469-477.

Mileham A.R., Jeffries, A.W., Owen, G.W., Pellegrin, F., 2000. The design and performance of walking fitter assembly lines in *Proceedings of the 16th International Conference on CAD/CAM, Robotics & Factories of the Future* (CARS & FOF 2000), Port of Spain, Trinidad and Tobago, West Indies, June 26-28.

Nakade, K., and Ohno, K., 2003. "Separate and carousel type allocations of workers in a U-shaped production line, *European Journal of Operational Research*, vol. 145, no. 2, pp. 403-424.

Ng. A.H.C., Dudas, C., Pehrsson, L. and Deb, K. (2012). Knowledge Discovery in Production Simulation by Interleaving Multi-Objective Optimization and Data Mining. In *Proceedings of the 5'th Swedish Production Symposium*, Linköping, Sweden, 6-8 November.

Ng, A.H.C., Urenda, M., Svensson J., Skoogh, A. And Johansson, B., 2007. FACTS Analyser: An innovative tool for factory conceptual design using simulation, In *Proceedings of Swedish Production Symposium*, Gothenburg 28-30 August.

Pehrsson, L., Ng, A.H.C. and Bernedixen, J., 2011. Multi-objective production system optimisation with investment and running costs. *In Evolutionary Multi-objective Optimization in Product Design and Manufacturing*, L. Wang, A. Ng, K. Deb (eds), Springer.

Pehrsson, L. and Ng, A.H.C., 2011. Applied framework for simulation-based multi-objective optimisation within production system development. In *Proceedings of the Industrial Simulation Conference*, Venice, Italy.

Sakamoto, S., 2010. Beyond *World-Class Productivity: Industrial Engineering Practice and Theory*. Springer, ISBN 978-1-84996-269-8.

Sarker, B.R, Pan, H., 1998. Designing a mixed-model assembly line to minimize the costs of idle and utility times in *Computers Industrial Engineering*. Vol 34, No. 3, pp 609-628.

Wang Q., Lassalle S., Mileham A.R., and Owen G.W., 2009. "Analysis of a linear walking worker line using a combination of computer simulation and mathematical modeling approaches," *Journal of Manufacturing Systems*, vol. 28, no 2-3, pp. 64-70, 2009.

Wang Q., Owen G. W. and Mileham A. R., 2005. "Comparison between fixed- and walking-worker assembly lines," *Proceedings of the Institution of Mechanical Engineers*, Part B: Journal of Engineering Manufacture, vol. 219, no 11, pp. 845-848, 2005.

Wang Q., Owen G., and Mileham A., "Determining numbers of workstations and operators for a linear walking-worker assembly line, *International Journal of Computer Integrated Manufacturing*, vol. 20, no. 1, pp. 1-10, 2007.

AUTHOR BIOGRAPHIES

LEIF PEHRSSON holds a M.Sc. degree in Automation Engineering and a B.Sc. degree in Mechanical Engineering, both from the University of Skövde. He has been working for more than 15 years in various positions, including production engineering manager, within automotive industry. In year 2000, he received the Henry Ford Technology Award for his work on production system development. Started from 2008, he is employed as a Volvo Industrial PhD Program (VIPP) candidate and as a Senior Advisor at Volvo Cars Engine, Skövde, simultaneously.

AMOS H.C. NG is an Associate Professor at the University of Skövde, Sweden. He holds a B.Eng. degree and a M.Phil. degree, both in Manufacturing Engineering from the City University of Hong Kong and a Ph.D. degree in Computing Sciences and Engineering from De Montfort University, Leicester, U.K. He is a member of the IET and a Chartered Engineer in the U.K. His research interests include modelling, simulation and multi-objective optimisation for production systems design and analysis.

JACOB BERNEDIXEN has a M.Sc. degree in Industrial Engineering and Management from the Institute of Technology at Linköping University, Sweden. Since his graduation in 2007, he has been working in industrial-based SBO research projects at the University of Skövde and is now an Industrial PhD student supported by Volvo Car Corporation.

DAVID J. STOCKTON is Director of the Centre for Manufacturing at the De Montfort University, UK, and leads the Lean Engineering Research Group within this centre. He has amassed over 20 years research, industrial and project management experience in the fields of cost modelling and the optimisation of manufacturing systems design. Professor Stockton has published 20 journal papers and over 50 conference papers and is a member of the Editorial Advisory Board of the Aircraft Engineering and Aerospace Technology journal and the Manufacturing Committee of the Association of Cost Engineers.

GENERIC SIMULATION MODEL FOR ASSEMBLY LINE SUPPLY

Veronique Limère
Kurt De Cock
El-Houssaine Aghezzaf
Department of Industrial Management
Ghent University
B 9052, Zwijnaarde,
Belgium
E-mail: Veronique.limere@ugent.be

KEYWORDS
Discrete Event Simulation, Materials Supply, Facility Logistics, Bulk Feeding, Kitting, Automatic Model Generation.

ABSTRACT

In this paper a simulation model for materials supply to mixed-model assembly lines is discussed. The model is created in FlexSim and demonstrates the in-plant logistical flows involved to supply parts from the warehouse to the use-points at the assembly line. The materials supply methods shown are bulk feeding (also referred to as line stocking) and kitting. The results of the simulation model are compared to the results of a deterministic mathematical cost model to study the impact of real production variations and dynamics. Testing is done based on a case study where 438 parts are supplied to 23 work stations o of which 146 are supplied in bulk, while the remaining parts are kitted. To create this specific model an automatic model generator is used. The findings of this study are reported.

INTRODUCTION

To have efficient assembly processes, the right parts need to be provided close to the use-points at the line at the right time. Different ways exist to supply parts to an assembly line. It can go from bulk feeding, where original supplier containers are transported to the line without any additional material handling, to kitting, where materials are already grouped in kit containers per end product before transport to the line takes place. Other line feeding methods exist which tend to be intermediate forms between bulk feeding and kitting.

Weighing the pros and cons of different line feeding methods is not an easy task and intuitive knowledge about the different systems is not enough to choose the best and most cost-effective solution. Therefore a mathematical model (Limère 2011, Limère et al. 2012) has been developed to guide the decision when choosing between bulk feeding and kitting. Limère et al. (2012) modeled the problem as a static and deterministic optimization model and take as such no stochastic effects into account. For this reason, a simulation model is now built to check if any stochastic effects would have an impact on the cost calculations and the optimal solution of the mathematical model.

LITERATURE REVIEW

In 2010, Hua and Johnson identified a number of research issues concerning the choice of kitting versus line stocking (Hua and Johnson, 2010). They confirmed that research on this topic has been sparse, and suggested further research to the impact of product and component characteristics on the choice to kit or not to kit. Before 2010, one important publication had already studied some of these aspects. Bozer and McGinnis (1992) developed a descriptive model for decision making. Multiple criteria were studied and thus facilitated a quantitative comparison between various kitting plans and line stocking.

To obtain an optimal decision for every part, Caputo and Pelagagge (2011) studied some hybrid policies based on an ABC classification. For each of the three classes A, B and C, a choice could be made between kitting, Kanban-JIT and line stocking. The different scenarios proposed are analyzed for a case study and multiple performance measures are given to compare the solutions and choose the most preferred one. This means that the different performance measures still will have to be weighted according to one's preference, before a solution can be selected.

Hanson and Brolin (2013) compare kitting and bulk feeding based on two case studies. After a detailed analysis of the case studies they give an overview of some comparisons between both systems according to four categories, namely man-hour consumption, product quality and assembly support, flexibility and inventory levels and space requirements. They confirm that generic guidelines for how in-plant materials supply systems should be designed should still be the focus of further research.

Limère et al. (2012) developed a mathematical optimization model for the assignment of each individual part to kitting or bulk feeding. The mixed integer linear programming model minimizes the overall in-plant logistics costs and finds an optimal assignment. Limère (2011) describes an extension of the model.

MATHEMATICAL MODEL

The simulation model will be used to validate the results of the mathematical cost model of Limère (2011). To understand this comparison we will briefly describe the mathematical model without representing the details of the many parameters involved. For further information we refer to the original paper.

The model is a linear mixed integer programming model and assigns for every part the value of a binary variable to zero if the part should be kitted in an optimal situation and to one if the part should be supplied in bulk to the line. The model has the following objective and constraints:

Minimize total in-plant logistics cost (1)

Subject to,

Weight constraint of kit container (2)
Volume constraint of kit container (3)
Space constraint at the border of line (4)

The total in-plant logistics cost (1) consists of four sub-costs:

a) The operator at the line needs to pick the parts for assembly from the border of line (i.e. material façade next to the line). This can be picking from bulk containers or from kits. For bulk containers we assume two different kinds of containers, namely boxes and pallets.
b) Parts need to be supplied from the warehouse or the supermarket to the border of line. Pallets are supplied by forklifts and boxes and kits by tugger trains doing a milk run tour at constant time intervals.
c) Kits need to be prepared. The kit assembly takes place in a supermarket area.
d) The supermarket needs to be replenished from the warehouse.

Constraint (2) and (3) make sure that the optimal assignment of binary variables takes into account that the kit container can only carry a certain maximum weight and fit a certain maximum volume of parts. If more parts are assigned to kitting that can fit into one kit container, the model will automatically create two kits at that work station and the additional cost of creating and transporting this second kit will also be taken into account.
Constraint (4) imposes that all stock at the line (boxes, pallets and kits) needs to be stored parallel to the line and the space along the line is limited to the width of the work station. As such, if there is not enough space to store all items in bulk at a work station, kitting will be imposed.

SIMULATION MODEL

The model described in the previous section (similar to all other existing models described in the literature review) is based on average performances. None of the models take into account stochastic effects. This is therefore the subject of this paper. The research objective is to build a simulation model for the optimal solution from the mathematical model introduced by Limère (2011) and compare the cost results of both models. This will give us insight in the impact of variability on the choice to kit or not to kit.

Aside from the impact of stochastic effects, the simulation model also allows us to model the flows more accurately. In the mathematical model we restrained ourselves to linear functions. Walking distances were therefore approximated by Manhattan distances although the operator would in reality always walk straight to the use-point next to the line (Euclidean distance). In the simulation the walking will be modeled more realistically.

The simulation model is created with the commercial software package FlexSim. Like other commercial discrete-event simulation software packages FlexSim supports the creation of customized simulation objects (Nordgren, 2002). For this model a 'workstation' object, a 'box' object, a 'pallet' object and a 'kit' object were developed. Based on these objects the assembly line and the supermarket, where kitting takes place, could be modeled. Tuggers and forklifts are used to model the different internal transportation flows. The objects are built with standard objects of the FlexSim library and scripting is used to get the functionality needed. The main purpose for the use of customized simulation objects was to minimize the usage of computer resources and maximize efficiency of the model.

Furthermore, the model itself is generated by scripts developed in flexscript (Nordgren, 2002). This is needed to limit development time; it allows generating new models in a flexible and cost-effective way (Govaert et al., 2009). The model holds more than a thousand interconnected objects. Doing this manually would take a few days and every important structural change would take another few days, while right now it takes only a few seconds to generate the model. If a change is needed, only the custom objects need to be changed and then regenerating the model takes again only a few seconds. Aside from the time gain, even more importantly, automatic model generation avoids human errors.

A print screen of the simulation model is shown in Figure 1. The central representation (within the frame) shows the complete model with in the lower part the assembly line, in the left upper corner the supermarket for kitting, and in-between the flow paths for the tuggers and the forklifts. Above the frame a part of the kitting supermarket is shown zoomed in. You can see how this is organized in parallel aisles for all the kits. Below the frame, part of the assembly line is represented. The work stations are located at both sides of the line and every workstation has a border of line organized with pallets, boxes and kits.

PRELIMINARY RESULTS

A preliminary study was done to compare the results of the mathematical model and the simulation model. A first important observation was a difference in the results because of optional parts. To explain this, we will first define what variant parts are, and how optional parts are a

special type of variant parts. Because of the variation in end products on a mixed-model assembly line, different parts need to be assembled in different end products. When a choice can be

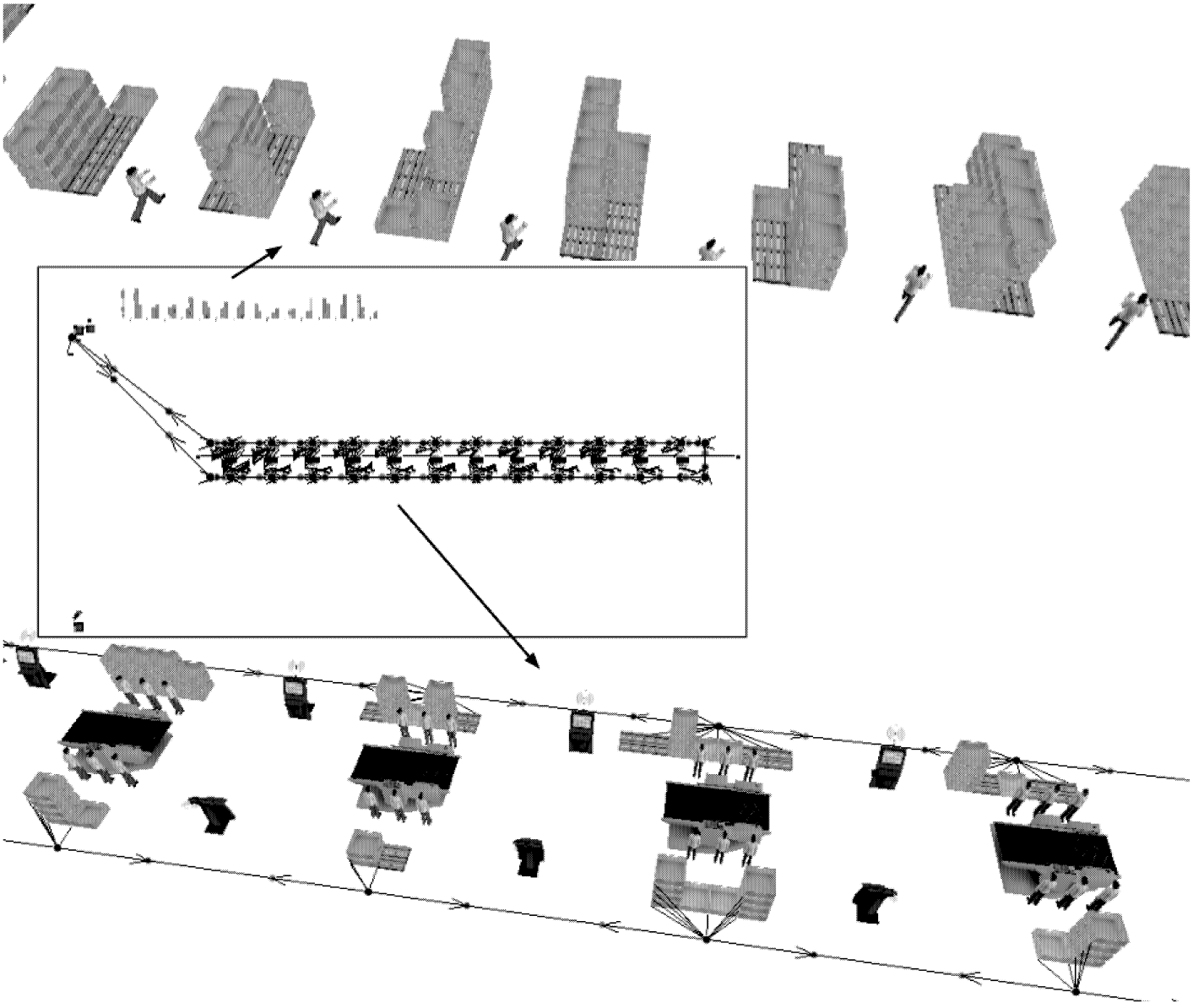


Fig. 1: Simulation model

made between a number of parts, e.g. different types of car radios, these parts are called *variant parts* of a part family. It is sure that you will never assemble two different parts of a part family, but you will make a choice. Usually, you would expect the frequency of occurrence of all variant parts of one part family to some up to 100%, e.g. either you choose radio 1, 2 , …, *n*. Nevertheless, this is not true for *optional parts*. In the case of *optional parts* it is possible that there are end products without that specific component, e.g. 5% of the cars do not have a radio installed.

To understand how the existence of optional parts leads to a difference in costs between the mathematical model and the simulation model, we need to understand how the kit preparation cost is calculated. The kit preparation cost consists of a fixed cost per kit and a variable cost depending on the parts in a kit. In the mathematical model the number of kits is calculated based on constraint (2) and (3). This means that for every part in the kit, a place is reserved with a specific volume and weight. If a part has multiple variants, only one place has to be reserved because every end product will only choose one variant of each part. In the case of optional parts a place is thus also reserved, even if that part is not needed in the specific end product in line. In the special case where only a few optional parts are consolidated in a kit, and all these parts are not required for the next end product in sequence, the mathematical cost model will incorrectly count the fixed cost for making this kit, although in reality no kit will be made.

Although this is an extraordinary case, we decided to further test the model without the inclusion of optional parts to avoid distorted results.

CASE STUDY

In the case study modeled, 438 parts need to be supplied to an assembly line of 23 work stations. A hybrid solution is found as the optimum from the mathematical model. 146 parts are supplied in bulk and the remaining 292 parts are grouped into kits. Table 1 gives an overview of the part dataset and the line feeding policies assigned.

Table 1: Case study – part data set

	Number of parts				
	Originally packaged in box		Originally packaged in pallet		
	Kit	Bulk	Kit	Bulk	Number of kits
Station 1	23	4	10	6	1
Station 2	0	15	0	0	0
Station 3	27	8	0	2	1
Station 4	17	4	8	4	2
Station 5	30	4	3	3	1
Station 6	0	2	0	0	0
Station 7	0	4	0	0	0
Station 8	0	7	0	0	0
Station 9	28	8	7	1	2
Station 10	27	7	2	2	1
Station 11	13	4	0	2	1
Station 12	0	2	0	0	0
Station 13	0	3	0	0	0
Station 14	0		0	1	0
Station 15	7	4	4	0	1
Station 16	0	3	0	4	0
Station 17	13	3	1	2	1
Station 18	10	7	9	0	1
Station 19	23	4	1	0	1
Station 20	14	8	15	0	2
Station 21	0	2	0	1	0
Station 22	0	5	0	0	0
Station 23	0	10	0	0	0

The frequencies of usage of the parts vary between occurrence in 5% of the end products till 100%. The histogram represented in Figure 2 shows the distribution of the frequencies over all parts. In the simulation model, the final assembly sequence is obtained by random sampling part usage information from Bernoulli distributions with the respective frequencies.

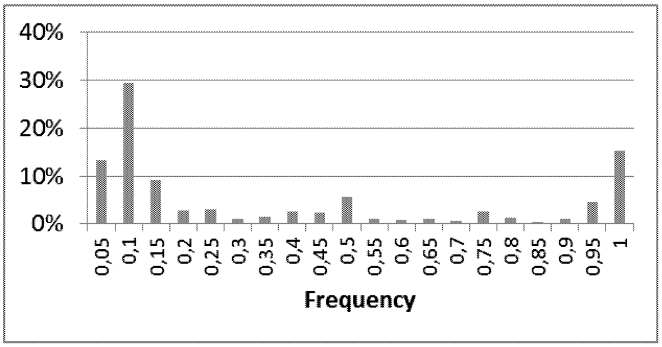


Fig. 2: Part frequencies

RESULTS

Picking at the line

To model the walking distances at the line the mathematical cost model uses a linear approximation of the real walking distances. The reason for this is keeping the CPU time for solving the problem acceptable. In Figure 3, the bold dotted lines represent the real walking paths towards the border of line. This is also how the walking paths are modeled in the simulation model. In the mathematical model, however, the distance is modeled as if the operator is first walking straight to the border of line (1.5m) and only then will walk along the border line to the required part container. This will give an over-estimation of the walking distances. To avoid this, the 1.5m to walk towards the border of line is set to 1m in the calculation.

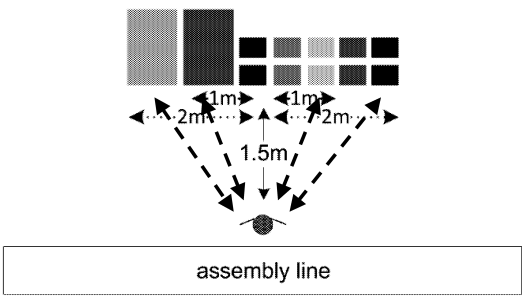


Fig. 3: Real and approximated walking path

Nevertheless, the approximation in the mathematical model will result into inaccurate walking distances. Figure 4 shows, for the case study, the walking distances towards all

parts in bulk for the two models. The simulation model seems to have longer distances than the mathematical model. Another reason for this is that the simulation takes into account that when there are kits, which are always positioned closest to the operator, they take up the best places at the border of line and the other part containers will be positioned further away. However, when we calculate the average difference in walking distances, we notice it is limited to 1.66%. Moreover, if we check the total cost for picking at the line, both from kits and from bulk, we notice that the costs differ by only 0.2%.

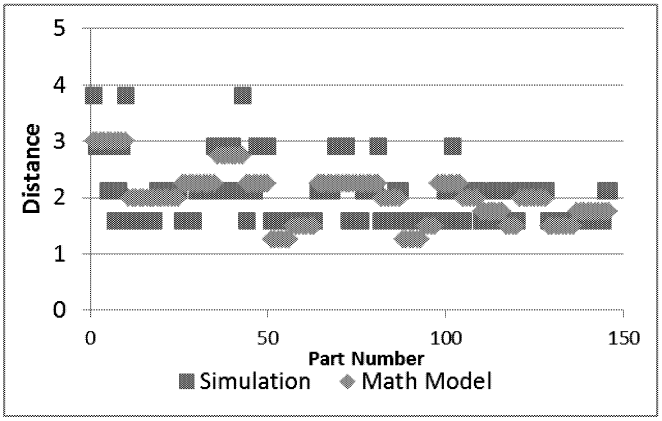


Fig. 4: Distribution of the walking distances

Yet, even though the average walking distances do not vary, the different distributions might have an impact on the choice for a particular line feeding method. It is therefore important to also consider the more accurate walking distances in the simulation. If we want to find an optimal line feeding solution in the future, which takes into account these accurate distances, a simulation-based optimization approach might be considered.

Internal transport

An important part of the model where the stochastic character of the part usage has an effect, is the internal tugger transport. We already mentioned that the tuggers drive a milk run tour at constant time intervals and a tugger has a capacity for pulling a certain amount of boxes or kits at once. Because of the variability of demand, it will not be possible to fully utilize the capacity of the tuggers.

For the kitting tugger there is no variability in demand since kits usage is perfectly predictable, it evolves synchronized with the takt time of the line. The predictability of demand is one of the major advantages of kitting. This is true especially since we are not considering optional parts. The kit tugger can drive at intervals for which the total capacity of the tugger is utilized.

However, for the tugger that transports the boxes to the line, there is a considerable variability in demand. Because of this reason, some tugger runs will be almost empty, while for peak loads the capacity of the tugger might be tight. To take this into account in the mathematical cost model, we used a parameter ρ representing the average utilization of the box tugger. This parameter needed to be estimated beforehand. Thanks to the simulation model we can now do some testing, with changing time intervals for the tugger, to find a valid capacity utilization. The mathematical model can be run again iteratively with the capacity utilization found until the optimal solution from the mathematical model does not change anymore. For the case study, the utilization for a tugger train, driving a milk run tour every 40 minutes, is found to be 56.6%.

Kitting cost

The kitting cost could be modeled accurately. Because of the use of Bernoulli distributions for the generation of an assembly sequence, kits can be prepared at almost constant efficiency. In future research, we would also like to test how different assembly sequencing, with some kind of leveling or batching of demand, could influence the efficiency of kit preparation when kits are assembled in batches.

CONCLUSION

We created a simulation model in FlexSim and an automatic model generator for the analysis of in-plant logistics systems. With the help of the automatic model generator different case studies can be modeled without a lot of additional effort or costs.

A case study is studied to compare the results of the stochastic simulation model with a deterministic mixed integer linear programming model from previous research. Results are reported. This study is still in a preliminary phase. Further investigations will be done to examine the impact of specific production variations, and different case studies. Furthermore, we want to check the impact of consolidating the replenishment of kits and boxes in one milk run tour that runs at smaller time intervals.

Finally, since both models, mathematical and simulation, have advantages, in the future we will also aim for an integration of both in a simulation-based optimization approach.

REFERENCES

Bozer, Y. A., Mcginnis, L. F., 1992. Kitting versus line stocking – a conceptual framework and a descriptive model. International Journal of Production Economics 28 (1), 1-19.

Caputo, A.C. and Pelagagge, P.M., 2011. A methodology for selecting assembly systems feeding policy. Industrial Management & Data Systems, 111(1), 84-112.

Govaert, T., Neirynck, S., Van Volsem, S., Van Landeghem, H., 2009. Combining scripting and commercial simulation software to simulate in-plant logistics. In: Proceedings of the 23th European Simulation and Modelling Conference, 26-28 October 2009, Leicester, UK, 114-118.

Hanson, R. and Brolin, A., 2013. A comparison of kitting and continuous supply in in-plant materials supply. International Journal of Production Research, 51(4), 979-992.

Hua, S.Y. and Johnson, D.J., 2010. Research issues on factors influencing the choice of kitting versus line stocking. International Journal of Production Research, 48 (3), 779-800. First published on: 17 November 2008 (iFirst).

Limère, V., 2011. To Kit or Not to Kit: Optimizing Part Feeding in the Automotive Assembly Industry. Ph.D. thesis, Ghent University.

Limère, V., Van Landeghem, H., Goetschalckx, M., Aghezzaf, E.-H. and McGinnis, L., 2012. Optimizing materials feeding in the automotive industry: kitting and line stocking. International Journal of Production Research, 50(15), 4046-4060.

Nordgren, W.B., 2002. Flexible simulation environment. In: Proceedings of the 2002 Winter Simulation Conference, 250-252.

A SCALABLE AGENT-BASED DECISION SUPPORT SYSTEM: CASE MANUFACTURING SUPPLY CHAIN

Lauri Lättilä
Lappeenranta University of Technology,
Kouvola Research Unit
Prikaatintie 9, FIN-45100 Kouvola, Finland

lauri.lattila@lut.fi

Samuli Kortelainen
Lappeenranta University of Technology,
Faculty of Technology Management
P.O. Box 20, FIN-53851 Lappeenranta,
Finland
samuli.kortelainen@lut.fi

KEYWORDS
Manufacturing, decision support systems, discrete simulation, Monte Carlo, Object-oriented

ABSTRACT

Supply chains are complex structures and require advanced decision support systems to come up with good solutions to difficult problems. Agent-based modelling is a relatively new simulation approach but most work on the issue has been done on a conceptual level. In this paper we utilize an agent-based approach to model supply chains. The model is part of a decision support system where the user can choose to make changes to the supply chain and the model analyzes the performance of the whole system under those settings. The model has been constructed for a manufacturing company with multiple manufacturing locations and central warehouses in Europe. In this paper we show the structure of the simulation model and the decision support system, as well as analyses conducted with the decision support system. The decision support system is able to scale the simulation model to analyze different kinds supply chain solution and as such provides a good platform to estimate possible supply chain alternatives.

INTRODUCTION

Supply chains are complex structures and globalization has made supply chains even more complex (Poon et al. 2009). This has lead to the requirement of having more advanced decision support systems (DSS) in the supply chain context (Hilletofth 2009). There are many different types of DSSs (Power 2002) and recently there has started to be more attention towards simulation based DSSs (Acar et al. 2010; Petering 2011). As the supply chains have become more complex, it can be argued that this creates a high amount of pressure for the DSS. According to Hilletofth et al. (2010a), a DSS needs to provide suitable and accurate information as well as predictions about the performance of the system with different types of structures.

Agent-based modelling (ABM) is a relatively new approach in simulations. Although the roots go back to von Neumann machines, it began to reach more interest in the 1990s as computing power has increased dramatically (Macal and North 2005). According to Macal and North (2006) there are four reasons behind the growing interest in ABM. Firstly, systems are becoming more complex in terms of interdependence. Secondly, many systems have been too complex to analyze with other methods. Thirdly, data is available at a finer level. And fourthly, increase in computational power has made ABM more feasible. These reasons are a good argument to use ABM also in a supply chain context. However, most work in ABM is still done on a conceptual level and only few empirical models exists (Davidsson et al. 2005; Hilletofth et al. 2010b).

In this research we use ABM to create a model to analyze manufacturing supply chains. The model has been done for a Finnish manufacturing company with multiple manufacturing units and central warehouses in Europe. The ABM is part of a DSS and the purpose of this research is to present the structure of the simulation model, as well as highlight the benefits of using ABM in a DSS. This paper provides both a conceptual framework how a multi-echelon supply chain can be modeled using ABM and an empirical case to analyze the feasibility of the proposed model.

This paper is structured as follows. The first section provides an overview of the paper, as well as presents the goal of the research. The second section consists of the literature review. We discuss the major decisions in manufacturing supply chains. In the third section we present the actual structure of the simulation model. Each agent is presented separately as well as the data requirements and the DSS structure. The fourth section discusses the actual analyses which can be conducted by using the DSS, while the last section concludes this research and provides further avenues of future research.

SUPPLY CHAIN MANAGEMENT IN MANUFACTURING ORGANIZATIONS

Logistics has integrated many functions due to improvements in information and communications technologies (Hesse and Rodrigue 2004). Different parts were individual entities earlier, but nowadays supply chain management can be seen to combine logistics with information technology, marketing, and strategic planning. Logistics began from different, separated functions but today it has become an integrated part of supply chain management. Supply chain management can be seen to be a part of a process called postponement and speculation. Initially this issue was analyzed in distribution channels by Buckling (1965). In speculation the actor chooses to shift risk on itself by speculating what is needed in the future. Postponement is the opposite of this. There are four different generic strategic for supply chains: manufacturing postponement, logistics postponement, full speculation, and full postponement (Pagh and Cooper 1998).

Manufacturing postponement regards the internal warehousing issues while logistics postponement will impact the structure of the whole distribution supply chain. Many issues (such as market and demand, product characteristics, manufacturing and logistics capability) impact what strategy should be chosen. However, it can be argued that one generic structure might not fit well for all products. In these cases a differentiated supply chain strategy needs to be developed (Hilletofth 2009). Hilletofth (2009) presented a model that consists of four phases: developing a segmentation model, understanding the market, understanding capabilities to serve the market, and finally developing necessary supply chain solutions.

THE SIMULATION MODEL AND THE DECISION SUPPORT SYSTEM

The simulation model construction process consisted of many phases, where the modeler met with two key personnel from the case company (supply chain development engineer and supply chain manager). During the construction process verification has been mainly done by using unit root testing and model logging. In addition to these, the model contains many stop conditions when it is not performing correctly (not ordering enough, too high inventory levels, etc.). The model also received validation when the engineer met with the key personnel and four supply chains were analyzed.

The simulation model can be said to be an agent-based model. The model consists of four different agents: distribution warehouses, central warehouses, machines, and products. Each of the separate agents will be presented and, as well as the decisions made by different agents during the simulation run. The simulation model was constructed by using Anylogic. Anylogic is a simulation tool which supports discrete-event simulation, ABM, as well as System Dynamics.

The agents and their connections in the model

Production machine

Production capability is modeled on the level of individual production machine. Most of the machines can be used to make many different products. However, there are significant efficiency differences in producing a product with different machines. Therefore, the production capacity over time is dependent on the combination of machine and product.

The machines are naturally geographically tied to a particular factory. Production machines within a factory replenish one particular central warehouse, which is physically situated at the factory. All the products produced by the firm flow through these central warehouses. It can be said that all of the machines in one central warehouse location represent a manufacturing unit.

Each machine has a specific amount of shifts and capacity per shift. The capacity depends on the amount of machines and overall equipment efficiency. Each machine keeps track of its current backlog of products and modifies its capacity accordingly. If the load on the machine is too high, new shifts will be included. On the other hand, when load is low enough, it is possible to remove shifts. This is presented in Formula (1). There is also a delay when the capacity is modified (management must inform production personnel one week before a new shift is included or removed).

$$MC_{m,t} = \begin{cases} 1 & CL_{m,t} > MaxC_m \\ -1 & CL_{m,t} < MinC_m \\ 0 & otherwise \end{cases} \quad (1)$$

where $MC_{m,t}$ = modify capacity for machine *m* during time period *t*,
$CL_{m,t}$ = current load at machine *m* during time period *t*,
$MaxC_m$ = maximum allowed load at machine *m*,
$MinC_m$ = minimum allowed load at machine *m*.

During each week each machine checks all of the orders from the central warehouses. The machine will first produce make-to-stock (MTS) products with pending distribution warehouse demand (a stock out), then make-to-order (MTO) products, and finally MTS products with enough stock in the central warehouses. Each product has a specific setup time for a batch and manufacturing time for each unit. The lot size calculations are presented in Formula (2). If no alarm level is set for the product, $A_{w,p}$ is assumed to be zero. Also, if there is no batch size for the product, $B_{w,p}$ is set to one. There are also backup machines for some products in which case the manufacturing unit might use the backup machine, if the load at the main machine is too high. The manufacturing units might also use another central warehouse as a backup source. The decision to utilize a backup machine (or country) is presented in Formula (3).

$$ls_{m,p,t} = B_{w,p} \times \left\lceil \frac{A_{w,p} - S_{w,p,t} + D_{w,p,t}}{B_{w,p}} \right\rceil \quad (2)$$

where $ls_{m,p,t}$ = lot size for machine *m*, product *p*, during time period *t*,
$B_{w,p}$ = batch size for product *p* at central warehouse *w*,
$S_{w,p,t}$ = stock level for product *p* at central warehouse *w* during time period *t*,
$D_{w,p,t}$ = demand for product *p* at central warehouse *w* during time period *t*,
$A_{w,p}$ = alarm level for product *p* at central warehouse *w*.

$$bu_{w,p,t} = \begin{cases} 1, & L_{m,t} > ML_m \wedge L_{bm_{m,p}} < MBL_m \\ 0, & otherwise \end{cases} \quad (3)$$

where $bu_{w,p,t}$ = use backup at warehouse *w* for product *p* during time period *t*,
$L_{m,t}$ = load at machine *m* during time period *t*,
ML_m = maximum load at machine m,
$L_{bmm,p}$ = load at backup machine *bm* for machine *m* for product *p*,
MBL_m = maximum backup load at machine *m*.

Central warehouses

After the products have been manufactured, they go to central warehouses. The central warehouse agent is responsible for sending orders to manufacturing and delivering products to distribution warehouses. The central warehouses handle also logistics between firm's central

warehouses. These internal transfers are used to compensate for high peaks of demand and to transfer products that are not manufactured in all factories.

There are two types of warehousing policies: MTO and MTS. For MTS products there is an alarm level, after which the central warehouse orders more goods. The MTO products can be represented by having an alarm level of zero. Formula (4) summarizes the ordering decision *ord* for each product *p* at each central warehouse *w*. With a MTO product, the alarm level is set to zero. If a product is ordered, the lot size calculations are totally similar to the manufacturing lot size decisions presented in Formula (2).

$$ord_{w,p,t} = \begin{cases} 1, & S_{w,p,t} - D_{w,p,t} < A_{w,p} \\ 0, & otherwise \end{cases} \quad (4)$$

where $ord_{w,p,t}$ = order product *p* for central warehouse *w* during time period *t*,
$S_{w,p,t}$ = stock level for product *p* at central warehouse *w* during time period *t*,
$D_{w,p,t}$ = demand for product *p* at central warehouse *w* during time period *t*,
$A_{w,p}$ = alarm level for product *p* at central warehouse *w*.

If goods need to be ordered, central warehouse will check whether the product is produced at a local manufacturing unit or should it be ordered from another central warehouse. If the product is ordered from another central warehouse, a delivery order is created. The delivery orders also keep track of unmet distribution warehouse demand, which will be used by the other central warehouse in decision making (treated as "current demand" in Formula (4)). The central warehouses also need to decide how much goods to send to various distribution warehouses if the amount of a specific product is too small compared to all demand. This is presented in Formula (5).

$$G_{w_1,w_2,p,t} = \frac{LS_{w_1,w_2,p,t}}{\sum_1^n LS_{w_1,n,p}} \quad (5)$$

where $G_{w1,w2,p,t}$ = how many product *p* to send from warehouse w_1 to w_2 during time period *t*,
$LS_{w1,w2,p}$ = lot size from warehouse w_1 to warehouse w_2 from product *p* during time period t,
n = total amount of warehouses.

Between some locations it is possible to include a minimum delivery size. This might be needed if the central warehouses are on different continents and it is too expensive to send small batches. If the size is too small, the central warehouse will wait for more delivery and distribution warehouse orders. This is presented in Formula (6)

$$del_{w_1w_2} = \begin{cases} 1, & \sum G_{w1,w2,p,t} < MS_{w1,w2} \\ 0, & otherwise \end{cases} \quad (6)$$

where: $del_{w1,w2}$ = send delivery from warehouse w_1 and w_2,
$G_{w1,w2,p,t}$ = how many product *p* to send from warehouse w_1 to w_2 during time period *t*.

Distribution warehouses

The distribution warehouses create the demand for the products in the model. In reality the distribution warehouses in the model are regional sales organizations, who themselves sell the products for final customers. Each market buys its products from designated central warehouses depending on the product. At the moment, they are not seen to be an integrated part of the supply chain, as the company does not use final distribution warehouse demand (purchases from the regional organization) in their production or warehousing planning. In the model, the performance of the supply chain does not impact how the regional organizations order goods from the larger distribution central warehouses and as such can be seen to be external to the system. The model does not simulate the stock levels of different products for the distribution warehouses, but only the demand for products is modeled.

Each distribution warehouse will order different products from different central warehouses. As such, a distribution warehouse will buy from multiple central warehouses during the simulation. The demand for each product is stochastic. We analyzed their historical demand data and most of the products followed an exponential distribution. There is a specific percentage whether a distribution warehouse will order during a week, and then the actual order will follow the exponential distribution. Two parameters (percentage of weeks with demand and average demand) are able to represent the distribution warehouse demand relatively well for each product. Figure 1 shows an example of generated demand in the model.

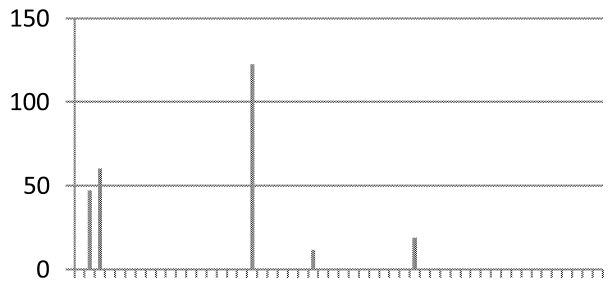


Figure 1: Example of demand

As the model is a weekly model, demand is generated for each product during each week for each distribution warehouse. The model randomized the arrival sequence for the demand and the central warehouses operate in a first-in-first-out policy, e.g. first goods are sent to the distribution warehouse who ordered first. If the central warehouse has enough goods, they are sent to the distribution warehouse. Otherwise a distribution warehouse order is generated to keep track how long the distribution warehouse has to wait for the goods.

Product

A product is in the core of the model, as they are tied to every part of the model. Machines produce the products, central warehouses are used to store them, and naturally

distribution warehouses order products. A high number of products can be explained by a high number of different product variations. These variations are significant as they control for the necessary production machines and the necessary manufacturing time. The high number of products causes that the complexity of the model increases. To cope with the complexity an ABM structure was a necessity. The products are more static parts in the models as they do not make any actual decisions. However, they tie up the whole supply chain together.

The decision support system

The DSS consists of the simulation model, an Access –file for the model structure, and a Graphical User Interface (GUI) as well as an Access -file for results. The simulation model uses the first Access -file to generate the supply chain. The data is inserted in the Access -file to various tables and each table represents one part of the supply chain. The user can modify the Access -file to analyze various types of supply chains and investigate the impact of modifying some parameters.
Table 1 shows the major tables in the Access -file used by the DSS.

Table 1: Data used in the DSS

Table name	Data to insert
Distribution warehouse demand	Used demand parameters for product at a distribution warehouse
Distribution warehouse serving locations	Distribution warehouse orders which products from which central warehouse
Delivery costs	Cost to deliver goods between warehouses
Delivery delay	Delivery delay between warehouses
Machinery	Production machines, capacity, location
Products	Product information (Segment, Group, Family, Weight), cost of raw materials in different manufacturing units
Central warehouses	Countries containing a central warehouse, cost of production in that country
Warehousing policy	Alarm levels, batch sizes, critical points and order locations for products at central warehouses

The actual simulation model is totally scalable using the Access -file. The GUI works as a layer between the decision-maker and the DSS. The user does not need to interact with the actual model at all as the Access -file creates the structure of the model. Any desired changes can be made through the Input Access -file and it is more user friendly to the users compared to the actual coding used in a simulation model.

Using the Decision Support System to Analyze Supply Chain Alternatives

The user can do various analyses using the DSS. The structure of the supply chain (amount of central warehouses, warehousing decisions) has a high impact on the costs of the whole supply chains, as well as the ability to serve distribution warehouses. Table 2 shows the potential options which the managers can modify by using the DSS.
The supply chain options come back to the postponement and speculation discussed in the literature review. Also, there are many delays in the model (manufacturing delay due backlog and capacity, delay due to delivery, delays due to minimum batch size) which are known to be difficult to comprehend in decision-making for manufacturing companies (Forrester 1958). There are also many different key performance indicators (KPI) which are of interest to the decision-makers. These are summarized in Table 3.

Table 2: Potential supply chain options to analyze with the model

Location	Supply chain options
Distribution warehouse	Impact of change in demand
Central warehouse	Which products are ordered from which central warehouses by distribution warehouses
Central warehouse	Size of partial delivery
Central warehouse	MTO or MTS, ordering batch size
Central warehouse	Order from other central warehouse or produced at local manufacturing unit
Central warehouse	Minimum delivery size to different countries
Manufacturing unit	Capacity per shift for each machine
Manufacturing unit	When to increase or decrease the amount of shifts
Manufacturing unit	Back up machinery or central warehouse, when to use back up

Table 3: Key Performance Indicators analyzed with the model

Key performance indicators	Description
Freight costs	Between warehouses and system total
Manufacturing costs	At each manufacturing unit and system total
Warehousing costs	At central warehouses and system total
Warehousing penetrations	For each product, family, group, and segment, and system total
Warehousing levels	For each product, family, group, segment, and all products; at the chosen analysis level (pieces, kg, Euros, critical point penetrations); which warehouse or system total
Manufacturing loads	For each machine, at the chosen level (load in week, utilization, amount of shifts)
Delivery times	For each product, family, group, segment, or all products; at the chosen central warehouse or system total

KPIs are the values which the simulation model presents through the GUI. The user chooses the desired KPI and analyses level, and then the GUI calculates these values from the simulation model's results. As there is stochastic demand in the model, the simulation model is usually run 100 times to analyze the performance. As such, most results are histograms showing the average values during the simulation run or cumulative values at the end of the simulation run. These can be analyzed in the GUI.
In the GUI the user can browse through the supply chain and its results. In addition to 100 run simulations, it is possible to analyze individual simulation runs. This might provide additional insights, as aggregation loses some information. For instance, some products might be too low level for long times in some cases and these do not show up that easily in aggregated values.

DISCUSSION AND CONCLUSIONS

ABM provides a good avenue to construct versatile models to enhance decision-making. By using an external file it is possible to make a totally scalable model which provides more possibilities for the actual user, compared to model with fixed amount of parameters. It is also good to separate the model from the users and use a GUI in between, as it is easier to approach simple controls than to use a simulation environment. ABM can also incorporate the major issues in supply chain management (Bucklin 1965; Pagh and Cooper 1998; Hilletofth 2009). It is possible to include both the speculation and postponement strategies in the supply chain, as well as create a differentiated strategy for different segments. It can be argued, that this type of DSS is a good way to test strategies created with other approaches (Hilletofth et al. 2010a). ABM is also able to grasp the complex interdependences which exist in supply chains (Macal and North 2006). This work is also based on an empirical case and gives further justification in usage of ABM is supply chain context (Davidsson et al. 2005; Hilletofth et al. 2010b). Also, ABM works very well inside the DSS (Power 2002, Shim et al. 2002). However, as the company has only just started to use the model it is still not possible to estimate the usefulness and impact of the DSS on the organizational performance (DeLone and Mclean 1992).

This work provides many managerial implications. The actual simulation model will help the case company in improving its operational performance. The simulation environment is a lot more advanced than any other methods the company has been using to analyze the problems. This work also provides guidelines to other simulation engineers to come up with similar kinds of systems. It is expected that these types of advanced, agent-based DSSs to become more prevalent in the future.

This work has some limitations. The manufacturing company operates in a batch production setting and it is possible that the model cannot properly represent a continuous process. The supply chain is also limited to its current operating procedures, e.g. no cross docking or other advanced methods to minimize logistical costs. It also does not take other production than final production into account, which limits the supply chain analysis to the distribution channels.

Further research includes expanding the model to include total production, as well as cross docking and other logistical options. These can be included in the model and it is possible to make it work through the input-file as well. This way the model could be expanded to analyze the whole supply chain. It would be interesting to see whether it is as "easy" to create a service supply chain model with a similar kind of structure. This research only shows one way how ABM can be used in DSSs. It seems to provide large possibilities in improving decision-making at individual companies. In the future we should see more usage of ABM and multi-method simulation models as part of DSSs and help managers make supply chains take the next step towards improved performance.

REFERENCES

Acar, Y.; Kadipasaoglu, S.; Schipperijn, P. (2010) "A decision support framework for global supply chain modelling: An assessment of the impact of demand, supply and lead-time uncertainties on performance", *International Journal of Production Research*, 48(11), pp. 3245-3268.

Bucklin, L.P. (1965), "Postponement, Speculation and the Structure of Distribution channels", *Journal of Marketing Research*, 2(1), pp. 26 – 31.

Davidson, P., Henesey, L., Ramstedt, L., Törnquist, J. and Wernstedt, F. (2005), "An analysis of agent-based approaches to transport logistics, Transportation Research Part C: Emerging Technologies", 13(4), pp. 255 – 271.

DeLone, W.H. and McLean, E.R. (1992), "Information systems success: The quest for the dependent variable", *Information Systems Research*, 3(1), pp. 60 – 95.

Forrester, JW. (1958), "Industrial Dynamics – A Major Breakthrough for Decision Makers", *Harvard Business Review*, 36(4), pp. 37 – 66.

Hesse, M and Rodrigue, K-P. (2004), "The transport geography of logistics and freight distribution", *Journal of Transport Geography*, 12(3), pp. 171 - 184

Hilletofth, P. (2009), "How to develop a differentiated supply chain strategy", *Industrial Management & Data Systems*, 109(1), pp.16 – 33.

Hilletofth, P.; Lättilä, L.; Ujvari, S.; Hilmola, O-P. (2010a) "Agent-based decision support for maintenance service provider", *International Journal of Services Sciences*, 3(2/3), pp. 194–215.

Hilletofth, P., Aslam, T., and Hilmola, O.-P. (2010b), "Multi-agent based supply chain management: Case study of requisites", *International Journal of Networking and Virtual Organisations*, 7(2/3), pp. 184–206.

Macal, C.M. and North, M.J. (2005), "Tutorial on agent-based modelling and simulation", in Kuhl, M.E., Steiger, N.M., Armostrong, F.B. and Joines, J.A. (Eds.), *Proceedings of the 2005 Winter simulation Conference*, pp. 2 – 16.

Macal, C.M., and North, M.J. (2006), "Tutorial on agent-based modeling and simulation. Part 2: How to model with agents", in L.F. Perrone, Wieland, F.P., Liu, J., Lawson, B.G., Nicol, D.M., and Fujimoto R.M. (Eds.), *Proceedings of the 2006 winter simulation conference*, pp. 73 – 83.

Pagh, J.D. and Cooper, M.C. (1998), "Supply chain postponement and speculation strategies: how to choose the right strategy", *Journal of Business Logistics*, 19(2), pp. 13 – 33.

Petering, M.E.H. (2011), "Decision support for yard capacity, fleet composition, truck substitutability, and scalability issues at seaport container terminals", *Transportation Research Part E*, 47(1), pp. 85-103.

Poon, T.C.; Choy, K.L.; Chow, H.K.H.; Lau, H.C.W.; Chan, F.T.S.; Ho, K.C. (2009), "A RFID case-based logistics resource management system for managing order-picking operations in warehouses", *Expert Systems with Applications*, 36(4), pp. 8277-8301.

Power, D.J. (2002), *Decision Support Systems: Concepts and resources for managers*. Quorum Books. Westport (USA).

ESTIMATING THE COST STRUCTURE OF WOOD CHIP SUPPLY CHAINS: A SIMULATION APPROACH

Lauri Lättilä
Lappeenranta University of Technology,
Kouvola Research Unit
Prikaatintie 9, FIN-45100 Kouvola,
Finland
lauri.lattila@lut.fi

Kalle Karttunen
Olli-Jussi Korpinen
Jarno Föhr
Tapio Ranta
Lappeenranta University of Technology,
Faculty of Technology
P.O. Box 20, FIN-53851 Lappeenranta,
Finland

KEYWORDS
Forestry, transportation, decision-making, hybrid simulation, system analysis

ABSTRACT
Bioenergy plays an important role in future energy production. In Finland the use of wood chips is increasing rapidly. However, as there is a high amount of competition in the wood industry in Finland and the collection strategy is an important strategic decision for power plants, it is important to make an informed decision using advanced tools. Due to high volumes even minor cost benefits can have large benefits. In this study a simulation model is provided which can be used to estimate the cost structures of various kinds of wood chip supply chains. The model uses data from a geographical information system, which enables accurate estimates about the cost structures. By utilizing advanced containers in wood chip supply chains it is possible to decrease costs by almost 20%.

INTRODUCTION

Wood is an important energy source in many countries. In Finland wood is the second largest energy source after oil. Oil, coal, and natural gas account to 45% of energy use in Finland (Official Statistics of Finland 2012) but Finland should still decrease national CO_2 emissions by 16% in 2020 compared to 2005 (Jalkanen et al. 2008). According to Chum et al. (2011), bioenergy plays a key-role in reducing global greenhouse gas emissions in the long term. According to Ylitalo (2012), 7.5 million solid cubic meters of wood chips were used in heat and energy production in 2011 in Finland. This amount is expected to increase to 12 – 13 million by 2020.

The costs of collecting wood chips using roadside chipping is about 18 € / MWh with a hauling of 50 kms (Karttunen et al. 2010; Bioenergia 2012). As the average price for heat production in a CHP-plant is about 40 € / MWh depending on the moisture content of the wood chips (Ihalainen and Niskanen 2010), the supply chain accounts to 40 % of the total costs. Improvements in the supply chain can have a large impact on the total costs of operations.

According to Laitila et al. (2010) there exists many different kinds of improvement areas in Finnish wood chip supply chains. Two of these are improving the efficiency of road transportation as well as improving the whole supply chain in long hauling utilizing multiple transports. Both of these issues are connected to a strategic decision made by power plant on how to collect the required wood chips (Ihalainen and Niskanen 2010).

The aim of this research is to provide a methodology which can be used to estimate the costs of different types of supply chains for wood chips collected using roadside chipping. The main research question of this study is "What are the main cost components of wood chip supply chains?". This manuscript provides a methodological contribution. A methodology to construct a simulation model for wood chip supply chains is presented. The same methodology can be used in other cases as well to estimate the cost structure of different kinds of supply chains. This paper will also look into advanced containers which can be used in wood chip supply chains. These type of containers can also utilize compression, where excess air is removed.

This manuscript is structured as follows. The first section provides an introduction. The second section contains the methodology of the paper. This includes the data sources as well as the actual simulation model. The third section uses the simulation model to analyze three different scenarios in a case study while the last section provides the discussion and conclusions of this paper.

METHODOLOGY

Structure and Data Sources

The simulation model was constructed during a larger research project and the simulation model benefitted from multiple sources of information. During the construction of the model an expert was interviewed. The expert worked as part of the research team but he had spent a considerable time participating in the operations of individual wood chip trucks. He also conducted demonstrations with the new type of technology to understand how the technology works in the actual operations. The data for the model comes from three sources. The first source is a survey (Karttunen et al. 2012). This survey concentrated on the currently used trucks as well as future expectations on volumes. The second source is an availability analysis conducted by utilizing Geographic Information Systems (Korpinen et al. 2012). This data is used to estimate the availability of wood chips next to a specific location. The final source of data was case studies which were used to estimate various delays.

The availability analysis is an important part of the model. The analysis takes into account land use, forest fuel

availability, and local competition. The software then minimizes distances to gather a specific amount of material (depending on the size of the production facility) and then exports data for the simulation model. The data contains information about the length and types of roads needed to reach specific wood sources. More information about the analysis can be found in Korpinen et al. (2012).

The model also utilizes a graphical user interface (GUI) which separates the user from the rest of the model. The user can set some parameters through the GUI, which then modifies the structure of the model according to the parameters. The simulation model then runs the analysis and provides the output data as more meaningful results to the user. The parameters allow the user to run different kind of scenarios. At the current state the user can choose between multiple scenarios (14 scenarios overall), which are meaningful from the perspective of the model. These modify the actual parameters, which are presented in Table 1.

Table 1: Parameters Available

Parameter	Values
Amount of trucks	0…N
Type of trucks	Container or traditional
Type of rotator	Mobile or fixed
Compression used	Yes or no
Satellite terminal	Yes or no
Size of plant	540 or 740 GWh

Scenarios were used instead of direct modifications as there are strong interdependencies in the system (for instance, when the amount of trucks is increased, they are able to gather more fuel, which would require a longer distance). Another important issue regarding the data sources is moisture data, which has an impact on the operations. The average moisture content in Finland is between 40 and 45% (Alakangas 2000), but in the daily operations the moisture content varies between 30 to 60%. For the simulation model data from a Finnish power plant was used to include monthly variation. This is presented in Figure 1.

The primary impact of the moisture content comes from two different sources. Firstly, as the trucks are mostly limited by the weight of the load, the higher the moisture content, the less material can be loaded into a vehicle (Figure 2). The type of truck has an impact on the size of the payload area. Secondly, the higher the moisture content, the less energy is generated at the power plant (Figure 3). This is simply due to the amount of energy which gets wasted when the wood chips burn.

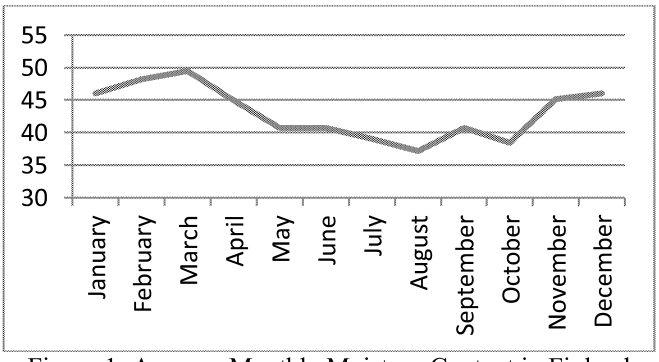


Figure 1: Average Monthly Moisture Content in Finland

The final issue from the perspective of data is the impact of moisture content on train containers, presented in Figure 4. Due to higher weight limits, trains are not that frequently limited by weight but instead of space.

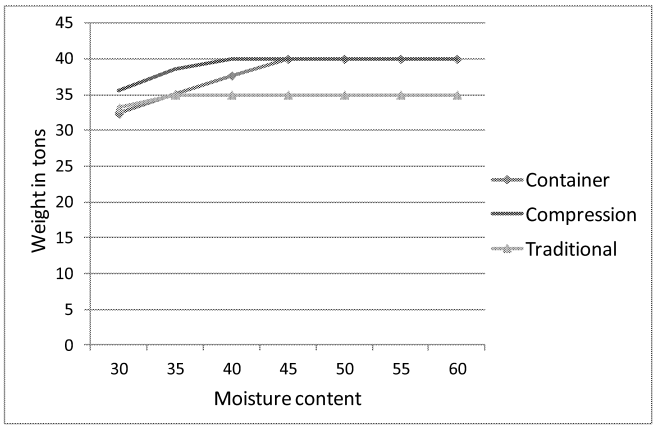


Figure 2: Impact of Moisture Content on Weight

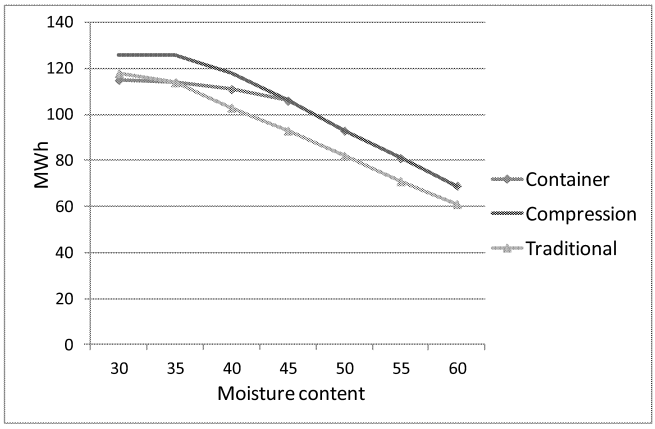


Figure 3: Impact of Moisture Content on Energy Content

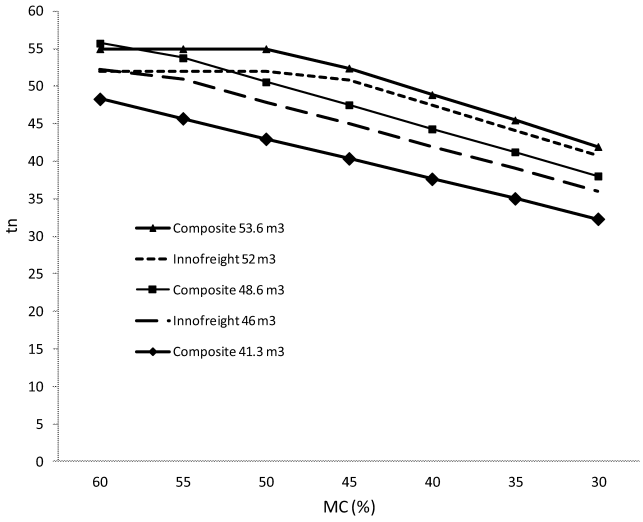


Figure 4: Impact of Moisture Content on Train Containers

Simulation Model

The simulation model utilizes both agent-based principles as well as discrete-event simulation. As the actual trucks play a central part in the supply chains, they are modeled as agents. However, the terminal can benefit from a discrete-event structure as it mainly consists of servers and queues. As multi-method simulation was used, the model was chosen to

be constructed using Anylogic. The truck agents will be first presented, thereafter the terminal operations and finally the trains. Statechart for the trucks is presented in Figure 5.

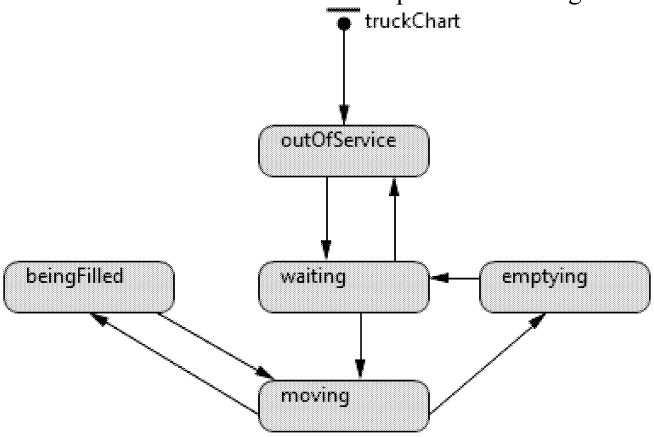


Figure 5: Statechart for the Truck-Agent

The trucks begin their simulation in the "outOfService" – state. This state indicates that the truck is not being used in operations. The trucks follow realistic work shifts. From January to April the trucks operate in two shifts, one shift in May and June, no shifts during July and August, one shift in September and October, and again two shifts in November and December. It is important to take the work shifts into account as fixed costs are relatively high and the moisture content varies during the year.

In the "waiting"-state trucks will wait the end of their shift if there are no appropriate loads available. When the truck enters the state, it will check whether a load exists, which it can collect and return during its working hours. In the real system there are also some potential loads near the power plants and the chippers choose an appropriate load. The model also includes potential overtime, as in many cases a small amount of overtime would allow the truck to pick up a load. When a load is available, the truck will change to the moving state. Depending on the load location (which is derived from the GIS data), the driving time will depend on the roads between the terminal and the load location (also derived from the GIS data).

When the truck reaches the location, it gets filled ("beingFilled"-state). The filling simply takes a random amount of time. After the filling operations, the truck moves back to the terminal. When the truck reaches the terminal, it enters the "emptying "-state, during which the truck moves the load to the terminal, which follows the discrete-event structure.

The operations at the terminal depend on the terminal location and whether the truck is a traditional truck or a container truck. If the terminal is the main terminal, the trucks will simply empty their load and look for a next one. The situation changes for the container trucks if the trucks are operating at a satellite terminal. The traditional trucks operate similarly at the satellite as they simply dump their load and then head for the next load. The container trucks, on the other hand, leave full containers at the terminal if the train also uses the same sized containers. If the train uses different sized on containers, the truck will empty the load as it if were at the main terminal.

The train operations depend on the size of containers used in the train. If the smallest container is used, the train operates on the principles of multi-modal supply chains, e.g. the train picks up the full containers and leaves the empty ones. In those cases the train is preferred before the trucks and gets first fully emptied and loaded before any truck gets emptied. The same is true at the main terminal as well. If the train utilizes larger containers, there is a separate machine which handles the filling of the containers. However, at the main terminal the operations are similar for both types of the train; the rotator prefers the train over trucks (in case of container trucks). The train follows a strict schedule and in the rare case that the train cannot be fully filled, the train will leave with partial load.

Parameters

The model contains many parameters which come mainly from the survey and the case studies. All of the parameters of the model are presented in Table 2. Parameters related to the rotators are based on a small number of initial experiments due to which they are still relatively unreliable compared to the parameters concentrating on the traditional trucks.

Table 2: Main Parameters of the Model

Parameter	Value	Scale	
Minimum filling time	Container	50	minutes / truck
	Traditional	50	minutes / truck
Maximum filling time	Container	80	minutes / truck
	Traditional	80	minutes / truck
Minimum emptying time	Fixed rotator	5	minutes / container
	Mobile rotator	5	minutes / container
	Traditional	20	minutes / truck
Maximum emptying time	Fixed rotator	7	minutes / container
	Mobile rotator	7	minutes / container
	Traditional	30	minutes / truck
Additional time	Container	10	minutes / truck
	Traditional	10	minutes / truck
Allowed overtime		120	minutes
Wage per shift		186	€ / day
Weight of truck and tare	Container	20	tn
	Traditional	25	tn

Costs also play an important part in the simulation model. The survey was used to estimate the costs. In addition to this, the research group has earlier developed cost models for wood chip supply chains. The costs have been divided to two main categories: fixed costs and variable costs. Due to the nature of simulations it is easy to separate the fixed costs from the variable costs during the simulation run. Spreadsheet based calculations need to make stronger assumption on many of the delays, which can simply be calculated in the simulation model during run time (for instance waiting time at the main terminal). All of the costs components are presented in Table 3.

In addition to the cost components it is assumed that the container trucks and train can be used to conduct other operations when they are not being used for wood chip supply chains. In Sweden similar trains spend summer months hauling trash, which could as well be the case in

Finland. The container trucks could also be used to transport gravel during summer months, when a high amount of construction is required. The traditional trucks are usually idle during the summer months. In the model the fixed costs are only calculated for the periods when the truck or train is operating in the wood chip supply chain.

Table 3: Cost Components

Component	Type	Value	
Chipper	Fixed	285000	€ / a
	Variable	3	€ / tn
Truck, traditional	Fixed	86000	€ / a
	Variable	0,01235	€ / TKm
Truck, container	Fixed	83000	€ / a
	Variable	0,01235	€ / TKm
Rotator, fixed	Fixed	57519	€ / a
	Variable	11,04	€ / rotation
Rotator, mobile	Fixed	67400	€ / a
	Variable	25,79	€ / rotation
Locomotive	Fixed	319039	€ / a
41,3 m3 container	Fixed	5400	€ / container
48,6 m3 container	Fixed	5797	€ / container
53,6 m3 container	Fixed	6195	€ / container
46 m3 container	Fixed	3036	€ / container
52 m3 container	Fixed	3036	€ / container
41,3 m3 container	Variable	1,6	€ / TKm
48,6 m3 container	Variable	1,3	€ / TKm
53,6 m3 container	Variable	1,2	€ / TKm
46 m3 container	Variable	1,4	€ / TKm
52 m3 container	Variable	1,3	€ / TKm
Terminal, traditional	Fixed	154202	€ / a
Terminal, container	Fixed	179023	€ / a

RESULTS

As the simulation model allows to study multiple scenarios, we will concentrate on three scenarios. In these scenarios the power plant requires 740 GWh of energy throughout the whole year. The main terminal is located in Jyväskylä while a potential satellite terminal with railway connection is located in Kontiomäki, located 363 km from Jyväskylä.

The first case to analyze concentrates only on the main terminal utilizing more traditional trucks. The material is collected nearby the main terminal and the average collection range is 83 kilometers. In the second and third case the main terminal is used to gather 540 GWh of material while the satellite supplies 200 GWh on material. The second case concentrates on a multimodal supply chain, where container trucks gather the wood chips both at the main terminal and the satellite terminal. In this case the collection range at the main terminal is 71 km while the collection range at the satellite is 48 km. In the third case traditional trucks gather material at both of the locations but the train utilizes large steel containers (Innofreight 52 m^3, used in Sweden). The results for these three scenarios are presented in Table 4.

The scenario with traditional trucks operating only at the main terminal is the “easiest” way to expand operations for a power plant. The trucks simply need to drive further away from the terminal to be able to collect enough wood chips, but there is no need to start making any kind of arrangements for the train. This scenario also has a clearly higher cost than the other two scenarios.

Table 4: Results of three different scenarios

	Traditional without satellite	Multimodal containers	Traditional with steel containers
Truck variable	0,909	0,758	0,985
Truck fixed	4,709	2,405	3,487
Rotator fixed	0	0,079	0,078
Rotator variable	0	0,322	0,07
Chipping fixed	3,462	2,345	2,755
Chipping variable	1,16	1,138	1,169
Train fixed	0	0,619	0,408
Train variable	0	0,459	0,339
Satellite terminal	0	0,245	0,21

In the second scenario multi-modal operations are utilized and both the main terminal and satellite terminal use container trucks. The fixed costs of the trucks are clearly smaller than in either of the first or third scenario. This is due to better utilization during the summer months. The container trucks are also able to make more trips during each day and due to this reason the fixed costs of the chippers is also smaller. The train contains 60 containers, as well as 60 additional containers to make sure that the satellite operations perform smoothly and do not become a bottleneck for the train.

In the third scenario only traditional trucks are used but the additional 200 GWh come from the satellite terminal using a train consisting of 45, 52 m^3 steel containers. These type of trains are used Sweden and would also be easy to implement in Finland as well. This scenario has smaller costs than the first scenario, but has higher costs than the multimodal scenario.

Even if the trucks and train could not be used in summer months in other operations, the multimodal supply chain is still the most competitive one. The cost impact is about 0,6 € / MWh, which increases the costs by 7 %.

DISCUSSION AND CONCLUSIONS

The supply chains for wood chips are relatively new and there still exists large potential for improvement (Laitila et al. 2010). Individual power plants need to make the strategic decision and choose the supply chain how to transport the appropriate fuel (Ihalainen and Niskanen 2010). Simulations are one possible tool to make these decisions (Power 2002). In this study a simulation model was constructed which was used to estimate the cost structures of various kinds of supply chains. The model was applied to estimate how a potential power plant expansion could be fueled. The convenient solution of simply increasing the amount of traditional trucks and expanding the collection range has clearly the highest costs. This might not be the case in all scenarios as competition has a high impact on the collection range. In Finland most power plants are located near major cities in Southern Finland which already are highly competitive in regards of wood collection. The satellite terminal can be located in more Northern Finland where competition is smaller and the collection can be conducted much closer to the actual terminal. A multimodal

supply chain also benefits from the possibility of using the trucks and containers in different operations during the summer months. In addition to this the container trucks are able to make more trips during the day, which decreases the relative fixed costs of the operations.

The main research question was: *"What are the main cost components of wood chip supply chains?"* Fixed costs for both of the truck as well as the chipper are the highest individual cost components in wood chip supply chains. If traditional trucks are used, the operational policies need to be efficient as they account to 80% of the costs of the supply chain. Even with supply chains with a satellite terminal these same items have the highest impact, but they are clearly smaller than in the extended, traditional truck supply chain. The costs related to the satellite terminal (train and additional satellite costs) are roughly 10% of the total costs.

Managers responsible for supply chain operations at power plants should carefully consider about using trains in their own supply chains. Due to high competition in the main collection areas (Southern Finland), the trucks need to gather the fuel much further than is possible in more Northern Finland. Similar supply chains are already being used in other countries so even without investing into specialized containers it is possible to decrease the total costs of the system by utilizing trains. The managers also need to increase the utilization of the traditional trucks if they want to have smaller cost structures. Many power plants have only one proper emptying area, which might cause excessive delays in the supply chain. Trains tend to have higher capacity, but a rail link needs to be constructed. Finally managers should look into new technological innovations. The innovations are becoming more prevalent in the industry and can provide advantages to the organization successfully implementing these.

This study has some limitations. The overtime of the drivers is simplified while in reality the drivers need to balance their working hours. This has a high impact on the results as fixed costs are high in all of the scenarios. Secondly, the model does not take into account more specific operational policies. The terminals operate in a simplified fashion, which might cause excessive delays in the model. The model also excludes other potential supply chain modes like barges, which are being used in these type of supply chains.

Further avenues of research include modifying the model to be chipper centric. Chippers work with two trucks while in the model it is assumed that a chipper is always available. The model should also include barge operations as they are being used both in Central Europe as well as in Finland. It would also be interesting to see whether a single train could provide material for multiple plants or how the train could collect wood chips from multiple terminals. This would allow studying more complex supply chain structures.

REFERENCES

Alakangas,E. (2000), "Properties of fuels used in Finland", (in Finnish, Suomessa käytettyjen polttoaineiden ominaisuuksia), *VTT Research notes 2045,* Finland: Valtion teknillinen tutkimuskeskus.

BioEnergia (2012), "Price of fuels", (in Finnish, Polttoaineen hinnat), Bio-energia 2 / 2012, Finland: Puuenergia ry.

Chum, H., A. Faaij, J. Moreira, G. Berndes, P. Dhamija, H. Dong, B. Gabrielle, A. Goss Eng, W. Lucht, M. Mapako, O. Masera Cerutti, T. McIntyre, T. Minowa, K. Pingoud, 2011: Bioenergy. In IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation [O. Edenhofer, R. Pichs-Madruga, Y. Sokona, K. Seyboth, P. Matschoss, S. Kadner, T. Zwickel, P. Eickemeier, G. Hansen, S. Schlomer, C. von Stechow (eds)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Ihalainen, T., and Niskanen, A. (2010), "The impact of cost factors in the bioenergy production value chain", (in Finnish, Kustannustekijöiden vaikutukset bioenergian tuotannon arvoketjuissa), *Working Papers of the Finnish Forest Research Institute 166*, Finland: Metsäntutkimuslaitos

Jalkanen, P., Laitinen, E., Kimari, R., Kuusisto, R., Laurikka, H., Ojala, J., Peuranen, E., Seppänen, Ar., Silfverber, L., Soveri, U-R., Wilde, J, and Rekola, A. (2008), "The long term climate and energy strategy", (in Finnish, Pitkän aikavälin ilmasto- ja energiastrategia), [e-publication], Ympäristöministeriön raportteja 19/2008, [referred: 17.5.2012], Access method: http://www.ymparisto.fi/download.asp?contentid=86191&lan=fi

Karttunen, K., Föhr, J., Ranta, T., Palojärvi, K. and Korpilahti, A. (2012), "The transportation fleet for wood chip and peat 2010", (in Finnish, Puupolttoaineiden ja polttoturpeen kuljetuskalusto 2010), Metsätehon tuloskalvosarja 2 / 2012, [e-document], From: http://www.metsateho.fi/files/metsateho/Tuloskalvosarja/Tuloskalvosarja_2012_02_Puupolttoaineiden_ja_polttoturpeen_kuljetuskalusto_ak_ym.pdf, [Retrieved May 5th, 2012]

Karttunen, K., Föhr. J. and Ranta, T. (2010), "Energywood from South Savo", (in Finnish, Energiapuuta Etelä-Savosta), *Lappeenranta University of Technology, LUT Energy, Research report 7*, Finland: Digipaino.

Korpinen, O.-J., Jäppinen, E. and Ranta, T. (2012),. "Advanced GIS-based model for estimating power plant's supply logistics in a region with intense competition of forest fuel". World Bioenergy 2012. 29–31 May 2012, in Jönköping, Sweden.

Laitila, J., Heikkilä, J. & Anttila, P. 2010. Harvesting alternatives, accumulation and procurement cost of small-diameter thinning wood for fuel in Central-Finland. *Silva Fennica* 44(3): 465-480.

Official Statistics of Finland (2012), "Energy supply and consumption" [e-publication], ISSN=1799-7976. Helsinki: Statistics Finland [referred: 17.5.2012], Access method: http://www.stat.fi/til/ehk/index_en.html.

Power, D.J. (2002), *Decision Support Systems: concepts and resources for managers*, USA: Quorum Books

Ylitalo, E. (2012), "Energy usage of wood in 2011", (in Finnish, Puun energiakäyttö 2011), *Metsätilastotiedote 16/2012*, Finland: Metsäntutkimuslaitos.

MULTIVARIATE INVENTORY AND SELECTED PROBLEMS USING MULTI STAGE MONTE CARLO OPTIMIZATION

William Conley
Austin E. Cofrin School of Business
University of Wisconsin at Green Bay
Green Bay, Wisconsin 54311-7001
U.S.A.
Conleyw@uwgb.edu

KEYWORDS
Inventory optimization, finite replenishment rate, economic order quantity models, additional problems, Monte Carlo, nonlinear

ABSTRACT

The multi stage Monte Carlo optimization (MSMCO) solution technique is used to solve a finite replenishment rate constrained inventory problem and, additionally, one in combination with economic order quantity ones. Then, some general purpose optimization problems are solved along with a nonlinear system of equations. A discussion on the wide variety of inventory problems in existence, along with the versatility of MSMCO, is presented. A partial geometric and statistical review of the multi stage technique is also mentioned along with the five examples presented in detail. Then a listing of various other applications areas for multi stage (MSMCO) is presented along with a section on the "Central Limit Theorem of Optimization Idea."

INTRODUCTION

There are probably as many multivariate inventory problems as there are companies and businesses in the world. Every business has inventory problems unique to its organization. However, two common ones are the so called economic order quantity problem, where delivery of the ordered quantity is assumed to be instantaneous and the finite replenishment rate problem where the ordered quantity arrives more slowly and/or is produced internally in the business at a set rate.

The first example presented here will be a combination of these two classic models. That will be followed by a 20 variable finite replenishment rate problem. Then two additional optimization problems of 25 and 30 variables will be presented. The final example will be a system of nonlinear equations that will be transformed into an optimization problem and solved by minimizing the function that is the sum of the absolute values of the differences between the left and right hand side of each equation.

Then the multi stage Monte Carlo approach will be mentioned in general terms in hopes of starting a discussion about a "central limit theorem of optimization."

SOME PRELIMINARIES

Let us have a brief review of the economic order quantity (EOQ) model and the finite replenishment rate inventory model.

First, the EOQ model attempts to balance the holding cost and ordering cost for an inventory item that a company needs to buy in bulk. As an example, suppose that a company needs 1,000,000 units of part A in the coming year. A unit of A costs $23 to store for a year in its warehouse and the ordering cost is $175 per order. The orders are assumed to arrive instantaneously after they have been placed. Therefore, Cost=holding cost + ordering cost. Cost = 23 Q/2. + 175·1000000/Q, where Q is the quantity ordered each time an order is placed. Dividing by 2 assumes a uniform usage rate.

However, with the finite replenishment rate model the units ordered do not arrive immediately. Therefore the holding cost is not so high. So using the EOQ model example numbers, let us assume that the supplier can produce 3,000,000 units a year at a uniform rate. Therefore, with the finite replenishment rate model the cost is lowered by the discount factor in parentheses

$$\text{Cost} = 23\left(1-\frac{1000000}{3000000}\right) Q/2 + 175.1000000/Q$$

A COMBINATION INVENTORY PROBLEM

Two standard inventory problems are the finite replenishment rate and economic order quantity problems where the reorder amount does not arrive immediately. The finite replenishment

rate units can be either produced internally for the company or an outside supplier delivers it at a certain rate. This results in the amount being stored for use to be smaller than with the regular economic order quantity model where it is assumed that the amount of an item ordered for inventory arrives instantaneously.

Consider Company A's following inventory problem. Items 1 through 5 follow the finite replenishment rate model with the following parameters:

for i=1, 2, 3, 4 and 5

Item	Amount needed in coming time period	Replenishment rate in coming time period	Ordering cost per order	Holding cost per unit per time period	Volume per unit	Weight per item
i	$30{,}000+5000_i$	$115{,}000+10{,}000_i$	$\$260-\$10_i$	$\$100+\$15_i$ cubic meters	15+i cubic meters	$25-1.5_i$ kilograms

However, inventory items 6-15 follow the standard economic order quantity (EOQ) model with the following parameters for i = 6, 7, 8, … 15.

Item	Amount needed in coming time period	Ordering cost per order	Holding cost per unit per time period	Volume per unit	Weight per unit
i	$20000 + 4000_i$	$\$240 + \$10_i$	$\$120 + \$3_i$	12 + i cubic meters	15 + i kilograms

Also, these 15 inventory items are stored in a company warehouse before use. It has manageable volume and weight limits of 5,000,000 cubic meters and 4,500,000 kilograms. Also, management has decided that not more than 5000 units of any particular item will be ordered at once.

So we seek to minimize

$$C = \sum_{i=1}^{5} (100 + 15_i)\, X_i \cdot (1 - (30{,}000 + 5{,}000_i)/(115000 + (10000_i))/2 + (260\text{-}10_i)\, (30000 + 5000_i)/x_i)$$

$$+ \sum_{i=6}^{15} (120 + 3_i)\, x_i/2 + (\$240 + \$10_i)\, (20000 + 4000_i)/x_i$$

subject to $x_i \geq 1$ for i=1, 2, 3, …15

$$\sum_{i=1}^{15} (15+i)x_i \leq 5{,}000{,}000 \text{ cubic meters}$$

and $\sum_{i=1}^{15} (25\text{-}15_i)\, x_i \leq 4{,}500{,}000$ Kgs (kilograms) and $1 \leq x_i \leq 5000$

with a multi stage Monte Carlo optimization (MSMCO) simulation optimization technique program adjusted for a 27 stage simulation (drawing 100,000 feasible solutions at each stage) always following the trail of better and better answers (less cost in this case).

Three runs of this program produced very similar answers in each case and a minimum overall cost of $1,032,949. The stages printout plus the three answers are given here.

The answers are in this pattern:

X_1	X_2	X_3	X_4	X_5
X_6	X_7	X_8	X_9	X_{10}
X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
460	458	455	450	445
437	459	481	501	522
541	560	579	597	615
460	458	455	451	445
437	459	481	501	522
541	560	579	597	615
460	458	455	450	445
437	459	481	501	522
541	560	579	597	615

Stage Number	Function Value
1	1466562
2	1277410
3	1210141
4	1118890
5	1083690
6	1061615
7	1056654
8	1048582

9	1039570
10	1036179
11	1034778
12	1033748
13	1033442
14	1033146
15	1033092
16	1033011
17	1032975
18	1032961
19	1032955
20	1032952
21	1032950
22	1032949
23	1032949
24	1032949
25	1032949
26	1032949
27	1032949

This should give the company at least a start on control of its 15 inventory costs under the given constraints.

A 20 VARIABLE PROBLEM

Company B is ordering 20 inventory items in bulk for storage in its warehouse that has a manageable weight constraint of 350,000 kilograms. All 20 of the items follow the finite replenishment rate model with the following parameters for i=1, 2, 3, … 20.

Item	Amount needed in coming year	Replenishment rates in coming year	Ordering costs per order	Holding cost per item per year	Weight per item
i	$15000 + 3000_i$	$74000 + 1000_i$	$\$150 - \$2_i$	$\$50 + \$_i$ per year	3 + 1.5i Kg

Therefore, we seek to minimize:

$$C = \sum_{i=1}^{20} (50+ i)\, X_i\, ((1-(15000 + 3000)/(7400 + 1000_i))/2.$$

$$+ (150 - 2_i)(15000 + 3000_i/X_i)$$

subject to all $X_i \geq 1$

and $\sum_{i=1}^{20} (3 + 1.5_i)\, X_i \leq 350{,}000$ kg

with an 18 stage multi stage Monte Carlo optimization (MSMCO) computer run drawing 100,000 potential solutions at each stage. A minimum cost of $38,963,616 is arrived at in stage 18. The stage values and amount of each inventory item to order are given as follows:

Stage Number	Function Value
1	48320988
2	43892604
3	41021812
4	39972388
5	39371372
6	39141184
7	39021220
8	38985252
9	38972832
10	38965924
11	38964144
12	38963808
13	38963664
14	38963686
15	38968624
16	38963620
17	38963620
18	38963616

$X_1 = 151$	$X_2 = 217$	$X_3 = 265$	$X_4 = 307$
$X_5 = 354$	$X_6 = 387$	$X_7 = 425$	$X_8 = 459$
$X_9 = 491$	$X_{10} = 530$	$X_{11} = 563$	$X_{12} = 602$
$X_{13} = 643$	$X_{14} = 672$	$X_{15} = 709$	$X_{16} = 756$
$X_{17} = 795$	$X_{18} = 851$	$X_{19} = 904$	$X_{20} = 966$

Therefore, $X_1 = 151$ of item one, $X_2 = 217$ of item 2 through $X_{20} = 966$ of item 20 should be reordered to obtain the minimum inventory cost of $38,963,615.

A 25 VARIABLE PROFIT PROBLEM

A consulting company helps clients by solving their optimization problems using multi stage (MSMCO). One of their clients emails them the following optimization problem:

Maximize PF = Rev – C = $[\sum_{i=1}^{25}(1400 + 15i) -$

$\sum_{i=1}^{25}.5(100 - 2_i)\, X_i + \sum_{i=1}^{25}(1000 + 10_i)(1.5 + .1_i)/X_i +$

$\sum_{i=1}^{25}(300 + 5_i)X_i]$

Subject to $X_i \geq 1_i = 1, 2, \ldots 25$

and $\sum_{i=1}^{25}(40 - i)\, X_i \leq 50000$

A 26 stage MSMCO simulation produces the solution in the 26th stage (drawing 100000 sample answers in each stage) of:

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	257
174	411	368	980	776

Therefore, X_1 through X_{19} are all one and $X_{20} = 257$, $X_{21} = 174$, $X_{22} = 411$, $X_{23} = 368$, $X_{24} = 980$, and $X_{25} = 776$ with a maximum profit of 6,595,436 Euro dollars. The stage values follow here.

Stage Number	Function Value
1	4550813
2	4550813
3	4643262
4	4644004
5	4670466
6	4907768
7	5116341
8	5448596
9	5662196
10	5794880
11	5944903
12	6154818
13	6328525
14	6387532
15	6443060
16	6470974
17	6488138
18	6508527
19	6531430
20	6545834
21	6564394
22	6595436
23	6595436
24	6595436
25	6595436
26	6595436

A 30 VARIABLE PROBLEM

The consulting company is then requested by another client to solve its problem.

Maximize PF = REV – C where REV = $\sum_{i=1}^{30}(89 - 2_i)X_i$

$C = \sum_{i=1}^{30}(17 + 2_i)X_i^{\,1.0 - .01i}$

Subject to $X_i \geq 0 = i = 1,2,3, \ldots 30$

and $\sum_{i=1}^{30}(20 - .25_i)X_i \leq$ 7,000 cubic meters with

$\sum_{i=1}^{30}(10 \text{ Kgs} - .1_i)\, X_i \leq 2500$ Kgs

A 23 stage Multi Stage Monte Carlo Optimization (MSMCO) drawing 200,000 sample answers at each stage produced the stage 23 optimal of $X_1 = 152$, $X_2 = 114$, $X_3 = 50$, $X_4 = 25$, $X_5 = 18$ and X_6 through X_{30} are all zero.

The stages printout is:

Stages	Function Value
1	375380
2	375380
3	386511
4	386511
5	387825
6	401659
7	404660
8	405322
9	421612
10	442874
11	453405
12	487604
13	558794
14	581633
15	609189
16	627851
17	642732
18	653799

19	660794
20	669364
21	677408
22	685624
23	685624 dollars

One of the criticisms of regular Monte Carlo and Multi Stage Monte Carlo Optimization can be trying to prove that the answer is optimal. Although business people are usually just looking for a cost effective approximate solution, let us address this criticism by solving a system of equations where the answer checks automatically.

A NONLINEAR SYSTEM OF EQUATIONS

Let us use MSMCO to find an all whole number solution in the range $0 \leq X_i \leq 125$ for i = 1, 2, 3, 4, 5, 6 and 7 to the system.

$$X_1^3 + X_3X_4 = 92413$$

$$X_2X_3X_5 + X_6X_7 = 557007$$

$$X_3X_4^2 + X_1X_2 = 77123$$

$$X_2X_4X_7 + X_5X_6 = 108200$$

$$X_5X_6X_7 + X_4X_5 = 338734$$

$$X_7X_6^2 + X_1X_7 = 149292$$

$$X_2X_6X_7 + X_3X_4 = 421201$$

$$X_6X_7^2 + X_1X_2 = 371946$$

$$X_1X_2X_3 + X_5X_7 = 283191$$

$$X_4X_5X_6 + X_1X_3 = 201079$$

by minimizing the sum of the absolute values of the differences between the left and right hand side of each equation with an eight stage simulation drawing 100,000 sample answers at each stage. The following two runs of the program produced the same answer (45, 111, 56, 23, 89, 97, and 39) in each case.

One run $\bar{Z}$ = 100000 at each stage

Stage Number	M Total Error
1	282062.0
2	169683.0
3	91912.0
4	49364.0
5	12274.0
6	0.0
7	0.0
8	0.0

$X_1 = 45$
$X_2 = 111$
$X_3 = 56$
$X_4 = 23$
$X_5 = 89$
$X_6 = 97$
$X_7 = 39$

Another run

Stage Number	M Total Error
1	242578.0
2	146563.0
3	81851.0
4	28997.0
5	24690.0
6	0.0
7	0.0
8	0.0

$X_1 = 45$
$X_2 = 111$
$X_3 = 56$
$X_4 = 23$
$X_5 = 89$
$X_6 = 97$
$X_7 = 39$

Notice that the total error of 0.0 in stages 6-8 means that the answer checks in all 10 equations.

A GEOMETRIC AND STATISTICAL VIEW OF MSMCO

Every optimization problem has a sampling distribution consisting of all of the feasible solutions. What makes multi stage Monte Carlo optimization (MSMCO) fairly effective is that these feasible solutions are interconnected in a way that the nearly optimal solutions are generally close to the true optimal which one is generally pursuing. Therefore, each stage of an MSMCO optimization simulation tends to follow a trail of better and better answers as the geometric shapes (please see Figures 1 & 2) get smaller and smaller as they track and subsequently find the optimal solution or a near

Figure 1: A Five Stage MSMCO Simulation Crossing the Sampling Distribution to the minimum

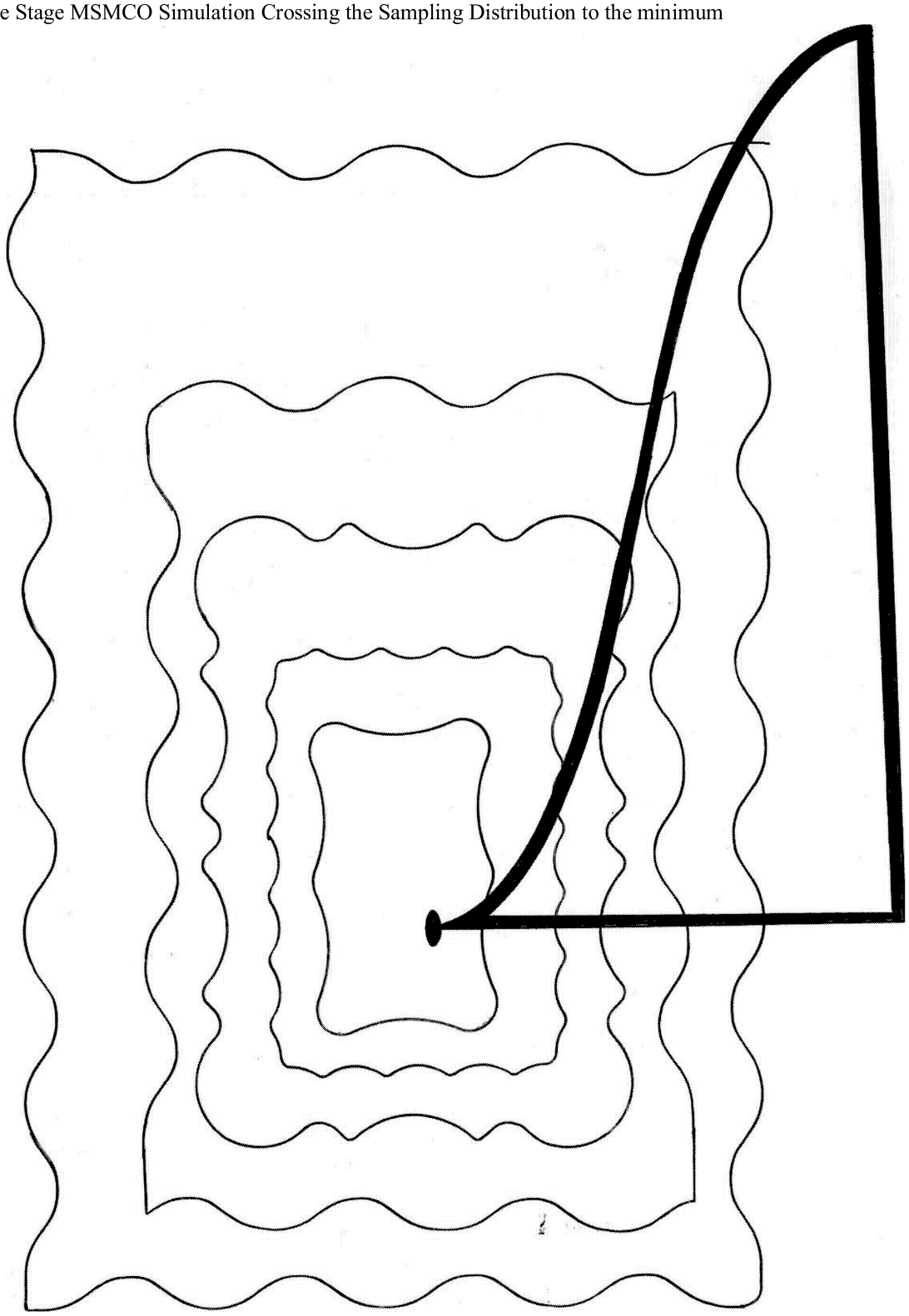
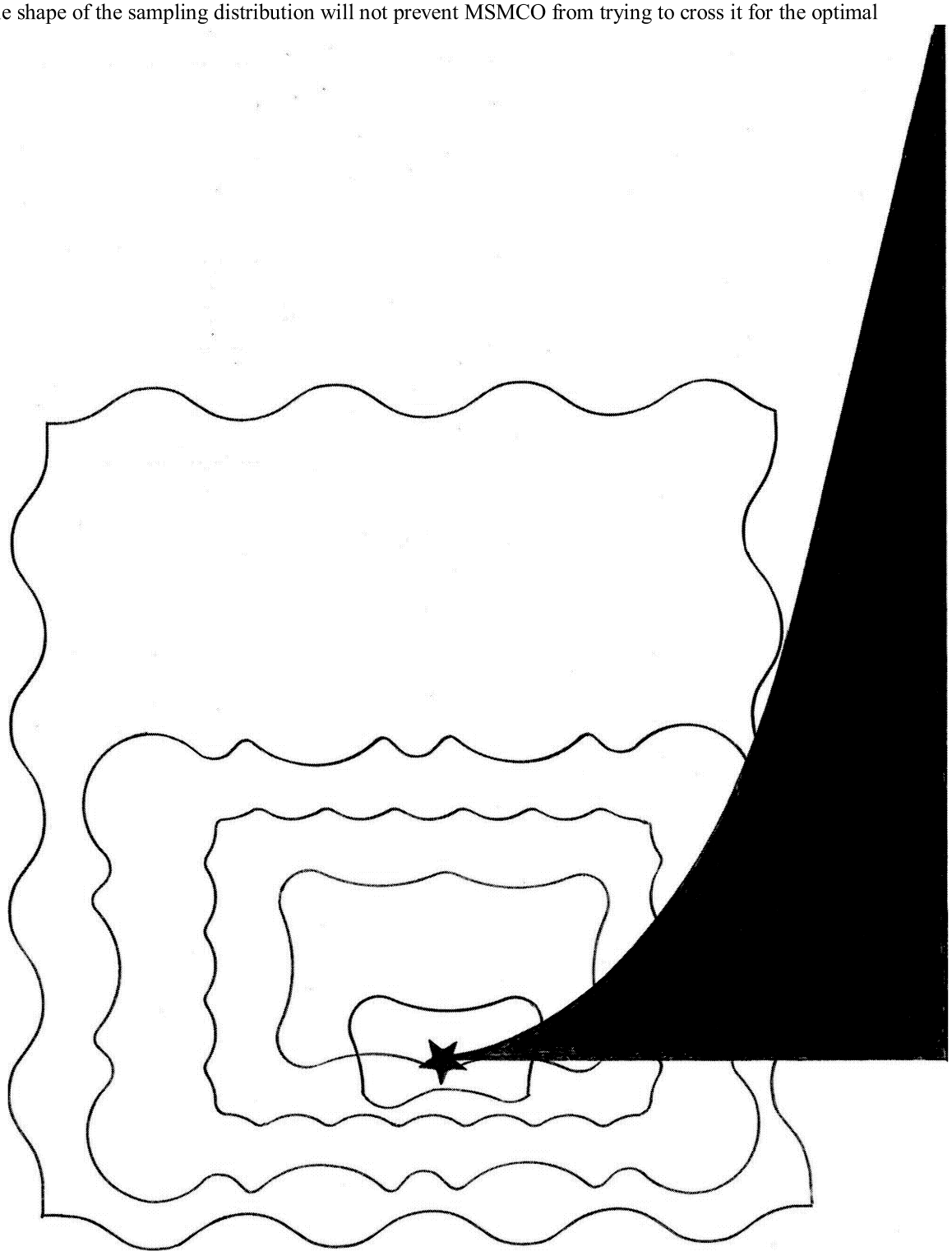


Figure 2: The shape of the sampling distribution will not prevent MSMCO from trying to cross it for the optimal



approximation. All of the examples presented here used 15, 20, 25, 30 and 7 dimensional "rectangles" to close in and find the optimal solution or a useful approximation. The so called n dimensional rectangles were used because they are easiest to program in the computer. However, as Figures 1 and 2 attempt to partially illustrate, any geometric "shapes" can be used as long as they get progressively smaller with each stage as they cross the sampling distribution of the feasible solutions to the optimal solution region.

Naturally, multivariate optimization problems become more difficult to solve as the number of variables (or dimensions) increases. However, with multi stage, large scale problems can be solved or at least approximate solutions can be produced with our 21st century powerful computers.

Many authors such as (Dongarra, 2011), (Kopytoff, 2011) and (Krantz and Swartz, 2011) have studied these computer speeds which seemed to be increasing at an enormous rate. Super computers can now approach speeds of 10^{15} calculations per second. Also, desk top and smaller computers (although not capable of 10^{15} calculations per second yet) are becoming very fast as their prices decline at the same time. These developments make multi stage Monte Carlo optimization (MSMCO) and other simulation optimization techniques increasingly viable and useful. The pioneering work done by (Dantzig, 1963) and (Dantzig and Ramser, 1959) on linear programming is still important. However, we can address nonlinear programming (Conley, 2012) and begin to take Jan Tinbergen's warning about not linearizing essential nonlinear problems, seriously (Tinbergen, 1978).

The trails across the sampling distributions of optimization problems are there (Figures 1 & 2). We can now follow them quickly and efficiently in the 21st century, even though many 20th century approaches (Mizrahi and Sullivan, 1993) remain useful.

The world has been searching for a fast, inexpensive way to do calculation for 5,000 years (Ifrah, 2000). The invention of the abacus those many thousands of years ago (still in use today in some places) shows how important our high speed, relatively inexpensive, computers can become. Professor Jack Dongarra (Dongarra, 2011) gives us an idea of how far we have progressed in speed of calculation by compiling a list of the 500 fastest computers in the world.

APPLICATION AREAS FOR MSMCO

Multi stage Monte Carlo optimization (MSMCO) is meant to be a general purpose optimization technique that can be applied to a wide variety of problems that occur in engineering, industry and science. Some examples of this are provided in electronics by (Conley, 1984), correlation analysis (Conley, 2007), ((Hayter, 2002) provides an engineering review of correlation), ecology (Conley, 2008), engineering (Szarkowicz, 1995) and general optimization (Wong, 1996). Also (Clark, 1984) addresses policy analysis from an operations research perspective using multi stage Monte Carlo optimization simulations.

THE CENTRAL LIMIT THEOREM OF OPTIMIZATION IDEA

After many thousands of runs of MSMCO on problems from one to one million variables (Conley, 2012) it has been fairly obvious that the MSMCO approximations are piling up around the true optimal on most difficult large scale nonlinear problems.

An easy illustration of this idea can be found on pages 147-159 in Chapter 7 of (Conley, 1984). Dozens of runs of MSMCO on a ten variable nonlinear optimization problem failed to produce the true optimal. However, the mean, median and modal coordinate wise averages of these "failed" solution attempts all yield the exact optimal solution.

Therefore (for larger and larger problems), continually averaging the approximate solutions to the problem at hand will speed the convergence of the MSMCO process to the optimal solution (or at least yield superior approximations).
It may even be possible for theoretical mathematicians to prove this idea under certain "regularity" conditions.

If one is thinking that this idea may not work if the optimal is at a "corner point" or along the "edge" of a constrained feasible solution region, that is a good question. However, the constraints will sometimes slow down the convergence, but not stop it in most cases. Again, in the future, theoretical mathematicians may be able to offer more precision to this idea through a series of theorems and proofs.

Another example of this averaging of the initial MSMCO runs closing in on and finding the true optimal is in (Conley, 1991). Two famous test problems that stood unsolved for years, called Problem 30 and Problem 32 in the mathematics literature (Lawler, Leenstra, Rinnooy Kan and Shmoys, 1985) of size 150 and 200 variables were tried with MSMCO and the modal averaging approach (mode being the most frequently occurring value of X_i). The true optimal was produced in both cases.

CONCLUSION

Fifteen variable and 20 variable inventory problems were presented and solved with MSMCO. The 15 variable problem involved a combination of the finite replenishment rate and economic order quantity approach to inventory management. The 20 variable problem was strictly a finite replenishment rate problem.

Then, 25 and 30 variable general maximization problems were presented. These were followed by using MSMCO to solve a nonlinear system of equations.

The geometry and statistics of the MSMCO approach were addressed, along with past and possibly future application areas for this versatile solution technique. Business people, engineers and scientists can make use of multi stage Monte Carlo optimization (MSMCO) now in our 21st century more than ever before. A good application area for MSMCO remains inventory problems because the industrial world has thousands of different inventory optimization situations.

REFERENCES

Clark, K. 1984. "Operations Research and Policy Problems: Simplified Computer Optimization Techniques for policy Analysis," *Policy and Information*, Vol. 8, No. 1, pp. 63-67.

Conley, W. 1984. Computer Optimization Techniques, Revised Edition, Petrocelli Books, Inc., Princeton, NJ.

Conley, W. 1984. "An RC Switching Network with Fixed Components," *International Journal of Electronics*, Vol. 3, pp. 433-438.

Conley, W. 1991. "Programming an Automated Punch or Drill," *International Journal of Systems Science*, Vol. 22 No. 11, pp. 2039-2056

Conley, W. 2007. "Simulation Optimization and Correlation with multi Stage Monte Carlo Optimization," *International Journal of Systems Science*, Vol. 38, No. 12, pp. 1013-1019.

Conley, W. 2008. "Ecological Optimization of Pollution Control Equipment and Planning From a Simulation Perspective," *International Journal of Systems Science*, Vol. 39, No. 1, pp. 1-7.

Dantzig, G. and Ramser, J. 1959. "The Truck Dispatch Problem," *Management Science*, Vol. 6, pp. 80-91.

Dantzig, G. 1963. Linear Programming and Extensions. Princeton University Press: Princeton, NJ.

Dongarra, J. June 2011. Personal communication.

Hayter, A. 2002. Probability and Statistics for Engineers and Scientists, 2nd Edition, Duxubry, Pacific Grove, CA.

Ifrah , G. 2000. The Universal History of Numbers. John Wiley & Sons, New York.

Kopytoff, V. 2011. "Japanese 'K' Computer is Ranked Most Powerful." New York Times. New York, June 20, 2011, p. B6.

Krantz, M. and Swartz, J. 2011. "IBM Shows Secret to Corporate Longevity: Big Blue Survives 100 Years by Always Adapting." USA Today. June 16, 2011, p. B1, Arlington, Virginia.

Lawler, E., Leenstra, J., Rinnooy Kan, A., and Shmoys, D. 1985. The Traveling Salesman Problem. John Wiley and Sons, New York.

Mizrahi, A and Sullivan, M. 1993. Mathematics for Business Science and Social Sciences. John Wiley & Sons, New York.

Szarkowicz, D. 1995. "Investigating the Brachistochrone with a Multi Stage Monte Carlo Method," *International Journal of Systems Science*, Vol. 26, pp. 233-243.

Tinbergen, J. 1978. "Effects of Computerization of Research," *Compstat 1978.* Physica Verlag Wurzburg and Vienna, pp. 10-20.

Wong, J. 1996. "A Note of Optimization in Integer," *International Journal of Mathematics Education in Science and Technology*, Vol. 27, pp. 865-874.

BIOGRAPHY

WILLIAM CONLEY is Professor Emeritus at the University of Wisconsin-Green Bay, U.S.A. He is the author of five books and more than 200 publications world-wide. He has a B.A., MA., M.Sc and Ph.D in Mathematics. He is a senior member of the Society for Computer Simulation, a fellow in the Institution of Electronic and Telecommunication Engineers, and a member of the American Chemical Society.

LOGISTICS SIMULATION

A FRAMEWORK FOR DISCRETE-EVENT SIMULATION OF PICK-UP-AND-DELIVERY PROBLEMS

Anna Syberfeldt
University of Skövde
Virtual Systems Research Center
Skövde, Sweden
anna.syberfeldt@his.se

KEYWORDS
Discrete-event simulation, pick-up-and delivery, waste collection.

ABSTRACT

So called pick-up-and-delivery problems are encountered in everyday life – for example parcel transportation, dial-a-ride services, and collection of mail from mail boxes. For modeling these problems, discrete-event simulation is a powerful technique that is able to handle complexties in a transportation system. In this paper, a new framework for discrete-event simulation of pick-up-and-deliviery problems are proposed. The framework includes a general and extensible library of simulation modeling components applicable to different types of pick-up-and-delivery problems. For evaluation, the proposed framework is used in the simulation-optimization on a real-world pick-up-and-delivery problem of waste collection.

INTRODUCTION

A common problem within the area of transportation systems is the pickup-and-delivery problem (PDP). The PDP basically involves a set of transportation requests and a fleet of vehicles. Generally, each request specifies the size of the goods to be transported and the pickup and delivery points. Each vehicle in the fleet has a given capacity, a start location, and an end location. The objective is to satisfy all transportation requests while at the same time minimize the total cost, which may include carbon dioxid emission or transportation times, amongst others. There are plenty of examples of PDPs encountered in everyday life – for example parcel transportation, dial-a-ride services, and collection of mail from mail boxes.

Real-world PDPs are often highly complex as a consequence of combinatorial relationships, uncertainty factors, and non-linearities in the transportation system. When traditional, analytical methods are used to model these problems, simplifications must usually be done since these methods do not provide means for modeling complex behaviour (Al-Nory & Brodsky, 2008). Furthermore, the concepts of order and timing of events are usually not explicit in analytical models, which makes it hard to achieve a correct representation of the system (Al-Nory & Brodsky, 2008). For a complex PDPs, discrete-event simulation is a powerful alternative to analytical methods for modeling the transportation system. Discrete-event simulation provides means for incorporating complex relations using easier methodologies and tools compared to mathematical programming (Law, 2007). A simulation model essentially consists of a set of system states and transitions between these states, which correspond to elements and event sequences in the real system.

The next two sections continue with describing the PDP and discrete-event simulation in further detail. Thereafter, a new framework for simulating PDPs are described, followed by an evaluation of the proposed framework on a real-world PDP problem of garbage collection. The paper ends with a discussion of results and conclusions.

THE PICK-UP-AND-DELIVERY PROBLEM

The general pick-up and delivery problem (GPDP) is about creating routes for carrying out a number of transportation requests using a fleet of vehicles (Savelsbergh & Sol, 1995). Each vehicle has a given maximum capacity, a starting point, and an endpoint. A transport request defines the quantity of goods to be transported and the locations where it is to be picked up and delivered. Each request is served by a single vehicle without intermediate transhipment. The pickup and delivery problem (PDP) is a specialization of GPDP in which all vehicles start and stop at a central depot. Furthermore, for each transport request only one location is defined for pick up and delivery, respectively.

Based on GPDP and PDP further problem specializations can be defined. These problems can be divided into to categories; static and dynamic. A problem is static if all transport requests are known when the routes are constructed (Savelsbergh & Sol, 1995). In a situation where the transport requests may be added while running routes, meaning that routes have to be changed on-the-fly, the problem is instead dynamic. Besides static and dynamic problems, it is also meaningful to distinguish between problems with one operating vehicle (single vehicle) and several operating vehicles (multi-vehicle). Furthermore, problems can also be categorized according to if they involve time constraints or not. As an example, vehicles are often only available during the drivers' working hours, or if transporting people their pickup must be performed at a specific time. Such restrictions are called time windows. Different problem specializations can also be combined into problem formulations such as ”static multi-vehicle pickup and delivery problem with time windows”.

DISCRETE-EVENT SIMULATION

Simulations in event-driven simulation models are driven by a set of events that take place at pre-defined, fixed time points (Banks, 1998). The events are either defined from the start, or generated as a result of other events during the simulation. The events are stored in an event list.

A simulation model contains a set of variables that holds the state of the system, the so-called state variables (Banks, 1998). Which state variables that are needed vary depending on what the simulation will be used for. Different applications require different output parameters, and hence different state variables to calculate the output parameters. Which state variables that are needed in the simulation model to simulate a particular system may vary depending on what the simulation should be used for. In an event-driven simulation model the values of the state variables may be changed only at the times when an event occurs. The opposite of this is so-called continuous models, where the state variables values are defined by mathematical equations and may change continuously over time (Banks, 1998).

The execution of an event-driven simulation model is controlled by the times points of the scheduled event. After an initialisation phase where the simulation clock is reset and the state variables, the statistical variables and the event list are initiated, the simulation cycle is started, that is the simulation main loop (Law & Kelton, 2000). In each iteration, the type of the next event in the event list is determined (ie the event which has the lowest start time). Subsequently, the simulation clock is advanced to the start time for this event and a routine for handling the current event type is run. An event routine first updates the state variables, then the statistical variables and, finally, future events are generated and added to the event list. The last thing that happens in the simulation cycle is to test whether the condition for when the simulation is considered to be completed is satisfied. If it is not, a new simulation cycle is run. If the condition is met, the simulation is terminated, and output parameters are calculated and reported.

In this study, a new framework for discrete-event simulation of PDPs is being prosed. This framework is described in the next section.

A DISCRETE-EVENT SIMULATION FRAMEWORK

This section describes the general and extensible library of simulation modeling components that is proposed for simulating PDPs. The library aims to be generally applicable to different types of PDPs. From an overall perspective, the library is divided into three subsystems: a) simulator, b) simulation data, and c) events. These subsystems are described below.

Simulator

The simulator subsystem (Figure 1) defines two main classes: `Simulator` and `Graph`. Simulator is an abstract class providing a general interface for PDPs. When simulating a specific variant of the general PDP (for example PDP with time windows), a subclass to Simulator is created by the user which defines all functionality needed for the specific implementation at hand. Two common operations are provided through the Simulator interface: Simulate and BuildRoute.

`Simulate` performs a simulation according to the given input parameters and returns the results from it. The input parameters are an ordered list of locations to be visited, the vehicle that is to be used, the vehicle operation hours, vehicle starting and ending points, and whether the simulation should be stopped prematurely if it is found to be invalid (this functionality can be used to save time). Note that in the optimization of multi-vehicle problems, the user must create a list of places to visit for each vehicle and call the simulation function once per vehicle.

`BuildRoute` perform a simulation under the given parameters while building a route object which is then returned. The route object contains a list of locations visited and the corresponding time of the visit. This object is built during the simulation based on subsequent locations visited in the simulation. The object also contains all the results returned from the simulator.

The motivation behind the use of two different functions (Simulate and BuildRoute) is that it is likely that the end user is interested in information about what time the vehicle will arrive at each location, while this information is not interesting for the optimization algorithm. Simulate is therefore intended to be used during the optimization, while BuildRoute is intended to be used afterwards to present detailed information to the user about which time different vehicles are estimated to arrive at each location.

The class Simulator also defines the enumeration type State, which is used during a simulation to determine what state the vehicle in the simulation currently are in. The state ON_THE_ROAD means that the vehicle left the previous location but not yet on the next spot. The state WAITING_TO_BEGIN_SERVICING_LOCATION means that the vehicle has arrived at a location but not yet started the service at the location. The state SERVICING_LOCATION means that the vehicle is currently undertaking the service.

The Simulator class also contains a graph object. The graph stores all the information about locations, distances, and travel times in the form of a neighboring array containing objects of the class Edge. The graph also holds an object of type EdgeDataReader. EdgeDataReader is an interface class used to load the edge data. An EdgeDataReader object is sent to the constructor for the simulator and then used to create the object graph used in Simulator.

Events

The subsystem Events (figure 2) provides functionality for event handling and is used internally in simulations. It includes two classes, `EventList` and `Event`, and the enumeration `EventType`. EventType is used to define an

event type. The events are modeled as Event objects and stored in a list of type EventList.

EventType contains five different types of events. LOCATION_ARRIVAL occurs when a vehicle arrives at a

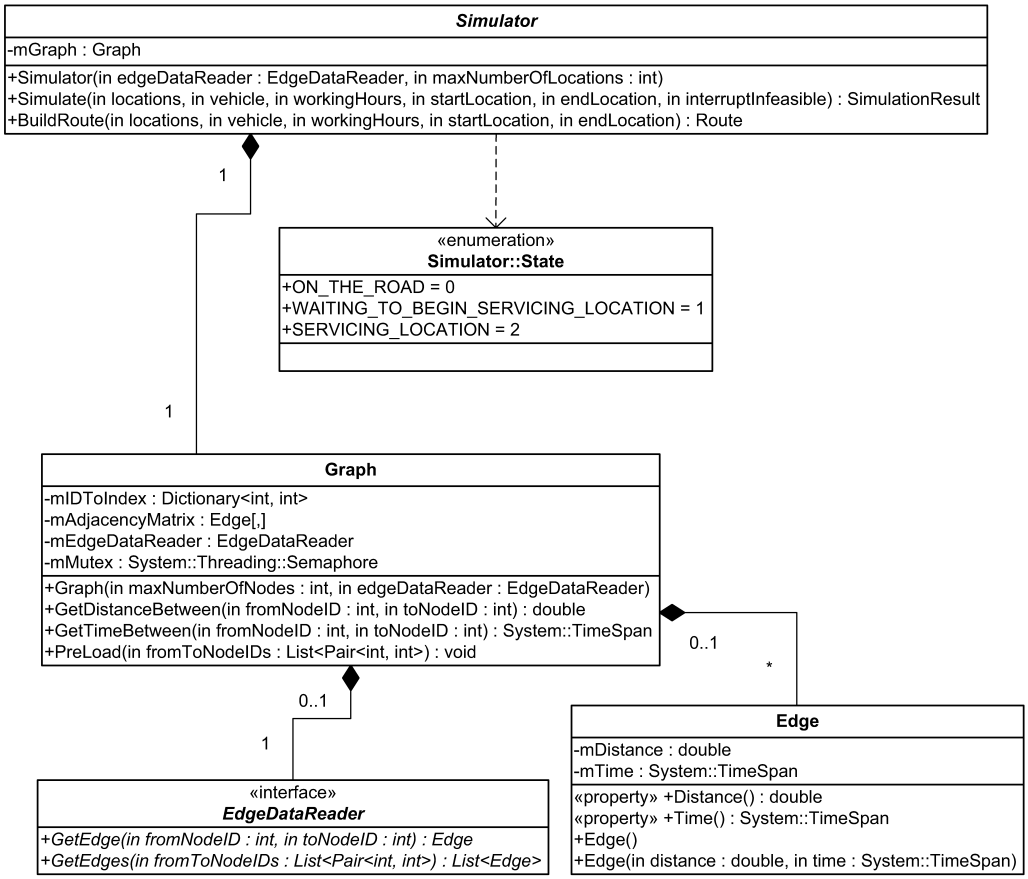


Figure 1: Components of subsystem Simulation

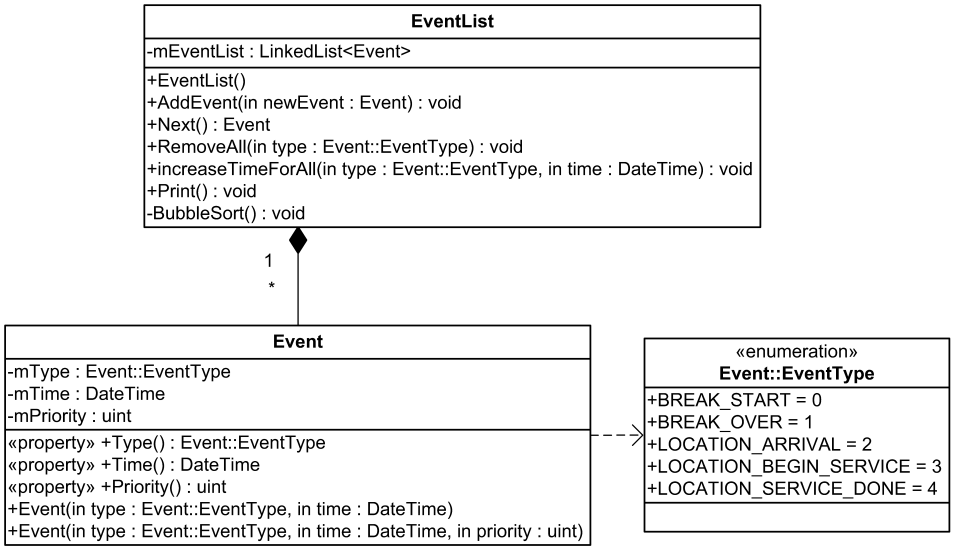


Figure 2: Components of subsystem Events

location. LOCATION_BEGIN_SERVICE occurs when a vehicle has arrived to a location and service is about to be started. LOCATION_SERVICE_DONE occurs when the servicing at a location is complete and the vehicle is ready to go on. BREAK_START occurs at the end of each session except the last. BREAK_OVER occurs at the start of each session except the first.

Each event has three attributes: type, time, and priority. Type is used by the simulator to determine what action should be taken when the event occurs. Time determines when the event occurs, and is used by the EventList in order to keep the events sorted, and also by the simulator in order to manage the simulation clock. Priority is used by the EventList to determine the order of events occurring at the same time.

The events are stored in a linked list in EventList. The list is kept sorted in ascending order according to time of occurrence, and in case two events have the same time of occurrence, according to priority. If two events both have the same time of occurrence and the same priority, the order is undefined. Four operations are possible on the event list: `AddEvent` (receives an event and adds it at the right place), `Next` (removes and returns the first event), `RemoveAll` (removes all events of a specific type), and `IncreaseTimeForAll` (increases the time with a given value for all events of a specific type).

Simulation Data

The subsystem Simulation Data (figure 3) provides classes for all non-common data that is provided for each simulation, and data to be returned from the simulator. This includes classes for representing vehicles, locations, time windows, routes and simulation results.

The class Location is a base class for places to be visited during a simulation. Its attributes are a list of time windows, a unique ID, and time consumption at the location. This is the minimal amount of information required for the simulator to be able to visit a location (the list of time windows may be blank, however). Two subclasses to Location are defined: `PickUpLocation` and `DeliveryLocation`. These subclasses both holds a Goods object. For a simulator to be able to perform more specialised activities at a location, further subclasses to Location need to be created.

The classes Route and SimulationResults are used to keep all data returned from the simulator's BuildRoute and Simulate functions. Route holds a list of pairs of locations and time points, which represent the locations to be visited, the order among them, and at what times. Route also contains a SimulationResult object. SimulationResult holds all information about performance of the system, including total time consumption, distance, travel time, end time, total transported weight, total transported volume, the total time all goods spend in the system, how much total maximum weight is exceeded by, how much total maximum volume has been exceeded by, the number of times a deliviery location has been visited before the corresponding pickup location, number of missed time windows, total delay with respect to time window, the total time deviation from the time window (in both early as the late arrival), and whether the simulation was interrupted prematurely.

The class Vehicle contains all relevant information about the vehicle used. Attributes of a vehicle includes maximum weight, volume, and velocity coefficient. The velocity coefficient is multiplied by the travel time between two locations that is returned from the graph. The purpose of having such coefficient is to allow vehicles of different velocities (for example, bicycles and trucks) to use the same underlying data for travel times between locations.

REAL-WORLD CASE STUDY

To evaluate the proposed framework, it is implemented on a real-world PDP concerning waste collection in a middle-sized municipality in Sweden. In general terms, the overall problem is to pick up waste on a number of predefined sites, and then leave it on landfill sites. The focus of the study is on larger waste containers located at companies and schools, rather than smaller household trash cans. Within the municipality, there are 182 such containers that the waste management organization is responsible for taking care of. Each of these containers hold either combustible waste, compost, or corrugated cardboard.

There is currently one truck dedicated only for collecting garbage from containers and the focus of the study is on this truck. The truck has a limited capacity in terms of both weight and volume, and when this capacity is reached must be emptying. The emptying is made either at a recycling central (compost and corrugated cardboard), or at a heat plant (combustible waste). Since different waste should be deposit at different places, different types of waste cannot be mixed in the truck.

Currently, each container is emptied once or several times during a 14-day period, according to a predefined emptying schedule. The laws specify that 14 days is a maximum limit on how long waste may be stored before it must be taken care of due to hygienic reasons. The 14-days' time limit means that time windows are present in the problem.

Today, the emptying schedule is defined by a transportation planner through a manual process based on domain knowledge. When the schedule is defined, the transportation planner must take into account the driver's working hours and breaks. Defining an efficient schedule is very hard to carry out manually, not at least since it is difficult to estimate how often or when the container needs to be emptied, which means that containers often only half full when the truck will or that they are overcrowded. Every day, the transportation planner provides the driver with a list of containers to be emptied for the specific day. The transportation planner does not specify which routes to take, but this must be decided by the driver itself. Performing a route optimization is very difficult to do manually due to the

high complexity of the transportation network and lack of supporting tools. Even greater difficulty is created by temporal constraints that must

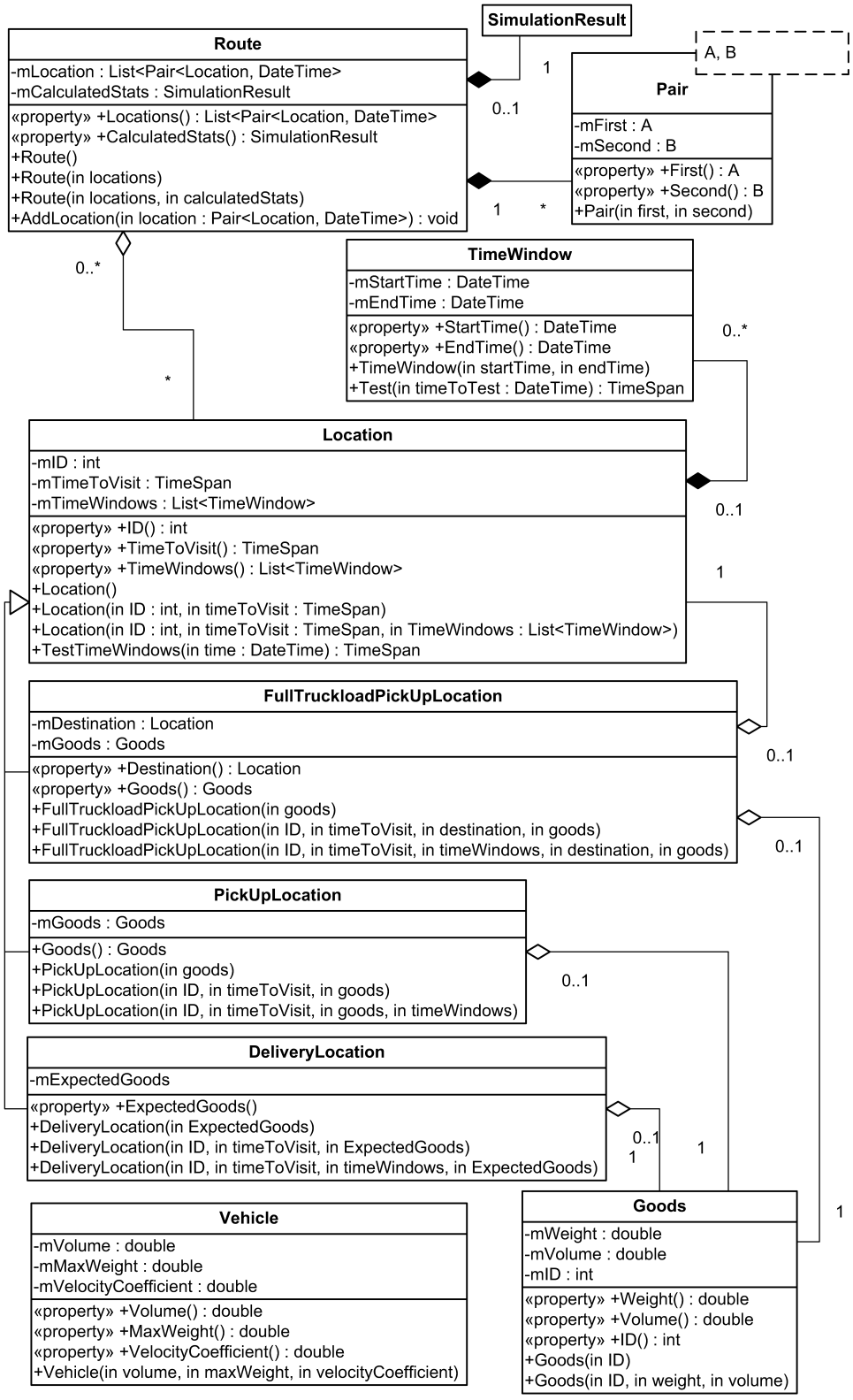


Figure 3: Components of subsystem Simulation Data

be taken into account, which means that some containers preferably should, or should not, be emptied at certain times (for example, containers at day care centers should be avoided when the kids are out playing).

Due to the high complexity of planning the waste collection and optimizing route, the waste management organization has expressed a need to improve the process by using simulation and optimization. The next section presents how this has been realized.

SIMULATION AND OPTIMIZATION

A simulation model of the waste collection procedure was built using the proposed simulation framework. Data for the model was collected during two months in corporation with the waste management organization. Validations of the simulation model indicates that it represents reality very well (average deviation from reality is 7%). For optimizing the problem, an optimization module has been developed and connected to the simulation. Two objectives are considered in the optimization, as specified by the waste management organization: 1) minimal carbon dioxide emissions (basically driving distance), and 2) minimal violations against temporal constraints. These two objectives are conflicting with each other, as considering temporal constraints negatively affects the driving distance.

The difficulty with conflicting optimization objectives is that there is usually no single optimal solution with respect to both objectives. Instead of a single optimum, there is a set of optimal trade-offs between the conflicting objectives, called Pareto-optimal solutions (Deb, 2004). In order to manage multiple objectives, specific optimization algorithms have been suggested. Instead of only seeking a single optimum, these algorithms maintain a set of Pareto-optimal solutions. One of the most efficient algorithms for Pareto optimization is the elitist non-dominated sorting genetic algorithm (NSGA-II) (Deb et. al, 2000). Due to the algorithm's proven efficiency, this is the optimization algorithm selected for the study.

NSGA-II is basically a genetic algorithm extended with features for handling multiple trade-off solutions. In this study, a solution consists of a collection of routes (a specification of which containers to visit a specific day). Like an ordinary genetic algorithm, NSGA-II maintains a population of solutions and refines these solutions through generations. In the refinement process, recombination (crossover) and mutation are being used to create offspring solutions that take part of the next generation. The basic steps involved in evolution are presented below (a complete description of genetic algorithms can be found in Deb 2004).

```
Initialize population
Evaluate the fitness of solutions in the
population
repeat
  Select solutions to reproduce
  Form a new generation of population
through
    crossover and mutation
  Evaluate the new solutions
until terminating condition
```

Contrary to an ordinary genetic algorithm, however, NSGA-II selects solutions for the next generation of the population based on Pareto ranks. Rank 1 includes the Pareto-optimal solutions in the complete population, and rank 2 the Pareto-optimal solutions identified when temporarily discarding all solutions of rank 1, and so on. More specifically, the selection of solutions for the next generation is done from the set R, which is the union of the parent population and the offspring population (both of size N). Pareto-based sorting (also called non-dominted sorting) is applied to R and the next generation of the population is formed by selecting solutions from one of the Pareto fronts at a time. The selection starts with solutions in the best Pareto front, then continues with solutions in the second best front, and so on, until N solutions have been selected. If there are more solutions in the last front than there are remaining to be selected, niching is applied to determine which solutions should be chosen. All the remaining solutions are discarded. The selection procedure is illustrated in Figure 4 (adopted from Deb, 2004).

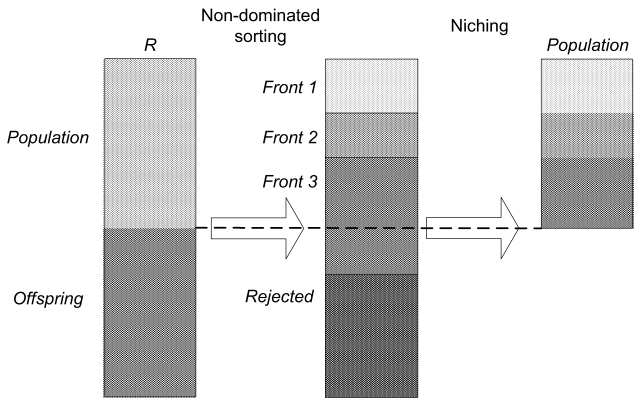


Figure 4: Selection procedure of NSGA-II.

The NSGA-II algorithm has been implemented in an optimization module that is connected to the simulation. The simulation-optimization process is an iterative loop in which the NSGA-II algorithm generates solutions that are evaluated by the simulation (figure 5).

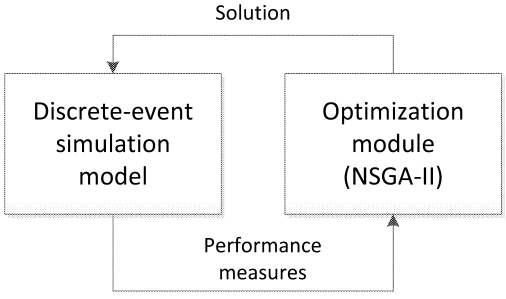


Figure 5: Simulation-optimization loop.

RESULTS

The simulation-optimization was evaluated using different real-world scenarios and the test results were carefully evaluated by the waste management organization. In comparison with solutions created manually by the transportation plannering, the simulation-optimization is able to find solutions with 13-28% lower carbon dioxide emissions (depending on scenario) while maintaining all temporal contraints. About 10 000 simulation runs are needed to reach these results, corresponding to a runtime of approximately 5 hours.

Compared to the manual approach of creating schedules used by the transportation planner, the simulation-optimization implemented has advantages not only with respect to the results, but also with respect to efficiency. Since no manual control or intervention in the simulation-optimization process is necessary, a lot of time and effort are saved for the transportation planner.

CONCLUSIONS

This study describes a new framework for simulating pick-up-and-delivery problems based on the discrete-event simulation paradigm. Compared to traditional, analytical methods, discrete-event simulation is powerful in modeling complex behavior and achieve a close representation of reality. The proposed framework includes a general and extensible library of simulation modeling components applicable to different types of pick-up-and-delivery problems. For evaluation, the proposed framework are used to model a real-world pick-up-and-delivery problem of waste collection. Validations of the simulation model indicates that it represents reality closely with an average error of only 7%.

To improve the waste collection process, the simulation model is connected to an optimization module. In the optimization, two objectives are considered: 1) minimal carbon dioxide emissions, and 2) minimal violations against temporal constraints. The two objectives conflict with each other, as improving performance on one of them deteriorates performance of the other. For handling the conflict, an optimization algorithm utilizing the Pareto concept is being used, called NSGA-II. Results from the study shows that the simulation-optimization is successful in improving the waste collection process, being able to find solutions with 13-28% lower carbon dioxide emissions while maintaining all temporal contraints.

To improve the optimization results further, future work includes studying how domain expert knowledge can be captured and incorporated in the optimization process. A human expert may have extensive knowledge valuable for the optimization, and incorporating this knowledge in the optimization strategy may be a way to obtain faster and more accurate optimization results. Combining the simulation with heuristics based on human knowledge is therefore an interesting topic for the future.

Another aspect that should be considered in future work is to reduce the time consumption of the simulation-optimization procedure. In the real-world case study, the simulation-optimization took about 5 hours to complete on a single PC. If more trucks and pick-up-locations are added to the problem, the runtime will be even longer. It can be noticed that the long runtime is not due to the simulation framework as such, but due to the enormous search space (e.g. a large number optimization iterations is needed), in combination with a heavy road database. Using the road database, which is provided by the Swedish Transport Administration, is necessary in order to obtain information about distances, speed limitations, traffic regulations, etc. A runtime of several hours is too long if the procedure is to be used on a regular basis (e.g. daily). Reducing the time consumption is, however, no easy task as real-world pick-up-and-delivery problems are highly complex. One way to address the problem might be to include a computationally efficient simulation metamodel in the proposed framework. A metamodel is essentially an approximation of the simulation and the underlying data. It has previously been shown that by adopting metamodels, the computational burden of the optimization process can be greatly reduced since the computational cost associated with running a metamodel is negligible compared to a simulation run (Jin et. al 2002). A metamodel could, however, never fully replace a simulation model as it is only an approximation.

REFERENCES

M. Al-Nory and A. Brodsky. “Unifying simulation and optimization of strategic sourcing and transportation”. *Proceedings of the 2008 Winter Simulation Conference*, Miami, Florida, pp. 2616-2624. 2008.

J. Banks. “Handbook of Simulation”. New York: John Wiley & Sons, Inc. 1998.

K. Deb, A. Pratap., S. Agarwal, and T. Meyarivan. “A fast and elitistmulti-objective genetic algorithm NSGA-II”. KanGAL Report 2000001, Indian Institute of Technology Kanpur, India. 2000.

K. Deb. “Multi-objective optimization using evolutionary algorithms”. Chichester: John Wiley & sons, Ltd. 2004.

A.M. Law “Simulation Modeling and Analysis”, Fourth Edition, McGraw-Hill, 2007.

Y. Jin, M. Olhofer, and B. Sendhoof, “A Framework for Evolutionary Optimization With Approximate Fitness Functions”, IEEE Transactions on Evolutionary Computation, Vol.6 No.5 (Oct), 2002, pp. 481-494.

M.P.W Savelsbergh, M. Sol. The general pickup and delivery problem. *Transportation Science*, 29(1), pp. 17-29. 1995.

BIOGRAPHY

ANNA SYBERFELDT is a senior researcher at the University of Skövde, Sweden. She holds a PhD in Computer Science from the De Montfort University, UK and a Master's degree in Computer Science from the University of Skövde, Sweden. Her research interests include simulation-based optimization, artificial intelligence, metaheuristics, and advanced information technology with applications in logistics and manufacturing. Her email address is anna.syberfeldt@his.se.

ACKNOWLEDGEMENTS

This work has been carried out within the SOL project which is partially financed by the Knowledge Foundation (KK Stiftelsen), Sweden. The author gratefully acknowledge the Knowledge Foundation for the provision of research funding and also the waste management organization Avfallshantering Östra Skaraborg for their support in this study. A special thanks also to the reviewers of the paper for providing helpful feedback.

SIMULATION-OPTIMIZATION OF WASTE COLLECTION BASED ON ULTRASONIC LEVEL MEASUREMENTS

Anna Syberfeldt
University of Skövde
Virtual Systems Research Center
Skövde, Sweden
anna.syberfeldt@his.se

KEYWORDS
Simulation-optimization, waste management, ultrasonic measurements.

ABSTRACT

This paper describes a study of improving the waste collection process through real-time simulation- optimization based on ultrasonic level measurements. In the study, ultrasonic level meters are placed inside waste containers and programmed to send measurement values to a computer server via GPRS at regular intervals. The measurement values are processed in a simulation-optimization program in order to determine which waste containers to visit a specific day. The program tries to minimize the number of collections over time while at the same time find the best routes to take. The aim of this approach is to minimize the driving distance and thereby minimize CO2 emissions.

INTRODUCTION

The rate at which waste containers fill up may vary depending on factors such as, for example, season, weather, holidays, feasts, and/or salary payment. Despite variations in fill levels, waste containers are usually emptied based on a static schedule with fixed collection dates and predetermined collection routes. Two potential drawbacks of such static schedules can be identified; a) the container is not completely filled (i.e. is emptied too soon), or b) the container gets overfilled (i.e. is emptied too late). In both cases the container is emptied at the wrong point in time and the collection process is not optimal when considering work effort, driving distance, and customer service. One way to overcome these drawbacks and improve the waste collection process is to schedule the collection dynamically (e.g. daily) based on the containers' fill levels. With such dynamic scheduling it is possible to ensure that the containers are emptied at the optimal point in time.

For realizing a dynamic scheduling, it must be possible to measure the containers' fill levels continuously in an efficient way. One way to do this is to mount ultrasonic level meters on the containers, which allows convenient measurements without physical contact. In short, an ultrasonic level meter reflects ultrasonic waves off the waste inside the container and converts the time between wave transmission and echo reception into distance. Measurements can be performed both in closed and opened containers, and liquids as well as solid substances (for example, paper, glass, metal, and plastics) can be measured.

In this study, ultrasonic level measurements are included in a simulation-optimization process for improving the waste collection process. More specifically, the aim of the simulation-optimization is to reduce fuel consumption, and thereby CO2 emissions, by avoiding visiting containers that are not completely filled. While several previous studies have investigated the use of remote measurement of waste containers (see for example [1-3]), we are not aware of any previous studies of simulation-optimization based on measurement values.

In the next section, ultrasonic level meters are presented in further detail. In Section III, the simulation-optimization process is described, followed by a presentation of the scenarios used for evaluating the process in section IV. Results from the study and conclusions are presented in Section V and VI, respectively.

ULTRASONIC LEVEL METERS

General description

An ultrasonic sensor emits high frequency (20 kHz to 200 kHz) acoustic waves and converts the time between wave transmission and echo reception into distance. The sensors are used for non-contact level measurements of everything from

highly viscous liquids to bulk solids. Measurements can be performed both in closed and opened containers. The sensors are also commonly used in water management applications for pump control and flow measurement.

Since measurements are affected by changing speed of sound (due to, for example, temperature variations or turbulence), correction algorithms are applied in order to ensure an acceptable accuracy of the measurements.

Other factors affecting the precision of the measurement are turbulence, foam, steam, vapors, and changes in the concentration of the process material. Turbulence and foam hinder the acoustic wave from being properly reflected back, steam and vapors distort the wave, and fluctuations in concentration affects the amount of energy in the acoustic wave that is reflected. To prevent these factors from disturbing the sensor, stilling wells and wave guides are used.

For precise measurements, careful mounting of the sensor is required in combination with a container that is relatively free of weldments, brackets, etc. Such things may cause false returns and erroneous response, although they are in most cases possible to detect and compensate for.

Device used in study

In the study, an ultrasonic level meter called UDM-G is used (Figure 1). Measurement data from the UDM-G meter is transmitted by GPRS via a GSM modem, which allows remote and wireless monitoring of the meter. The UDM-G level is programmed to send two measurements reports per 24 hours.

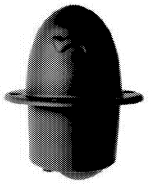


Figure 1: The UDM-G ultrasonic level meter (real height is 195 mm).

The meter is mounted on the top of the container and has a measuring range of 0.05 - 3 meters. The measurement accuracy is less than 0.02 meters. The UDM-G meter can used in various types of containers and used to measure compact, liquid, solid and transparent materials. In this study the focus is on measuring waste – including for example glass, paper, aluminum, and biological waste.

An example of measurement values from an UDM-G level meter mounted on a waste container is presented in Figure 2. As can be seen in the figure, during the selected time period the fill level of the container increases from about 30% to 80%.



Figure 2: Measurement data from waste container.

The measurement values can not only be used to monitor the historical and current fill levels, but also to forecast the quantity of the waste. Containers equipped with ultrasonic level meters transmit measurement values by GPRS via a GSM modem. Values are sampled continuously and transmitted at regular intervals with a default setting of two reports per 24 hours. The reports are received by a computer server that stores the values in a database.

SIMULATION-OPTIMIZATION PROCESS

Overall system

Data sent from the level meter are received by a computer server and used in a simulation-optimization process in order to find the most efficient waste collection routes (Figure 3).

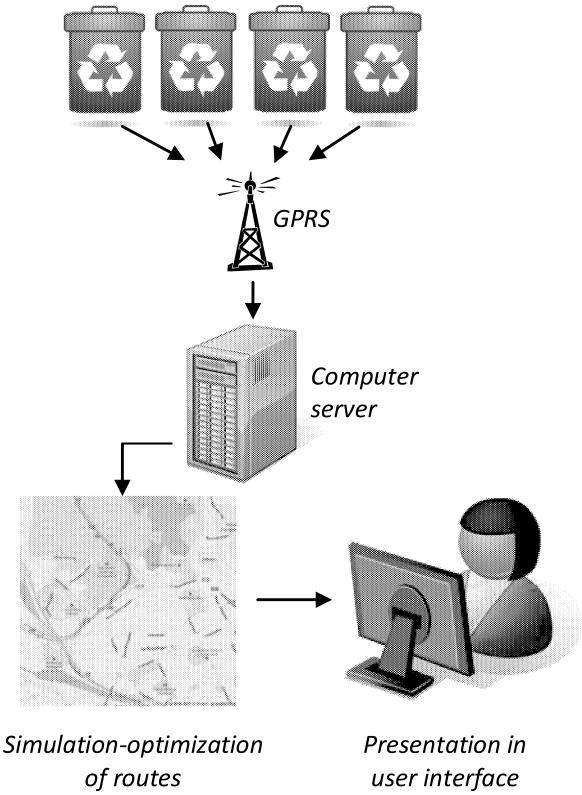


Figure 3: Overall system.

The routes resulting from the simulation-optimization process are presented to the user on a virtual map on a web page, according to Figure 4.

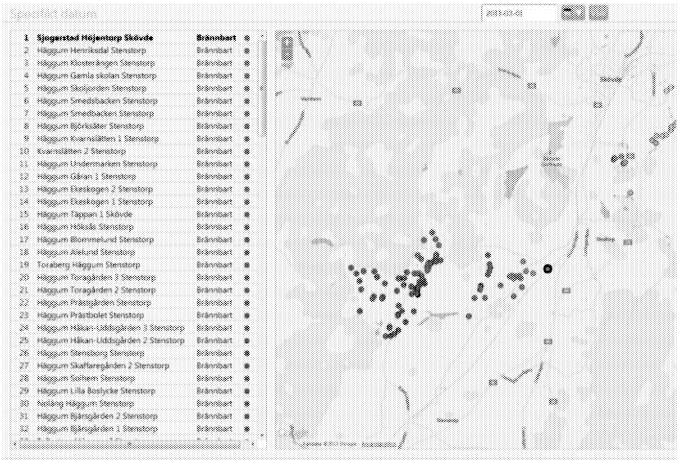


Figure 4: Containers presented on a virtual map.

Simulation

For simulating the waste collection process, a discrete-event simulation model has been constructed using the C# programming language. The reason for using the discrete-event simulation paradigm is that it provides a mean for incorporating complex relations using easier methodologies and tools compared to mathematical programming [4]. The discrete-event simulation model essentially consists of a set of system states and transitions between these states, which correspond to elements and event sequences in the real waste collection process.

When performing optimizations, the simulation is used by an optimization algorithm to evaluate the performance of the system based on given parameter values (Figure 5). So called simulation-based optimization is an iterative process; the optimization algorithm generates a set of routes, and feeds them to the simulation model, which computes CO2 emissions. Based on the evaluation feedback obtained from the simulation model, the optimization algorithm generates a new set of parameter values and the generation-evaluation process continues until a user-defined stopping criterion is satisfied. Such a criterion may, for example, be that a certain amount of time has passed, or that specific objective values have been achieved.

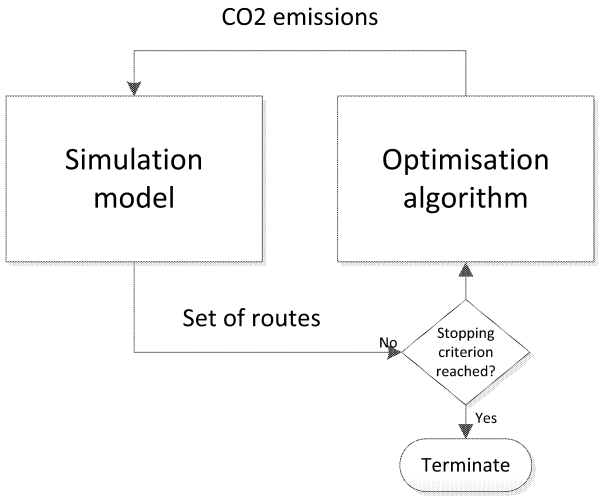


Figure 5: Simulation-optimization process.

Optimization

Finding the shortest route between a number of points (in this case represented by waste containers) might seem simple, but is classified as a "NP-hard" problem in its simplest form [5]. This means that the time it takes to solve the problem grows exponentially with the problem size (that is, the number of nodes). In a real-world routing problem, there are often a large number of parameters to consider. For example, there might be a fleet of heterogeneous vehicles with different weight, volume, and speed capacities. Furthermore, the vehicles may also have different operating costs and special working hours. When undertaking route planning, all the different parameters must be considered simultaneously.

Finding the optimal route is possible, but might take very long time because *all* possible routes must be evaluated in order to find the best one. Even for small-sized problems there are a huge number of possible routes. As an example, for a problem with only 15 nodes to visit, there are 6 227 020 800 possible routes. Due to the general complexity of routing problems, optimization techniques that are not guaranteed to find the optimal solution, but a sufficiently good one in a short time, are often used rather than exact methods. This is also the strategy selected in this study.

Genetic algorithms are a class of such inexact optimization techniques that can be used to approach routing problems [6], and this is also the technique used in this study. Genetic algorithms are essentially based on Darwin's theories about "survival of the fittest" [7]. According to this theory, in a population of individuals (solutions), the ones having the most desirable characteristics will be given the best opportunities to mate and carry on their genes. In this way, Darwin argued, good genes will propagate through generations and the population increasingly improves over time.

In evolving a population of solutions, evolutionary algorithms apply biologically inspired operations for selection, crossover and mutation. The operators are applied in a loop, and an iteration of the loop is called a generation. In the pseudo code below, the basic steps involved in this evolutionary process are presented.

```
Initialize population
Evaluate the fitness of solutions in the
population
repeat
  Select solutions to reproduce
  Form a new generation of population through
    crossover and mutation
  Evaluate the new solutions
until terminating condition
```

In this study, a solution consists of a collection of routes (a specification of which containers to visit a specific day). For the initial population, solutions are generated randomly. During each generation, a proportion of the solutions in the population is selected to breed offspring for the next generation of the population (that is, create new solutions). Solutions are selected based on their fitness, representing a quantification of their optimality, in this case the CO emissions of the route (basically a transformation of driving distance). From the solutions selected, new solutions are created to form the next generation of the population. For to creation of each new solution, two parent solutions are chosen and through crossover an offspring is produced. Occasionally, the new solution can undergo a small mutation in order to keep the diversity of the population large and avoid local minima. A mutation involves changing an arbitrary part of a solution with a certain probability, in this case changing the order of two of the collection sites to visit. When the new population is formed, the average fitness will have generally increased and after evolving populations during some time near-optimal collection of routes have been found. At the end, the route collection involving the shortest driving distance is selected. This route usually involves as few collections as possible, that is, no container is emptied before it is sufficiently filled.

EVALUATION SCENARIOS

For evaluating the applicability and the gain of the simulation-optimization process two real-world scenarios have been identified and used for evaluation. The first scenario represents a larger city in Sweden in which the waste containers are equipped with level meters. The second scenario represents a country-side city in Sweden in which level meters are currently not used, but in which the authorities are interesting in evaluating the potential of the technique. In other words, the first scenario aims to evaluate the *actual* gain of using simulation-optimization based on level measurements, while the other scenario aims to evaluate the *potential* gain. The scenarios are described in further detail below.

Scenario 1: Evaluation of actual gain

Scenario 1 includes 20 waste containers located in a densely populated city. Each of the containers is equipped with an UDM-G meter (see Section II-B) and programmed to send measurement reports twice per 24-hour period. Currently, the containers are emptied at fixed collection dates with either seven or fourteen days in between.

For building the simulation model, measurement values from the containers were first collected for three full months. The collected values clearly revealed three categories of containers with respect to fill rate, and these have been used in the simulation:

a) containers that are more than 70% filled up at collection (in total 12),
b) containers that are filled to approximately 65-70% at collection (in total 6), and
c) containers that are filled to approximately 50% at collection (in total 2).

To ensure that a container is never overfilled, the optimization has been configured so that all containers are always emptied before they get filled to 85%.

Scenario 2: Evalution of potential gain

Scenario 2 aims to evaluate the potential gain of using simulation-optimization based on level measurements in a country-side city. In the city, the waste containers are distributed over a relatively large geographical area. For the study, containers having a volume of 500 liters or more are selected. The majority of these containers are owned by companies or industries which produce relatively large amounts of waste. The current cost of the ultrasonic level meter and the belonging GSM subscription do not motivate to consider smaller containers (which are usually owned by private individuals).

For evaluating the scenario, the simulation model was configured to reflect the settings of the city and the included containers. To mimic level meters in the simulation, predicted filling rates of each container were used as specified by experienced staff members of the waste collection organization.

The next section continues with presenting the results from the scenarios.

RESULTS

In Scenario 1 (actual gain), the results show that in total 85 collections can be saved per year without risking any container to be overfilled. To empty a container usually takes 5-15 minutes, so in average approximately 0.6x86≈56 work hours can be saved. With respect to driving distance, only a few kilometers can be saved each year. The reason for this is

the geographical location of the containers; the containers are located in clusters with only a few hundred meters in between. An example of one cluster, include half of all containers, are given in Figure 6. When the containers are located this way, avoiding collecting one or a few containers does not affect the driving distance since the location has to be visited any way.



Figure 6. Example of container cluster in real-world scenario.

In Scenario 2 (potential gain), comparative simulation-optimizations with and without level measurements were performed in order to estimate the gain of using level meters. The results show that a 5-35% reduction of CO2 emissions, as well as working effort, can be achieved when using ultrasonic level meters in the waste collection process. It is clear from the results that the largest gain of level meters is achieved for collection routes at the country side, where the distance between containers is relatively large.

SUMMARY AND CONCLUSIONS

Today, waste collection is usually performed according to static schedules although the rate at which waste containers fill up may vary depending over the year depending on season, weather, holidays, feasts, salary payment, etc. A consequence of static schedules is that the container is seldom emptied at the exact right day – it might either be emptied when it is not yet full, or when it is overfull. When either of this happens, the collection process is not optimal.

In this study, we investigate the possible gains of schedule waste collections dynamically (e.g. daily) instead of statically by equipping the containers with ultrasonic level meters and utilizing simulation-optimization. The level meters are placed inside the waste containers and programmed to send measurement values to a computer server via GPRS. The measurement values are processed in a simulation-optimization program in order to determine which waste containers to visit a specific day. The program tries to minimize the number of collections over time while at the same time find the best route to take. The aim of this approach is to minimize the driving distance and thereby minimize CO2 emissions.

In evaluating the simulation-optimization, two real-world scenarios are being used. The first scenario covers a larger city in Sweden in which the containers are equipped with level meters. In this scenario, the results show no significant savings with respect to CO2 emissions. An analysis show that this is due to the geographical location of the containers; the containers are located in clusters and avoiding collecting one or a few containers inside a cluster virtually does not affect the driving distance.

The second scenario used in the evaluation represents a country-side city in Sweden in which the containers are currently not equipped with level meters. Results from this scenario show that a quite significant (5-35%) reduction of CO2 emissions can be achieved when using level meters compared to when not using level meters.

In general, the results indicate that the largest gain of simulation-optimization based on level measurements is achieved at the country side where the distance between containers is relatively large. In these settings, the most significant reduction of CO2 emissions can be noticed, and also the best "pay-off" when considering the cost of the level meters in relation to the economic savings (including mainly labor costs and fuel costs).

Regarding installation costs, it is important to be aware of the relatively large investment related to purchasing and mounting ultrasonic level meters. Widespread use of level meters in the waste collection process is probably realistic first at a point where the cost of the sensors is reduced. Studies on how to produce cheaper, yet robust, sensors for the purpose are therefore needed.

Robustness is a feature that is of great importance for the sensor readings to be useful. Measurement values that are stable and reliable over time are critical for an efficient simulation-optimization process. A general and effective algorithm for eliminating noisy measurement values inside waste containers is therefore also an interesting topic for future studies and also an aspect that would improve the gain of the simulation-optimization.

In this study, a genetic algorithm was used to optimize the collection process. Another approach that should be evaluated in the future is a rule-based system using a combination of due dates and filling percentage to determine collection routes. Whenever the truck is unloaded at the garbage dump, the system could check which containers that are close in due date, and combine this information with filling percentages and distances in order to determine the next containers to visit. With this approach, even better results may be obtained.

In the future it would also be interesting to study more complex scenarios with a mix of containers that are equipped with level meters and those that are not (in this study it was assumed that all containers included in the simulation were equipped with level meters). It would also be interesting to study what happens if measurements reports are from the level meters are sent only once per 24 hours (instead of twice) for energy saving reasons.

REFERENCES

[1] F. Vicentinia, A. Giustia, A. Rovettaa, X. Fanb, Q. Heb, M. Zhub, B. Liub, "Sensorized waste collection container for content estimation and collection optimization," Waste management, vol. 29 no. 5, 2009.

[2] A. Rovettaa, F. Xiuminb, F. Vicentinia, Z. Minghuab, A. Giustia, H. Qichangb, "Early detection and evaluation of waste through sensorized containers for a collection monitoring application," Waste management, vol. 29 no. 12, pp. 2939-49, 2009.

[3] P.J. de Groot, G.J. Postma, W.J. Melssen, L.M.C. Buydens, "Validation of Remote, On-line, Near-InfraRed Measurements for the Classification of Demolition Waste", Analytica Chimica Acta 453, pp. 117-124, 2002.

[4] A.M. Law "Simulation Modeling and Analysis", Fourth Edition, McGraw-Hill, 2007.

[5] N. Jozefowiez, F. Semet, E.G. Talbi, "Mulit-objective vehicle routing problems," European journal of operational research, 189, pp. 293-309, 2008.

[6] C. Darwin, "On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life" 1st ed., London: John Murray, 1859.

BIOGRAPHY

ANNA SYBERFELDT is a senior researcher at the University of Skövde, Sweden. She holds a PhD in Computer Science from the De Montfort University, UK and a Master's degree in Computer Science from the University of Skövde, Sweden. Her research interests include simulation-based optimization, artificial intelligence, metaheuristics, and advanced information technology with applications in logistics and manufacturing. Her email address is anna.syberfeldt@his.se.

ACKNOWLEDGEMENTS

This work has been carried out within the SOL project which is partially financed by the Knowledge Foundation (KK Stiftelsen), Sweden. The author gratefully acknowledge the Knowledge Foundation for the provision of research funding and also the waste management organization Avfallshantering Östra Skaraborg for their support in this study. A special thanks also to the reviewers of the paper for providing helpful feedback.

TRAFFIC SIMULATION

ANALYSIS OF POSSIBLE SPEEDUP OF THE MICROSCOPIC ROAD TRAFFIC SIMULATION USING AGGREGATE MOVEMENT OF VEHICLES

Tomas Potuzak
Department of Computer Science and Engineering, Faculty of Applied Sciences
University of West Bohemia
306 14, Plzen,
Czech Republic
E-mail: tpotuzak@kiv.zcu.cz

KEYWORDS
Simulation speedup, hybrid simulation, nanoscopic simulation, microscopic simulation, parallel/distributed simulation.

ABSTRACT

In this paper, we describe the microscopic/nanoscopic Parallel/Distributed Hybrid Urban Traffic Simulator (P/D-HUTS), which is currently under development. The P/D-HUTS system incorporates three different microscopic traffic models for the simulation within the traffic lanes and a nanoscopic traffic model for a more detailed simulation within the crossroads. In the description, we are focusing mainly on the microscopic simulation utilized in the P/D-HUTS system. We discuss the possibility of the speedup of the simulation using the aggregate movement of the particular vehicles in the traffic lanes. The aggregate movement can be used in all three traffic models, which are at our disposal.

INTRODUCTION

The computer simulation of road traffic is an important tool for control and analysis of existing or designed road traffic networks. The detail of the simulation can vary from a simple modeling of traffic flows (Lighill and Whitman 1955) in particular traffic lanes to modeling of many aspects of the particular vehicles and their interactions including individual behavior of the drivers (Ni 2003). Naturally, the more complex and detailed the simulations are, the higher is their demand for computing power. On the other hand, the highly detailed simulations are able to model the real traffic situations more accurately. Hence, the speedup of the detailed simulations is a subject to the active research.

There are two main approaches to the speedup of the simulation. The first approach lies in the utilization of a distributed computing environment where combined power of multiple computers (nodes) connected via a computer network is utilized (Nagel and Rickert 2001; Gonnet 2001). With the current advancement and spreading of the multi-core computers, it is also possible to utilize a parallel or a parallel/distributed computation (Potuzak 2012). The second approach lies in simulating of different parts of the road traffic networks at different levels of detail (Hartman and Herout 2007; Ni 2003; Ni 2006; Huang et al. 2008). In such a hybrid simulation, some areas are simulated at a higher level of detail while other areas are simulated at a lower level of detail (Potuzak 2013).

In our current research, we are working on a combination of both approaches. The expected result is a Parallel/Distributed Hybrid Urban Traffic Simulator (P/D-HUTS), currently in an early stage of development at Department of Computer Science and Engineering of University of West Bohemia (DCSE UWB). In the system, we plan to utilize the simulations developed in our previous research (Potuzak 2009).

However, in order to achieve the maximal speedup of the resulting simulation, it would be convenient to optimize these simulations. In this paper, we describe the basic concepts of the P/D-HUTS system. We focus mainly on the possible speedup of one of the simulations of the P/D-HUTS system using the aggregate movement of vehicles.

SIMULATION OF ROAD TRAFFIC

Before we proceed with the analysis, it is convenient to briefly describe the simulation of road traffic.

Time-Flow Mechanism

One of the most important aspects of every computer simulation (and the simulation of road traffic as well) is the way the simulation time is advanced. There are two commonly used time-flow mechanisms – the *event-driven* and the *time-stepped* time-flow mechanism (Fujimoto 2000).

Using the event-driven time-flow mechanism, the simulation advances in time by interpreting of events. With each event, an action and a time stamp is associated. The action is an incremental change to the simulation state and the time stamp determines when this change shall happen. The events are queued in a structure called calendar and are processed sequentially. New events can be added to the calendar as a consequence of an action of a different event. So, the simulation advances in time irregularly from one time stamp to another (Fujimoto 2000). The simulation ends if a preset simulation time is reached or there are no more events to be processed.

Using the time-stepped time flow mechanism, the entire simulation time is divided into sequence of equally-size time steps. Often, the simulation step length is set to one second (Potuzak 2009). In each time step, the entire simulation state is recomputed. So, the simulation advances in time regularly step by step (Fujitmoto 2000). The simulation ends if a preset number of time-steps is performed.

Level of Detail

As it was said before, the simulation of road traffic can be of different level of detail. There are four types of simulations, which are commonly used – the *macroscopic*, *mesoscopic*, *microscopic*, and *nanoscopic* simulations.

The macroscopic simulation deals only with traffic flows in particular traffic lanes. The flows are described by a set of parameters (e.g. mean speed, vehicle density, etc.), which are periodically recomputed. Since there are no vehicles modeled, the macroscopic simulation can be very fast. These simulations are the oldest ones (Lighill and Whitman 1955) and exist in many modifications (Braban-Ledoux 2000). Both described time-flow mechanisms are commonly used.

In the mesoscopic simulation, there is a representation of the vehicles, but their interactions are modeled at a low level of detail (Burghout 2004). There are several traffic models, which are considered mesoscopic, for example *gas-kinetic models* (Nagatani 1996) or *queuing networks* (Nizzard 2002). Due to the presence of the vehicles, the simulation is generally slower than the macroscopic simulation. Again, both described time-flow mechanisms are commonly used.

In the microscopic simulation, each vehicle is modeled as an object with its own position, direction, speed, and acceleration. The positions of the vehicles are periodically recomputed based on their current speed and the traffic conditions in their vicinity. There are two commonly used microscopic traffic models – the *cellular automaton* (Nagel and Schreckenberg 1992) and the *car-following model* (Gipps 1981). The microscopic simulations are much slower than the macroscopic or mesoscopic simulations due to the high number of vehicles' interactions. Usually, the time-stepped time flow mechanism is utilized (Potuzak 2009). A microscopic simulation is used in the P/D-HUTS system.

The nanoscopic simulation is based on the microscopic simulation, but models many additional aspects of the vehicles and their interactions. For example, it considers the individual behavior of the drivers of the vehicles, various features of the vehicles, additional features of the simulated traffic network, and so on (Vanderschuren 2007; Ni 2006). Due to the high detail, the nanoscopic simulation is even slower than the already slow microscopic simulation. Similarly to the microscopic simulation, the time-stepped time flow mechanism is usually used. A nanoscopic simulation is used in the P/D-HUTS system.

Microscopic Traffic Models

Using the cellular automaton model, the traffic lanes and crossroads are divided into equally-sized traffic cells (see Fig. 1). Each cell can be occupied by a single vehicle or can be empty. The vehicle cannot be placed between the traffic cells. When the vehicle moves, it "jumps" from a cell to another based on its current speed. The current speed of the vehicle is calculated and the movement is performed using four rules – *acceleration*, *deceleration*, *randomization*, and *movement*. First rule represents the tendency of the vehicle to speed up to its maximal speed. Second rule represents the necessary slow down if there is an obstacle in the traffic lane or crossroad (e.g. a slower vehicle). The third rule represents natural fluctuations of the speed of the vehicle due to the driver's behavior and actual conditions of the traffic lane, in which the vehicle is driving. It is implemented as a random slow down of the vehicle. The fourth rule is the movement of the vehicle based on its speed calculated using first three rules (Nagel and Schreckenberg 1992).

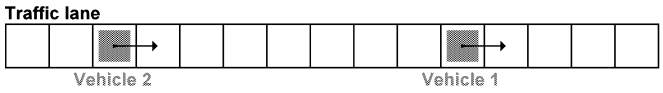


Figure 1: Traffic lane in a cellular automaton model

Using the car-following model, there are no traffic cells. So, the vehicle can be placed anywhere in the traffic lane or crossroad (see Fig. 2). The basic concept of the movement of a vehicle is its response to a stimulus. The response of the vehicle is the acceleration or deceleration depending on the stimulus. The stimulus for a vehicle can be represented by another vehicle, which is moving in front it. So, each vehicle adapts its speed according to the vehicle in front of it in order to achieve its maximal speed and contemporarily to avoid collision. If there is no other vehicle in front of the vehicle, it speeds up to its maximal speed (Gipps 1981).

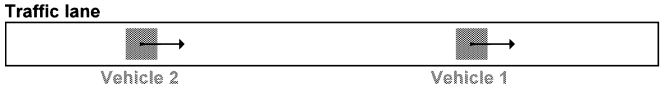


Figure 2: Traffic lane in a car-following model

P/D-HUTS

As it was mentioned, the P/D-HUTS system is currently in a preliminary stage of development at DCSE UWB.

Basic Concept

The P/D-HUTS system is a hybrid microscopic/nanoscopic simulator of road traffic, which can utilize the parallel, distributed, or parallel/distributed computing environment for the speedup of the simulation.

For a hybrid simulation of road traffic, it is important to decide, which parts of the simulated road traffic network will be simulated at lower detail (i.e. at microscopic level) and which parts will be simulated at higher detail (i.e. at nanoscopic level). In the P/D-HUTS system, a systematic approach is used where the nanoscopic simulation is used for crossroads and the microscopic simulation is used for traffic lanes (Potuzak 2013). The reason is that, in the crossroads, there are many complex interactions, which are neglected by a microscopic simulation (e.g. influence of the pedestrians moving on the crosswalks, different reaction times of the drivers, etc.). In the traffic lanes, the interactions of the vehicles are less complex. The speed of the vehicles is relatively steady, the vehicles in one traffic lane are moving in one direction, there are usually no pedestrians, and so on (Potuzak 2013). Hence, for the traffic lanes, the microscopic simulation is sufficiently detailed and, in comparison to the nanoscopic simulation, much faster (Potuzak 2013).

Another important issue is the conversion of the traffic flows on the edges of the parts of road traffic network simulated at different levels of detail. In the P/D-HUTS system, these edges are situated between the traffic lanes and the crossroads. Since both simulations consider the vehicles, the conversion is not problematic. The object of the vehicle can incorporate parameters necessary for both simulations for the whole time the vehicle is moving in the simulated road traffic network. The nanoscopic parameters can be then used when the vehicle is moving in a nanoscopic part of the road traffic network. Similarly, the microscopic parameters can be used when the vehicle is moving in a microscopic part. When a vehicle is moving from a nanoscopic to a microscopic part, the microscopic parameters are initialized using the last calculated nanoscopic parameters. Vice versa, the nanoscopic parameters are initialized using the last calculated microscopic parameters. The nanoscopic parameters, which do not have their microscopic counterparts, are initialized using their last known values from the previous nanoscopic part or are set to default values if the vehicle has not visited a nanoscopic part yet (Potuzak 2013).

The P/D-HUTS system is designed for a parallel/distributed computing environment. It means that it can be running on a distributed computer – a set of ordinary computers (nodes) interconnected by means of a computer network (e.g. Ethernet). An example of such a distributed computer can be the computers in a classroom of a university. In this case, on each node, a simulation process simulating a part of the road traffic network (a traffic sub-network) is running. If the particular nodes incorporate multi-core processors, the P/D-HUTS system can utilize all the processor cores using multithreaded processes. Both pure distributed and parallel/distributed approaches were tested on the Distributed Urban Traffic Simulator (DUTS), developed at DCSE UWB. Based on the tests, the parallel/distributed approach gives better results than the pure distributed approach (Potuzak 2012). Hence, it will be used in the P/D-HUTS system as well.

Microscopic Simulation Utilized in P/D-HUTS

The DUTS system incorporates three different traffic models inspired by three existing simulators. These models are utilized in the P/D-HUTS system as well, but only for the simulation of the traffic lanes. For the crossroads, a nanoscopic traffic model is used (see next section).

The AIMSUN-based model is inspired by the AIMSUN simulator (Barceló and Casas 2002), which is based on a modified Gipps' car-following model (Gipps 1981). The vehicles have various lengths from 2.5 to 15 meters (see Fig 3). The time-stepped mechanism with one second long time step is utilized for the advancement of the simulation in time (Potuzak 2009).

The JUTS-based model is inspired by the JUTS simulator, developed at DCSE UWB as well. It is based on a modified Nagel-Schreckenberg's cellular automaton model (Hartman 2004; Nagel and Schreckenberg 1992). The length of the traffic cells is set to 2.5 meters (see Fig. 3). A vehicle can occupy more than one cell, since various lengths of the vehicles (from 2.5 to 15 meters – i.e. from one to six traffic cells) are considered (Potuzak 2009). The speed of the vehicles is calculated using the rules of the original Nagel-Schreckenberg's model (Nagel and Schreckenberg 1992). Again, the time-stepped mechanism with one second long time step is used (Potuzak 2009).

The TRANSIMS-based model is inspired by the TRANSIMS simulator (Nagel and Rickert 2001), which is based on a Nagel-Schreckenberg's cellular automaton model (Nagel and Schreckenberg 1992). The model is very similar to the JUTS-based model, but the length of the traffic cells is set to 7.5 meters and each vehicle can occupy only one cell (Potuzak 2009).

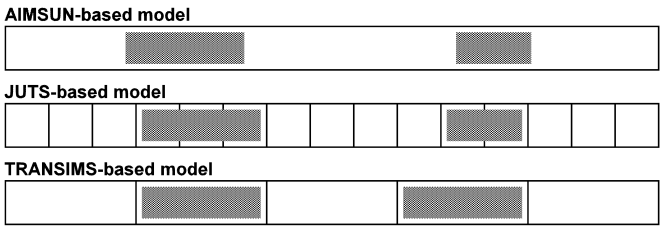


Figure 3: Comparison of the microscopic traffic models

Nanoscopic Simulation Utilized in P/D-HUTS

The nanoscopic simulation utilized in the P/D-HUTS system came from a nanoscopic road traffic simulator called DUTS Nanosuite, currently under development at DCSE UWB. There is only one nanoscopic traffic model, which is loosely based on the Gipps' car following model (Gipps 1981). It is used solely for the crossroads in the P/D-HUTS system (see Fig 4).

The vehicles have more parameters than their microscopic counterparts (e.g. various lengths, widths, accelerations, reaction times of the drivers, etc). The modeling of the individual behavior of the vehicles is currently under development and will be included in the P/D-HUTS system. This individual behavior has two aspects – the driving style (e.g. fluent, slow, aggressive etc.) and the routing (i.e. where the vehicles are heading). Since the nanoscopic traffic model is used only in crossroads, the routing influences, to which target traffic exit the vehicle will be heading. The driving style then influences the reaction to the green or red traffic signal or the carefulness by yielding (Potuzak 2013).

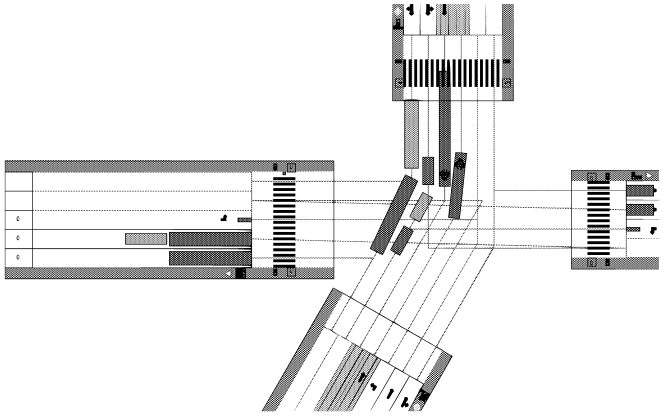


Figure 4: Screenshot of the nanoscopic simulation

Besides the vehicles, the pedestrians are included in the nanoscopic traffic model as well. The reason is that the number of pedestrians trying to cross the crosswalks can significantly influence the fluentness of the traffic (Potuzak 2013). Not only in the crossroads without traffic lights, but in the crossroads with traffic lights as well, since green signal for the vehicles turning to the right are often (willingly) in conflict with the green signal for the pedestrians. In that case, the vehicles must stop in front of the crosswalk and wait until the pedestrians cross the crosswalk.

Similar to the microscopic traffic models, the time-stepped time-flow mechanism is used for the advancement in time. However, in order to model the movement of the vehicles and pedestrians more fluently, the time step length is adjustable and is preset to 0.1 second. That means that, for each time step of the microscopic simulation, ten time steps of the nanoscopic simulation is performed (Potuzak 2013).

SPEEDUP OF THE MICROSCOPIC SIMULATION

We can now focus on the optimization of the microscopic traffic models in order to achieve speedup of the simulation.

Basic Idea

In all three microscopic traffic models utilized in the P/D-HUTS system, the calculation of the speed and the consequent position is performed for each vehicle in each time step (Potuzak 2013). This is very time-consuming and presents the majority of the computations of the entire microscopic simulation. Nevertheless, the microscopic simulation can be sped up using the aggregate movement of the vehicles regardless the utilized traffic model. It means that the movements of a vehicle from several time steps can be aggregated into one long movement. Since the movement of each vehicle consists of several separate operations (acceleration to maximal speed, deceleration due to an obstacle in the road, and the randomization of the movement), this can be achieved without the loss of precision (Potuzak 2013).

The described idea can be illustrated by an example. Assume now that there is a vehicle at the beginning of a traffic lane, which is 100 meters long. The maximal speed of the vehicle is 10 m/s (i.e. 36 km/h) and the vehicle is moving at this speed. In that case, the vehicle will require roughly ten steps to reach the end of the traffic lane. This means roughly ten calculations of the speed of the vehicle and its consequent position. Using the aggregate movement, the described movement of the vehicle can be interpreted as a single jump to the end of the traffic lane. This means only one calculation of the speed of the vehicle and its consequent position (see Fig. 5).

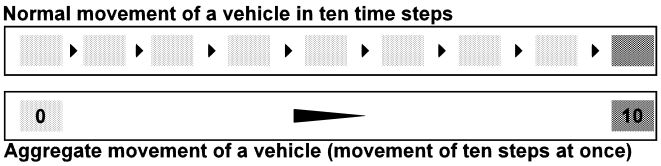


Figure 5: Comparison of normal and aggregate movement

However, the described substitution is possible only in case that there are no vehicles blocking the road. Otherwise, the movement can be aggregated only to the nearest vehicle in front of the vehicle, for which the movement is aggregated. Still, it is possible that the movement from several steps can be aggregated, although not the entire movement to the end of the traffic lane. The consequence is that, for various vehicles, the movement from different numbers of time steps can be aggregated. Hence, each vehicle must incorporate a counter containing the number of time steps the vehicle is ahead in time (see Fig. 5). The value of the counter is set once the aggregate movement is performed. In the following time steps, the vehicle does not perform the movement (i.e. it is standing) and only the counter is decremented. Once the counter reaches zero, the vehicle can resume its normal movement or perform the aggregate movement depending on the free space in front of the vehicle in the traffic lane.

Implementation in the Car-Following Model

Using the car-following model (i.e. the AIMSUN-based model), all vehicles in one traffic lane are stored in a list, where there are ordered according to their positions in the lane. The current distance of the vehicle from the end of the traffic lane (to which the vehicle is heading) is stored in each vehicle.

During the computation of the speed of a vehicle, which is performed in each time step, the vehicle must determine its position in the list and calculate the distance to the nearest vehicle in front of it (if there is any) or to the end of the traffic lane. Then, the new current speed of the vehicle is calculated using its speed from the previous time step and its acceleration. The vehicle then moves by the minimum from the distance corresponding to the calculated speed and to the length of the time step (one second) and from the distance to the nearest vehicle in front of it or to the end of the traffic lane.

Using the aggregate movement of vehicles, the described calculation is very similar. For the vehicle, the distance to the nearest vehicle in front of it or to the end of traffic lane is determined in the same manner. The change is that the position of the vehicle is changed using this distance. Simultaneously, the distance and the current speed of the vehicle from the previous step are used for the calculation of the vehicle speed at the end of the aggregate movement and for the calculation of the value of the vehicle's counter.

Implementation in the Cellular Automaton Models

Using the cellular automaton models (i.e. JUTS- and TRANSIMS-based models), the traffic lanes consist of traffic cells, in which the vehicles are moving. Each vehicle knows its current cell (or current set of traffic cells in case of longer vehicles in the JUTS-based model). From this cell, the vehicle can investigate the cells in front of it to determine the distance to the nearest vehicle in front of it or to the end of the traffic lane.

During the computation of the speed of the vehicle, the distance to the nearest vehicle in front of it is determined.

Then, the new current speed of the vehicle is calculated using its speed from the previous time step and its acceleration. Again, the vehicle then moves by the minimum from the distance corresponding to the calculated speed and to the length of the time step (one second) and from the distance to the nearest vehicle in front of it or to the end of the traffic lane. So, the only difference in comparison to the car-following model is the manner the distance to the nearest vehicle is determined.

Also, the calculation of the aggregate movement is very similar to the car-following model, only the manner the distance to the nearest vehicle is determined is different. However, since the investigation of the free space in front of the vehicle is performed traffic cell by traffic cell, the larger the free space is, the more traffic cells must be investigated and the calculation is more computation-intensive. Consequently, the speedup of the simulation using the aggregate movement in the cellular automaton models is likely to be smaller than in the car-following models.

Estimation of Possible Speedup

For the estimation of possible speedup using the aggregate movement of vehicles, it is necessary to know the numbers of vehicles moving in the road traffic network and the average amount of the free space in front of the vehicles. Precise information can be obtained by testing. Nevertheless, some conclusions can be made without testing as well.

It is obvious that the increasing number of vehicles moving in the road traffic network is negatively influencing the usability of the aggregate movement of vehicles. The reason is that there is a lower amount of the free space in the road traffic network, which could be efficiently used for the aggregate movement.

Another observation is that the aggregate movement of vehicles is more efficient for the car-following model than for the cellular automaton models. The reason is the more computation-intensive calculation of the amount of free space in front of the vehicles using the cellular automaton models.

The last observation is that the simulation consumes more memory due to the aggregate movement of vehicles, since its implementation requires a counter in each vehicle. However, this is not a big issue, since one byte per counter is sufficient for the implementation. This means only several kilobytes for several thousands of vehicles moving in the traffic network.

CONCLUSION

In this paper, we described the basic concepts of the microscopic/nanoscopic Parallel/Distributed Hybrid Urban Traffic Simulator (P/D-HUTS), which is currently under development at DCSE UWB. We focused mainly on the microscopic simulation, which will be utilized for the traffic lanes in the P/D-HUTS. We discussed its possible speedup using the aggregate movement of the vehicles from several time steps.

In our future work, we will continue with development of the P/D-HUTS system. In the nearest future, we will focus on the implementation of the aggregate movement of vehicles and its testing.

REFERENCES

Barceló J. and J. Casas. 2002. “Dynamic network simulation with AIMSUN”, In *Proceedings of the International Symposium on Transport Simulation*, Yokohama.

Braban-Ledoux C. 2000. METACOR – *A Macroscopic Modelling Tool for Corridor Application to the Stockholm Test Site*. Final Report CRT2000:05, Center for Traffic Engineering & Traffic Simulation, Kungl Tekniska Högskolan, Valhallavägen.

Burghout W. 2004. *Hybrid microscopic-mesoscopic traffic simulation*. Doctoral thesis, Royal Institute of Technology, Stockholm.

Fujimoto R.M. 2000. *Parallel and Distributed Simulation Systems*. John Wiley & Sons, New York.

Gipps P.G. 1981. “A behavioural car following model for computer simulation”. In *Transp. Res. Board*, 15-B(2), 403-414.

Gonnet P.G. 2001. *A Queue-based Distributed Traffic Micro-Simulation*. Technical Report.

Hartman D. 2004. “Leading Head Algorithm for Urban Traffic Model”. In *Proceedings of the 16th International European Simulation Symposium ESS*, Budapest, 297-302.

Hartman D. and P. Herout. 2007. “Construction of a Hybrid Traffic Model Based on JUTS Cellular Model”. In *Proceedings of the 6th EUROSIM Congress on Modelling and Simulation*, Ljubljana.

Huang W., J. Sun, and K. Li. 2008. “A Multiple Representation Entity-Based Approach to Hybrid Traffic Simulation Model”. In *2008 International Conference on Information Management, Innovation Management and Industrial Engineering*, Taipei, 7-12.

Lighill M.H. and G.B. Whitman. 1955. “On kinematic waves II: A theory of traffic flow on long crowed roads”. In *Proceedings of the Royal Society of London*, s. A, 229, London, 317-345.

Nagatani T. 1996. “Gas Kinetic Approach to Two-Dimensional Traffic Flow”. In *J. Phys Soc Jap*, 65 (10), 3150-3152.

Nagel K. and M. Rickert. 2001. “Parallel Implementation of the TRANSIMS Micro-Simulation”. In *Parallel Computing*, 27 (12), 1611-1639.

Nagel K. and M. Schreckenberg. 1992. “A Cellular Automaton Model for Freeway Traffic”. In *Journal de Physique I*, 2, 2221-2229.

Ni D. 2003. “2DSIM: A Prototype of Nanoscopic Traffic Simulation”. In *Proceedings of Intelligent Vehicles Symposium*, Columbus, 47-52.

Ni D. 2006, “A Framework for New Generation Transportation Simulation”. In *Proceedings of the 2006 Winter Simulation Conference*, Monterey, 1508-1514.

Nizzard L. 2002. *Combining Microscopic and Mesoscopic Traffic Simulators*. Raport de stage d’option scientifique, Ecole Polytechnique, Paris.

Potuzak T. 2009. *Methods for Reduction of Interprocess Communication in Distributed Simulation of Road Traffic*. Doctoral thesis, University of West Bohemia, Pilsen.

Potuzak T. 2012. “Distributed-Parallel Road Traffic Simulator for Clusters of Multi-core Computers”. In *2012 IEEE/ACM 16th International Symposium on Distributed Simulation and Real Time Applications – DS-RT 2012*, Dublin, 195-201.

Potuzak T. 2013. “Issues of Parallel Hybrid Nanoscopic/ Microscopic Road Traffic Simulation”. In *Proceedings of Eurocon 2013*, Zagreb, to be published.

Vanderschuren M. 2007. “Calibrating Microscopic Simulation Models”. In *Proceedings of the 26th Southern African Transport Conference*, Pretoria, 610-620.

MICROSCOPIC SIMULATION OF A MULTIMODAL URBAN TRAFFIC NETWORK

Ivana Ćavar
Marjana Petrović
Miroslav Vujić
Faculty of transport and traffic sciences
University of Zagreb
Vukelićeva 4
HR 10000 Zagreb
Croatia
E-mail: ivana.cavar@fpz.hr; marjana.petrovic@fpz.hr, miroslav.vujic@fpz.hr

KEYWORDS
Microscopic simulations, multimodal urban traffic, traffic planning and modeling, urban transport.

ABSTRACT

The interest in complex systems where several individuals are interacting in a nonlinear manner has seen a remarkable increase in recent years. One of such systems is traffic system with the high number of elements and the complex interactions among them. To model complex road traffic interactions, a number and/or combination of techniques (microscopic, macroscopic, mesoscopic etc.) is required. The paper describes a microscopic modeling and simulation of a multimodal urban traffic systems (Case Study City of Zagreb, Zvonimirova road). This plays an important role in real world decision making process, allowing different scenarios to be applied in a simulation model and best solution to be selected for implementation.

INTRODUCTION

Since 1980s computing power stopped being limiting element and application of traffic simulations started to be essential tool for transport planners. This allowed more attention to be given to development of complex traffic models. From then till now different approaches have been tested for application in diverse traffic conditions, planning phases and with various aims (Sadat Hoseini 2011).
There are three basic levels of traffic simulation models: microscopic, mesoscopic and macroscopic. A macroscopic traffic simulation model describes the entities and their activities at a low level of detail (Lieberman and Rathi 1997). The approach is used to capture traffic dynamics of large networks covering big geographic area. These models are often based on the deterministic relationships between the traffic entities like network level traffic conditions, route choices, roadway environment, traffic flow etc. Microscopic simulations are high fidelity models capable of representing the individual characteristics of the traffic elements at a high level of detail, while mesoscopic traffic simulation models use a mixture of these techniques.

Paper present research focused on microscopic modeling of urban traffic in area of Zvonimirova street, Zagreb, Croatia.

SITE CONTEXT

Zagreb is the capital and largest city of the Republic of Croatia. According to the results of the last official census that was held in 2011 the population of the City of Zagreb is 792875 inhabitants (Statistical Information 2011). As many capital cities, Zagreb is having problem with the traffic congestion due to increase of motorization level (from 2002. till 2007. average annual growth of motorization was 20 vehicles per 1000 inhabitants (Statistical Information 2011)). Consequently, it is necessary to shift users from private transport to public transport in order to solve this problem.
In Zagreb, there are 15 tram lines and more than a hundred bus lines (including bus lines that operate in Zagreb County). According to the results of modal split, in the City of Zagreb 22% of all journeys are made with the tram (Civitas elan 2011). The main objective of this research is to build simulation scenario that would contribute to increase of average speed, primary on tram lines. The biggest issue in reducing tram travel times is the fact that public transport (PT) vehicles share lanes with private cars. This problem is especially emphasized at approaches to major intersections when cars block the passage to trams through the intersection. For this research we chose 5,6 km long Kralja Zvonimira Street. One of main reasons for this was the fact that on this corridor tram infrastructure is physicaly separated from private cars infrastructure, allowing mesures to be evaluated without intervening in regulation (but possibly giving good motivation for it).

SIMULATION MODEL DESCRIPTION

Tram lines number 1, 9 and 17 are passing the entire length of Kralja Zvonimira Street while tram line number 13 passes only partially. Tram lines 5 and 7 and bus lines 215, 216, 217, 237 and 279 intercept corridor. On the west end of Kralja Zvonimira street there is a roundabout and on the east end there is the public transport terminal where tram lines 1, 9 and 17 terminate. Along corridor, there are 8 signalized

intersections and 8 tram stops for each tram line direction, mainly located before or after the intersection.

The traffic network model has been developed in VISSIM software (Vissim software 2013) (Figure 1 and 2). For this purpose, the necessary data were collected in order to create a high-quality model that will simulate the behavior of private car users and urban public transport. Running speed and dwell times were obtained by GPS devices and traffic count data were collected at each intersection.

The basic data necessary to develop the simulation model include: vehicle acceleration and deceleration functions, distribution of the desired speed, distribution of dwell times at stops, type, class and category of vehicle and drivers behavior. A total of 8 signaling devices with fixed exchange of signaling programs have been modeled in VISSIG tool.

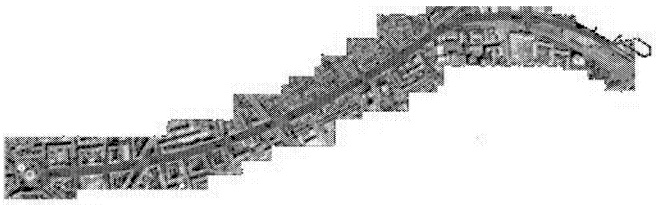


Figure 1 Kralja Zvonimira Network Modeled in VISSIM

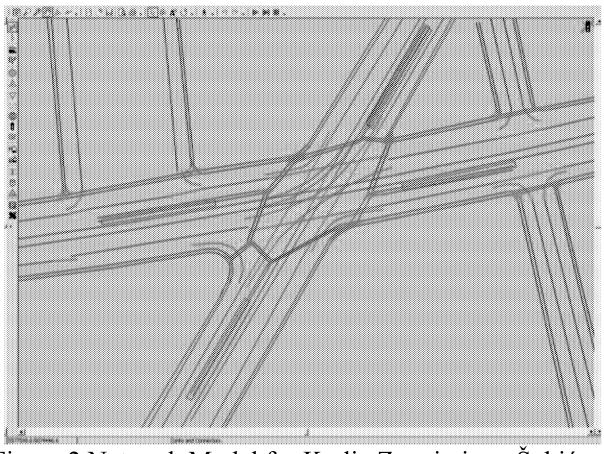


Figure 2 Network Model for Kralja Zvonimira – Šubićeva Intersection

After the modeling process, the next step was model calibration, referring to comparison of real life measured data and data obtained by the simulation. The calibration can be considered right when the determination coefficient between model and collected data is close to unity. Parameters used for models calibration are listed hereafter.

Average waiting time:

• not signalized intersections (Highway Capacity Manual 2010):

$$D_i = \frac{3600}{Ci} + 900 \cdot T \cdot \left[\frac{Q_i}{C_i} - 1 + \sqrt{\left(\frac{Q_i}{C_i} - 1\right)^2 + \frac{\left(\frac{3600}{C_i}\right) \cdot \left(\frac{Q_i}{C_i}\right)}{450 \cdot T}} \right] + 5 \text{ [sec/veic]} \quad (1)$$

Where: D_i represents the vehicle delay due to the presence of control [s/vehicle]; Q_i is the traffic flow [vehicles/hour]; C_i the capacity [vehicles/hour]; T is the duration of the period of analysis [min/60min];

• signalized intersections (Highway Capacity Manual 2010) average delay for vehicle "d" is determined as the sum of 3 elements: d_1 (uniform delay, delay in operation of uniform arrivals), d_2 (incremental delay, delay in operation of random arrivals), d_3 (initial queue delay, delay due to the queue of the previous period) with:

$$d = d1(PF) + d2 + d3 \quad (2)$$

$$d_1 = \frac{0.5C\left(1 - \frac{g}{C}\right)^2}{1 - \left[\min(1, X)\frac{g}{C}\right]} \quad (3)$$

$$d_2 = 900T\left[(X-1) + \sqrt{(X-1)^2 + \frac{8klX}{cT}}\right] \quad (4)$$

$$d_3 = \frac{1800Q_b(1+u)t}{cT} \quad (5)$$

Where: C is the cycle time, g is the actual green, X is the level of saturation, l and k are corrective coefficients, c and T are the same as for not traffic lights situation.

Average queue length:

• not signalized intersections (Highway Capacity Manual 2010). The formula used is as follows (which provides the number of motor vehicles, then multiplied by the average length of 5.6 and divided by the number of lanes):

$$N_{95,i} = 900 \cdot T \cdot \left[\frac{Q_i}{C_i} - 1 + \sqrt{\left(\frac{Q_i}{C_i} - 1\right)^2 + \frac{\left(\frac{3600}{C_i}\right) \cdot \left(\frac{Q_i}{C_i}\right)}{150 \cdot T}} \right] \cdot \left(\frac{C_i}{3600}\right) \quad (6)$$

Where: N_{95}, the 95th percentile represents the number of vehicles in queue, Q_i traffic flow [vehicles/hour]; C_i the capacity [vehicles/hour]; T the duration of the period of analysis [time] (T = 0.25 for a period of 15min);

• signalized intersections procedure involves the calculation of two different situations and the adoption, for the next steps, of the highest value among the two:

The first takes into account the queue which is also created during the green light:

$$n_1 = f\left(\frac{C - v}{2} + r\right) \quad (7)$$

Where f = incoming traffic flow, C = cycle time of traffic light, v = effective green, r = red time;

The second takes into account the only red light time period:

$$n_2 = f(C - v) \quad (8)$$

Where f = incoming traffic flow, C = cycle time of traffic light, v = effective green.

To obtain the queue length in meters Max (n_1; n_2) has to be multiplied by the average length of the vehicle, fixed equal to 5.6 m and then divided by the number of lanes.

For model calibration it is necessary to determine those parameters of the model that can be analyzed and compared with the collected data. It is important to emphasize that data used for the calibration cannot be used as input data for the model. After calibration the model is translated from general form into specific form adjusted to local conditions. For this model two sets of parameters are necessary, the first one is about private transport and the second is about public transport. The model was calibrated on the basis of queue

length on each intersection (Figure 3). Compared, calculated and simulated, values gave R^2=0,9435 and RMSE (root mean square error) of 2.88 veh / line.
For the public transport it is necessary to collect data about timetable, average speed and dwell time on each stop. Average speed and dwell time were collected using GPS logger and data about timetable are publicly available. The model calibration was based on the travel time for each tram line on the corridor with R^2 equal to 0,902 and RMSE of 6,84 sec/vehicle. Consequently, we can consider the model as affordable due to the values of R^2 and RMSE obtained.

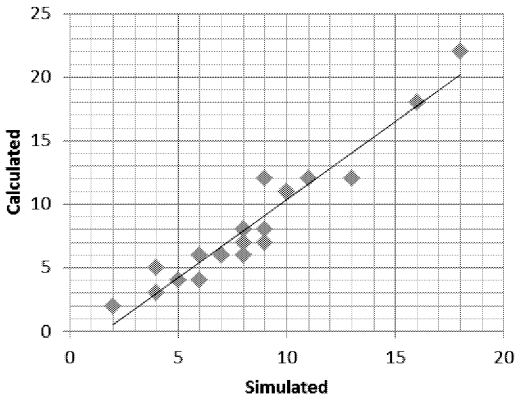


Figure 3: Calibration Parameter Queue Length [veh/line]

RESULTS ANALYSIS

After the data input, modeling process and calibration, microscopic model was set to simulate traffic conditions for corridors afternoon peak period. Based on GPS tracks analysis afternoon peak period for Kralja Zvonimira street is considered to be between 15:30 and 17:00h (Ćavar et al. 2011).
Based on simulation results Level of Service (LoS) was determined for each intersection (signalized and unsignalized) according to HCM (Highway Capacity Manual 2010). Due to very high correspondence between average waiting times obtained by the simulation and evaluated on real data, to evaluate the LoS we used a weighted average, where F_i is the real flow at the intersection i, and $\sum T_{mi}$ is the sum of simulated and calculated average waiting times at the intersection i.

$$LoS = \frac{\sum_i T_{mi} \cdot F_i}{\sum_i F_i} \quad (9)$$

Based on the simulation results it is visible that public transport speed is at an inadequate level for public transport to be competitive to the private one (Table 1) and adequate measures should be taken into account.

Table 1 Average Traveling Speeds for Private Cars and Public Transport During the Peak Periods [km/h]

Direction East -West		**Direction West-East**	
Car	Tram	Car	Tram
18.242	9.4	24.29	10.9

Reason for this lies in sequence of signalized and unsignalized intersections where PT vehicles are situated at the same level as private vehicles. To solve this issue implementation of public transport priority at signalised intersections was considered. The goal was to improve or keep existing LOS for private vehicles while improving (shortening) travel time for trams. For passive traffic signal priority average tram travel time was reduced by 4%, while for active traffic signal priority scenario this wos up to 11%.

CONCLUSION

When considering improvements of traffic network it is justified to use simulation approach to test all suggested scenarios. When doing this it is necessary to consider all relevant elements as interaction between vehicle, transportation network, signalization etc. Also, all traffic modes that share same infrastructure and have complex interaction should be included in modeling process. Considering dynamic interaction and large number of elements special attention should be given to model calibration ensuring that the model represents real world as best as possible. The level of detail used in simulation needs to be justified in regard with the simulation goal. Microscopic scale involves computational complexity problems related to the number of interacting entities resulting in fluctuation errors that cannot be avoided. Despite of this, application of traffic simulations gives the best overview of possible scenarios and aids decision makeing process at highly efficient way.

REFERENCES

Ćavar, I., Kavran, Z., Petrovic, M. 2011."Hybrid approach for urban roads classification based on GPS tracks and road subsegments data", *PROMET – Traffic&Transportation*, No. 4,Vol. 23, 289-296

Highway Capacity Manual. 2010. TRB Publications, 2010.

Kvaliteta usluge u javnom gradskom prijevozu. 2011. Civitas elan document, Fakultet prometnih znanosti, Sveučilište u Zagrebu (in Croatian)

Lieberman E. and Rathi A.K.1997. "Traffic simulation' in Traffic flow theory", Gartner and Messer, eds, Oak Ridge National Laboratory. 1997.

Statistical Information. 2011. The Croatian Bureau of Statistics

Sadat Hoseini, S.M.. 2011. "Comparison of microscopic drivers' probabilistic lane-changing models with real traffic microscopic data" ", *PROMET – Traffic&Transportation*, No. 4,Vol. 23, 241-251

WEB REFERENCES

PTV Vissim software, http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/functions/ (March 2013)

AUTHOR LISTING

AUTHOR LISTING

Aghezzaf E.-H. **192**
Alupei-Cojocariu O.-D. . **155**
Amaral R.R. **139**
Anthonis J. **67**

Bernedixen J. **187**
Bilegan R.M. **88**
Brahmadeep **127**

Călinoiu C. **75**
Castells M.P. **177**
Ćavar I. **240**
Celda A.G. **177**
Conley W. **13/208**
Croitorescu V. **67**

Decock K. **192**
Derka I. **95/101**
Drăgoi C. **83**
Dubovská R. **119**

Feher S. **155**
Föhr J. **203**

Gábor Lencse G. **101**
Gheorghe (Nicolici) C. .. **88**
Ghosh T. **51**
Gimeno R.V.C. **177**

Hellfeld S. **28**
Hoppen M. **147**
Huitema G.B. **23**

Irimia P.-C. **59/83**
Ittoo A. **23**

Karttunen K. **203**
Klumpp M. **5/160**
Korpinen O.-J. **203**
Kortelainen S. **198**
Kuchshaus V. **160**
Kvasnica I. **119**
Kvasnica P. **119**

Lättilä L. **198/203**
Lättilä L. **203**
Limère V. **139/192**
Lokshina I. **108**

Mitroi M. **75**
Muka L. **95**

Ng A.H.C. **171/187**
Nicolici S. **88**

Ottoy P. **139**
Overmeyer L. **165**

Pehrsson L. **171/187**
Petrović M. **240**
Pîrăianu V.F. **83**
Plumé J.M. **177**
Potuzak T. **235**
Prisecaru I. **88**
Puhalschi R. **75**

Quiocho L.J. **51**

Ramirez-Martinez A. **132**
Ranta T. **203**
Reuter M. **114**
Romanin-Jacur G. **34**
Rossmann J. **147**
Roudot A.-C. **132**

Sauer J. **165**
Schachmanow J. **165**
Schiele E. **108**
Schluse M. **147**
Schwarz C. **165**
Segala F. **34**
Służalec A. **39**
Stockton D. **187**
Studzinski J. **44**
Syberfeldt A. **219/227**
Szirbik N.B. **23**

Thomassey S. **127**

Ullmann G. **165**

Vasiliu D. **59**
Vasiliu G.C. **59**
Vasiliu N. **59/67/75/155**
Vujić M. **240**

Wesolek N. **132**
Wortmann J.C. **23**

Ziółkowski A. **39**