

**7<sup>th</sup> INTERNATIONAL INDUSTRIAL  
SIMULATION CONFERENCE  
2009**

**ISC'2009**

**EDITED BY  
Diganta Bhusan Das  
Vahid Nassehi  
and  
Lipika Deka**

**JUNE 1-3, 2009  
LOUGHBOROUGH, UNITED KINGDOM**

**A Publication of EUROSIS-ETI**





7<sup>th</sup> Industrial Simulation Conference 2009

LOUGHBOROUGH, UNITED KINGDOM

JUNE 1-3, 2009

Organised by

ETI- The European Technology Institute

In cooperation with

Loughborough University

I-Chem

Sponsored by

EUROSIS - The European Simulation Society

Co-Sponsored by

AEKI

CREAX

ENSAIT

Ghent University

UPV

Hosted by

Compass Hotel

Loughborough, United Kingdom

## EXECUTIVE EDITOR

**PHILIPPE GERIL  
(BELGIUM)**

## EDITORS

### General Conference Chairs

Diganta Das, Loughborough University, Loughborough, United Kingdom  
Vahid Nassehi, Loughborough University, Loughborough, United Kingdom  
Lipika Deka, Loughborough University, Loughborough, United Kingdom

### LOCAL COMMITTEE

Prof. Paul Chung, Loughborough University, Loughborough, United Kingdom  
Dr. Navraj Hanspal, Manchester University, Manchester, United Kingdom  
Prof. W. Malalasekera, Loughborough University, Loughborough, United Kingdom  
Prof. Jim McGuirk, Loughborough University, Loughborough, United Kingdom  
Prof. Vahid Nassehi, Loughborough University, Loughborough, United Kingdom  
Prof. Dani Or, Swiss Federal Institute of Technology, Zurich, Switzerland  
Prof. Chris Rielly, Loughborough University, Loughborough, United Kingdom  
Prof. Victor Starov, Loughborough University, Loughborough, United Kingdom

### INTERNATIONAL PROGRAMME COMMITTEE

#### Discrete Simulation Methodology, Languages and Tools

Matthias Becker, University Hannover, Hannover, Germany  
Carmen Bobeanu, Ghent University, Ghent, Belgium .  
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  
Panagiotis Katsaros, Aristotle University of Thessaloniki, Thessaloniki, Greece  
Sarka Kvetonova, Brno University of Technology, Brno, Czech Republic  
Jan Lemeire, VUB, Brussels, Belgium  
Goreti Marreiros, Polytechnic of Porto, Porto, Portugal  
Stefano Marrone, Seconda Università di Napoli, Naples, Italy  
Maria do Rosário Moreira, University of Porto, Porto, Portugal  
Aziz Naamane, ISIS-EPUM, Marseille, France  
Jose Neves, Universidade do Minho, Braga, Portugal  
Paulo Novais, Universidade do Minho, Braga, Portugal.  
Werner Pohlmann, Universitaet Salzburg, Salzburg, Austria.  
Jiri Safarik, University of West Bohemia, Plzen, Czech Republic  
Werner Sandmann, TU-Clausthal, Clausthal, Germany.  
Antonella Zanzi, Universita' degli Studi dell'Insubria, Varese, Italy

#### Simulation in Manufacturing

Reza Azadegan, Urmia University, Urmia, Iran  
Peter Byrne, Dublin City University, Dublin, Ireland  
Ana Camacho, UNED, Madrid, Spain  
Luis Camarinha-Matos, Universidade Nova de Lisboa, Lisbon, Portugal  
Eduardo Castellano, IKERLAN Technological Research Centre, Mondragón-Arrasate, Spain  
Remy Dupas, Universite Bordeaux 1, Bordeaux, France.  
Alexander Felfernig, University of Klagensfurt, Klagensfurt, Austria  
Pouria Homayonifar, REC Scanwafer, Prossgrunn, Norway  
Habtom Mebrahtu, APU, Chelmsford, United Kingdom  
Carlo Meloni, Politecnico di Bari - DEE, Bari, Italy  
Pascal Meyer, Forschungszentrum Karlsruhe, Karlsruhe, Germany  
Marcos A. Rodrigues, Sheffield Hallam University, Sheffield, United Kingdom  
Marina Valles, UPV, Valencia, Spain .  
Markus Vorderwinkler, PROFACTOR GmbH, Steyr, Austria

## INTERNATIONAL PROGRAMME COMMITTEE

### **Simulation in Steel Manufacturing**

Brian Hollocks, Bournemouth University, Bournemouth, United Kingdom

### **Simulation in Automotive Systems**

Naoufel Cheikhrouhou, EPFL, Lausanne, Switzerland

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

### **Simulation in Electronics, Computer and Telecommunications**

Christos Bouras, University of Patras, Patras, Greece

Boguslaw Butrylo, Bialystok Technical University, Bialystok, Poland

Tom Dhaene, Ghent University, INTEC, Ghent, Belgium

Domenico Giunta, ESA-ESTEC, Noordwijk, The Netherlands

Paulo Maciel, University of Pernambuco, Recife, Brazil

Silvia Mirri, University of Bologna, Bologna, Italy

Maurizio Palesi, Università di Catania, Catania, Italy

Marco Rocchetti, University of Bologna, Bologna, Italy

Fernando Boronat Segui, UPV, Valencia, Spain

Renate Sitte, Griffith University, Gold Coast, Australia

### **Simulation of Complex Multiprocessor Systems**

Orhan Gemikonakli, Middlesex University, London, United Kingdom

### **Simulation in Computer Science**

Paul Chung, Loughborough University, Loughborough, United Kingdom .

Lipika Deka, Tezpur University and IIT Guwahati, India

### **Simulation in Logistics, Transport and Harbour Simulation**

El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium

Christian Almeder, University of Vienna, Vienna, Austria

Isabel Garcia Gutierrez, Univ. Carlos III de Madrid, Madrid, Spain

Martin Grunow, Technical University of Denmark, Lyngby, Denmark

Thomas Hanne, University of Applied Sciences Northwestern Switzerland, Olten, Switzerland

Peter Lawrence, Monash University, Clayton, Australia

Marie-Ange Manier, UTBM, Belfort, France

Roberto Montemanni, IDSIA, Manno-Lugano, Switzerland

Jaap Ottjes, TU Delft, Delft, The Netherlands.

Selwyn Piramuthu, University of Florida, Gainesville, USA

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

Antonio Abelha, Universidade do Minho, Braga, Portugal

Jose Machado, University of Minho, Braga, Portugal

## INTERNATIONAL PROGRAMME COMMITTEE

### Complex Systems Modelling

Frantisek Capkovic, Slovak Academy of Sciences, Bratislava, Slovak Republic  
Christophe Claramunt, Naval Academy Research Institute, Brest, France  
Gabiella Dellino, University of Siena, Siena, Italy  
Werner Dubitzky, University of Ulster, Coleraine, United Kingdom  
Krzysztof Fajarewicz, Silesian Technical University, Gliwice, Poland  
László Gulyás, AITIA International Inc., Budapest, Hungary  
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

Cyril Ray, Ecole Navale, Brest, France

### Simulation in Industrial Design and Product Design

Chiara Catalano, IMATI-CNR, Genoa, Italy  
Yan Luo, NIST, Gaithersburg, USA  
Jorge Perez Mateos, Industria de Turbopropulsores SA, Madrid, Spain  
Catarina Rizzi, University of Bergamo, Bergamo, Italy

### Simulation in Engineering Processes

Chrissanti Angeli, Technological Institute of Piraeus, Athens, Greece  
Blazej Balasz, Technical University of Koszalin, Koszalin, Poland  
Alejandra Gomez Padilla, University of Guadalajara, Mexico  
Fouad Riane, CREGI-FUCAM, Mons, Belgium  
Christopher Schlick, RWTH Aachen, Aachen, Germany  
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  
Ashu Jain, IIT Kanpur, India  
Graham Saunders, Loughborough University, Loughborough, United Kingdom

### Simulation in Energy and Power Systems

Jean-Frederic Charpentier, Ecole Navale, Brest, France  
Manuel Cordeiro, CETAV/UTAD, Vila Real, Portugal  
Navraj Hanspal, Manchester University, Manchester, United Kingdom  
Janos-Sebestyen Janosy, KFKI Atomic Energy Research Institute, Budapest, Hungary  
Gustaf Olsson, Lund University, Lund, Sweden

### Simulation in Multibody Systems

Track Chair: Bernd Schaefer, DLR, Wessling, Germany  
Ignacio García-Fernández, University of Valencia, Valencia, Spain  
Jose Diaz Lopez, Dynasim AB, Lund, Sweden

### Simulation in Chemical, Petroleum and Mining Engineering

Rene Banares-Alcantara, Oxford University, Oxford, United Kingdom  
Diganta Das, Loughborough University, Loughborough, United Kingdom  
Vahid Nassehi, Loughborough University, Loughborough, United Kingdom  
Mohamad R. Riazi, Kuwait University, Kuwait  
Raj Sharma, India

## INTERNATIONAL PROGRAMME COMMITTEE

### **Simulation in Military and Defense**

Roberto de Beauclair Seixas, IMPA, Rio de Janeiro, Brazil  
Carlos Palau, UPV, Valencia, Spain  
Matthias Reuter, CUTECH 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

### **Verification, Validation and Accreditation**

Mokhtar Beldjehem, École Polytechnique de Montréal, 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 Modelling and Simulation in the Textile Industry**

Track Chair: Vladan Koncar, ENSAIT, Roubaix, France  
Antonela Curteza, "Gh. Asachi" Technical University of Iasi, Iasi, Romania  
Jiri Militky, Technical University Liberec, Liberec, Czech Republic

### **Workshop on Intelligent Transport Systems**

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, Purdue University, West Lafayette, 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**

El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium  
Hendrik Van Landeghem, Ghent University, Ghent, Belgium



# **INDUSTRIAL SIMULATION 2009**

© 2009 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.

All articles published in these Proceedings have been peer reviewed

EUROSIS-ETI Publications are ISI-Thomson and INSPEC referenced

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

EUROSIS-ETI Publication  
**ISBN: 978-90-77381-4-89**  
**EAN: 978-90-77381-4-89**



## PREFACE

Phenomenal advances in computer technology together with progress in the areas of mathematical modelling in recent decades have made simulation procedures very powerful tools for the analysis and design of almost all types of industrial and economic processes. Accurate and reliable predictions about the outcome of complex natural transport processes and performance of novel designs for industrial equipments are routinely made using modern simulation methodologies. The annually held international Industrial Simulation Conferences (ISC), organised and run by EUROSIS in conjunction with various organisations, provides an important forum for the presentation and exchange of new ideas related to the development and application of computer simulation techniques in a very diverse and wide ranging area of industrial relevance.

We are very pleased to be able to co-organise the 7<sup>th</sup> ISC at Loughborough University with EUROSIS. In continuation with its rich tradition, the ISC'2009 will provide an excellent overview of industrial simulations and related research within the European Community and the rest of the world. The varied simulation work presented in the conference relate to a wide ranging problems of industrial importance. These include robotics and automotive industry, sensors and process control, traffic simulation, multibody simulation, biomedical engineering, data mining, manufacturing processes, chemical engineering, etc. More information on the areas of the conference can be found in the programme. Although various engineering disciplines are covered as separate themes, the focus of the conference remains interdisciplinary. This allows one from a particular area to learn from the simulation methods in another area and apply them in his/her areas of research.

The present proceedings consists of the papers which were selected after a rigorous refereeing process for presentation at ISC'2009. We received 76 papers for the conference out of which 69 were accepted. We are pleased to see a high standard of the papers originating from many countries in the world. The conference programme will be further enhanced by keynote lectures by three renowned experts, namely, Prof Allison Noble (Oxford University, UK), Prof Adel Sharif (University of Surrey, UK) and Professor Philip Eames (Loughborough University, UK), an invited lecture by Mr Duncan Forgan (University of Edinburgh, UK), software demonstration and a number of visits to various laboratories at Loughborough University.

Loughborough is a town in Leicestershire, central England with a population of more than 57,000 as of 2004. It is about 90 minutes away from London by train and only a short distance away from major cities of Leicester, Nottingham and Derby. It is the administrative centre for the Charnwood district and home to Loughborough University. Loughborough University, which is celebrating its centenary in 2009, has been at the core of many scientific and technological advances and continues to excel as a premier research and teaching institute in the UK. For example, the Sunday Times University Guide awarded the university the coveted 'University of the Year' title for the year 2008-2009 in recognition of its excellent all-round performance among all the higher education institutes in the UK

On behalf of the Organising and International Programme Committees of ISC' 2009, we would like to welcome you to Loughborough. We are certain that ISC'2009 will provide an important platform for the development of new opportunities, collaborations, ideas and inspirations.

Diganta Bhusan Das, Vahid Nassehi, Lipika Deka  
Editors, Conference Proceedings, ISC 2009, Loughborough, UK



## CONTENTS

|                                   |            |
|-----------------------------------|------------|
| <b>Preface .....</b>              | <b>XI</b>  |
| <b>Scientific Programme .....</b> | <b>1</b>   |
| <b>Author Listing.....</b>        | <b>359</b> |

## DISTRIBUTED AND PARALLEL SIMULATION

|  |           |
|--|-----------|
| <b>PARMAX – A Predictor for Distributed Time Series</b><br>Dan Stefanoiu and Janetta Culita .....                          | <b>5</b>  |
| <b>The Use of Refined Descriptive Sampling in Parallel Simulation</b><br>Abdelouhab Aloui and Megdouda Ourbih ex Tari..... | <b>13</b> |

## VALIDATION, VERIFICATION AND OPTIMIZATION

|  |           |
|--|-----------|
| <b>Parameter Validation using Constraint Optimization for Modeling and Simulation</b><br>Guodong Shao, Charles McLean, Alexander Brodsky and Paul Ammann .....                                 | <b>21</b> |
| <b>Applying a Documentation Guideline for Verification and Validation of Simulation Models and Applications: An Industrial Case Study</b><br>Zhongshi Wang, Heike Kißner and Martin Siems..... | <b>26</b> |
| <b>Performances evaluation and optimization of a failure-prone manufacturing system with random delivery time and random demand</b><br>Sadok Turki, Sophie Hennequin and Nathalie Sauer .....  | <b>31</b> |

## PRODUCTION AND PRODUCT ANALYSIS

|  |           |
|--|-----------|
| <b>Linear and Nonlinear Input Output Analysis</b><br>William Conley .....  | <b>39</b> |
| <b>Latent Class Mixture models for analyzing rating responses</b><br>Liberato Camilleri .....  | <b>42</b> |
| <b>Predicting Post Machining Distortions via FE Simulations at Design Phase</b><br>Waqas Saleem and Fan Yuqing .....                               | <b>47</b> |
| <b>On-Line Fault Detection of a Close Loop Circuit Hydraulic System with a Variable Displacement Pump</b><br>C. Angeli and A. Chatzinikolaou ..... | <b>52</b> |

## CONTENTS

|  |    |
|--|----|
| <b>Comparison of Steel, Aluminum and Composite Bonnet Concerning Adult/Child Pedestrian Protection</b><br>Abolfazl Masoumi and Amir Najibi ..... | 57 |
|--|----|

## DATA MINING

|  |    |
|--|----|
| <b>Information Extraction from Solution Set of Simulation-based Multi-objective Optimisation using Data Mining</b><br>Catarina Dudas, Amos Ng and Henrik Boström ..... | 65 |
| <b>Using Ant Colony in optimizing Associative Classification</b><br>Kifaya Qaddoum .....   | 70 |

## AI APPLIED TO MANUFACTURING

|   |     |
|---|-----|
| <b>A Hierarchical Object-Oriented Software Structure for Manufacturing Simulation</b><br>Lina A. M. Huertas Quintero, Andrew A. West, Diana M. Segura Velandia, Paul P. Conway and Radmehr Monfared .....       | 79  |
| <b>Knowledge Management of Sheet Hydroforming Process</b><br>Antonio Del Prete, Alfredo Elia and Gabriella Romano .....   | 86  |
| <b>Memory Based Immune Network for Multi-Robot Cooperation</b><br>Sazalinsyah Razali, Qinggang Meng and Shuang-Hua Yang .....   | 91  |
| <b>A fuzzy randomness approach for preventive and corrective maintenance</b><br>F. Aguirre, S. Hennequin and N.Rezg .....   | 97  |
| <b>A threaded Java concurrent implementation of the Monte-Carlo Metropolis Ising model</b><br>Carlos Castañeda-Marroquín, Alfonso Ortega de la Puente, Manuel Alfonseca, James A. Glazier and Maciej Swat ..... | 103 |

## SIMULATION IN CIM

|   |     |
|---|-----|
| <b>New SysML based approach for integrated system design</b><br>G. Auriol and C. Baron .....  | 111 |
| <b>A Virtual Enterprise Application for the Textile and Apparel Industry</b><br>Antonela Curteza, Daniela Farima, Ovidiu Stoica and Liviu Lupu..... | 116 |

## CONTENTS

### **Embedding Simulation in Lean Projects**

Yan Jia and Terrence Perera ..... 122

### **Software Tools Integration for the Design of Manufacturing Systems**

Pavel Vik, Guilherme Pereira and Luís Dias..... 127

### **Optimization of Industrial Robot Work Cells using the Automated Variant Simulation**

Jürgen Roßmann, Roland Wischnewski and Patrick Lenk..... 133

## **PRODUCTION AND TRANSACTIONS LOGISTICS**

### **Understanding Supply Chain Performance by Simulation**

Alejandra Gomez-Padilla ..... 141

### **Simulation of a Milk Run Material transportation System in the Semiconductors Industry**

Ricardo Raposo, Guilherme Pereira and Luís Dias..... 144

### **Implementation Steps for Development of a simulation Application Software for Production Scheduling and Door-to-Door Logistics Management of a Multi-Assembly-Lines Automotive Vehicle Manufacturing Plants**

Hossein Hosseinzadeh, Mohsen Bahrami and Alireza Hosseinzadeh..... 152

### **A study on transaction scheduling in a real-time distributed system**

Y. Jayanta Singh, Hasan Al-Saedy and S.C. Mehrotra ..... 157

## **URBAN AND INTERMODAL TRAFFIC SYSTEMS**

### **Analyzing in Flight Air Hostess Tasks through Discrete Event Simulation**

Mehmet Savsar and Esra E. Aleisa ..... 165

### **A Software Environment for Microscopic Pedestrian Simulation**

Edgar F. Esteves, Rosaldo J. F. Rossetti and Eugénio C. Oliveira..... 173

### **Using Heuristic Search for Initiating the Genetic Population in Simulation Based Optimization of Vehicle Routing Problems**

Anna Syberfeldt and Lars Persson ..... 178

### **Towards a Microscopic Traffic Simulation Framework to Assess Vehicle-to-Vehicle Networks**

João F. B. Gonçalves, Rosaldo J. F. Rossetti, Edgar F. Esteves and Eugénio C. Oliveira ..... 183

## CONTENTS

|   |     |
|---|-----|
| <b>Development and Evaluation of Traffic Management Strategies for Personal Rapid Transit</b><br>Pengjun Zheng, David Jeffery and Mike McDonald ..... | 191 |
| <b>Microsimulation Models in Intermodal Container Terminals ordinary and perturbed Conditions</b><br>Vincenzo Assumma and Antonino Vitetta.....       | 196 |

## HOSPITAL INFRASTRUCTURE LOGISTICS

|   |     |
|---|-----|
| <b>Quality of Service in Transplantation via the Electronic Medical Record</b><br>David Belo, Miguel Miranda, Antonio Abelha, Jose Machado and Jose Neves .....   | 203 |
| <b>System Dynamics Approach for Modelling Complex Healthcare Systems</b><br>Ruby Wai Chung Hughes and Terrence Perera .....   | 209 |
| <b>Paediatric palliative care planning: models and simulation</b><br>Giada Aspergh, Paola Facchin, Anna Ferrante, Laura Visonà Dalla Pozza and Giorgio Romanin Jacur .....  | 214 |
| <b>Birth and perinatal assistance network on the territory: model and simulation of service dynamical behaviour</b><br>Paola Facchin, Anna Ferrante, Elena Rizzato, Laura Salmaso and Giorgio Romanin-Jacur ..... | 219 |
| <b>Critical newborn assistance in intensive care units: model and simulation</b><br>Monica Da Frè, Paola Facchin, Elena Rizzato, Laura Salmaso, Laura Visonà Dalla Pozza and Giorgio Romanin-Jacur .....          | 222 |

## MULTIBODY SIMULATION

|   |     |
|---|-----|
| <b>Methodology for Flexible Modeling of Escalator Multibody Systems</b><br>Juan D. Cano-Moreno, José M <sup>a</sup> Cabanellas Becerra, Carlos Labajo Tirado and Jesús Félez Mindán ..... | 227 |
| <b>Spatial Kinematics of Gears in Absolute Coordinates</b><br>Dmitry Vlasenko and Roland Kasper .....   | 232 |
| <b>Modeling and Simulation of Gait for <i>MERO</i> Modular Walking Robot crossing an unarranged Terrain</b><br>Ion Ion, Alexandru Marin, Cristian Doicin and Constantin Chirita .....     | 237 |
| <b>Modeling, Simulation and Control of Rotary Wing Platforms in a Computer Generated Forces Toolkit</b><br>Semuel Franko, Seniha Köksal and Mehmet Haklıdır.....                          | 243 |

## SENSORS AND PROCESS CONTROL

|   |            |
|---|------------|
| <b>Multivariate Control Structure Synthesis for Fluid Catalytic Cracking Unit Using Relative Gain Array</b><br>Redah Alsabei, Zoltan K. Nagy and Vahid Nassehi .....                        | <b>251</b> |
| <b>Soft-sensor for on-line mass of crystals measurement: application to cane sugar crystallization</b><br>Cédric Damour, Michel Benne, Brigitte Grondin-Perez and Jean-Pierre Chabriat..... | <b>257</b> |
| <b>Design General Fuzzy Controller Implemented in VHDL and Synthesized using FPGA</b><br>Rasha E. Maged and Basil Sh. Mahmood.....  | <b>262</b> |
| <b>Adjustment of conformity parameters of PID-type compensators using simulation by AMESim</b><br>Teodor Costinel Popescu, Iulian Dutu, Catalin Vasiliu and Marius Mitroi .....             | <b>269</b> |

## FLUID FLOW SIMULATION IN ENGINEERING APPLICATIONS

|  |            |
|--|------------|
| <b>Notes on the Implementation of Brownian motion in Mesoscopic Fluid Particle Models</b><br>Massimo Lai and Dimitris Drikakis .....   | <b>277</b> |
| <b>Estimation of Longitudinal Aerodynamics of a Transport Aircraft Using FDR Data through Neural Networks</b><br>Ming-Hao Yang, Cjing-Shun Ho, Fei-Bin Hsiao, C. Edward Lan and Chien-Chun Kung..... | <b>280</b> |

## FLUID FLOW SIMULATION IN DESIGN

|   |            |
|---|------------|
| <b>Numerical Simulation of Refrigerant Flow Through Adiabatic Capillary Tube: A Comparison between Homogeneous Equilibrium Model (HEM), Delayed Equilibrium Model (DEM) and Improved Delayed Equilibrium Model (IDEM)</b><br>Abdul Qaiyum Shaik and Henk Versteeg ..... | <b>287</b> |
| <b>Simulation Based Optimization of Microneedle Geometry to Improve Drug Permeability in Skin</b><br>Ololade Olatunji, Diganta B Das and Barrak Al-Qallaf .....   | <b>293</b> |

## CONTENTS

### **Modelling of Flashing Propellant Flow in Pressurised Metered Dose Inhalers**

Henk Versteeg and Abdul Qaiyum Shaik ..... 301

## **POROUS MEDIA AND CHEMICAL ENGINEERING SIMULATION**

### **Modelling of three dimensional Darcy flow through soils**

Bahareh Kaveh-Baghbaderani, Vahid Nassehi and Abhijeet Kulkarni..... 309

### **Mathematical modelling of Biofilm Growth**

Yi-Ping Lo, John Ward, Felipe Iza, Rob Seager and Michael Kong..... 315

### **Modelling, Simulation and Control of a Semi-Batch Industrial Polymerization Reactor**

Nadja Hvala, Teodora Miteva and Dolores Kukanja..... 318

### **Solving Challenging Problems in the Oil Industry using Artificial Intelligence Based Tools**

Ana Vila Fernandes, Henrique Vicente and José Neves..... 325

### **Multi Project Scheduling in the Chemical Industry using a Genetic Algorithm**

Sven Tackenberg, Sebastian Schneider and Christopher M. Schlick..... 332

## **TEXTILE SIMULATION**

### **Statistical Analysis of the Results of Some Technologies for Cotton Fireproofing**

Vasilica Popescu, Liliana–Rozemarie Manea and Gabriel Popescu..... 347

### **Objective Hierarchization of Some Technologies for Cotton Fireproofing**

Vasilica Popescu, Liliana–Rozemarie Manea and Gabriel Popescu..... 352



# **SCIENTIFIC PROGRAMME**



# **DISTRIBUTED AND PARALLEL SIMULATION**



# PARMAX – A PREDICTOR FOR DISTRIBUTED TIME SERIES

Dan Stefanoiu, Janetta Culita  
Dept. of Automatic Control and Computer Science  
„Politehnica” University of Bucharest,  
313 Splaiul Independentei 060032 Bucharest, ROMANIA  
E-mails: danny@indinf.pub.ro, jculita@yahoo.com

## KEYWORDS

Prediction, MIMO-ARMAX models, distributed signals, prediction quality, Particle Swarm Optimization.

## ABSTRACT

Within this article, the problem of multi-variable signals prediction is approached. Such signals are provided especially by natural phenomena with geographical distribution. Presumably, the set of time series generated by the same source are more or less correlated each other. Instead of processing each time series independently, they can be seen as measurable outputs provided by some open MIMO system. Therefore, modeling and forecasting of system behavior can rely on multi-dimensional ARMAX models. The research based on this idea led us to construction of PARMAX predictor. Optimal structural indices of MIMO-ARMAX model can be found by means of an evolutionary strategy: Particle Swarm Optimization.

## 1. INTRODUCTION

Evolution of many natural phenomena can be monitored with higher accuracy when the observation and measurement are accounting not only the time variation of some parameter, but also its distribution over a geographical area. Take for example the snow thickness monitoring application on White Valley from French Alps (see Figure 1).

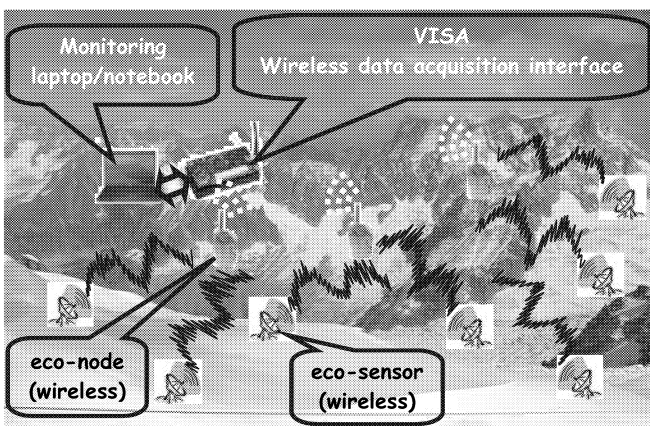


Figure 1: Snow Thickness Measuring on White Valley

The monitoring system consists of 3 layers. The lowest level layer includes on-field wireless sensors (*eco-sensors*). Signals are first gathered by *eco-nodes*, then labeled and transmitted at middle level layer – a versatile acquisition interface referred to as *VISA* [Stefanoiu 2008a]. *VISA* performs primary filtering and data organizing, before sending them to the highest level layer – a laptop in charge

with monitoring. In this context, by monitoring, one understands data prediction mainly.

When using a stand alone sensor, the main problem is to locate it in such a vast environment. As Figure 1 is suggesting, the thickness varies both in time and space. A network of wireless sensors is seemingly more suitable to perform monitoring than isolated sensors. Several *time series* (*ts*), from different locations are thus provided. Data coming from different channels are in general more or less correlated. For example, the snow melting on top may increase the thickness at the bottom of the valley (especially when avalanches are produced). Assume that the monitoring goal is to predict the snow thickness in order to prevent avalanches. It is very likely that better prediction results be obtained when considering the collection of all data sets, rather than when building independent prediction models for each of them.

A quick approach to the prediction problem is to use data fusion techniques [Wang 2005], followed by prediction. After fusion, a unique *ts* results [Stefanoiu 2008a] and, thus, on one hand, some prediction model can easily be implemented. On the other hand, after fusing the data, much of the prediction information provided by each sensor is lost, which strongly limits the prediction accuracy.

A different approach is described within this paper. One considers that sensors provide data as result of unknown colored noises that stimulate an open and quasi-ubiquitous system. The system has in fact a continuous collection of variable states. Placing a finite set of sensors at different locations, in order to perform measuring, is equivalent to sampling the system both in time and space. While the time sampling is governed by Shannon-Nyquist rules, sampling the system over space is by far more empirical. To the best of our knowledge, no general rules of space sampling are currently available. However, apparently, simulations have shown that the prediction quality is less affected by the sensors location than by their number.

This paper is structured as follows. Section 2 succinctly describes the construction of MIMO-ARMAX predictors. Section 3 is devoted to *Particle Swarm Optimisation (PSO)* [Kennedy 1995] – an evolutionary technique to find optimal structural indices of prediction model. The performance of PARMAX predictor is demonstrated within the last section, within an application of temperature monitoring. A conclusion and the references list complete the article. (More details can be found in [Stefanoiu 2008c].)

## 2. KERNEL OF PARMAX: MIMO-ARMAX MODELS

In a previous research [Stefanoiu 2008c], a comparison between classical prediction methods (based on ARMA modeling) and a wavelet-based method has been realized. But the analyzed ts were mono-variable. In case of multi-variable ts, the SISO-ARMA model can be extended to a MIMO-ARMAX model [Söderström 1989], according to the rationale described in this section.

Let  $\mathcal{Y} = \{y_j\}_{j \in \overline{1, ny}}$  be the set of  $ny$  ts provided by some distributed set of sensors and denote the maximum number of acquired data by  $N_y$ . Before any other operation, data of  $\mathcal{Y}$  have to be synchronized. Synchronization may involve interpolation (e.g. spline type) and re-sampling, such that each data set includes exactly  $N_y$  samples. This allows construction of output vector  $\mathbf{y} = [y_1 \cdots y_{ny}]^T$ , which belongs to some MIMO system. Thus, a general MIMO-ARMAX model can be assigned to data set  $\mathcal{Y}$ :

$$\mathbf{A}(q^{-1})\mathbf{y} \equiv \mathbf{B}(q^{-1})\mathbf{u} + \mathbf{C}(q^{-1})\mathbf{e}. \quad (1)$$

In equation (1),  $\mathbf{A} \in \mathbb{R}^{ny \times ny}(q^{-1})$ ,  $\mathbf{B} \in \mathbb{R}^{ny \times nu}(q^{-1})$  and  $\mathbf{C} \in \mathbb{R}^{ny \times ny}(q^{-1})$  are matrix polynomials,  $\mathbf{u} \in \mathbb{R}^{nu}$  is the input vector ( $nu$  channels) and  $\mathbf{e} \in \mathbb{R}^{ny}$  is the vector of Gaussian uncorrelated white noises. Two main problems have to be solved, to identify the model (1) from data  $\mathcal{Y}$ : specifying the model structure and estimating the input signals. If all polynomials of matrices  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  are non null, then identification is extremely difficult to perform. Firstly, because the number of parameters to estimate is quite large (even for small number of input-output channels). Secondly, because the output channels are mixed within the AR part (the left side of equation (1)). Therefore, some simplifying hypotheses are necessary. One of the most natural assumptions has been adopted within MATLAB environment: each output vary independently on the other outputs, solely as result of stimulation with inputs and noises. This means the matrix  $\mathbf{A}$  is diagonal. The second assumption regards the white noises: since they are uncorrelated each other, only one noise per output channel can be kept. Thus, the matrix  $\mathbf{C}$  is diagonal as well. As direct consequence of the two hypotheses, the MIMO-ARMAX model can be split into  $ny$  MISO-ARMAX models (with natural notations):

$$A_j(q^{-1})y_j \equiv B_j(q^{-1})u + C_j(q^{-1})e_j, \quad \forall j \in \overline{1, ny}. \quad (2)$$

This time,  $B_j \in \mathbb{R}^{1 \times nu}(q^{-1})$  is a row vector of polynomials. Identification of MIMO-ARMAX model reduces now to estimation of parameters for each MISO-ARMAX model, separately. Although models (1) and (2) are not equivalent (in the second one, correlation between outputs was removed), identification is now affordable. Despite output channels isolation, as it will be shown, prediction can be performed only when solving all equations (2) at each step. Thus, correlations between outputs are indirectly encoded by the collection of MISO-ARMAX models.

Obviously, since only the output data set  $\mathcal{Y}$  is available, it is necessary to derive a method for estimating the inputs  $\mathbf{u}$ .

The main idea is to come back to the unknown perturbations that stimulate the system to evolve. Since the hidden correlations between output channels have been removed, they can partially be reconsidered by selecting colored noises of measured data as inputs. More specifically, let  $\{e_j\}_{j \in \overline{1, ny}}$

be the white noises corrupting the measured data. By filtering, they are producing colored noises that corrupted output channels. Since the perturbation of channel  $j$  has already been accounted by the MA part of equations (2) (the second right term), only noises of the other output channels,  $\{e_i\}_{i \in \overline{1, ny}, i \neq j}$ , are affecting the data through input channels.

Consequently,  $nu = ny - 1$  and  $u_i \equiv e_i, \quad \forall i \in \overline{1, ny}, i \neq j$ . Practically, matrix  $\mathbf{B}$  (of (1)) has null diagonal, in this case. The colored noises are then:  $\mathbf{v} \equiv \mathbf{B}\mathbf{e}$ .

Since the filters are unknown, estimating inputs means estimating filters parameters. Fortunately, the measured data encode the information regarding colored noises. This allows identification of filters with some accuracy. Given the equations of model (2), it is natural to consider that filters belong to the same models class. A simple assumption yields expressing the filters equations: each output data set is a colored noise produced by the corresponding white noise. Consequently, data for each channel can roughly be modeled by means of a (SISO-)ARMA filter:

$$A_j^s(q^{-1})y_j \equiv C_j^s(q^{-1})e_j, \quad \forall j \in \overline{1, ny}, \quad (3)$$

where  $A_j^s$  and  $C_j^s$  are (singleton) polynomials for each output channel  $j \in \overline{1, ny}$ . Identification of models (3) is simply performed via Minimum Prediction Error Method (MPEM) [Söderström 1989]. Moreover, the procedure based on MPEM returns estimates of white noise values. (By convention, estimated parameters or signals carry a hat over their notations.) After identification of models (3), the prediction error on each channel can roughly be estimated. This error actually stands for the input on some channel:

$$\hat{u}_i \equiv \hat{e}_i \equiv \hat{A}_i^s(q^{-1})y_i + [1 - \hat{C}_i^s(q^{-1})]\hat{e}_i, \quad \forall i \in \overline{1, ny}, \quad (4)$$

Equations (4) allow not only estimating the inputs of model (2), but also predicting them with sufficient accuracy. The final running MIMO-ARMAX models are then:

$$A_j(q^{-1})y_j \equiv B_j(q^{-1})\hat{\mathbf{u}} + C_j(q^{-1})e_j, \quad \forall j \in \overline{1, ny}, \quad (5)$$

where  $\hat{\mathbf{u}} \in \mathbb{R}^{ny}$  gathers all rough estimates of white noises and  $B_{j,j} \equiv 0, \quad \forall j \in \overline{1, ny}$ . After estimation of inputs, identification of model parameters and white noise values relies on MPEM as well. Actually, approximating ARX models return more refined estimated values of white noises:

$$\hat{e}_j \equiv \hat{A}_j^{n\alpha}(q^{-1})y_j - \sum_{i=1, i \neq j}^{ny} \hat{B}_{j,i}^{n\beta}(q^{-1})\hat{u}_i, \quad \forall j \in \overline{1, ny}, \quad (6)$$

where  $A_j^{n\alpha}$  and  $B_{j,i}^{n\beta}$  are polynomials with degrees  $n\alpha$  and  $n\beta$ , respectively. (Usually,  $n\alpha$  and  $n\beta$  are sensibly larger than the degrees of polynomials in (5).)

After estimation of parameters, the optimal predictor can be enabled. The following equations show how the output predicted data are recursively computed, on a prediction horizon of length  $K \geq 1$  (for any  $k \in \overline{N_y + 1, N_y + K}$ ):

$$\begin{aligned}
\hat{y}_j[k|N_y] &= -\hat{a}_{j,1}\hat{y}_j[k-1|N_y] - \dots - \hat{a}_{j,na_j}\hat{y}_j[k-na_j|N_y] + \\
&+ \sum_{i=1, i \neq j}^{ny} [\hat{b}_{j,i,1}\hat{u}_i[k-1] + \dots + \hat{b}_{j,i,nb_{j,i}}\hat{u}_i[k-nb_{j,i}]] + \\
&+ \hat{c}_{j,1}\hat{e}_j[k-1] + \dots + \hat{c}_{j,nc_j}\hat{e}_j[k-nc_j]; \quad (7) \\
\hat{e}_j[k] &= \hat{y}_j[k|N_y] + \hat{\alpha}_{j,1}\hat{y}_j[k-1|N_y] + \dots + \hat{\alpha}_{j,k-N_y-1}\hat{y}_j[N_y+1|N_y] + \\
&+ \hat{\alpha}_{j,k-N_y}\hat{y}_j[N_y] + \dots + \hat{\alpha}_{j,na}\hat{y}_j[k-na] - \\
&- \sum_{i=1, i \neq j}^{ny} [\hat{\beta}_{j,i,1}\hat{u}_i[k-1] + \dots + \hat{\beta}_{j,i,nb}\hat{u}_i[k-nb]] = \hat{u}_j[k]. \quad (8)
\end{aligned}$$

By definition, predicted outputs are equal to measured outputs on measuring horizon  $\overline{1, N_y}$ . Obviously, as the last equation shows, predicted inputs are equal to predicted white noises on prediction horizon  $\overline{N_y+1, N_y+K}$ . Equations (7) and (8) have to be iterated successively for any prediction instant  $k$ . Moreover, next predicted values cannot be estimated unless all output channels have previously been predicted (since all inputs must first be predicted). So, before approaching the next prediction step, all channels have to be predicted at the current step. This mechanism expresses the attempt to consider correlations between different outputs, even though the MIMO-ARMAX model is composed by several MISO-ARMAX models. (In case of isolated SISO-ARMA models, each channel can independently be predicted on the whole prediction horizon, without waiting for the other channels.)

The performance of PARMAX predictor can be assessed by means of *prediction quality* (PQ) criterion, which is defined as follows, for any channel  $j \in \overline{1, ny}$ :

$$PQ_j = \frac{100}{1 + \frac{\sqrt{\sum_{k=1}^K \hat{\sigma}_{j,k}^2}}{\hat{\lambda}_{m,j} \sqrt{\text{SNR}_{m,j}} \sqrt{\text{SNR}_{p,j}}}} \quad [\%], \quad (9)$$

where:  $\text{SNR}_{m,j}$  and  $\text{SNR}_{p,j}$  are signal-to-noise ratios of data on channel  $j$ , during the measuring horizon and prediction horizon, respectively;  $\hat{\sigma}_{j,k}$  is the estimated standard deviation of current prediction error, on channel  $j$ ;  $\hat{\lambda}_{m,j}$  is the estimated standard deviation of noise  $e_j$  on measuring horizon. The SNRs are defined as follows:

$$\text{SNR}_{m,j} = \frac{\sigma_{m,y_j}^2}{\hat{\lambda}_{m,j}^2} \quad \& \quad \text{SNR}_{p,j} = \frac{\sigma_{p,y_j}^2}{\hat{\lambda}_{p,j}^2}, \quad (10)$$

where:  $\sigma_{m,y_j}$  is the standard deviation of measured data,  $\sigma_{p,y_j}$  is the standard deviation of data on prediction horizon and  $\hat{\lambda}_{p,j}^2$  is the estimated variance of overall prediction error. The latter is simply expressed below:

$$\hat{\lambda}_{p,j}^2 = \frac{1}{K} \sum_{k=1}^K [y_j[N_y+k] - \hat{y}_j[N_y+k|N_y]]^2. \quad (11)$$

As of  $\hat{\sigma}_{j,k}$ , the following recursive equation has been proven:

$$\hat{\sigma}_{j,k}^2 = \hat{\sigma}_{j,k-1}^2 + \hat{\lambda}_{m,j}^2 \hat{\lambda}_{j,k-1}^2, \quad \forall k \in \overline{N_y+1, N_y+K}, \quad (12)$$

where  $\hat{\sigma}_{j,N_y}^2 = 0$  and  $\{\hat{\gamma}_{j,k}\}_{k \in \mathbb{N}}$  are the coefficients of endless division between polynomials  $\hat{C}_j$  and  $\hat{A}_j$  (starting from their free terms). Obviously,

$$\hat{\lambda}_{m,j}^2 = \frac{1}{N_y} \sum_{k=1}^{N_y} (\hat{e}_j[k])^2 = \frac{1}{N_y} \sum_{k=1}^{N_y} (y_j[k] - \hat{y}_j[k|N_y])^2. \quad (13)$$

The bigger the norm of vector  $\mathbf{PQ} = [\text{PQ}_1 \dots \text{PQ}_{ny}]^T$ , the better the performance of PARMAX algorithm.

### 3. SELECTING OPTIMAL PREDICTORS VIA PSO

Despite the simplification applied to MIMO-ARMAX models, their complexity could still be high. Even in case of networks with small number of sensors, accurate prediction requires large number of parameters to be identified. For example, if  $ny = 4$ , and the maximum values of structural indices  $na$  (degree of A polynomials),  $nb$  (degree of B polynomials), and  $nc$  (degree of C polynomials) is 30, then the best predictor (in terms of  $\|\mathbf{PQ}\|$ ) has to be selected among more than  $30^{28} \cong 2.29 \cdot 10^{41}$  identification models (when accounting the rough SISO-ARMA models as well). One copes thus with an NP hard optimization problem.

The problem can be approached by means of an evolutionary strategy, based on some heuristic. From all such strategies, the PSO [Kennedy 1995] has been preferred, for the convenient trade off between convergence speed and algorithm complexity it exhibits.

The principle of PSO is illustrated in Figure 2.

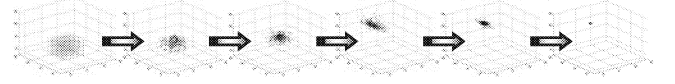


Figure 2: Principle of Particle Swarm Optimization

A population of entities referred to as *particles* is initiated to run on some trajectories, starting from initial positions. Trajectories are adaptively adjusted on the run, according to some fitness function (that has to be optimized). The goal is to concentrate the population around some optimum of fitness function and to increase the chance that the optimum be global over the searching space.

The population of particles is denoted by  $\mathcal{P} = \{\mathbf{x}_p\}_{p \in \overline{1, P}}$ .

Positions of particles are actually accounted in  $\mathcal{P}$  (as vectors of some Euclidean space). Usually, the number of particles ( $P$ ) varies from 50 to 500. The population evolves towards optimum through *generations*. At every generation, particles take new positions, which can be updated by means of additive corrections, like below:

$$\mathbf{x}_p^{m+1} = \mathbf{x}_p^m + \Delta \mathbf{x}_p^{m+1}, \quad \forall p \in \overline{1, P}, \quad \forall m \geq 0, \quad (14)$$

where  $m$  is the *generation index*. Corrections are computed as displacements of particles with some speeds with  $\Delta T$  delay, i.e.:  $\Delta \mathbf{x}_p^{m+1} = \mathbf{v}_p^{m+1} \Delta T$ . Speeds can also be updated:

$$\mathbf{v}_p^{m+1} = \mu_p^m \cdot \mathbf{v}_p^m + \lambda_{p,c}^m \cdot \frac{\mathbf{x}_{p,0}^m - \mathbf{x}_p^m}{\tau_{p,c}^{m+1}} + \lambda_s^m \cdot \frac{\mathbf{x}_{p,0}^m - \mathbf{x}_p^m}{\tau_s^{m+1}}, \quad \forall p \in \overline{1, P}, \quad \forall m \geq 0, \quad (15)$$

where:

- $\mu_p^m \in [0,1]$  is the particle *mobility factor*;
- $\mathbf{x}_{p,0}^m$  is the best position of particle, on its way from the beginning to the current generation;
- $\mathbf{x}_{p,0}^m$  is the best position reached by a particle of the whole population (from the beginning to the current generation);
- $\lambda_{p,c}^m \in [0,2]$  is the normalized *cognitive variance* of particle positions with respect to its best position;
- $\lambda_s^m \in [0,2]$  is the normalized *social variance* of all particles positions with respect to the best population position;
- $\tau_{p,c}^{m+1} \geq \Delta T$  is the delay of particle transition from current position to its best position;
- $\tau_s^{m+1} \geq \Delta T$  is the delay of population transition from current positions of its particles to the best population position.

The PSO algorithm has been modified from its traditional design, in order to match PARMAX. Thus, its parameters are adaptively set, according to  $\|\mathbf{PQ}\|$  criterion. The particles are split in 4 parts, depending on the nature of structural indices in PARMAX (which are degrees of polynomials, in fact):

$$\mathbf{x} = \begin{bmatrix} nt & \mathbf{x}_{ARMA}^T & \mathbf{x}_{ARMAX}^T & \mathbf{x}_{am}^T \end{bmatrix}^T \in \mathbb{N}^{2ny+5}, \quad (16)$$

where  $nt \in [0,5]$  is the degree of polynomial trend [Stefanoiu 2008c]. The other compounds in (16) are:

$$\mathbf{x}_{ARMA} = \begin{bmatrix} na_0 & nc_0 \end{bmatrix}^T \in \mathbb{N}^2 \text{ (for rough ARMA models); } (17)$$

$$\mathbf{x}_{ARMAX} = \begin{bmatrix} na & nb_1 & \dots & nb_{ny-1} & nc \end{bmatrix}^T \in \mathbb{N}^{ny+1} \text{ (for ARMAX models); } (18)$$

$$\mathbf{x}_{am} = \begin{bmatrix} n\alpha_0 & n\alpha & n\beta_1 & \dots & n\beta_{ny-1} \end{bmatrix}^T \in \mathbb{N}^{ny+1} \text{ (for AR and ARX approximating models); } (19)$$

For example, if the number of channels is 4 ( $ny = 4$ ), then every particle accounts 13 structural indices. Their values are integer, so that the corrections in (14) have to be rounded to the nearest integer.

The following definitions are given for arbitrarily set  $m \geq 0$  and  $p \in \overline{1, P}$ , in order to express the variables in (15). The mobility factor is:

$$\mu_p^m = \frac{\|\mathbf{PQ}[\mathbf{x}_{p,0}^m]\| - \|\mathbf{PQ}[\mathbf{x}_p^m]\|}{\|\mathbf{PQ}[\mathbf{x}_{p,0}^m]\| - \|\mathbf{PQ}[\mathbf{x}_{p,1}^m]\|}, \quad (20)$$

where  $\mathbf{x}_{p,1}^m$  is the worst position of particle, on its way from the beginning to the current generation. The relative cognitive variance is computed in two steps. First, the absolute variance is evaluated:

$$\sigma_{p,c}^m = \sqrt{\frac{1}{m+1} \sum_{i=0}^m \left( \|\mathbf{x}_{p,0}^m\|^2 - \|\mathbf{x}_p^i\|^2 \right)}. \quad (21)$$

Then the variance is normalized with respect to minimum and maximum values, after updating them with (21):

$$\lambda_{p,c}^m \stackrel{\text{def}}{=} 2 \frac{\sigma_{p,c}^m - \sigma_{p,c,\min}^m}{\sigma_{p,c,\max}^m - \sigma_{p,c,\min}^m}. \quad (22)$$

In a similar manner, the relative social variance can be evaluated. The absolute value leads to the relative value:

$$\sigma_s^m \stackrel{\text{def}}{=} \sqrt{\frac{1}{P} \sum_{p=1}^P \|\mathbf{x}_p^m - \mathbf{x}_{p,0}^m\|^2} \Rightarrow \lambda_s^m \stackrel{\text{def}}{=} 2 \frac{\sigma_s^m - \sigma_{s,\min}^m}{\sigma_{s,\max}^m - \sigma_{s,\min}^m}. \quad (23)$$

It is practically impossible to detect the transition delays. Therefore, they are selected at random. More specifically, the *transition frequencies*  $v_{p,c}^{m+1} = 1/\tau_{p,c}^{m+1}$  (*cognitive*) and  $v_s^{m+1} = 1/\tau_s^{m+1}$  (*social*) are randomly generated in  $(0, 1/\Delta T]$ .

Equation (15) is the searching engine of PSO strategy. Three terms are contributing to update the speed. The first one is based on the current speed. The weight applied by the mobility factor shows that the former speed is strongly attenuated when the particle is close to its best position. This means the particle is rather attracted by such a position. The second term quantifies the cognitive motivation of particle to move, in terms of driven speed. As the population evolves, every particle acquires some experience that can determine its future behavior. As result of particle experience, the *cognitive speed* becomes important when the trajectory is drifted away with respect to its best position. Finally, the third term expresses the influence of population on each particle behavior, naturally referred to as *social speed*. Depending on particle position versus population best position and variance, the social speed increases either when the population is dispersed or the particle is far away from the best position.

The evolutionary character of PSO algorithm is involved by the randomly selected transition frequencies. Population can thus escape from the capture of some local optimum, by jumping onto another zone. Like most of the evolutionary procedures based on populations, the trade-off between diversity and convergence is important. Diversity of population allows one to check for optimum in many zones of searching space and, thus, to increase the chance of global optimum finding. On the contrary, convergence (or *particles swarm*) is necessary to avoid searching oscillations. The key parameter of this trade-off monitoring is the absolute social variance. Three populations can evolve in parallel: the current one, the elite (that keeps all the best positions) and the anti-elite (that keeps all the worst positions). Every time the product between social variances of current generation and elite overflows or underflows some bounds, the diversity-convergence trade-off is unbalanced. Over-floating points to abnormal increase of population diversity, i.e. to oscillatory behavior. To reduce it, crossover between current generation and elite can be performed. Under-floating is caused by concentration of both populations around some optimum, which may decrease the chance to reach for the global optimum. In this case, diversity is increased by performing crossover between current generation and anti-elite. By crossing over two particles  $\mathbf{x}_1$  and  $\mathbf{x}_2$  with speeds  $\mathbf{v}_1$  and  $\mathbf{v}_2$ , respectively, two offspring are obtained:

$$\tilde{\mathbf{x}}_1 = \gamma \mathbf{x}_1 + (1-\gamma) \mathbf{x}_2 \quad \& \quad \tilde{\mathbf{x}}_2 = \gamma \mathbf{x}_2 + (1-\gamma) \mathbf{x}_1, \quad (24)$$

with corresponding speed values:

$$\tilde{\mathbf{v}}_1 = \gamma \mathbf{v}_1 + (1-\gamma) \mathbf{v}_2 \quad \& \quad \tilde{\mathbf{v}}_2 = \gamma \mathbf{v}_2 + (1-\gamma) \mathbf{v}_1, \quad (25)$$

where  $\gamma \in [0,1]$  is selected at random (uniformly). When crossover is applied on current population and elite or anti-



elite, all of them of size  $P$ , the offspring population consists of  $2P$  particles. Only the most fitted ( $P-1$ ) offspring are selected to complete the current population, after removing all its particles, but the most fitted one. (Maybe this one is the global optimum and must be kept.)

The PSO algorithm is initiated to run starting from some initial population (usually, uniformly distributed inside the searching space). To stop the procedure, several tests can be performed. Two very effective tests are the following: the maximum number of iterations was touched (say 100) or the most fitted particle succeeded to survive within the elite population for several generations (say at least 5).

The overall strategy of PARMAX algorithm consists of the following main steps:

1. Perform data acquisition on several channels.
2. Estimate the optimal SISO-ARMA prediction model (3) for every channel (independently on the other channels). In this case, the particle includes only the structural indices of ARMA models, i.e.:  $nt$ ,  $na_0$ ,  $nc_0$  and  $na_0$ .
3. Perform prediction, like in (7) and (8). In this case, there are no inputs, so the corresponding terms have to be removed in equations (7) and (8).
4. Estimate the input signals (4) on measuring horizon, by means of SISO-ARMA prediction models.
5. Estimate the optimal MISO-ARMAX models (5) from output acquired data and input estimated data. This time, the particle configuration is given by definition (16), i.e. it includes the ARMA structural indices as well. (ARMA indices previously estimated in step 2 are cancelled.)
6. Perform prediction, according to equations (7) and (8). Recall that predicted inputs are actually estimated white noises.

Steps 2, 3 and 4 are not really necessary for PARMAX. (They belong to PARMA.) Step 5 actually includes them. Repeating them only for SISO-ARMA models allows comparing the performance of both predictors. Note however that optimal SISO-ARMA models in step 2 are usually different from optimal SISO-ARMA models estimated inside step 5, because the PQ criterion operates with different predicted data.

#### 4. SIMULATION RESULTS AND DISCUSSION

The PARMAX procedure has been implemented within MATLAB environment. There are many implementation details that cannot be described within this paper (but can be found in [Stefanoiu 2008c]). One can only mention here the strategy of optimal structural indices selection (for both PARMA and PARMAX predictors). Obviously, the PQ fitness function based on definitions (9) cannot be evaluated unless data on prediction horizon are acquired as well. It follows that all ARMA(X) models have to be identified from  $N_y - K_f$  data, instead of  $N_y$  data. The last  $K_f$  data are preserved as test horizon for the PSO algorithm. More specifically, after identification of some ARMA(X) model, its prediction performance is tested on the last  $K_f$  data (which have not been involved in identification). Usually,  $K_f \in \max\{3, K/2\}, \bar{K}$ . After selecting the best predictor on

the test data set, the optimal structural indices are just employed to identify a new model based on the whole data set (i.e. including the test data). Usually, the prediction performance decreases on the prediction horizon, comparing to the test horizon.

The goal of implementation is to perform a comparison between PARMA and PARMAX algorithms in terms of prediction quality. If data from different channels are inter-correlated, PARMAX is expected to perform better than PARMA.

The application consists of temperatures monitoring in two cities, which are located at 60 km each other, slightly below the latitude of 45°. For each city, minimum and maximum temperatures are daily acquired, starting from November 23, 2007, until March 24, 2009. Data look like in Figure 3 and were transmitted via internet to the central processing unit.

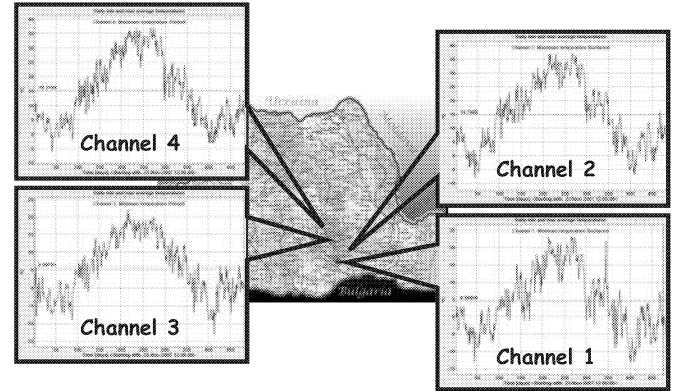


Figure 3: Minimum and maximum temperatures of two cities

Given the short distance between cities and their geographical position (in a middle of a plane), one can easily notice correlations between the four channels. Beside this feature, some other characteristics of data can also be outlined. For example, noises are corrupting in a different manner the channels. Therefore, data are, at the same time, apparently independent. Another property is related to the seasonal behavior. Normally, the measuring horizon is too short to detect the main period (the acquisition period should be almost twice). It follows that PQ will increase whenever the prediction model is able to detect the right periodic variation of each ts.

One of the most important advantages of PSO algorithm is the possibility to perform the search either on a parallel machine or on several computers of a network at the same time. When using a single regular computer (e.g. laptop/notebook), finding the optimal structure of prediction model might be a lengthy operation. But ecological or meteorological data are provided by slow systems. For example, the temperature ts above are collected daily. For an amount of 500 data per channel and a population of 10 particles per channel, PARMA runs about 30', while PARMAX needs about one hour on a regular computer. On a parallel machine, the delays are reduced approximately 8 times. In this research, the simulations were performed on 8 regular computers in a network. The population size ( $P$ ) varied in range 5:15 particles per channel. Thus, practically, the overall population included at least 20 particles per computer, which means at least 160 particles. The other configuration parameters have been set as follows:

- prediction horizon length ( $K$ ): 5;
- maximum number of generations ( $M$ ): 100;
- minimum survival factor ( $S$ ): 5;
- maximum degree of trend ( $NT$ ): 5;
- maximum structural index for AR part ( $NA$ ):  $N_y/3$ ;
- maximum structural index for MA part ( $NC$ ):  $N_y/3$ ;
- maximum structural index for X part ( $NB$ ):  $N_y/2$ ;
- maximum structural index for AR part of approximating models ( $N\alpha$ ):  $N_y/2$ ;
- maximum structural index for X part of approximating models ( $N\beta$ ):  $N_y/2$ .

This leaves a large searching space for optimal structural indices. The last  $K$  data have been removed completely. They are accounted only after the final prediction data have been estimated, in order to compute PQ. (Thus, when searching for optimal structure of prediction model, only  $N_y - K_t - K$  data are employed.)

After running PARMA on the temperature ts (with  $N_y = 482$  data), the best results are illustrated in Figures 4, 6, 8 and 10 (for each channel). In parallel, Figures 5, 7, 9 and 11 display the best results of PARMAX. Each of which includes 3 variations: the original ts together with its optimal trend (top), the estimated white noise on measuring horizon (middle) and the zoom on prediction horizon (bottom). The PQ values are depicted on the latter. One can see that:

$$\begin{aligned} \mathbf{PQ}_{ARMA}^T &= [67.83 \ 53.66 \ 49.11 \ 66.52] \Rightarrow \|\mathbf{PQ}_{ARMA}\| \cong 119.65; \\ \mathbf{PQ}_{ARMAX}^T &= [69.97 \ 72.37 \ 59.71 \ 71.82] \Rightarrow \|\mathbf{PQ}_{ARMAX}\| \cong 137.32. \end{aligned} \quad (26)$$

Since the correlation between channels is so obvious, the superior performance of PARMAX versus PARMA is an expected result. Beside this general remark, the simulation results yield several insights.

The fact the measuring horizon is not large enough to allow detection of seasonal variation decreases the performance of both predictors. As one can see from Figures 4, 5 and 9, the trend tried to decode data periodicity by increasing the polynomial degree to its maximum value. However, for the other models the trend degree is bounded by 3. Normally, the trend should not include the periodic variation, because polynomials are unbounded in vicinity of infinity and, thus, the prediction can fail very easily. In this application, higher degree trend is accepted because the prediction horizon is quite short, comparing to the measuring horizon. PARMA detected a period on channel 3, since the ts is closer here to symmetry than for the other channels. The period is almost half of  $N_y$ . Simulations have proven that, without seasonal compound, PARMA systematically fails on channel 3. Oppositely, PARMAX removed any seasonal variation and tried to replace its contribution by colored noises.

As already mentioned, data are affected by different noises on different channels. The second variation of all figures shows that the estimated SNR varies along channels. In general, the higher PQ, the higher SNR and the smaller white noise variance.

The third variations of all figures are grouped in couples (one for each channel). In order to focus on details, the vertical axis has been scaled differently from a variation to another. Actually, scaling is uniquely determined by the confidence tube. Large aperture of tube usually involves small value of PQ, even though the predicted values might be close each other. Prediction on channels 2 and 3 is quite remarkable (even at eastern end of prediction horizon, where, usually, predicted values are less accurate). This is somehow surprising, since the ts belong to different cities and, moreover, are representing opposite temperatures (maximum and minimum, respectively). But the correlations between data are mostly hidden and cannot simply be “read” on time variation, regardless the user’s experience.

For data with less correlation than the above ones, the superior performance of PARMAX over PARMA is not so obvious. Usually, PARMAX performs better on some channels only and its fitness (the norm of PQ) can even fall below the PARMA fitness. In this case, PARMA should be preferred, due to its smaller complexity and higher speed.

## 5. CONCLUSION

Lately, there is an increasing interest in handling and monitoring distributed data, on different purposes. In case of ecological or meteorological data, prediction is an important goal. This article introduced a method of prediction based on multi-dimensional identification models from ARMAX class. An evolutionary strategy (PSO) was employed to speed up the searching for optimum prediction model. The resulting algorithm can be implemented on a mobile computer. Prediction can thus directly be performed inside the system that provides the data. Whenever the measuring channels are correlated, multi-dimensional prediction models should be constructed instead of singleton models for each channel in isolation.

## REFERENCES

- Kennedy J., Eberhart R. 1995. „Particle Swarm Optimization”, *IEEE International Conference on Neural Networks*, Piscataway – N.J., USA, pp. 1942-1948.
- Söderström T., Stoica P. 1989. *System Identification*, Prentice Hall, London, UK.
- Stefanoiu D., Petrescu C. 2008a. „VISA – A Versatile Interface of Signal Acquisition and Wavelet Based Data Fusion”, *Proc. of the 5-th Europe-Asia Mechatronics Congress*, Annecy, France, Paper no. 167.
- Stefanoiu D., Culita J., Ionescu F. 2008b. „FORWAVER – A Wavelet-Based Predictor for Non Stationary Signals”, *Proc. of EUROSIS-ISC-2008*, Lyon, France, pp. 377-381.
- Stefanoiu D., Petrescu C. et al. 2008c. „Numerical Models and Fast Methods of Ecological Phenomena Prediction”, Res. Rep. CNMP.UPB-P4.31050-2007.II/DS.CP. AD.JC-12.2008, “Politehnica” University of Bucharest, Romania.
- Wang Z., Ziou D., et al. 2005. „A Comparative Analysis of Image Fusion Methods”, *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 43, No. 6, pp. 1391-1402.

## ACKNOWLEDGMENTS

*This research was developed with the support of National Center for Programs Management (Romania), in the framework of grant #31050/2007. The authors wish to address its team grateful and special thanks.*

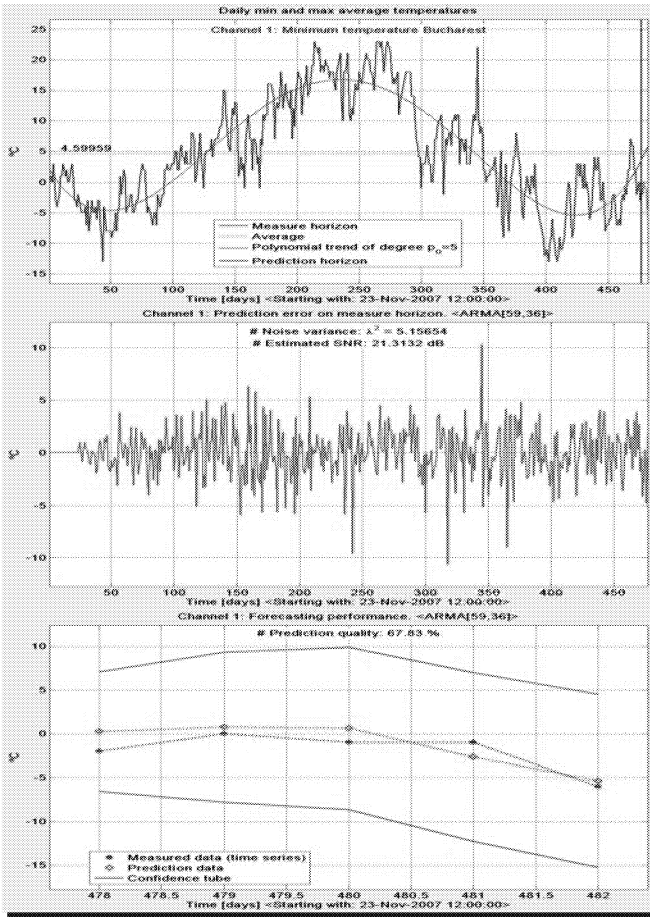


Figure 4: PARMA performance on channel 1

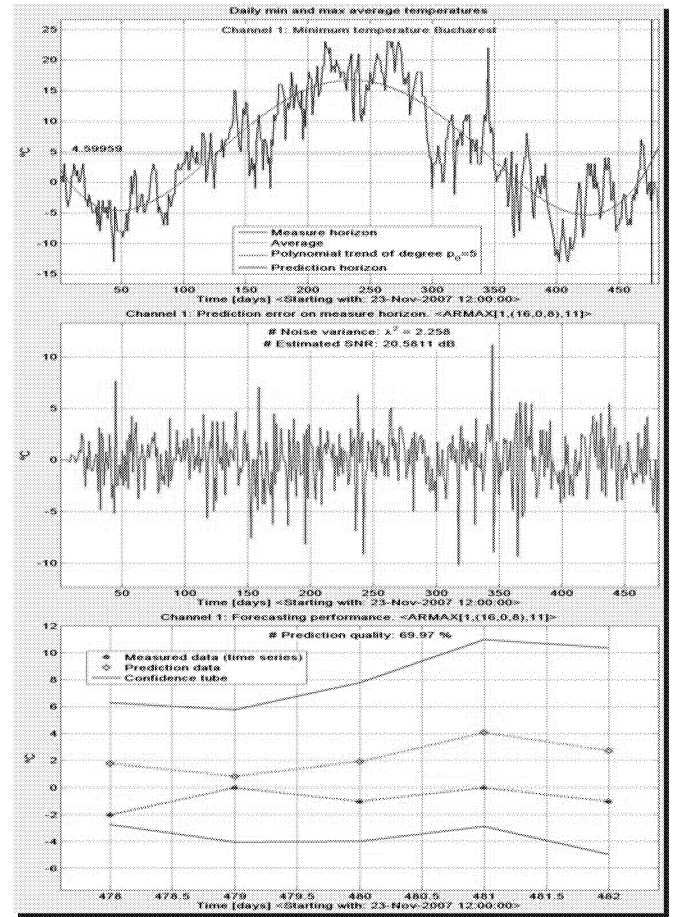


Figure 5: PARMAX performance on channel 1

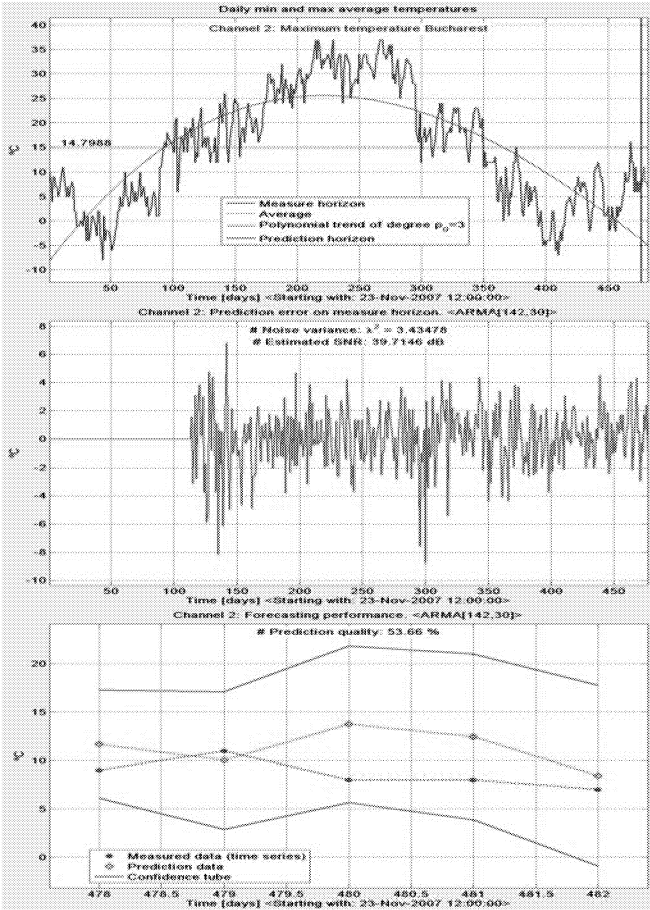


Figure 6: PARMA performance on channel 2

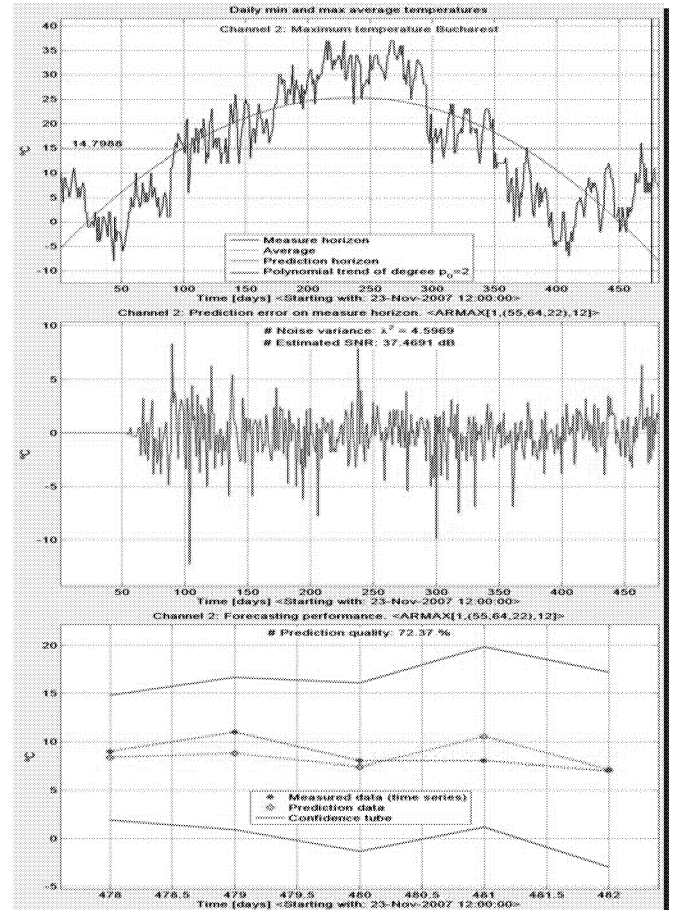


Figure 7: PARMAX performance on channel 2

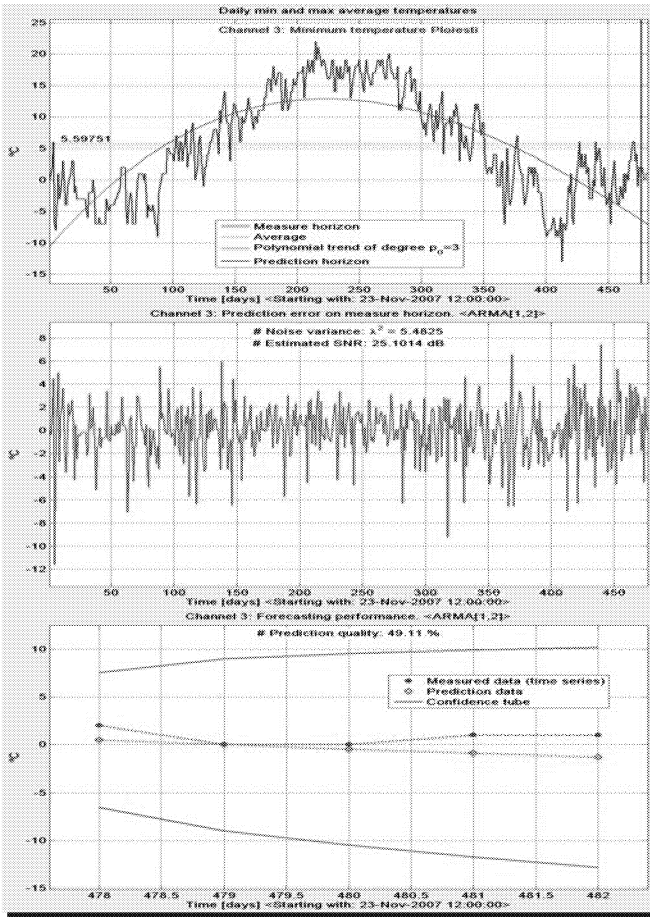


Figure 8: PARMA performance on channel 3

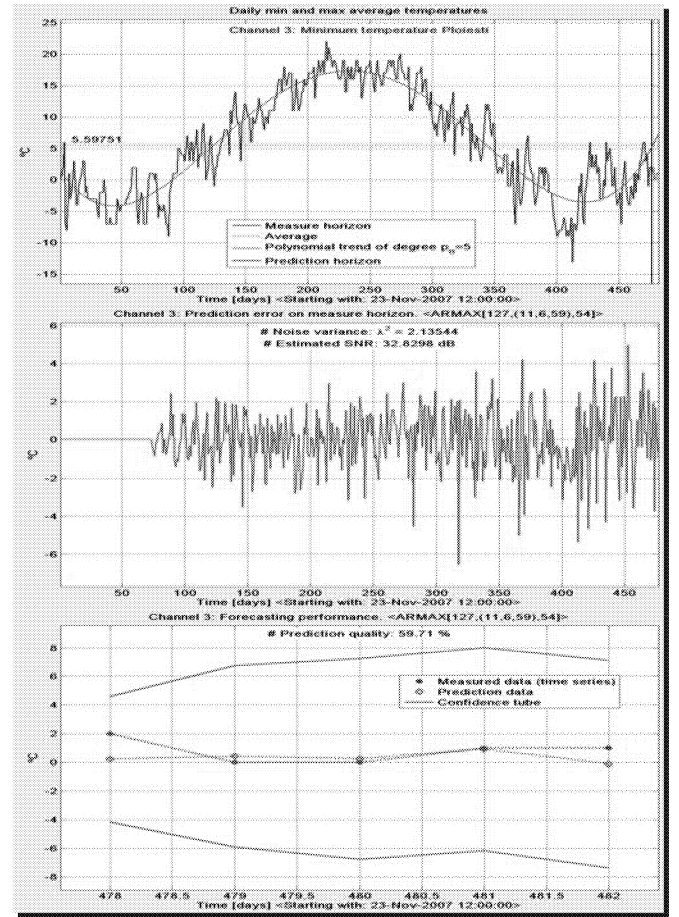


Figure 9: PARMAX performance on channel 3

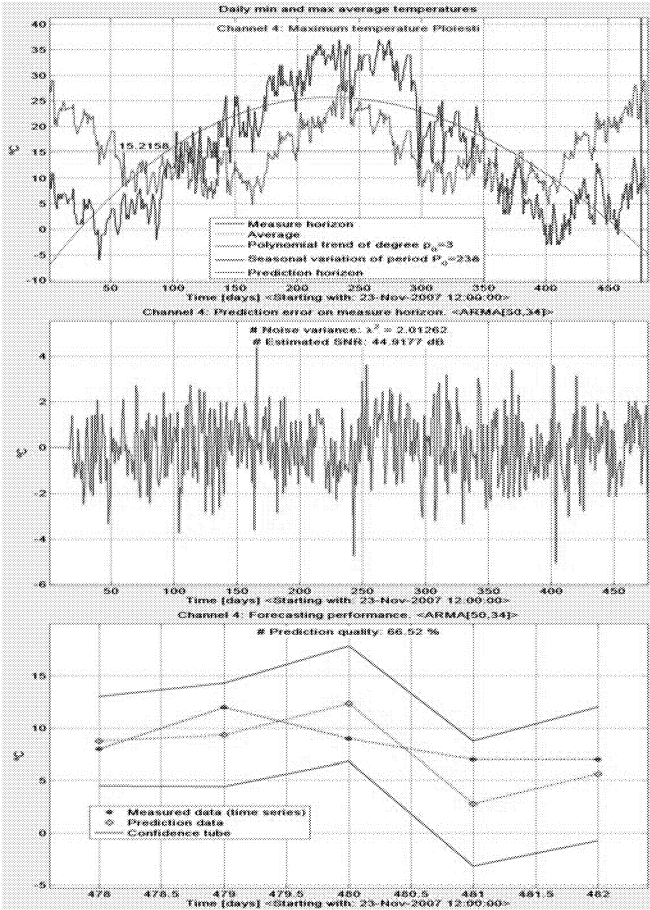


Figure 10: PARMA performance on channel 4

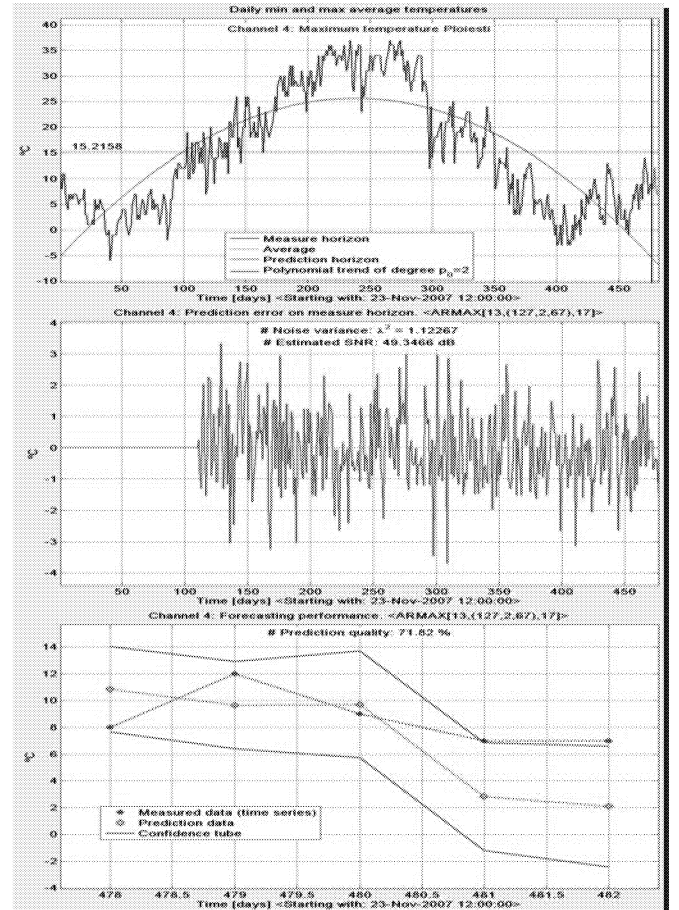




Figure 11: PARMAX performance on channel 4

# THE USE OF REFINED DESCRIPTIVE SAMPLING IN PARALLEL SIMULATION

Abdelouhab ALOUI and Megdouda OURBIH ex TARI  
Laboratory of Applied Mathematics  
Bejaia University  
Algeria

E-mail:  aaloui\_abdel |  megtari @yahoo.fr

## KEYWORDS

Monte Carlo, Deterministic, Parallel simulation, Pert network, Inventory.

## ABSTRACT

Refined descriptive sampling is designed to improve upon the descriptive sampling method for experimentation in simulation. The former reduces significantly the risk of sampling bias generated by descriptive sampling and eliminates its problem related to the sample size. In this paper, we propose an optimal parallel Monte Carlo simulation algorithm using refined descriptive sampling and evaluate in parallel architecture, performance measures of a Pert network and the Newsboy problem.

## INTRODUCTION

Monte Carlo method also called random sampling (RS) can solve a large variety of problems, but both kinds of variations (set effect and sequence effect) are present in a randomly generated sample that provides an imprecise evaluation of each simulation estimates (which is a function of the input values). Given the relationship between the lack of precision and the method itself, it was born a remedial new paradigm. The latter says that is not always necessary to generate sample values randomly to describe a stochastic behaviour. Then, new deterministic methods, like Quasi Monte Carlo (Owen 1998), Latin hypercube sampling (Loh 1996), Descriptive Sampling (DS) (Saliby 1990) as well as Refined Descriptive Sampling (RDS) (Tari-Ourbih 2005) were derived from this paradigm.

By the way, we can find several work on the parallelization of Monte Carlo methods but all kinds of variations are still present in a randomly generated sample and simulation estimates remain affected by such sampling errors through Monte Carlo parallel algorithms already proposed for example in (Alme et al.1998; Okten et al. 2006; Tuffin 2000).

In this paper, the best sampling procedure: RDS is selected to improve upon the accuracy of simulation estimates and its safe implementation is suggested in parallel architecture to reduce the cost of running simulation experiments.

RDS is based on a deterministic selection of the input sample values. Subsets of regular numbers are first generated, then, randomly shuffled and finally observations

of the input random variables are generated as required by the simulation since the dimension of subsets are prime numbers generated randomly. This method was proposed by (Tari and Dahmani 2005a) to make DS safe, efficient and convenient. It is safe by reducing substantially the risk of sampling bias, efficient by producing estimates with lower variances and convenient by removing the need to determine in advance the sample size. (Tari and Dahmani 2005b, 2005c) compare the efficiency of DS and RDS on a flow shop system and a production system showing that RDS produces better results than DS and RS.

However, RDS procedure is regarded as an algorithm of dependent instructions capable to be executed just in sequential way. The only instruction able to be parallelized is related to the loop of the array filling up with regular numbers but it is not well-suited to parallelization because of the high communication cost that may be generated. However, MC simulation using RDS involves repetition and this feature is specially well-suited to parallelization. Then, to avoid such communication cost, parallelism of the number of replicated runs is proposed using RDS to generate input distributions. We have then carried out a parallel Monte Carlo simulation program running only on machines with distributed memory. Accordingly, the Single Program Multiple Data (SPMD) model has been selected for a better programming related to the replicated runs which consists on multiple processing of the same algorithm over samples of different data. The Message Passing Interface (MPI) library has been suitably chosen to be used with such model. The selected library supports both the language C/C++ and FORTRAN and allows the data exchange between the processors. The selected library fits well our requirements; nevertheless we can find other libraries like Parallel Virtual Machine (PVM) (Gest et al. 1994), Portable Programs for Parallel Processors (P4) (Butler and Lusk 1994) and Open Multi Processing (OpenMP) (Chergui and Lavallée 2006).

The use of SPMD model provides an optimal efficiency of the proposed parallel algorithm; this paper also evaluates in parallel architecture performance measures of both problems, Pert network and Newsboy problem. In the designed parallel software (The simulation programme of each application is written by using the C++ compiler and run on Pentium 4 under the operating system Linux), we used RDS algorithm given in (Tari and Dahmani 2006) to generate input refined descriptive sample values and the



built in generator of the computing environment to generate both prime numbers and integers to permute the regular points related to RDS method (The built in pseudo random number generator used is the rand() function of the C++ compiler under Linux).

## THE USE OF REFINED DESCRIPTIVE SAMPLING

RDS is concerned with a block that must be situated inside a generator aiming to distribute regular samples of prime number sizes when required by the simulation. We stop the process when the simulation terminates.

In this procedure, each run is determined by a block of different prime numbers. The values of refined descriptive sample vary between different runs since the sample size is a prime number randomly selected. The generation process of the sample values is deterministic whereas the generation of the prime numbers and the sequence of the sample values are random.

Without loss of generality, we assume that one input random variable  $X$  drives the simulation and one output random variable with  $k$  parameters  $\theta_j$ ,  $j = 1, 2, \dots, k$  to be estimated is observed through simulation.

### The refined descriptive samples

Suppose that  $m$  prime numbers have been used in a simulation run. Formally, in RDS, regular sample values are generated for any input random variable  $X$  as required by the simulation using the inverse transform method such as

$$(xd)_{i+\sum_{l=1}^{q-1} p_l}^i = H^{-1}\left(r_{i+\sum_{l=1}^{q-1} p_l}^i\right)$$

for  $i = 1, 2, \dots, p_q$  and  $q = 1, 2, \dots, m$

where  $\left\{r_{1+\sum_{l=1}^{q-1} p_l}^1, r_{2+\sum_{l=1}^{q-1} p_l}^2, \dots, r_{\sum_{l=1}^q p_l}^{p_q}\right\}$

for  $q = 1, 2, \dots, m$  are considered as subsets of dependent regular points of prime size  $p_1, p_2, \dots, p_m$  uniformly distributed between  $[0, 1]$  and are obtained by the following formula

$$r_{i+\sum_{l=1}^{q-1} p_l}^i = \frac{i - 0.5}{p_q}$$

for  $i = 1, 2, \dots, p_q$  and  $q = 1, 2, \dots, m$ .

and  $H^{-1}$  is the inverse cumulative function of the input random variable  $X$ .

The refined descriptive samples of prime number size  $p_q$  are obtained by:

- 1) Generating the subsets of regular points of size a randomly chosen prime number  $p_q$ ,  $q = 1, 2, \dots, m$
- 2) Randomizing their sequence for any  $p_q$
- 3) Computing the refined descriptive samples values of the input variable as required by the simulation.

### RDS simulation estimates

In a simulation study, a logical model is built and used as a vehicle for experimentation. So, experiments are carried out on the model and producing in a given run, the following  $m$  estimates of the parameters  $\theta_j$ ,  $j = 1, 2, \dots, k$

$$\begin{aligned} (Yr)_j^1 &= F_j(r_1^1, r_2^2, \dots, r_{p_1}^{p_1}) \quad \text{for } j = 1, 2, \dots, k \\ (Yr)_j^2 &= F_j(r_{1+p_1}^1, r_{2+p_1}^2, \dots, r_{p_2+p_1}^{p_2}) \quad \text{for } j = 1, 2, \dots, k \\ &\vdots \\ (Yr)_j^m &= F_j\left(r_{1+\sum_{l=1}^{m-1} p_l}^1, r_{2+\sum_{l=1}^{m-1} p_l}^2, \dots, r_{\sum_{l=1}^m p_l}^{p_m}\right) \quad \text{for } j = 1, 2, \dots, k \end{aligned}$$

such as

$$\sum_{i=1}^m p_i \geq n$$

and conventionally

$$\sum_{i=1}^0 p_i = 0.$$

where  $F_j$ ,  $j = 1, 2, \dots, k$  is a simulation function usually defined by a program that relates the input variables with each estimator.

Therefore, in a given run, the use of RDS method leads to the following sampling estimates of  $\theta_j$ ,  $j = 1, 2, \dots, k$ , defined by the average of these estimates

$$(Yr)_j = \frac{1}{m} \sum_{i=1}^m (Yr)_j^i \quad \text{for } j = 1, 2, \dots, k$$

If  $N$  replicated runs are needed to run a simulation experiment, it is then, necessary to consider  $N$  blocks of  $m_1, m_2, \dots, m_N$  refined descriptive samples.

### Prime number generator

We first generate an unspecified integer number. Then, we multiply it by two and we add one in order to make the generated number, an odd number since all prime numbers are odd except number 2 which is ignored. Finally, we test if the odd number is a prime number by successive division on odd numbers, starting from three with a step of two. If it is a prime number we use it in running simulation by generating observations of input random variables, otherwise we go back to generate another unspecified integer. In this manner, we reduce the time allocated to the generation of prime numbers by 75%, we then reduce the time of sequential running simulation experiments and by the way, we reduce again the parallel time by reducing the computation time.

### Motivation of parallelism

Simulation modelling involves both repetition and iteration. In statistic, it is well known that replicated runs is compulsory to have statistical results close to the optimum, such as, each run generates output parameters of the

simulator and at the end of the simulation, we compute averages for each considered parameter. Such situation requires a large computing load and a significant amount of memory for a computer. It is then a time-consuming approach. Even (Taylor et al. 2002) explore the use of net conferencing during simulation studies. It is obvious that a simulation study is expensive in time, memory and more generally in necessary resources (Ahn and Danzig 1996). Indeed, the traditional simulation tools could be unusable when complex systems are studied. We are then confronted with the problem of the best compromise between the precision of simulation estimates and the cost of the experiment. The latter of course must be balanced against the benefits that can be gained from the use of simulation, by the same way, the precision of the estimates that are often an order of magnitude greater than the cost. Unfortunately, the limited means of available sequential computation make this goal difficult to achieve. Therefore, a parallel simulation procedure is compulsory for solving the above problem. The aim of this work is to introduce parallelism to take advantage of computing power. Parallelism is regarded as a means of reducing the cost in time and memory of a program solving a complex problem.

### THE PROPOSED PARALLEL SIMULATION

In this section, we propose a parallel Monte Carlo simulation using refined descriptive sampling method. Let us suppose  $Q+1$  processors taking part in the simulation experiment. We appoint one processor as a master and the remaining  $Q$  by the slaves' processors.

In order to carry out a simulation experiment, the master distributes the program of each developing simulation model between the available slaves processors, such as each processor carries out the same copy of such program. Indeed, in parallelism, the program distribution is done in turn between the various processors but in the proposed parallelization, the master distributes to each processor both the program and the number of replicated runs, only once. The reception of simulation results by the master is done in the same manner. This sort of distribution reduces the communication cost.

It is well known that the parallel processing of a program by more than one task is independent with each task being able to execute the same or different statement at the same moment. Therefore, the proposed parallelization is straightforwardly a replication of the whole RDS program on various processors. Given the regular number generation of RDS, its parallelization is easy and simplified regarded to the load balancing. The latter is one of the most significant concepts in the process of parallelism. Given a simulation experiment of  $N$  simulation runs, the load balancing can be introduced if  $N$  is greater than  $Q$ . We can pass around the load balancing if the considered application is regular that generates a number of runs multiple of a slave processor number. Consequently, instead of carrying out  $N$  replicated runs as in the case of a sequential program, each slave processor will carry out just  $\frac{N}{Q}$  runs. At the end of the simulation, each processor computes and sends its

simulation results to the master that computes, in its turn, the final results for each parameter under study. Consequently, these  $\frac{N}{Q}$  runs are regarded as one simulation

experiment and we deduce that the developed simulation model is regular and the load balancing is then not taken into account. This sort of parallelization is called distribution of computation (Lin 1994). In this manner, we reduce the time of parallel running simulation experiments by reducing the computation time.

In the proposed parallelization, the number of messages is reduced from two  $N$  to two  $Q$  then the influence of the communication cost is highly controlled and therefore it is regarded as an insignificant time compared to the computation time. Then,

$$T_{parallel} = \text{computation time}$$

To ensure the independence of the generated prime numbers related to the sampling method itself, we suppose independent prime numbers generators on each processor.

### SAMPLE PROBLEMS

In this section, we evaluate in parallel simulation, performance measures of a pert network and the newsboy problem. Before each simulation run, we generate randomly a prime number  $p$  using the prime number generator given in subsection 2.3. A value of the sample is then generated to be used in the simulation experiment. We stop generating sample values when the simulation terminates. In each sample problem, we choose the same input parameters for all carried out experiments.

We summarize each experiment by computing the mean and variance of the estimates of the output random variables parameters.

#### The parallel algorithm

In this sub-section, we define the main steps of the proposed and implemented parallel algorithm.

Begin

1. The Master reads the simulator parameters

$N$  : Number of replicated runs

$T$  : Simulation period

And other parameters related to the studied problem

2. The Master distributes the parameters to each slave processor

For  $k$  from 0 to  $Q-1$

Send the simulator parameters to the processor  $k$

3. Each slave processor carries out  $\frac{N}{Q}$  simulation runs of the sample problem using RDS algorithm given in (Tari and Dahmani 2006) to generate input refined descriptive samples

4. Gathering together the simulation results obtained on each slave processor.

For  $k$  from 0 to  $Q-1$  send all computed estimates of the  $\frac{N}{Q}$  runs to the Master

5. The Master computes the overall mean of each studied

parameter based on all runs

6. Display the final results by the Master

End

## THE PERT NETWORK

This problem concerns a simple pert network already studied by (Kleindorfer 1971). As shown in figure 1, this network has eight activities. All activities durations are independent and identically distributed random variables following a discrete uniform distribution defined by

$$f(d) = 0.2 \quad d = 1, \dots, 5$$

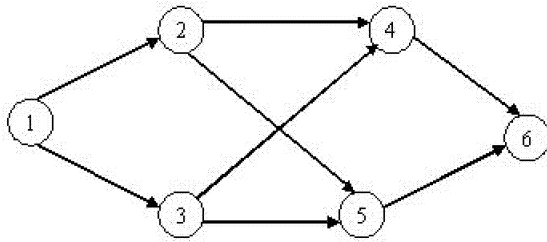


Figure1 : The simulated Pert network

In this problem, there is one response variable with two parameters the mean and the standard deviation to be estimated. The observed response variable is the total project duration. Its parameters are both estimated by  $\overline{DT}$  and  $S_{DT}$ .

Four simulation experiments of different replicated runs were carried out on the studied problem, where each run was defined by 50 observations of the total project duration. We take  $Q = 5$  in each experiment and for different number of runs  $N$ , the observed results of the mean and variance of both estimates  $\overline{DT}$  and  $S_{DT}$  are given in table 1 below.

| Estimate \ N    | 50    |      | 200   |      |
|-----------------|-------|------|-------|------|
|                 | Mean  | Var  | Mean  | Var  |
| $\overline{DT}$ | 10.13 | 0.36 | 10.13 | 0.41 |
| $S_{DT}$        | 2.24  | 1.05 | 2.3   | 0.92 |

| Estimate \ N    | 250   |      | 300   |      |
|-----------------|-------|------|-------|------|
|                 | Mean  | Var  | Mean  | Var  |
| $\overline{DT}$ | 10.11 | 0.42 | 10.13 | 0.41 |
| $S_{DT}$        | 2.3   | 0.95 | 2.29  | 0.95 |

Table1: Empirical results showing the parallelization efficiency for different experiment

We can see from table 1 that the growing up of the number of replicated runs does not affect the observed results.

## THE NEWSBOY PROBLEM

### Problem description

This problem concerns a simple inventory problem already studied by (Kaufmann and Faure 1975). Daily, a newsboy buys  $B$  issues of a newspaper at 0.5£ each. The selling price is 1£. At the end of the day, all remaining issues are restored and the salvage price for surplus newspapers is 0.2£ each. Daily demand ( $D$ ) is independent and identically distributed according to the following statistical distribution where  $f$  stand for frequency and  $Cf$  stand for cumulative frequency

| D  | f | Cf | D  | f | Cf | D  | f | Cf | D    | f | Cf  |
|----|---|----|----|---|----|----|---|----|------|---|-----|
| 0  | 0 | 0  | 13 | 1 | 17 | 26 | 4 | 64 | 39   | 1 | 94  |
| 1  | 0 | 0  | 14 | 3 | 20 | 27 | 3 | 67 | 40   | 2 | 96  |
| 2  | 1 | 1  | 15 | 3 | 23 | 28 | 3 | 70 | 41   | 0 | 96  |
| 3  | 1 | 2  | 16 | 3 | 26 | 29 | 4 | 74 | 42   | 1 | 97  |
| 4  | 1 | 3  | 17 | 4 | 30 | 30 | 2 | 76 | 43   | 1 | 98  |
| 5  | 2 | 5  | 18 | 3 | 33 | 31 | 3 | 79 | 44   | 0 | 98  |
| 6  | 1 | 6  | 19 | 4 | 37 | 32 | 3 | 82 | 45   | 0 | 98  |
| 7  | 1 | 7  | 20 | 3 | 40 | 33 | 2 | 84 | 46   | 1 | 99  |
| 8  | 1 | 8  | 21 | 4 | 44 | 34 | 2 | 86 | 47   | 0 | 99  |
| 9  | 2 | 10 | 22 | 5 | 49 | 35 | 2 | 88 | 48   | 0 | 99  |
| 10 | 2 | 12 | 23 | 4 | 53 | 36 | 1 | 89 | 49   | 1 | 100 |
| 11 | 1 | 13 | 24 | 4 | 57 | 37 | 2 | 91 | 50   | 0 | 100 |
| 12 | 3 | 16 | 25 | 3 | 60 | 38 | 2 | 93 | > 50 | 0 | 100 |

Table2: Cumulative frequency of daily demand

The newsboy wishes to know the quantity  $B$  of newspapers to purchase in order to maximize its daily profit. Then, the simulation purpose is to study the daily profit distribution  $P$  varying the quantity to be bought until 50 and then, searching the optimal profit value and the corresponding quantity to be purchased. In this problem, there is one response variable with the mean parameter to be estimated by  $\overline{P}$ .

### The Newsboy Algorithm

Begin

For 1,  $B = 1$  until 50 do

For 2,  $j = 1$  until 100 do

Simulate daily demand  $D$  with RDS algorithm and compute the expected profit,

If  $B < D$  then expected profit =  $B \times 0.5$

Otherwise expected profit =  $D \times 0.5 - (B - D) \times 0.3$

End if

End for 2

Save the daily expected profit value

Compute the mean of the daily expected profit over 100 working days

Save the expected profit mean value

End for 1

For 3,  $i = 1$  until 50

Look for the optimal expected profit

Look for the corresponding quantity  $B$  of newspapers to purchase



End for 3  
 Display the optimal profit and the optimal quantity to purchase  
 End

### Empirical results

Six simulation experiments of different replicated runs were carried out on the newsboy problem, where each run was defined by 100 days for each bought quantity. We simulate the daily demand distribution according to table 2 and the newsboy problem according to the above algorithm. The summarized results of the mean and variance of the estimator  $\bar{P}$  are given below together with the number of replicated runs and the optimal quantity of newspaper to purchase.

| $N$             | 20    | 40    | 70   | 100  | 200  | 350  |
|-----------------|-------|-------|------|------|------|------|
| $Mean(\bar{P})$ | 8.01  | 7.91  | 8.06 | 7.75 | 7.89 | 7.53 |
| $B$             | 22    | 23    | 26   | 23   | 22   | 21   |
| $Var(\bar{P})$  | 30.83 | 13.15 | 8.03 | 4.91 | 2.58 | 1.1  |

Table3: Empirical results showing the parallelization efficiency for different experiment

We can see from the simulation results of table 3 that the minimum variance is obtained with 350 replicated runs. We then, suggest to the Newsboy to purchase a quantity of newspaper equal to 21 in order to make a profit of 7.53£.

### CONCLUSIONS AND REMARKS

It is well known that the parallel time of running experiments is inversely proportional to the number of available processors. Then, for a given number of runs, whatever the number of processors is, the parallel simulation results are similar to those obtained sequentially, but the parallel time of running simulation experiments is different and for a given sequential time, it decreases when  $Q$  increases according to the following formula,

$$T_{parallel} = \frac{T_{sequentiel}}{Q}$$

The principal interest of the proposed MC parallelism using the best sampling RDS procedure is its running speed while preserving the assets of the sampling method, its efficiency and safety. Since this method works by replication, then it was naturally parallelized. Given the scalability advantage of the SPMD model, the proposed parallelization is more appropriate with a large number of replicated runs and it is related to both, the algorithm of RDS procedure and the resolution algorithm of a complex problem. A Pert network and the newsboy problem were simulated in parallel architecture to confirm this fact, but any other application can be solved by the proposed parallel algorithm.

### REFERENCES

Ahn, J. S. and Danzig, P. 1996. " Packet: network simulation: Speedup and accuracy versus timing granularity." *IEEE/ACM*

- Transactions on Networking*, 4, 5, 743—757.
- Alme, H. J. et al. 1998. "Domain decomposition for parallel Laser-Tissue Models with Monte Carlo Transport". In H. Niederreiter and J. Spanier: Monte Carlo and Quasi Monte Carlo Methods. Springer, 86-97.
- Butler, R. M. and Lusk, E. L. 1994. "Monitors, message and clusters: The p4 parallel programming system." *Parallel Computing*, 20, 4, 547-564.
- Chergui, J. and Lavallée, P. F. 2006. "OpenMP Multi tasks parallelism for shared memory machines." (in French). Institut du développement et des ressources en Informatique scientifique.
- Gest, A. et al. 1994. "PVM: Parallel Virtual Machine, a user Guide and Tutorial for network". *Parallel Computing*. The MIT press.
- Kaufmann, A; and Faure, R. 1975. *Invitation to operational research*. (in French). Bordas, Dunod entreprise, Paris.
- Kleindorfer, G. B. 1971. "Bounding distributions for a stochastic acyclic network." *Journal of the Operational Research Society*, 19, 1586-1601.
- Lin, Y. B. 1994. "Parallel independent replicated simulation on networks of workstations". In *Proceedings of the 8<sup>th</sup> Workshop on Parallel and Distributed Simulation*, PADS'94 (edited par I. C. S. Press) 71-81.
- Loh, W. L. 1996. "On Latin hypercube sampling", *The annals of statistics*, 24, 2058-2080.
- Okten, G. et al. 2006. "A central limit theorem and improved error bounds for a hybrid-Monte Carlo sequence with applications in computational finance". *Journal of Complexity*, 22(4), 435-458.
- Owen, A. B. 1998. "Monte Carlo, Quasi Monte Carlo, and randomized Quasi Monte Carlo". In H. Niederreiter and J. Spanier: Monte Carlo and Quasi Monte Carlo Methods. Springer, 86-97.
- Saliby, E. 1990. "Descriptive Sampling: A better approach to Monte Carlo simulation". *Journal of the Operational Research Society*, 41, 12, 1133-1142.
- Tari-Ourbih, M. 2005. "Improvement of descriptive sampling: Application to a production and a scheduling workshop problems" (in French). Doctorat en Science, Bejaia University, Algeria.
- Tari, M. and Dahmani, A. 2005. "The refining of descriptive sampling." *International Journal of Applied Mathematics & Statistics*, 3, M05(march), 41-68.
- Tari, M. and Dahmani, A. 2005. "Flowshop simulator using different sampling methods." *Operational Research: An International Journal*, 5, 2, 261-272.
- Tari, M. and Dahmani, A. 2005. "The three phase discrete event simulation using some sampling methods" *International Journal of Applied Mathematics & Statistics*, 3, D05 (Dec) 37-48.
- Tari, M. and Dahmani, A. 2006. "Refined descriptive sampling : A better approach to Monte Carlo simulation." *Simulation Modelling Practice and Theory*, 14, 143-160.
- Taylor, S. J. E. et al. 2002. "GroupSim: Investigating issues in collaborative simulation modelling". *Operational Reaserch Society Simulation Workshop 2002*. 11-18.
- Tuffin, B. and Le Ny, L. M. 2000. " Parallelization of a combination of Monte Carlo and quasi-Monte Carlo methods networks and application on queuing networks." (in French), *RAIRO Operations Research*, 34, 85-98.



# **VALIDATION VERIFICATION AND OPTIMIZATION**



# PARAMETER VALIDATION USING CONSTRAINT OPTIMIZATION FOR MODELING AND SIMULATION

Guodong Shao, Charles McLean  
Manufacturing Simulation and Modeling Group  
National Institute of Standards and Technology (NIST)  
100 Bureau Drive, MS 8260  
Gaithersburg, Maryland 20899-8260, U.S.A.  
{gshao, mclean}@nist.gov

Alexander Brodsky, Paul Ammann  
*George Mason University*  
Fairfax, VA 22030, U.S.A.  
{Brodsky, Pammann}@gmu.edu

## KEYWORDS

Modeling and Simulation (M&S), Verification and Validation (V&V), Model Parameter, Optimization.

## ABSTRACT

Modeling and simulation (M&S) techniques are increasingly being used to solve problems and aid decision making in many different fields. Results of simulations are expected to provide reliable information for decision makers. But potential errors may be introduced during the M&S development lifecycle. It is critical to ensure to build the right model and the model is built right. M&S community has had intensive Verification and Validation (V&V) research. But V&V activities are often not formally performed in most of the cases. For those who perform V&V activities, they normally wait until development of the simulation modeling is finished. Practical and solid validation techniques are hence needed. In this paper, the authors propose a validation methodology that allows parallel simulation development and model parameter validation, i.e. first the simulation model can be built with unknown parameters included; and then, those parameters can be estimated using a built-in constraint optimizer. Finally the initially unknown parameters are replaced with the found optimal values. The model is then ready for future output prediction. As an example application, a simple supply chain cost simulation model was discussed using the proposed methodology.

## INTRODUCTION

In order to perform the study of the real world problem scientifically, we often have to make a set of assumptions about how it works. These assumptions, which usually take the form of mathematical or logical relationships, constitute a model that is used to gain some understanding of how the corresponding system behaves. If the relationships are simple enough, one may just use an analytic solution that is a mathematical function to express it and obtain exact information on questions of interests. Unfortunately, most real-world problems we are trying to solve are too complex for an exact mathematic function to represent and there may be many parameters that are unknown. They must be studied by means of simulation. That's why simulation is regarded as second only to "math programming" among 13 operations-research techniques (Law and Kelton 2000).

M&S is the process of constructing a model of a system that contains a problem and conducting experiments with the model for a specific purpose of solving the problem and

aiding in decision-making. M&S is particularly valuable for Department of Homeland Security (DHS) applications and manufacturing applications, because it can provide a non-destructive and non-invasive method of observing a system and also provide a way to test multiple inputs and evaluate various outputs (Jain and McLean 2006). For example, in DHS applications, simulations allow users to reconstruct a comprehensive representation of real-world features during disaster response. Simulation models can help the decision makers determine staff and resource levels in hypothetical terrorist attack scenarios (Shao and McLean 2008) (Shao and Lee 2007). But for the developers and users of the simulation models, the decision makers using the results of these models, and individuals affected by decisions based on such models are all concerned with whether a model and the simulation results are correct (Sargent 2007). Even though M&S community has had intensive V&V research (DOD 2001), V&V activities are often not formally performed in most of the cases. Validation efforts have often been limited to the use of less rigorous techniques, such as face validation and traceability assessment (Sargent, et al., 2000). Practical and solid validation techniques are hence needed to make sure the simulation model is validated and the simulation results are credible.

The contribution of this paper is to propose a novel validation methodology that allows parallel simulation development and model parameter validation. The technique integrates constraint optimizer that performs the parameter validation for M&S. First the simulation model can be built with unknown parameters included; and then, those parameters can be estimated using a built-in constraint optimizer. Finally the initially unknown parameters are replaced with the optimal values. After validating the simulation results using corresponding set of input data, the model is ready for future output prediction. The constraint optimizer uses Constraint Optimization Regression in Java (CoReJava) that implements Regression Analysis (RA) to estimate the parameters based on a training data that could either be historical data or experimentation data (Brodsky, et al. 2008).

The rest of the paper is organized as follows: next section identifies the V&V needs and issues. Then related work and technologies are discussed. The parameter validation technique is presented. A methodology to validate the technique, and a simple supply chain example modeled using CoReJava and a simulation tool are introduced. Finally the paper is concluded with future works and discussion.

## VERIFICATION AND VALIDATION NEEDS AND ISSUES

During the development lifecycle of M&S, risks associated with potential errors in creating the model (programming errors) and inadequate fidelity (errors in accuracy when compared to real-world results) may be introduced (Cook and Skinner 2005). To guarantee that you have a valid model and simulation that produces correct results, V&V of the model and data used for the simulation must be employed throughout the life cycle of an M&S application.

Balci (Balci 2007) defines the model V&V as follows:

“Model validation is substantiating that the model, within its domain of applicability, behaves with satisfactory accuracy consistent with the study objectives. Model validation deals with building the right model. It is conducted by running the model under the “same” input condition that drive the system and by comparing model behavior with the system behavior. Model verification is substantiating that the model is transformed from one form into another, as intended, with sufficient accuracy. The accuracy of transforming a problem formulation into a model specification or the accuracy of converting a model representation in micro flowchart into an executable computer program is evaluated in model verification.”

Figure 1 is the Sargent’s circle - a simplified version of the M&S process (Sargent 2007). The problem entity shown in the figure could be a real or proposed system, idea, situation, policy, or phenomena to be modelled.

1. Conceptual model validity should answer the questions: Is the description of the system sufficient and correct? Is it valid for the intended use?
2. Computerized model verification deal with the questions: Is the numerical implementation of the model correct? Are the numerical algorithms employed correct and fully converged?
3. Operational validity answers the questions: Are we able to predict the experiment(s) in sufficient detail? How do we formulate quantitative validation metrics given a specific application?
4. Data validity answers the questions: Is the experimental data used in the comparisons a sufficiently accurate description of reality? How do experimental uncertainties affect predictive performance? Are the experiments used in the validation exercise appropriate?

To perform a complete validation of the model, appropriate validation techniques need to be applied to each step. Figure 2 shows the simulation modeling process steps and each of them may be a source of errors that will influence the validity of the model. For example, incorrect conceptual modeling will make the model invalid for the intended use. Lack of calibration of the parameters will not produce an accurate function that sufficiently describes the problem. Implementation error will make the simulation model invalid even if the conceptual model is valid. The use of poor quality input data will increase the risk of providing incorrect results to the users, i.e. trash in will have trash out. Unsatisfied operational conditions will cause wrong estimates. Results comparison is to compare the observed

data and the simulation outputs. A large number of data are needed to have a meaningful evaluation of model performance in statistical terms. We should note that the model predictions and measured data will never match exactly; trends over time are one of the most useful tools to evaluate model performance (Donatelli and Stockle 1999).

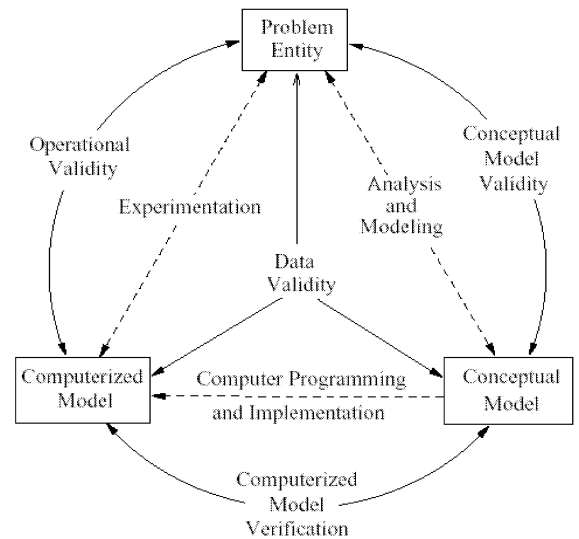


Figure 1. Simulation modeling and validation process (Sargent’s circle) (Sargent 2007)

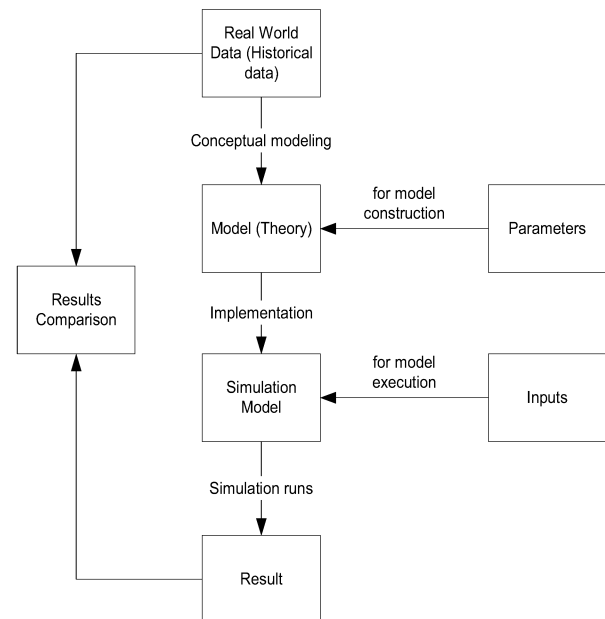


Figure 2. Simulation modeling and executing steps

## RELATED WORK AND TECHNIQUES

This section discusses the related research work, techniques, and tools for the proposed methodology.

Law and Kelton suggested that quantitative techniques should be used whenever possible to test the validity of various components of the overall model (Law and Kelton 2000). An important technique for determining which model factors have a significant impact on the desired measures of performance is sensitivity analysis, the factors may be the value of the parameters, or the choice of a distribution, etc. One approach to determine the sensitivity of the factors is to use statistical experimental design. In experimental-design terminology, the input parameters and structural assumptions composing a model are called factors; the output performance measures are called responses. Factors can be quantitative or qualitative, controllable or uncontrollable (Law and Kelton 2000).

In (Doebling and Hemez 2007), model validation is also defined as “The process of assessing and improving confidence in the usefulness of computational predictions for a particular application” and “Solving the right equations.” Model validation is an application-specific process. Fidelity of the model relates to agreement with real world/test data, validity relates to suitability for the specific application. Validity of a model is defined over a region of the parameter space.

Model validation supporting technologies include:

- Metamodeling - simplified relationship between model parameters and response features,
- Design of Experiments - generate metamodels & plan validation tests,
- Parameter Optimization - quantify unmeasured variables, calibrate surrogate mechanics models, and
- Data Compression - extracting features from simulation and test data.

Machine learning techniques enable us to estimate the system parameters with specified confidence intervals using historical data, predict the outcome by given a new input, identify adjustment to the system parameters to meet performance requirements (Zabaras 2003).

Brodsky, Luo and Nash proposed and implemented the language CoReJava, which extends the programming language Java with Regression Analysis (RA), i.e. the capability to perform parameter estimation for a function. In a Java program, some parameters are not a priori known, but can be learned from training sets provided as input. Existing RA software typically requires inputting a data structure that describes the parametric functional form, or assumes this data structure to be fixed. The problem, however, is that in many applications, a functional form is not explicitly available. CoReJava allows the user to encode complex computational processes in Java, in which some parameters used are not a priori known. Unknown parameters can be learned from a training data set. The CoReJava compiler analyzes the structure of the learning function method to automatically generate a constraint optimization problem, in which constraint variables correspond to parameters that need to be learned. The objective function to be minimized is the summation of squares of errors with respect to the training set, and then solves the optimization problem using the non-linear optimization solver - A Mathematical Programming Language (AMPL)/SNOPT. (Brodsky, et al.

2008) provides detail descriptions of language syntax, use, and semantics of CoReJava.

AMPL is a comprehensive and powerful algebraic modeling language for linear and nonlinear optimization problems, in discrete or continuous variables. A few solvers such as CPLEX 11, SNOPT, and MINOS are free to download from (Bell Lab 2008).

## PROPOSED PARAMETER CALIBRATION AND VALIDATION METHODOLOGY

The proposed novel validation methodology will focus on the parameter calibration and validation as shown in Figure 2. A validated set of system parameters for a specific problem function will allow users to properly characterize the system under study. There is no universal model that would work with an unaltered set of parameters for all conditions. Adjustment of parameter values must be done within the range known for the parameters. Calibration assumes the availability of observed data to adjust model parameters in order to match model outputs to measured data.

Depicted in Figure 3 is the model parameter calibration and validation approach. In practice, most of the validation processes do not start until the completion of M&S development. In our proposed methodology, the validation of the simulation model can be parallel to the development. First we build the simulation template with unknown key parameters included. We can think of it as a black-box model. To do this, we need to first analyze the data available and decide what we should measure and what parameters we do not know and need to be estimated. Then we start the parameter validation process, that is to find out the “correct” model parameters. This can be done by applying an RA tool. The RA module will find the best estimate of the unknown parameters by learning from the available training data sets. The training data sets may be either real world historical data or experiments data. RA is one of the metamodeling techniques for investigating and modeling the relationship between variables. As input to RA, a parametric functional form can be either linear or non-linear, e.g.,  $f(x_1, x_2, x_3) = p_1x_1 + p_2x_2 + p_3x_3$ , and a set of training examples, e.g., tuples of the form  $(x_1, x_2, x_3, f)$ , where  $f$  is an experimental observation of the function  $f$  value for an input  $(x_1, x_2, x_3)$ . The goal of RA is to find the unknown parameters, e.g.,  $p_1, p_2, p_3$  that “best approximate” the training set. Once we find the optimal set of the parameters, we replace the variables in the simulation template with the parameters values. The simulation model then becomes a white-box model. By feeding in new inputs, the simulation can produce and predict valid outputs. When compare the simulation results to the historical data, we hope that the simulation results will be as closely as possible to the collected data within a confident interval. If the comparison results are not satisfied, simulation template needs to be verified, or the parameters needs to be re-estimated using more data or new data, even the simulation model execution scenarios, conditions need to be checked to make sure sufficient runs are performed. After iteratively comparing

with the training data and modifying the model, we can eventually obtain a more validated model over that particular parameter domain and valid data range.

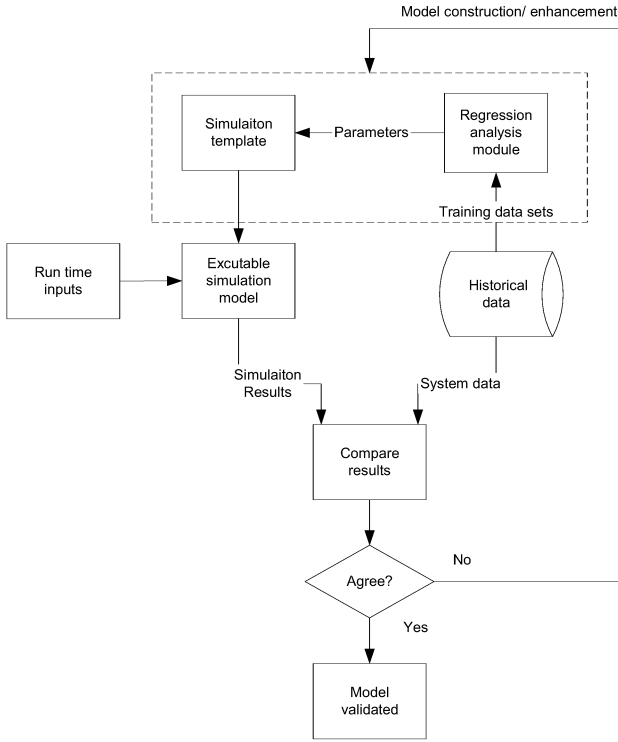


Figure 3. Simulation model parameter validation approach

## VALIDATION METHOD FOR THE PROPOSED TECHNIQUE

In order to validate the proposed technique, we need large sets of historical data. Based on the different category of the collected data, we can divide the data into at least two groups, for example, use one set of data to train and next set of data to validate the result. We should use one of the groups as the RA training data, once we derived the best estimate model parameters values, the other data sets can be used to verify the result. We can check if the differences between the simulation outputs and the real collected data are within a confident interval.

We use a simple example to explain the proposed methodology. Figure 4 depicts a simplified supply chain cost model, a functional form may be given that computes the total cost of manufacturing, given three products to be produced. This functional form may have unknown parameters, e.g., the unit costs and the required quantities of component materials to produce specific products.

The manufacturer produces three products using three components. The quantities of component materials needed are functions of the required quantities of products. The cost of the produced products is the total cost of the required components. Thus, the cost of manufacturing is a function of the required quantities of products. However, the coefficients of this function, i.e. the unit cost of each component and the amount of each component material to

produce 1 unit of each product may be unknown and subject to RA, which is provided by CoReJava.

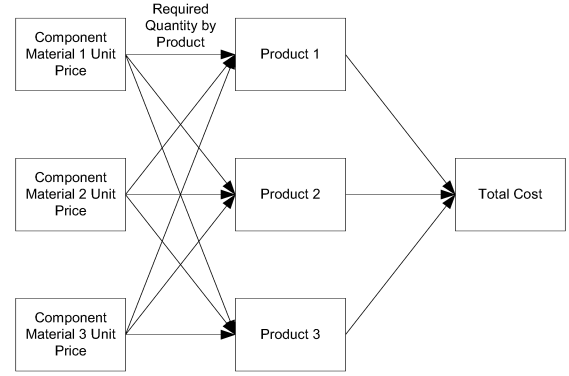


Figure 4. A Simple supply chain cost example

The example historical data as training data set is listed in Table 1. Each row is a learning set. Each learning set includes three product quantities and the actual total cost recorded. The size of the table is decided by the information the user has collected (Brodsky, et al. 2008).

**Table 1.** Input learning set

| Product 1 Quantity | Product 2 Quantity | Product 3 Quantity | Actual Total Cost |
|--------------------|--------------------|--------------------|-------------------|
| 8                  | 5                  | 9                  | 20                |
| 9                  | 7                  | 6                  | 18                |
| 7                  | 6                  | 14                 | 25                |
| 10                 | 11                 | 12                 | 29                |
| 5.5                | 12.1               | 9.8                | 31.1              |
| ...                | ...                | ...                | ...               |
| 11.2               | 9.6                | 6.5                | 25.33             |

Figure 5 shows the simulation result using CoReJava for a product set of (64.2, 50.4, 35.5) as inputs, i.e. the product1 quantity is 64.2, product2 quantity is 50.4, and product3 quantity is 35.5. We need to determine the actual total cost for this set of products. In Figure 5, the two-dimensional array reqMatQty represents the quantity of each required material to produce one unit of every product. The array matUnitCost represents the unit cost of every material. The data includes the values of coefficients (matUnitsCost and reqMatQty arrays) and the total cost (Brodsky, et al. 2008).

For comparison purpose, a simulation model of the same problem was also developed using a simulation software, which does not have a RA module. Data in Table 1 alone is not sufficient for constructing the model. Important data such as component unit price, and amount of components material needed is not available. A triangular distribution is typically used in a model for a source of randomness when no system data are available. T (0.3, 1, 3) is chosen for that, we used the min\_Bound value in CoReJava example as the minimum value of the triangular distribution, which is 0.3 and used the max\_Bound value in CoReJava example as the maximum value, which is 0.3, for the triangular distribution, then arbitrarily choose a mode value as 1. From the results showed in Figure 6, we can see that the results are far apart from the results of CoReJava. This is because the model did not incorporate the existing data sets. From the results



comparison, we can see this model did not accurately describe the problem. By using the proposed technique, the parameters are ensured to be learned based on the past historical data. The model will properly represent the problem. The results will be within a confident interval of the existing data. That makes the proposed model more valid for the specific application.

```

[echo] ##### 4. Interpreting the optimized results...
[java] objective: 5533.03269354264
[java] matUnitCost[0]:0.3
[java] matUnitCost[1]:0.3
[java] matUnitCost[2]:0.3
[java] reqMatQty[0][0]:0.741492
[java] reqMatQty[0][1]:0.3
[java] reqMatQty[0][2]:0.45798
[java] reqMatQty[1][0]:0.741492
[java] reqMatQty[1][1]:0.3
[java] reqMatQty[1][2]:0.45798
[java] reqMatQty[2][0]:0.868433
[java] reqMatQty[2][1]:0.3
[java] reqMatQty[2][2]:1.03503
[java] The cost to produce a set of products:19.067327489999997

```

Figure 5. Simulation result using CoReJava.

| Report for products_v2 (Avg. of 999 replications)    |               |                           |               |               |               |           |
|--|---------------|---------------------------|---------------|---------------|---------------|-----------|
| Variables for products_v2 (Avg. of 999 replications) |               |                           |               |               |               |           |
| Name   | Total Changes | Avg Time Per Change (MIN) | Minimum Value | Maximum Value | Current Value | Avg Value |
| total cost   | 0.00          | 0.00                      | 0.00          | 0.00          | 0.00          | 0.00      |
| number of product1                                   | 0.00          | 0.00                      | 64.20         | 64.20         | 64.20         | 64.20     |
| number of product2                                   | 0.00          | 0.00                      | 50.40         | 50.40         | 50.40         | 50.40     |
| number of product3                                   | 0.00          | 0.00                      | 35.50         | 35.50         | 35.50         | 35.50     |
| product1 cost  | 1.00          | 0.00                      | 0.00          | 6.09          | 6.09          | 6.09      |
| product2 cost  | 1.00          | 0.00                      | 0.00          | 6.13          | 6.13          | 6.13      |
| product3 cost  | 1.00          | 0.00                      | 0.00          | 6.23          | 6.23          | 6.23      |
| total cost   | 1.00          | 0.00                      | 0.00          | 920.96        | 920.96        | 920.96    |

Figure 6. Same example modeled using distribution

# CONCLUSION

M&S techniques are increasingly used to solve problems and aid decision making in many different fields. Results of simulations are expected to provide reliable information for the decision makers. However potential errors may be introduced in the process of the M&S development lifecycle. It is critical to make sure to build the right model and that the model is built right.

This paper demonstrated a novel approach of unknown parameters calibration and validation through constraint optimization based on training data sets within a data range. The technique proposed a parallel process of developing and validating the simulation model, before every parameter is known. A simulation template can be built with unknown variables in it. By using a built-in RA module that learns the training data sets, the best estimates of the variable values can be derived. This will help to ensure the accurate relationship between independent inputs and dependent output. Future work may include the development generic UML model for the RA module. That will provide a software independent design of the validation technique. Any interested simulation software vendor can implement it as a module of their product. Also more implementation for real world problems may be needed to validate the technique.

# REFERENCES

Sargent, R. 2007. "Verification and validation of simulation models," In *Proceedings of the 2007 Winter Simulation*

*Conference*, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton. 124--137.

Law, A. and W. Kelton. 2000. *Simulation Modeling and Analysis*, McGraw-Hill Higher Education, Boston.

Balci, O. 2007. "Validation, Verification, and Testing Techniques throughout the Life Cycle of A Simulation Study," *Annals of Operation Research*, 53 (1994), New York, pp. 121-173.

Jain, S. and C. R. McLean. 2006. "An Integrating Framework for Modeling and Simulation for Incident Management," *Journal of Homeland Security and Emergency Management*, Vol. 3, Iss. 1, Article 9. Available at: <http://www.bepress.com/jhsem/vol3/iss1/9>.

Cook, D. and J. Skinner. 2005. "How to Perform Credible Verification, Validation, and Accreditation for Modeling and Simulation."

Sargent, R. P. Glasow, J. Kleijnen, A. Law, I. McGregor, and S. Youngblood. 2000. "Strategic Directions in Verification, Validation, and Accreditation Research," In *Proceedings of the 2000 Winter Simulation Conference*. J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

Brodsky, A. J. Luo, and H. Nash. 2008. "CoReJava: Learning Functions Expressed as Object-Oriented Programs," International Conference on Machine Learning and Applications (ICLMA) Forum 2008.

Zabaras, N. 2003. "Simulation matching and model estimation using statistical learning techniques," *Advanced Mechanical Technologies*, GE Global Research Center, NY.

Bell Laboratories. 2008. AMPL. [Http://www.ampl.com](http://www.ampl.com)

Department of Defense. 2001. "Department of Defense Verification, Validation and Accreditation (VV&A) Recommended Practices Guide," Defense Modeling and Simulation Office, Alexandria, VA. (Coauthored by: O. Balci, P. A. Glasow, P. Muessig, E. H. Page, J. Sikora, S. Solick, and S. Youngblood.) [http://vva.dmsomil/Ref\\_Docs/VVTechniques/vvtechniques.htm](http://vva.dmsomil/Ref_Docs/VVTechniques/vvtechniques.htm)

Doebeling, S. and F. Hemez. 2008. "Finite Element Model Validation, Updating and Uncertainty Quantification, A 2-day short course." [http://www.la-dynamics.com/courses/validation\\_sample\\_vg\\_0302.pdf](http://www.la-dynamics.com/courses/validation_sample_vg_0302.pdf)

Shao, G. and C. McLean, 2008. "Emergency Room Simulation Prototypes for Incident Management Training," In *Proceeding of Industrial Simulation Conference 2008*, pp. 323-327.

Shao, G. and Y. Lee. 2007. "Applying Software Product Line Technology to Simulation Modeling of Emergency Response Facility," *Journal of Defense Modeling and Simulation*, The Society for Modeling and Simulation International, Vol. 4(4), October 2007. <http://www.scs.org/pubs/jdms/vol4num4/vol4num4.html>.

Donatelli, and Stockle. 1999. "Model calibration and validation, short course on the model cropsyst," Middle East Technical University. - Dept. of Economics, Ankara, Turkey. [http://www.bsyse.wsu.edu/cropsyst/Documentation/lectures/crobot/lect2\\_Calibration.PDF](http://www.bsyse.wsu.edu/cropsyst/Documentation/lectures/crobot/lect2_Calibration.PDF)

# APPLYING A DOCUMENTATION GUIDELINE FOR VERIFICATION AND VALIDATION OF SIMULATION MODELS AND APPLICATIONS: AN INDUSTRIAL CASE STUDY

Zhongshi Wang<sup>1</sup>, Heike Kißner<sup>2</sup>, Martin Siems<sup>2</sup>

<sup>1</sup> Institut für Technik Intelligenter Systeme (ITIS)  
Universität der Bundeswehr München  
D-85577 Neubiberg, Germany  
zhongshi.wang@unibw.de

<sup>2</sup> szenaris GmbH, Otto-Lilienthal-Str.1  
D-28199 Bremen, Germany  
{heike.kissner, martin.siems}@szenaris.com

## KEYWORDS

Documentation of Model Development, Documentation of Model Verification and Validation, Case Study

## ABSTRACT

Documentation in a structured and well-defined way becomes an urgent requirement not only for developing simulation models and applications (M&S), but also for conducting their verification and validation (V&V). Since meaningful model documentation is always associated with more effort than usual, a reasonable balance between quality and efficiency should be achieved in all practical applications. This work presents an industrial case study of applying a documentation guideline to a small simulation development project. The main objectives of this case study were firstly to estimate the effort involved in the documentation of M&S work products and their V&V in this project context, and in addition, to investigate the feedback for improving the proposed documentation guideline. Based on the gathered findings, a further case study for a more complex simulation project is planned.

## INTRODUCTION

That the documentation of modelling and simulation (M&S) applications is an essential issue for a successful simulation project is hardly controversial. However, under pressure, time and cost constraints, model documentation in practical applications is often sacrificed first (Balci 1997), or conducted only in an arbitrary and informal manner (Gass 1984, Kleijnen 1995). Suchlike problems of documentation lead not only to loss of operational efficiency, continued use and further development of simulation models (or their components), but also to increasing risks of using improper results in terms of credibility assessment.

Faced with this practical challenge, the Institute for Technology of Intelligent Systems (ITIS) has developed a model documentation guideline (Lehmann et al. 2005), which enables structured and well-defined documentation for developing M&S applications as well as conducting their verification and validation (V&V). Furthermore, a tailoring concept is also proposed, by means of which project-specific model documentation can be determined with respect to actual cost, time and application constraints. For the purpose of concept refinement, empirical investigations are necessary to learn more about the transformation and application of this guideline in practice. Hence a series of case studies were planned, which enable the estimation of this guideline in different project contexts, on the one hand, and the close cooperation of ITIS with a selected simulation provider, namely the company szenaris GmbH which was the simulation and training division of Ray Sono AG at the time when the studies were planned and partially conducted, to gather empirical findings and exchange their knowledge, on the other hand.

This paper presents the first effort of the planned investigations, in which the proposed documentation guideline was applied to a relatively small and manageable simulation development project, the so-called KoCUA FSB (Ray Sono AG 2008). However, due to some unexpected limitations related to the project execution, the complete documentation of the simulation model and its V&V was no more possible in this regard, and some kind of adaptation had to be arranged. Thus, the purposes of this case study were modified as follows: (1) estimating the effort expended in the documentation activities for a rough cost-benefit analysis; and (2) investigating the feedback for further development of the proposed guideline.

The remainder of this paper is structured as follows: a brief overview of the applied documentation guideline is introduced in the following section. Then the simu-

lation project selected for this case study is described. After that, the execution and some relevant results of this case study are presented. Based on these results, lessons learned are discussed, and the last section concludes with thoughts on further work.

## The APPLIED DOCUMENTATION GUIDELINE

The proposed model documentation guideline defines detailed requirements, useful and applicable templates for documenting an M&S project and its associated V&V efforts throughout the entire model development life cycle, including the following essential components (as shown in Figure 1):

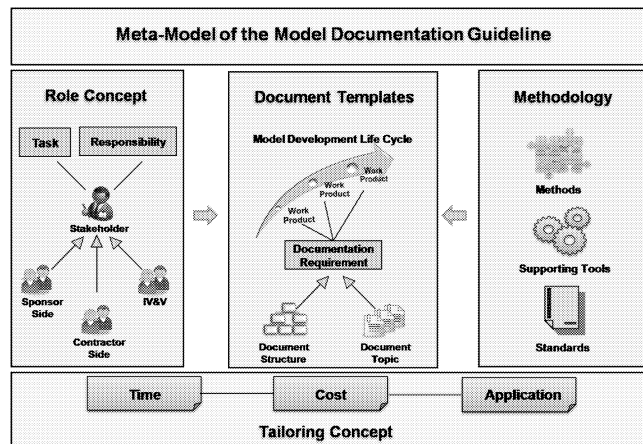


Figure 1: Overview of the Applied Guideline for Model Documentation

- **Meta models.** The fundamental structuring of the guideline is illustrated in the form of meta models, which serve as a common basis to describe the introduced core elements, such as the role concept, the documentation templates, the methodological supporting and the tailoring concept, and relationships between these elements as well as their semantic. In addition, all meta models are specified and organized in different degrees of abstraction to meet the specific needs of each user group of this guideline.
- **Roles.** With the intention of defining functional tasks and competencies for persons, who will participate in the model development and documentation activities, a well-structured role concept is introduced. During the process of project planning, roles are assigned to concrete persons or organizational units with certain responsibilities, e.g. contributory or responsible. While several roles may support in the creation of a product, only one responsible role can be assigned to each work prod-

uct. Since an M&S project should normally involve three kinds of parties (Wang and Lehmann 2007), namely participants from the sponsor side, the contractor side, and the independent V&V (Arthur and Nance 2000), this role concept has been developed in a manner which offers organization-independent orientation for project management.

- **Document templates.** To facilitate the model documentation activities, concrete structure and contents requirements for each work product created during the M&S life cycle are specified in the form of document templates. As the reference process, a structured process for developing M&S applications and conducting their V&V (Brade 2000) is employed. With respect to the compatibility with current M&S (quasi) standards (Balci and Saadi 2002, Sargent 2003), this process has been mapped to several selected international M&S V&V processes (Brade et al. 2005). All of the proposed templates can be directly applied to an M&S project or adapted as required before use.
- **Methodological supporting.** With the aid of some selected practical methods, supporting tools, and standardized processes (The V-Modell XT 2005) specified in this guideline, it is possible for each potential user to find a more efficient way of preparing model documentation and keeping consistent information across all created documents.
- **Tailoring.** Because of different characteristics of organization structures and project environments, this model documentation guideline needs some kind of adaptation or tailoring prior to application. For this purpose, a multistage tailoring concept is established, which enables the project-specific selection of essential products, documents, and activities for developing M&S applications and conducting their V&V according to specified cost, time and application constraints. By applying this tailoring mechanism, only relevant M&S products are taken into account during model development process. This implies a reduction of project complexity.

## CASE STUDY CONTEXT

In order to gain first experiences in practical application, this guideline was used to document the simulation model being developed by the company szenaris in the context of the project KoCUA FSB (Ray Sono AG 2008). As Figure 2 shows, the project considered in this case study was intended to build a training simulator for cooperative training in teams (KoCUA) of an Improved Ribbon Bridge (FSB), which was to be used in training of military personnel for the exercise scenarios like transporting, launching, and retrieving interior and ramp bays of a float bridge, moving and connecting

bridge bays on the water etc. As an extended system component, the developed simulator had to be then integrated into an existing computer aided training system environment, and cooperate with other components.



Figure 2: Team Training with the Simulator

The case study of using the documentation guideline was defined as a separate project in addition to the main development project, and had to be conducted cooperatively by the company szenaris and the research institute ITIS. While szenaris had to implement the proposed guideline and prepare model documents for each work product created during the development of the training simulator, ITIS provided support of coaching when required and evaluated the quality of completed model documents in form and content.

Unfortunately, the planning of this case study was behind schedule with the simulator development project. Starting already in November 2007, the development of the training simulator was planned to be exactly one year. The kickoff meeting of the case study project, however, was held only in March 2008. This meant that the completely continuous model documentation in parallel with each phase of the M&S life cycle was no more possible. Since the model development was under a company specific modelling process already in the programming phase at that time, and the documentation so far was only performed fractionally in an informal way, some kind of “post-documentation” had to be arranged and conducted. Additionally, due to the somewhat short duration of the case study project, only a continuous subset of the developed simulation model was selected from the sponsor needs, the M&S requirements specification, the conceptual model to the executable model, and then consistently documented. Thus, the main purposes of this case study were limited to:

- on the part of szenaris
  - using the proposed tailoring concept to select a subset of the model being developed for documentation

- preparing model documents for each M&S development phase according to the guideline
- preliminary cost-benefit analysis for application of the guideline
- on the part of ITIS
  - supporting the model documentation as well as the tailoring activities
  - if possible, checking correctness and consistency of the subset of the model based on the prepared model documents (mainly for estimation of V&V effort)
  - gathering feedback for the purpose of further development

## EXECUTION AND RESULTS OF THE CASE STUDY

After selecting a subset of the sponsor needs with respect to the actual project conditions supported by ITIS, the team members from szenaris began backwards to prepare the associated model documents of each already performed M&S phase for the selected requirements, and regularly sent their finished versions as email attachments. The received documents were then inspected by the personnel from ITIS. The inspection focused on (1) whether the prepared model documents met the requirements of the proposed guideline formally; and (2) whether they were correct and consistent with respect to the selected subset of the M&S requirements. The objective of the inspection was not to conduct V&V of the overall simulation model being developed officially, but rather to acquire a preliminary effort estimation of applying this documentation guideline.

To facilitate communication between the two working teams from the different locations (Bremen and Munich, Germany), the solution of video conferencing was applied. Depending on requirements, open questions and different opinions were exchanged through video conferencing weekly or biweekly. Additionally, four workshops were arranged for the purpose of face-to-face discussion, and distributed evenly over the project period from March to November 2008.

Table 1: Model Documents Prepared in the Case Study

| Predefined work products            | Versions |
|-------------------------------------|----------|
| Structured Problem Definition (SPD) | 5        |
| Conceptual Model (CM)               | 7        |
| Formal Model (FM)                   | 2        |
| Executable Model (EM)               | 7        |
| Simulation Result (SR)              | 1        |

Following the instructions of the applied guideline, model documentation in this case study was conducted

in an iterative manner. Hence, as shown in Table 1, there exist different but consecutive model documents for each developed M&S phase. Compared to seven document versions of CM, only two documents of FM were prepared. This was due to the fact that the simulation environment “Virtools” (Virtools 2008) was applied to the development of the training simulator. Since the physically correct behaviors and interactions of all the objects considered in the simulation study were realized by means of the “Havocs Physics Engine” which is integrated in the simulation environment of Virtools, and its associated mechanisms, however just like a black box, are not available to the users, an explicit description of the formal model developed in this way was extremely difficult, and therefore was not considered as a part of model documentation in this case study. In the present two documents of FM, the reasons for this tailoring decision were documented. The single document of SR included the operation protocols and results of different test scenarios.

Table 2: Estimated Effort on the part of szenaris

| Work expended                    | Pers. days |
|----------------------------------|------------|
| 1. Guideline introduction        | 16.5       |
| 2. Preparing the model documents | 58.5       |
| 3. Quality assurance             | 10         |
| 4. Workshops and meetings        | 15         |
| <b>Total</b>                     | <b>100</b> |

Table 3: Estimated Effort on the part of ITIS

| Work expended                        | Pers. days |
|--------------------------------------|------------|
| 1. Coaching                          | 8          |
| 2. Inspection of the model documents | 20         |
| 3. Workshops                         | 5          |
| <b>Total</b>                         | <b>33</b>  |

Table 2 and 3 show the respective estimates of the expended effort directly related to the model documentation activities of szenaris and ITIS. Other expenditure such as project management, business trips, or preparation of the final report is not taken into account in this context. Since only a fractional part of the developed simulation model was considered in the case study and the calculation made in Table 2 was just based on the selected subset, the real expenditure required for a complete model documentation on the part of szenaris could be incurred approximately five times higher.

## LESSONS LEARNED

Based on the findings achieved in this case study, some relevant lessons learned are summarized, which could serve as expertise for both theoretical concept refine-

ment and practice applications. They are largely independent of the current project constellation and as such can be applied to other M&S projects as well.

- **The applied guideline was perceived as beneficial by the team members of szenaris.** For instance, in the course of the model development, it became apparent that the original interaction concept, which had been developed in the context of an early simulation project and was applied as a finished component to the conceptual model (CM) in this project, was in some respects incomplete and thus caused confusing ambiguity and misunderstanding within the model development team. In this regard, the precisely conducted model documentation according to this guideline was an appropriate remedy.
- **Standardized formalization methods are indispensable for model documentation.** Conducting a simulation project always requires multifaceted knowledge in diverse disciplines. A development team alone includes different project roles with individual expertise. Hence, using standardized formalization methods enables a common platform for understanding and exchange of opinions not only within a model development team, but also between the three participating parties of sponsors, contractors and IV&V.
- **Adaptability is one of the key criteria for application of a model documentation guideline.** Since there exist no two organizations with the same structural properties, and furthermore, every simulation project differs in terms of objectives, scale, scope and technical challenges, a model documentation guideline could be never well suited for all possible circumstances. Therefore, the potential of an M&S documentation guideline to be adapted to different project environments is a crucial issue for its acceptance in practice.
- **Know-how protection is still an essential topic concerning model documentation and (independent) model V&V.** From the V&V’s point of view, precise and detailed model description significantly facilitates V&V efforts. On the other side, however, a simulation provider usually expresses great concerns about the possible loss of his estimable know-how, when all key parts of models have to be available to an independent third party by reason of V&V. Therefore, additional investigations should be carried out for the purpose of conducting model V&V more effectively and keeping company’s know-how safe contractually and legally.
- **The expenditure of model documentation and model V&V should be regarded as early**

as tendering of an M&S project. In order to prevent that model documentation and V&V become an easy prey because of time and cost pressures occurred during project execution, sponsors should require the separate calculation of planned documentation and V&V costs in request for proposal, and all potential contractors must give a realistic cost estimate in their project offers.

## CONCLUSION AND FUTURE WORK

This paper presents an empirical investigation of applying a documentation guideline to a practical M&S project. Due to some inevitable limitations, however, the documentation of the complete simulation model being developed and its associated V&V activities could not be conducted in this case study. Hence, a consistent subset of the model was determined by means of the proposed tailoring concept according to the actual project conditions. Additionally some kind of “post-documentation” had to be partially conducted. According to the selected modelling aspects, the model documents were prepared in an iterative manner for each development phase of the M&S life cycle.

The applied documentation guideline, especially due to its well-defined structure and adaptability for different project constellations, was appreciated as understandable and profitable by the development team of szenaris. The experiences gained from this case study can be used to improve the model development process of szenaris as well as to refine the proposed documentation guideline. Additionally, several lessons learned provide insights into general issues concerning model documentation and model V&V.

Since the calculation of the effort expanded in this investigation was based on the partially implemented model documentation, the impact of introduction and application of this guideline on the overall project costs could only be estimated roughly on a large scale. Thus, in the context of this case study, a reasonable cost-benefit analysis is infeasible. Our future work is intended to conduct a further industrial case study, in which this documentation guideline will be applied to a more complex simulation project, and model documentation will be conducted completely and continuously with model development. Additionally, the effort required for a systematic model V&V will be also investigated.

**Acknowledgment.** The authors would like to thank the Bundeswehr Transformation Centre (ZTransfBw) and the German Federal Office of Defense Technology and Procurement (BWB) for financial support of this research.

## REFERENCES

- Arthur J.D. and Nance R.E., 2000. *Verification and Validation without Independence: A Recipe for Failure*. In: Proceeding of the 2000 Winter Simulation Conference.
- Balci O., 1997. *Verification, Validation and Accreditation of Simulation Models*. In: Proceeding of the 1997 Winter Simulation Conference.
- Balci O. and Saadi S.D., 2002. *Proposed Standard Processes for Certification of Modeling and Simulation Applications*. In: Proceeding of the 2002 Winter Simulation Conference.
- Brade D., 2000. *Enhancing Modeling and Simulation Accreditation by Structuring Verification and Validation Results*. In: Proceeding of the 2000 Winter Simulation Conference.
- Brade D.; Jaquart R.; Voogd J.; and Yi C., 2005. *Final State of the REVVA Methodology*. In: Proceeding of the 2005 Spring Simulation Interoperability Workshop.
- Gass S.I., 1984. *Documenting a Computer-Based Model*. Interfaces 14 (3): pp. 84-93.
- Kleijnen J.P., 1995. *Verification and Validation of Simulation Models*. European Journal of Operational Research, 82, no. 1, pp. 145-162.
- Lehmann A.; Bel-Haj-Saad S.; Best M.; Köster A.; Pohl S.; Qian J.; Waldner C.; Wang Z.; and Xu Z., 2005. *Leitfaden für Modelldokumentation*. Abschlussbericht, ITIS e.V. an der Universität der Bundeswehr München.
- Ray Sono AG, 2008. *Studie V&V Modelldokumentation an der “Kooperativen Computerunterstützten Ausbildung Faltschwimmbrücke (KoCUA FSB)”*. Abschlussbericht, Bremen, Germany.
- Sargent R.G., 2003. *Verification and Validation of Simulation Models*. In: Proceeding of the 2003 Winter Simulation Conference.
- The V-Modell XT, 2005. *Documentation Edition 1.01, 2005*. Online: <http://www.v-modell-xt.de>.
- Virtools, 2008. Online: <http://www.virttools.com>.
- Wang Z. and Lehmann A., 2007. *A Framework for Verification and Validation of Simulation Models and Applications*. In J.W. Park; T.G. Kim; and Y.B. Kim (Eds.), *AsiaSim 2007*, Springer-Verlag, Berlin Heidelberg, Communications in Computer and Information Science. pp. 237-246.

# Performances evaluation and optimization of a failure-prone manufacturing system with random delivery time and random demand

Turki Sadok<sup>1</sup>  
Hennequin Sophie<sup>2</sup>  
Sauer Nathalie<sup>1</sup>

<sup>1</sup>COSTEAM- INRIA Nancy Grand Est /LGIPM – Université de Metz, Ile du Saulcy, 57045 Metz Cedex, France  
{sadok.turki | nathalie.sauer}@univ-metz.fr

<sup>2</sup>COSTEAM- INRIA Nancy Grand Est /LGIPM-ENIM Ile du Saulcy, 57045 Metz Cedex, France  
(hennequin@enim.fr)

## KEYWORDS

*Failure-prone manufacturing system, discrete flow model, random delivery time, performances evaluation, optimization, perturbation analysis.*

## ABSTRACT

In this paper, a manufacturing system composed by a single machine, a buffer and a stochastic demand is considered. A discrete flow model is proposed taken into account stochastic delivery times. The goal of this paper is to evaluate the optimal buffer level used in hedging point policy taken into account random delivery times, machine failures and random demands. This optimal buffer permits to minimize the total cost which is the sum of inventory cost, backlog cost, transportation cost, early delivery cost and delayed delivery cost. These two last costs correspond to products which will not arrive at exact expected time. A simulation based on perturbation analysis is then proposed for performances evaluation and optimization and results are compared with a discrete events system based simulation.

## INTRODUCTION

Many manufacturers are working to reduce transportation delays such as the delivery time, which is the period of time that the material takes between a warehouse and a customer. These manufacturers are then using this planned delivery time as a marketing strategy to attract customers. Realizing the importance of delays, Van Ryzin et al. (1991) considered the impact of delays for optimal flow control of job shops in order to minimize the discount and infinite-horizon average cost. Mourani et al. extended the model of Van Ryzin and proposed a continuous Petri net model with delays for performances modeling and optimization of transfer lines (Mourani et al. 2005; 2008). Dolgui and Aly Ould-Louly (2002) presented a model for supply planning under lead time uncertainty; and proposed a method to determine the optimal value of the planned lead time under lead time uncertainty.

In this paper, we consider the delivery times as a random variable, the producer proposes for customers a fixed delivery time called planned delivery time. If the materials arrive to the customer before the end of the planned delivery time (early delivery materials), the early delivery materials are accepted by the customer, but are penalized with a early delivery cost, in other side, when the materials ar-

rive to the customer after the end of the planned delivery time (delayed delivery materials), the delayed delivery materials are accepted by the customer, but are penalized with a delayed delivery cost. In this paper we evaluate the optimal buffer level taken into account random delivery times and machine failures, in order to minimize the inventory, transportation, backlog, delayed delivery and early delivery costs. Under this model, we will formulate a stochastic optimization problem which aims to minimize the total expected cost via a simulation based optimization.

The method proposed in this paper is the Perturbation Analysis (PA) (Yu and Cassandras 2004). PA is a technique which permits to obtain sample path derivatives of a random variable with respect to some parameters of interest. Panayiotou and Cassandras (2006) determined the optimal capacities (or hedging points) of the finished goods and work-in-process buffers to minimize a cost function. They estimated the gradient of the cost function with respect to hedging points. Paschalidis et al. (2004) combined the large deviation (LD) and Perturbation Analysis (PA) technique; and explored synergies between LD and PA. Markou and Panayiotou (2007) investigated the implementation of various IPA estimates that have been derived based on a stochastic fluid model for the optimization of parameters (buffer size) of a discrete event system. However, systems with transport delays combined with PA is a very recent topic, and the few existing results indicate that the problem may become challenging, as the PA derivative are more complicated than those that would be obtained for the system without delays. Turki et al. (2009a) studied the discrete case with constant delivery time and prove that for their model the discrete setting are statistically unbiased. Turki et al. (2009b) proposed a simulation algorithm based on perturbation analysis to determine the optimal buffer level in the presence of constant delivery time.

The goal of this paper is to propose a PA based simulation for performances evaluation and optimization of a discrete flow model taking into account stochastic delivery times, stochastic demands and machine failures. The proposed method will be compared with a discrete events system based simulation to show the interest of our method.

The paper is organized as follows. The discrete-flow model with random delivery time is presented in section 2. In section 3, the perturbation analysis is studied. In section 4, the numeral results are presented. Finally, the last section

concludes the paper and gives some perspectives to our work.

## MODEL AND PROBLEM FORMULATION

The studied manufacturing system is composed by a machine  $M$ , a buffer  $B$  and a customer who demands at every time  $t$  a quantity of material  $D(t)$ . The materials outgoing from buffer  $B$ , denoted as  $y(t)$ , are transported to the customer and take a delivery time, denoted as  $\tau(t)$ , to arrive to the customer. The producer proposes for the customer a constant planned delivery time denoted as  $\tau_{ex}$ . The transported materials, can arrive to the customer exactly in time (i.e.  $\tau(t) = \tau_{ex}$ ), or can arrive to the customer before the end of  $\tau_{ex}$  (i.e.  $\tau(t) < \tau_{ex}$ ), or can arrive to the customer after the end of  $\tau_{ex}$  (i.e.  $\tau(t) > \tau_{ex}$ ). The materials arriving to the customer at time  $t$ , are denoted as  $y^A(t)$  (see Figure. 1).

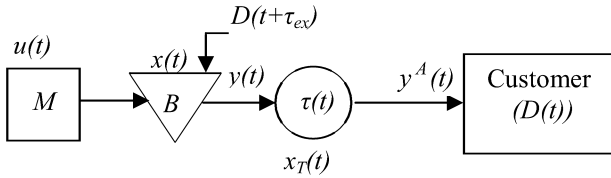


Figure 1: Studied manufacturing system.

We assume that the machine is never starved. The machine  $M$  is either up or down. The state of the machine at time  $t$ , denoted  $\alpha(t)$ , if the machine is up  $\alpha(t) = 1$ , else  $\alpha(t) = 0$ . When the machine is up, the production rate of  $M$ , denoted by  $u(t)$ , could take any integer value between 0 and  $U$ , where  $U$  is the maximal production rate (the machine capacity). The times to failure and times to repair are random. The failure/repair process is an independent random process. It does not depend on the system parameters. The delivery time  $\tau(t)$  is random and strictly greater than zero ( $\tau(t) > 0$ ). Furthermore, we assume in this paper that: The maximal production rate permits to satisfy the demand, i.e.  $U > D(t)$ . If the demand is unsatisfied, it encounters a backlog cost. At time  $t = 0$ , we suppose that we have enough parts in the buffer to satisfy the first demand, i.e.  $x(0) \geq D(\tau_{ex})$ .

We consider the case of infinite buffer capacity, where the buffer level or production surplus at time  $t$  is given by:

$$x(t) = x(t-dt) + u(t-dt) - y(t). \quad (1)$$

The material outgoing from buffer  $B$ , denoted by  $y(t)$ , at time  $t$  needs a period of time  $\tau(t)$  to be transported to the customer, so the demand  $D(t+\tau_{ex})$  must leave the buffer  $B$  at time  $t$ . The value of  $y(t)$  can be described as follows:

$$y(t) = \begin{cases} D(t+\tau_{ex}) & \text{if } x(t-dt) \geq D(t+\tau_{ex}) \\ 0 & \text{if } x(t-dt) = 0 \\ x(t-dt) & \text{if } x(t-dt) < D(t+\tau_{ex}) \end{cases} \quad (2)$$

The number of materials arriving to the customer at time  $t$ , denoted as  $y^A(t)$  is equal to the sum of materials outgoing

from buffer  $B$  at time  $t^i$  and arriving at time  $t$  (i.e.  $t = t^i + \tau(t^i)$ ),  $y^A(t)$  is defined as follows:

$$y^A(t) = \sum_{t^i / t^i + \tau(t^i) = t} y(t^i) \quad \text{with } (i > 0). \quad (3)$$

The number of materials transported at time  $t$  is given by:

$$X_T(t) = X_T(t-dt) + y(t) - y^A(t). \quad (4)$$

The number of early delivery materials is strictly greater than zero, when the delivery time  $\tau(t)$  is strictly smaller than the planned delivery time  $\tau_{ex}$ , in this case the materials arrive before the expected time  $t + \tau_{ex}$ . The number of early delivery materials at time  $t$  is indicated as  $s(t)$  and given by:

$$s(t) = \sum_{t^i / t^i + \tau(t^i) = t} 1. [\tau(t) < \tau_{ex}] \times y(t), \quad (5)$$

$$\text{with } 1. [\tau(t) < \tau_{ex}] = \begin{cases} 1 & \text{if } \tau(t) < \tau_{ex} \\ 0 & \text{if } \tau(t) \geq \tau_{ex} \end{cases}$$

The number of delayed delivery materials at time  $t$  is indicated as  $r(t)$  and given by:

$$r(t) = \sum_{t^i / t^i + \tau(t^i) = t} 1. [\tau(t) > \tau_{ex}] \times y(t), \quad (6)$$

The control policy is defined as follows:

$$u(t) = \begin{cases} U & \text{if } \alpha(t) = 1 \text{ and } x(t) < h \\ D(t+\tau_{ex}) & \text{if } \alpha(t) = 1 \text{ and } x(t) = h \\ D(t+\tau_{ex}) - (x(t) - h) & \text{if } \alpha(t) = 1 \text{ and } x(t) > h \\ 0 & \text{if } \alpha(t) = 0. \end{cases} \quad (7)$$

This control policy is a hedging point policy (Mokou and Porter. 2005) which ensures that the material does not exceed a given number of parts, denoted by  $h$ .

The possible events at every time  $t$  are: machine production (UM), failure (PM), repair (RM), buffer full i.e.  $x(t) = h$  (SS), buffer empty i.e.  $x(t) = 0$  (SV) and demand (DE).

*Assumption:* priority between events is assigned in a decreasing order as follows: (1) Buffer event (SS or SV), (2) Machine event (UM, PM or RM), (3) Demand event (DE).

The cost function  $C(t)$ , at time  $t$ , which depends on the number of products in the buffer  $B$ , the number of transported parts, the number of delayed delivery materials, the number of early delivery materials and the lost demands, is given by:

$$C(t) = cs \times x(t) + cs^- \times (D(t+\tau_{ex}) - y(t)) + ct \times X_T(t) + ctr \times r(t) + cts \times s(t). \\ C(t) = cs \times x(t) + cs^- \cdot D^0(t+\tau_{ex}) + ct \cdot X_T(t) + ctr \cdot r(t) + cts \cdot s(t). \quad (8)$$

Where:  $ct$  is the unit transportation cost,  $cs$  is the unit inventory cost,  $cs^-$  is the unit backlog cost,  $cts$  is the unit



early delivery cost,  $ctr$  is the unit delayed delivery cost,  $D^0(t+\tau_{ex})$  is the number of unsatisfied demands ( $D^0(t+\tau_{ex}) = D(t+\tau_{ex}) - y(t)$ ).

### PERTURBATION ANALYSIS

PA is an approach intended to estimate gradients of performances metric with respect to some parameters of interest, in this case, the optimal buffer level  $h$ . This method consists on observing and analyzing two sample paths, one is the nominal sample path ( $x(t)$ ), and the other is the perturbed sample path ( $x^\delta(t)$ ) (see Figures 2 and 3). We assumed that the optimal buffer level is increased by a perturbation, denoted by  $\delta$ . In this paper, we consider  $\delta > 0$  and we evaluate the resulting changes in the cost function using geometric arguments. The optimal buffer level of the perturbed sample path ( $x^\delta(t)$ ) is  $h + \delta$ .

#### Study of trajectories

The following assumptions are considered: Same distribution of random variables (*time to failure, time to repair, delivery time and demand customer*) for both sample paths. The maximal production is the same for both sample paths. The perturbation value of  $h$  is  $\delta = 1$ .

Let  $[0, T]$  be a finite horizon, we consider the trajectories of  $x(t)$  and  $x^\delta(t)$  over  $t \in [0, T]$ . The interval  $[0, T]$  is divided on two alternating periods: the firsts when  $t \in [t_{ssj}^\delta, t_j]$  and the others when  $t \in [t_j, t_{ss(j+1)}^\delta]$  (see Figures 2 and 3).

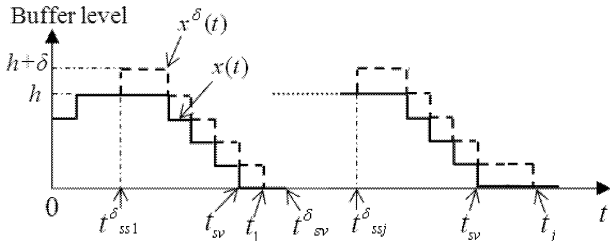


Figure 2: Buffer level of the perturbed and nominal paths (case A).

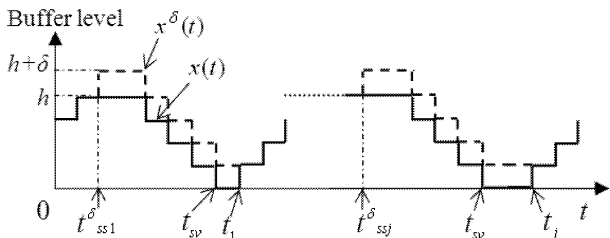


Figure 3: Buffer level of the perturbed and nominal paths (case B).

#### PA based optimization

The total cost function corresponds to the nominal path, denoted by  $CT(h)$  depending on  $h$  and given by:

$$CT(h) = \sum_{t=0}^{t=T} C(t) \quad (9)$$

Then:  $CT(h) = cs.X(h) + ct.XT(h) +$

$$cs^-.B(h) + cts.S(h) + ctr.R(h).$$

$$\text{with } X(h) = \sum_{t=0}^{t=T} x(t), \quad XT(h) = \sum_{t=0}^{t=T} X_T(t)$$

$$B(h) = \sum_{t=0}^{t=T} D^0(t + \tau_{ex}), \quad R(h) = \sum_{t=0}^{t=T} r(t) \text{ and}$$

$$S(h) = \sum_{t=0}^{t=T} s(t).$$

The total cost function corresponding to the perturbed path, denoted by  $CT^\delta(h)$  depending on  $h$  is given by :

$$CT^\delta(h) = \sum_{t=0}^{t=T} C^\delta(t) \quad \forall t \in [0, T] \quad (10)$$

with  $C^\delta(t)$  the cost function corresponding to the perturbed path at time  $t$ .

Then  $CT^\delta(h) = cs.X^\delta(h) + ct.XT^\delta(h) +$

$$cs^-.B^\delta(h) + cts.S^\delta(h) + ctr.R^\delta(h)$$

$$\text{with } X^\delta(h) = \sum_{t=0}^{t=T} x^\delta(t), \quad XT^\delta(h) = \sum_{t=0}^{t=T} X_T^\delta(t),$$

$$B^\delta(h) = \sum_{t=0}^{t=T} D^{0\delta}(t + \tau_{ex}), \quad R^\delta(h) = \sum_{t=0}^{t=T} r^\delta(t) \text{ and}$$

$$S^\delta(h) = \sum_{t=0}^{t=T} s^\delta(t).$$

The difference between the perturbed path and the nominal path is given by:

$$\begin{aligned} CT^\delta(h) - CT(h) = & cs.(X^\delta(h) - X(h)) + \\ & ct.(XT^\delta(h) - XT(h)) + cs^-.(B^\delta(h) - B(h)) + \\ & cts.(S^\delta(h) - S(h)) + ctr.(R^\delta(h) - R(h)) \end{aligned}$$

We assume that in the interval  $[0, T]$  they are  $n$  intervals  $[t_{ssj}^\delta, t_j]$ ,  $l$  intervals  $[t_j, t_{ss(j+1)}^\delta]$  and  $m$  instants  $t_j + \tau_{ex}$  ( $l \geq m$ ). Let  $TS$ ,  $C1$ ,  $C2$ ,  $C3$  and  $C4$  be PA estimators, these estimators are given by:

$$X^\delta(h) - X(h) = \sum_{j=1}^{j=n} \frac{(t_j - t_{ssj}^\delta)}{dt} . \delta = TS.$$

$$XT^\delta(h) - XT(h) = \sum_{j=1}^{j=l} \left( \frac{\tau(t_j)}{dt} \right) . \delta = C1.$$

$$B^\delta(h) - B(h) = \sum_{p=1}^{p=l} \delta = l . \delta = C2.$$

$$S^\delta(h) - S(h) = \sum_{j=1}^{j=m} 1. [\tau(t_j) < \tau_{ex}] \frac{(\tau_{ex} - \tau(t_j))}{dt} . \delta = C3.$$

$$R^\delta(h) - R(h) = \sum_{j=1}^{j=m} 1_{[\tau(t_j) > \tau_{ex}]} \frac{(\tau(t_j) - \tau_{ex})}{dt} \cdot \delta = C4.$$

The PA estimators are determined in the following algorithm (*PA estimation algorithm*):

$C1=0, C2=0, C3=0, C4=0$  TS=0,  $t=0$  // Initialization

Do

- $t = t + dt$
- If  $t = t_{ssj}^\delta$  then,  $C2 = \delta + C2$ .
- If  $t = t_j$  then  $((t_j - t_{ssj}^\delta) \cdot \delta) / dt$ ,  
and  $C1 = C1 + (\frac{\tau(t_j)}{dt}) \cdot \delta$ .
- If  $t = t_j + \tau_{ex}$  and  $\tau(t_j) < \tau_{ex}$   
then  $C3 = C3 + \frac{(\tau_{ex} - \tau(t_j))}{dt} \cdot \delta$ .
- If  $t = t_j + \tau_{ex}$  and  $\tau(t_j) > \tau_{ex}$   
then  $C4 = C4 + \frac{(\tau(t_j) - \tau_{ex})}{dt} \cdot \delta$ .
- $t = t + dt$ .

While  $t < T$ .

In the following, numerical results are presented to show the impact of the random delivery time and the impact of the unit costs ( $cs$ ,  $cs^-$ ,  $ct$ ,  $cts$  and  $ctr$ ) on the value of  $h$ , and to determine the optimal buffer level  $h$ .

## NUMERICAL RESULTS

The following parameters are used for the simulation:

$U = 1$  part/time unit, the demand which is Boolean is given by a Bernoulli distribution with probability  $p=0.8$ , the times to failure or repair are given by exponential laws with  $1/\lambda = 2$  and  $1/\mu = 1$ , the total simulation time is equal to  $10E+06$ . The delivery time  $\tau(t)$  is generated according to a truncated Normal distribution, the Normal distribution gives same value out the interval  $[\tau(min), \tau(max)]$ , the truncated Normal distribution gives bounded value between  $\tau(min)$  and  $\tau(max)$ .

### Performances Evaluation

In what follows, we will study the impact of the unit costs on the value of  $h$ .

*Impact of the unit costs  $cs$ ,  $ct$ ,  $cs^-$ ,  $cts$  and  $ctr$  on  $h$ .*

Such as the total cost  $CT(h)$  depends of the value of the unit costs  $cs$ ,  $ct$ ,  $cs^-$ ,  $cts$  and  $ctr$ , the value of  $h$  which minimizes  $CT$  will certainly also depends on these unit costs. Examples of simulation results are presented in table 1. The random delivery time  $\tau(t)$  can be 1, 2 or 3. The other parameters are the same as given before (parameters presented in the beginning of the numerical results section).

Table 1: Impact of unit costs  $cs$ ,  $ct$ ,  $cs^-$ ,  $cts$  and  $ctr$  on  $h$

| $cs$ | $ct$ | $cs^-$ | $cts$ | $ctr$ | Optimal buffer level ( $h$ ) |
|------|------|--------|-------|-------|------------------------------|
| 1    | 1    | 30     | 1     | 1     | 5                            |

|   |   |    |    |    |   |
|---|---|----|----|----|---|
| 1 | 1 | 30 | 1  | 10 | 4 |
| 1 | 1 | 30 | 10 | 1  | 4 |
| 1 | 1 | 30 | 10 | 10 | 4 |
| 1 | 1 | 30 | 25 | 30 | 3 |
| 1 | 1 | 30 | 30 | 25 | 3 |
| 1 | 1 | 20 | 1  | 1  | 4 |
| 1 | 1 | 20 | 1  | 10 | 3 |
| 1 | 1 | 20 | 10 | 1  | 3 |
| 1 | 1 | 20 | 10 | 10 | 2 |

We see that more the unit costs  $cts$  and  $ctr$  become important comparing to  $cs^-$ ; more the delivery cost increases and normally the total cost increases, then the value of  $h$ , which minimizes the total cost, of course decreases as the unit costs  $cts$  and  $ctr$  increase.

*Impact of planned delivery time on  $h$ .*

We vary the standard deviation  $\sigma(\tau(t))$  of the random delivery time  $\tau(t)$  and the program determines an optimal buffer level  $h$ , which minimizes the total cost function, in function of the standard deviation  $\sigma(\tau(t))$ .

Simulation results are presented in table 2 with  $cs = 1$ ,  $ct=1$ ,  $cs^- = 30$ ,  $ctr=1$  and  $cts=1$ . The other parameters are the same as given before.

Table 2: Impact of planned delivery time and the standard deviation  $\sigma(\tau(t))$  on  $h$

| $\tau(t)$ | $\tau_{ex}$ | $\sigma(\tau(t))$ | Optimal buffer level ( $h$ ) |
|-----------|-------------|-------------------|------------------------------|
| [1,3]     | 2           | 0.69              | 5                            |
| [1,7]     | 4           | 1.53              | 4                            |
| [7,9]     | 8           | 0.69              | 4                            |
| [1,15]    | 8           | 3.39              | 4                            |
| [9,11]    | 10          | 0.69              | 3                            |
| [1,19]    | 10          | 4.12              | 3                            |
| [1,29]    | 15          | 6.52              | 2                            |

For our study, we can see that planned delivery time  $\tau_{ex}$  has an impact on the value of  $h$ . normally when  $\tau_{ex}$  becomes large, the total delivery cost ( $S(h).cts$  and  $R(h).ctr$ ) increases. Then the value of  $h$ , which minimizes the total cost, of course decreases as the planned delivery time  $\tau_{ex}$  increases. The standard deviation  $\sigma(\tau(t))$  has not an impact on the value of  $h$  when the distribution of  $\tau(t)$  is symmetric, besides The standard deviation  $\sigma(\tau(t))$  has an impact on the value of  $h$ , when the distribution of  $\tau(t)$  is asymmetric (for example  $\tau_{ex}=5$  and  $\tau(t)=[1,14]$ ).

### Optimization

*Optimization algorithm*

The optimization algorithm will allow us to determine the optimal level buffer  $h$  and is given by:

$h_{new}=2$  ( $h>1$ )

$cout2 = CT(h_{new})$

Do

- $h=h_{new}$
- $cout1 = cout2$

- Determine the estimators  $C1$ ,  $C2$ ,  $C3$ ,  $C4$  and  $TS$  by using the PA estimation algorithm.

- Determine the difference estimators  $H(h)$

$$H(h) = \left[ (cs^- \times C2) - (cs \times TS) - (ct \times C1) - (cts \times C3) - (ctr \times C4) \right] / T.$$

- If  $H(h) > 0$  then  $h_{new} = h + 1$ .
- Else if  $H(h) < 0$  then  $h_{new} = h - 1$ .

- $Cout2 = CT(h)$

While ( $cout2 < cout1$ )

Total cost function  $CT = f(h)$ .

We are interested to find the value of  $h$  which corresponds to the lowest total cost  $CT$ , by using two different simulation, simulation based on DES and simulation based on PA. The simulations are done with  $cs = 1$ ,  $ct = 1$ ,  $cs^- = 35$ ,  $ctr = 4$ ,  $cts = 2$  and  $\tau(t)$  is random and can be 1, 2 or 3. The other parameters are the same as given before. The results are given in Figure 4.

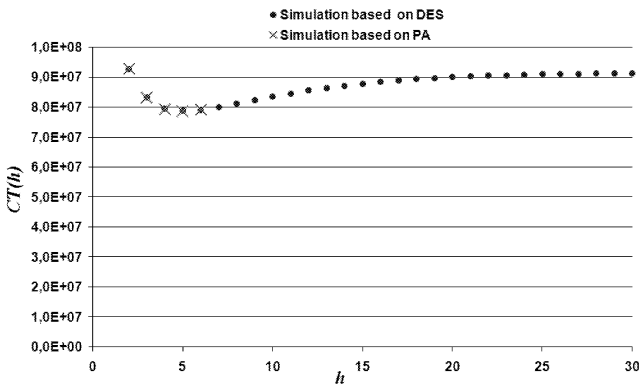


Figure 4: Evolution of  $CT$  in function of  $h$

By simulations, and for our example, the lowest total cost  $CT$  is obtained for  $h = 5$ . We see that the simulation based on PA gives of course the same results as the one given by the simulation based on DES, the only difference is the number of simulation runs. Indeed, the simulation based on DES needs to compute for every  $h$  the corresponding  $CT$ , but the simulation based on PA only needs to compute the  $CT$  corresponding to five  $h$ . This difference is explained by the fact that the optimization based on PA computes the difference estimators  $H(h)$  at every time when the  $h$  value is changed, the change of  $H(h)$  gives information about when the simulation must stops (i.e. the optimal value of  $h$  is found). The simulation based on DES must calculate for every  $h$  the corresponding  $CT$  and then, the optimal  $h$  which corresponds to the minimal value of  $CT$  is obtained. Therefore, the optimization based on PA permits to reduce the simulation time comparing to a simulation based on DES.

## CONCLUSION

In this paper, a manufacturing system composed by a failure-prone machine, a buffer and a stochastic demand is considered. The times to failure and times to repair are random variables. The demand is lost when it is unsatisfied. A discrete flow model with explicit modeling of the random delivery time between the buffer and the customer is pro-

posed. A planned delivery time  $\tau_{ex}$  with early delivery cost and delayed delivery cost are proposed for modeling the random delivery time. The buffer level trajectories is studied and analyzed. The optimal value of  $h$  is determined by using the optimization algorithm based on PA. The optimal value of  $h$  increases, as the unit costs  $cts$  and  $ctr$  decrease. The unit costs  $cts$  and  $ctr$  have the same impact on  $h$ , and become important comparing to  $cs^-$ . The optimal value of  $h$  decreases, when  $\tau_{ex}$  increases. The standard deviation  $\sigma(\tau(t))$  has an impact on the value of  $h$ , when the distribution of  $\tau(t)$  is asymmetric.

Our manufacturing system is simple and composed by a failure-prone machine, a buffer and a stochastic demand, For future research, we will consider a more complex multistage system. We will consider also more than one customer with random demands for each one. We will determine the optimal planned delivery time, which minimizes the total cost. We will apply a continuous model to our system, and then we can compare the discrete model to the continuous model applied in our system.

## REFERENCES

- Dolgui, A. and M. Aly Ould-Louly. 2002. "A model for supply planning under lead time uncertainty", *International journal on production economics*, 145–152.
- Markou, M. and C. G. Panayiotou. 2007. "Optimization of Discrete Event System Parameters using SFM-Based Infinitesimal Perturbation Analysis Estimates", *46th IEEE Conference on Decision and Control*, New Orleans, USA, 5068-5073.
- Mok, P.Y. and B. Porter. 2005. "Evolutionary optimization of hedging points for unreliable manufacturing systems", *International Journal Adv Manufacturing Technologies*, 205–214.
- Mourani, I.; S. Hennequin; and X. Xie. 2005. "Continuous Petri nets with delays for performances evaluation of transfer lines", *International Conference on Robotics and Automation*, Barcelona, Spain, 3721-3726.
- Mourani, I.; S. Hennequin; and X. Xie. 2008. "Simulation-based optimization of a single-stage failure-prone manufacturing system with transportation delay", *International Journal of Production Economics*, Vol.12, No.1, 26-36.
- Panayiotou, C.G. and C.G. Cassandras. 2006. "Infinitesimal Perturbation Analysis and Optimization for Make-to-Stock Manufacturing Systems Based on Stochastic Fluid Models", *Discrete Event Dynamic System*, Vol.16, No.1, 109–142.
- Paschalidis, I.C.; Y. Liu; C.G. Cassandras; and C.G. Panayiotou. 2004. "Inventory control for supply chains with service level constraints: A synergy between large deviation and perturbation analysis", *Annals of Operations Research*, 231-258.
- Turki, S.; S. Hennequin; and N. Sauer. 2009a. "Performances evaluation of a failure-prone manufacturing system with time to delivery and stochastic demand", (*accepted in March 5 2009*), *INCOM 09, 13th IFAC Symposium On Information Control Problems in Manufacturing*. Moscow, Russia.
- Turki, S.; S. Hennequin; and N. Sauer. 2009b. "Perturbation analysis based-optimization for a failure-prone manufacturing system with constant delivery time and stochastic demand", (*accepted in March 9 2009*), *39th CIE39 International Conference on Computer and Industrial Engineering*, Troyes, France.
- Van Ryzin, G. J.; S. X. C. Lou; and S. B. Gershwin. 1991. "Scheduling job shops with delay". *Int. J. Prod. Res.*, Vol.29, No.7, 1407-1422.
- Yu, H. and C.G. Cassandras. 2004. "Perturbation analysis for production control and optimization of manufacturing systems", *Automatica*, Vol.40, No.6, 945–956.



# **PRODUCTION AND PRODUCT ANALYSIS**



# LINEAR AND NONLINEAR INPUT OUTPUT ANALYSIS

William Conley  
Department of Business Administration and Mathematics  
University of Wisconsin at Green Bay  
Green Bay, Wisconsin 54311  
U.S.A.  
E-mail: [Conleyw@uwgb.edu](mailto:Conleyw@uwgb.edu)

**KEYWORDS:** Monte Carlo, Optimization, Corporate planning

## ABSTRACT

Wassily Leontief did pioneering work on industrial input output analysis many decades ago. Much of his work in this area was on linear input output problems and using matrix linear algebra to solve the attendant linear systems of equations. Presented here will be a computer simulation solution technique (statistical optimization) approach to solving the input output analysis problem. We will start with a three industry linear example. The multi-stage Monte Carlo optimization (statistical optimization by simulation) approach will then be employed to solve this system even though the linear algebra technique developed long ago would work too.

However, for the second example, a seven industry nonlinear input output analysis problem, multi stage Monte Carlo optimization (MSMCO) will still be used as a solution technique, because the linear algebra approach would not be appropriate.

Then a third general nonlinear system of equations, requiring an all integer solution, will be presented to feature the general purpose nature of statistical optimization in our computer age.

## INTRODUCTION

Let us look at the following three industry linear input output analysis example. The variable values for the total output for our three industries ( $x_1$ ,  $x_2$ ,  $x_3$ ) are in units of millions of U.S. dollars. The actual final demands (for our time period) for three industries are 525.7, 2857.25 and 2262.5 units for industries 1, 2, and 3 respectively.

However, industry one requires .08, .11 and .16 units from industries 1, 2 and 3 respectively to produce a unit for industry one. Also, industry two requires .19, .15, and .21 units from industries 1, 2, and 3 respectively to produce a unit for industry two, while industry three

requires .15, .12, and .07 units from industries 1, 2, and 3 to produce a unit for industry three. Therefore, letting  $x_1$ ,  $x_2$ ,  $x_3$  be the total output of units from industries 1, 2, and 3 respectively, we have the following input output linear system of equations to solve.

$$x_1 = .08x_1 + .19x_2 + .15x_3 + 525.7 \quad (1)$$

$$x_2 = .11x_1 + .15x_2 + .12x_3 + 2857.25 \quad (2)$$

$$x_3 = .16x_1 + .21x_2 + .07x_3 + 2262.5 \quad (3)$$

Instead of the usual matrix inversion and linear system of equations analysis, we solve this system by minimizing  $f(x_1, x_2, x_3) = E_1 + E_2 + E_3$  where the  $E_i$ 's are the absolute values of the differences between the left and right hand sides of equations (1), (2) and (3).

## THE 3X3 INPUT OUTPUT SOLUTION

We used a fourteen stage Monte Carlo simulation (statistical optimization) drawing five thousand feasible solutions at each of the fourteen stages and storing the best answer (lowest  $f(x_1, x_2, x_3)$  value) after each solution try and subsequently re-centering the search about each newest best answer so far. Additionally, after each of the fourteen stages the search region is reduced. This allows the one page statistical optimization computer simulation to "funnel" into the solution. The fourteen stage printout is:

| Stages | $x_1$ | $x_2$ | $x_3$ | Total Error |
|--------|-------|-------|-------|-------------|
| 1      | 1709  | 4026  | 3522  | 398.830     |
| 2      | 2216  | 4319  | 3871  | 292.910     |
| 3      | 1919  | 4140  | 3725  | 134.090     |
| 4      | 2087  | 4120  | 3708  | 98.750      |
| 5      | 2076  | 4154  | 3730  | 40.120      |
| 6      | 2040  | 4144  | 3717  | 14.260      |
| 7      | 2028  | 4150  | 3716  | 9.240       |
| 8      | 2039  | 4151  | 3720  | 4.380       |
| 9      | 2036  | 4151  | 3721  | 1.310       |
| 10     | 2035  | 4150  | 3720  | .45         |
| 11     | 2035  | 4150  | 3720  | .45         |
| 12     | 2035  | 4150  | 3720  | .45         |
| 13     | 2035  | 4150  | 3720  | .45         |
| 14     | 2035  | 4150  | 3720  | .45         |

Therefore, industry 1 should produce  $x_1=2035$  units and industry 2,  $x_2=4150$  units and industry 3,  $x_3=3720$  units of

output to satisfy internal demand and actual final demand for the three industries with a total error of .45 units caused by requiring whole number  $x_i$  values. A real valued multistage simulation would reduce the error to zero.

## A 7X7 NONLINEAR PROBLEM

Let us look at a seven industry input output analysis problem that has returns to scale in its industrial output. Let  $x_i$  (for  $i = 1, 2, \dots, 7$ ) be the unit output (in billions of Euros) for industry  $i$ . The reasoning and structure of the system of equations is the same as for the previous 3x3 analysis problem except for the returns to scale. As an example, an internal demand of .09 of a unit of output from industry one is required to make a unit for industry one. However, as more units are produced the cost goes down as expressed by  $.09x_1^{.98}$ . There are  $7 \times 7 = 49$  of these cost terms and along with final demands of 530, 432, 212, 358, 205, 688 and 663 respectively, for our industries 1, 2, 3, 4, 5, 6 and 7 respectively we have the following input output system:

$$x_1 = .09x_1^{.98} + .03x_2^{.91} + .07x_3^{.92} + .04x_4^{.96} + .05x_5^{.91} + .07x_6^{.95} + .03x_7^{.90} + 530 \quad (4)$$

$$x_2 = .11x_1^{.97} + .04x_2^{.93} + .12x_3^{.87} + .07x_4^{.97} + .07x_5^{.88} + .06x_6^{.94} + .07x_7^{.92} + 432 \quad (5)$$

$$x_3 = .03x_1^{.95} + .06x_2^{.94} + .11x_3^{.89} + .09x_4^{.90} + .13x_5^{.90} + .04x_6^{.98} + .08x_7^{.91} + 212 \quad (6)$$

$$x_4 = .07x_1^{.90} + .05x_2^{.88} + .05x_3^{.85} + .11x_4^{.89} + .14x_5^{.91} + .02x_6^{.97} + .10x_7^{.89} + 358 \quad (7)$$

$$x_5 = .04x_1^{.87} + .09x_2^{.92} + .07x_3^{.89} + .10x_4^{.96} + .08x_5^{.95} + .11x_6^{.90} + .03x_7^{.95} + 205 \quad (8)$$

$$x_6 = .06x_1^{.93} + .10x_2^{.97} + .06x_3^{.93} + .04x_4^{.97} + .03x_5^{.96} + .15x_6^{.85} + .06x_7^{.96} + 688 \quad (9)$$

$$x_7 = .08x_1^{.94} + .08x_2^{.93} + .05x_3^{.94} + .03x_4^{.95} + .02x_5^{.94} + .06x_6^{.94} + .04x_7^{.94} + 663 \quad (10)$$

Therefore, we use a ten stage multi stage Monte Carlo statistical optimization to minimize  $g(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = E_1 + E_2 + E_3 + E_4 + E_5 + E_6 + E_7$  where  $E_i$  is again the absolute value of the difference between the left and right hand side of each equation. We draw 15,000 feasible solutions at each of the ten stages. The computer printout of the stages is:

| Stage | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $x_7$ | Total Error |
|-------|-------|-------|-------|-------|-------|-------|-------|-------------|
| 1     | 664   | 648   | 476   | 570   | 392   | 888   | 960   | 342.529     |
| 2     | 707   | 599   | 433   | 461   | 365   | 886   | 803   | 161.979     |
| 3     | 725   | 610   | 404   | 506   | 430   | 904   | 834   | 104.696     |
| 4     | 699   | 605   | 395   | 541   | 385   | 892   | 836   | 57.576      |
| 5     | 698   | 611   | 408   | 516   | 397   | 891   | 818   | 24.786      |
| 6     | 695   | 609   | 395   | 524   | 393   | 895   | 820   | 13.620      |
| 7     | 699   | 609   | 395   | 519   | 395   | 892   | 821   | 8.075       |
| 8     | 697   | 611   | 395   | 521   | 395   | 893   | 821   | 3.167       |
| 9     | 697   | 612   | 395   | 520   | 396   | 893   | 820   | 1.697       |
| 10    | 697   | 612   | 395   | 521   | 396   | 893   | 820   | 1.655       |

Therefore, (in units of billions of Euros over our extended time period) industries 1 through 7 should budget for production of  $x_1=697$ ,  $x_2=612$ ,  $x_3=395$ ,  $x_4=521$ ,  $x_5=396$ ,  $x_6=893$  and  $x_7=820$  units to approximately satisfy all internal, external, final and total outputs.

## GENERAL NONLINEAR SYSTEMS

The dominance of the linear algebra and linear simplex optimization techniques developed before the computer age, forced many business people, scientists and engineers to develop linear systems in their applications work because they would frequently be solvable. However, now the 21<sup>st</sup> century's powerful desk top PC's allow new solution techniques, like statistical optimization (multi stage Monte Carlo optimization or MSMCO) to be attempted on difficult nonlinear systems too.

Let us use MSMCO to try to find an all integer solution to the following 13 variable 16 equation nonlinear systems in the range  $0 \leq x_i \leq 100$  for  $i=1, 2, 3, \dots, 13$ .

$$x_1^3 + x_5 + x_9 + x_{10} + x_{13} = 658,726 \quad (11)$$

$$x_2^3 + x_1 + x_5 + x_8 + x_{12} = 250,261 \quad (12)$$

$$x_4^3 + x_2 + x_4 + x_{10} + x_{11} = 7,076 \quad (13)$$

$$x_3^3 + x_3 + x_6 + x_{12} + x_{13} = 358,095 \quad (14)$$

$$x_7^3 + x_4 + x_7 + x_{10} + x_{12} = 531,635 \quad (15)$$

$$x_6^3 + x_2 + x_8 + x_9 + x_{10} = 43,108 \quad (16)$$

$$x_5^3 + x_1 + x_7 + x_{11} + x_{13} = 132,938 \quad (17)$$

$$x_8^3 + x_3 + x_4 + x_9 + x_{11} = 104,041 \quad (18)$$

$$x_{10}^3 + x_3 + x_7 + x_{10} + x_{13} = 274,891 \quad (19)$$

$$x_9^3 + x_2 + x_5 + x_8 + x_9 = 195,331 \quad (20)$$

$$x_{11}^3 + x_1 + x_6 + x_7 + x_{13} = 343,252 \quad (21)$$

$$x_1^3 + x_2 + x_4 + x_{10} + x_{12} = 658,679 \quad (22)$$

$$x_1^3 + x_2 + x_6 + x_{11} + x_{13} = 24,606 \quad (23)$$

$$x_2^3 + x_2 + x_5 + x_9 + x_{10} = 250,284 \quad (24)$$

$$x_{13}^3 + x_1 + x_6 + x_{11} + x_{13} = 117,890 \quad (25)$$

$$x_4^3 + x_3 + x_5 + x_9 + x_{12} = 7,068 \quad (26)$$

However, if we seek to minimize  $f(x_1, x_2, \dots, x_{13}) = \sum E_i$  subject to  $0 \leq x_i \leq 100$  and all integer values for  $i = 1, 2, 3, \dots, 13$  with a ten stage Monte Carlo simulation drawing 10,000 feasible solutions at each stage, the simulation will



probably not pin down all 13  $x_i$  values exactly because of the “random” nature of the multi stage Monte Carlo approach. Therefore, we deal with this problem with outer controlling loops. We twice do ten runs of the above mentioned ten stage simulation. After the first ten runs, any  $x_i$  that occurred at least five times in the final stage output (modal averaging) is fixed at that value for the second stage run. This greatly reduces the search region for the second ten runs and easily produces the exact solution of  $x_1 = 87$ ,  $x_2 = 63$ ,  $x_3 = 71$ ,  $x_4 = 19$ ,  $x_5 = 51$ ,  $x_6 = 35$ ,  $x_7 = 81$ ,  $x_8 = 47$ ,  $x_9 = 58$ ,  $x_{10} = 65$ ,  $x_{11} = 70$ ,  $x_{12} = 29$  and  $x_{13} = 49$ .

## MULTISTAGE MONTE CARLO AND EDUCATION

The early years of the computer age saw simulation techniques being criticized by mathematics educators, because they were too slow or could not guarantee an exact solution with the random process nature of simulation. The speed, capacity and affordability of twenty-first century PC's along with the advancement of simulation techniques have really made those past criticisms of simulation theory and practice almost irrelevant. Similar techniques (MSMCO, statistics, other genetic-based algorithms and general simulation) should now begin to form a large part of the applied mathematics course work for business, engineering and science majors at colleges and universities throughout the world.

Computer time is cheap and available to anyone within reach of a PC. Also, it will be even less expensive when the current university students proceed with their future careers. This does not mean that because one can draw a sample of ten million feasible solutions on a desktop PC in a few seconds that rigor and logic are suddenly unimportant in applied mathematics. However, it does mean that simulation based applied mathematics techniques should begin to take their rightful place as a key component of the mathematics curriculum for our twenty-first century business, science and engineering students.

## CONCLUSION

Many mathematics books (Barnett and Ziegler, 1994) view input output analysis from a linear perspective. This was the natural approach given the power and sophistication of linear algebra and subsequent matrix inversion techniques for solving multivariate linear systems in the twentieth century. However, the computer age allows one to bring statistical optimization (multi stage Monte Carlo simulations) to bear on difficult nonlinear systems and optimization problems. Selected examples are in (Conley, 2008) and (Wong, 1996). More examples using the modal averaging approach to large scale

problems are featured in (Conley, 2006) and (Conley, 2007).

## REFERENCES

- Barnett, R.A. and Ziegler, M. R., 1994. Applied Mathematics for Business, Economics, Life Sciences and Social Sciences. 5<sup>th</sup> Edition. MacMillan, New York:
- Conley, W. C. 2006. “Discovering Relationships that are not Functions.” In Proceedings of 4<sup>th</sup> International Industrial Simulation Conference, ISC 2006 (Palermo, Italy, June 5-7), EUROSIS-ETI, Ghent, Belgium, 163-167.
- Conley, W. C. 2007. “Finding Relationships that are Functions.” In Proceedings of 5<sup>th</sup> International Industrial Simulation Conference, ISC 2007 (Delft, The Netherlands, June 11-13, EUROSIS-ETI, Ghent, Belgium, 61-63).
- Conley, W.C. 2008. “Ecological Optimization of Pollution Control Equipment and Planning from a Simulation Perspective.” International Journal of Systems Science, Vol. 39, No. 1, 1-7.
- Wong, J. 1996. “A Note on Optimization in Integers,” International Journal of Mathematical Education in Science and Technology, Vol. 27, No. 6, 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 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 195 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.

# Latent Class Mixture models for analyzing rating responses

Liberato Camilleri  
Department of Statistics and Operations Research  
University of Malta  
Msida (MSD 06)  
Malta  
E-mail: liberato.camilleri@um.edu.mt

## KEYWORDS

Proportional Odds Model, Latent Class Mixture Model, EM algorithm, Maximum Likelihood Estimation

## ABSTRACT

Latent class methodology has been used extensively in market research. In this approach, segment membership and parameter estimates for each derived segment are estimated simultaneously. A popular approach for fitting latent class models to rating responses is to assume mixtures of multivariate conditional normal distributions. An alternative approach is to assume a Proportional Odds model. These two approaches are compared empirically in a Monte Carlo study, assessing segment membership and parameter recovery, goodness of fit and predictive accuracy.

## INTRODUCTION

Latent class segmentation models are used extensively in various fields of application to identify latent segments that can explain unobserved heterogeneity in the data. Traditionally, segmentation procedures were carried out using a two-stage approach in which estimation and clustering were conducted consecutively. In the first step, individual-level parameter estimates were derived from a Normal regression model. In the second step, individuals were clustered on the basis of similarity of the estimated parameters using a clustering algorithm. Typically, latent class regression analysis comprises the following three simultaneous steps – identify hidden segments; classify each individual in an appropriate class and estimate a regression model for each segment. Latent class models have been utilized for various types of responses, mainly rating responses. Multivariate normal latent class models have been applied by DeSarbo, et al. (1992) to analyze rating responses in a conjoint study that examines the design of a remote automobile entry device. Helsén, et al. (1993) used a similar model to classify countries into homogeneous groups having similar patterns of diffusion of durable goods. Ramaswamy, et al. (1993) applied the mixture regression model to cross-sectional time-series data. Camilleri and Green (2003) combined a latent class model with the proportional odds model to analyze rating responses in a study related to car preferences.

## LATENT CLASS MODELS FOR RATING DATA

One approach of utilizing latent class models to rating responses is to assume that respondents perceive scale spacing so that preferences are used as metric data. A latent class methodology can then be employed by using mixtures of multivariate conditional normal distributions combined with the EM algorithm to estimate parameters of these mixtures. The conditional multivariate density function of the response vectors  $\mathbf{y}_n = (y_{nj})$  for  $j = 1, \dots, J$  replications, given that the  $n^{th}$  respondent belongs to the  $k^{th}$  segment is:

$$f_{n|k}(\mathbf{y}_n; \boldsymbol{\beta}_k) = (2\pi)^{-J/2} |\boldsymbol{\Sigma}_k|^{-1/2} \exp \left[ -\frac{1}{2} (\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta}_k)' \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_n - \mathbf{X}\boldsymbol{\beta}_k) \right]$$

where  $\boldsymbol{\Sigma}_k$  is the variance-covariance matrix of  $\mathbf{y}_n$  given segment  $k$ . The unconditional density function is:

$$f_n(\mathbf{y}_n; \boldsymbol{\pi}, \boldsymbol{\beta}) = \sum_{k=1}^K \pi_k \cdot f_{n|k}(\mathbf{y}_n; \boldsymbol{\beta}_k)$$

The likelihood approach is very often used for estimation of finite mixtures because maximum likelihood estimates have been found to be superior to other methods. The log likelihood function can be formulated as:

$$\ln L(\boldsymbol{\pi}, \boldsymbol{\beta}) = \ln \prod_{n=1}^N f_n(\mathbf{y}_n; \boldsymbol{\pi}, \boldsymbol{\beta}) = \sum_{n=1}^N \ln \sum_{k=1}^K \pi_k \cdot f_{n|k}(\mathbf{y}_n; \boldsymbol{\beta}_k)$$

The derivatives of the expected log-likelihood function  $E[\ln L(\boldsymbol{\pi}, \boldsymbol{\beta})]$  with respect to the parameters  $\boldsymbol{\beta}$  and  $\boldsymbol{\pi}$  are not straightforward. An effective procedure that fits a latent class model with  $K$  segments is to maximize the expected complete log-likelihood function using the EM algorithm. The EM algorithm augments the observed data by introducing unobserved 0-1 indicators  $\lambda_{nk}$ , where  $\lambda_{nk}$  indicates whether the  $n^{th}$  respondent belongs to the  $k^{th}$  segment. Given the matrix  $\boldsymbol{\Lambda} = (\lambda_{nk})$  of unobserved data, the complete log-likelihood function is:

$$\ln L(\boldsymbol{\pi}, \boldsymbol{\beta} | \boldsymbol{\Lambda}) = \sum_{n=1}^N \sum_{k=1}^K \lambda_{nk} \cdot \ln f_{n|k}(\mathbf{y}_n; \boldsymbol{\beta}_k) + \sum_{n=1}^N \sum_{k=1}^K \lambda_{nk} \cdot \ln(\pi_k)$$

$\ln L(\boldsymbol{\pi}, \boldsymbol{\beta} | \boldsymbol{\Lambda})$  has a simpler form than  $\ln L(\boldsymbol{\pi}, \boldsymbol{\beta})$  and is easy to differentiate. Once the parameter estimates for  $\boldsymbol{\beta}_k$  and  $\boldsymbol{\pi}_k$  are obtained, an estimate for the posterior probability  $\hat{p}_{nk} = E(\lambda_{nk})$  can be calculated using Bayes' theorem.

$$\hat{p}_{nk} = E(\lambda_{nk}) = \frac{\hat{\pi}_k \cdot f_{n|k}(\mathbf{y}_n | \hat{\boldsymbol{\beta}}_k)}{\sum_{i=1}^K \hat{\pi}_i \cdot f_{n|i}(\mathbf{y}_n | \hat{\boldsymbol{\beta}}_i)} \quad \text{where} \quad \sum_{i=1}^K \hat{p}_{ni} = 1$$

Each iteration is composed of an E-step and an M-step. In the E-step, the expected log-likelihood function is calculated by replacing  $E(\lambda_{nk})$  by  $\hat{p}_{nk}$ .

$$E[\ln L(\boldsymbol{\pi}, \boldsymbol{\beta} | \boldsymbol{\Lambda})] = \sum_{n=1}^N \sum_{k=1}^K \hat{p}_{nk} \cdot \ln f_{n|k}(\mathbf{y}_n | \boldsymbol{\beta}_k) + \sum_{n=1}^N \sum_{k=1}^K \hat{p}_{nk} \cdot \ln(\pi_k)$$

In the M-step the two terms of  $E[\ln L(\boldsymbol{\pi}, \boldsymbol{\beta} | \boldsymbol{\Lambda})]$  are maximized separately with respect to the parameters. Maximizing the first term with respect to  $\boldsymbol{\beta}_k$  leads to independently solving each of the  $K$  expressions

$$\sum_{n=1}^N \hat{p}_{nk} \cdot \frac{\partial}{\partial \boldsymbol{\beta}_k} \ln f_{n|k}(\mathbf{y}_n | \boldsymbol{\beta}_k) \quad \text{for} \quad k = 1, \dots, K$$

Maximizing the second term with respect to  $\pi_k$ , subject to the constraint  $\sum_{i=1}^K \pi_i = 1$ , yields

$$\hat{\pi}_k = \frac{1}{N} \sum_{n=1}^N \hat{p}_{nk} \quad \text{for} \quad k = 1, \dots, K$$

The iterative procedure is initiated by setting random values to  $\hat{p}_{nk}$ . The algorithm then alternately updates the parameters  $\hat{\boldsymbol{\beta}}_k, \hat{\pi}_k$  and the posterior probabilities  $\hat{p}_{nk}$  until the process converges. Individuals are then assigned to the segment with the highest posterior probability  $\hat{p}_{nk}$ .

An alternative method to use ordered response categories is by forming models of cumulative probabilities. The Proportional Odds model described by McCullagh (1980) is a cumulative model that preserves the discrete ordinal nature of the rating responses.

Let  $y_j$  be the rating response for the  $j^{\text{th}}$  item using an  $R$ -point Likert scale and let  $P(y_j \leq r)$  be the  $r^{\text{th}}$  cumulative probability of this item. Cumulative probabilities reflect the ordering since

$$P(y_j \leq 1) \leq P(y_j \leq 2) \leq \dots \leq P(y_j \leq R) = 1$$

To include the effects of explanatory variables we use the model:

$$P(y_j \leq r) = F(\alpha_r + \mathbf{x}_j' \boldsymbol{\beta}) \quad \text{for} \quad r = 1, \dots, R-1$$

$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{R-1})$  is a vector of cutpoint parameters such that  $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{R-1}$ ,  $\alpha_0 = -\infty$  and  $\alpha_R = \infty$ .  $\boldsymbol{\beta}$  is a parameter vector that contains the regression coefficients for the covariate vector  $\mathbf{x}_j$ .  $F$  is a cumulative distribution function which includes the logistic, normal or the extreme value distributions which respectively lead to the logit, probit and complementary log-log link functions.  $P(y_j = r)$  is just the difference of the  $r^{\text{th}}$  and  $(r-1)^{\text{th}}$  cumulative probabilities.

$$P(y_j = r) = F(\alpha_r + \mathbf{x}_j' \boldsymbol{\beta}) - F(\alpha_{r-1} + \mathbf{x}_j' \boldsymbol{\beta})$$

The link function  $F^{-1}$  is a strictly monotonic function in the range  $[0, 1]$  onto the real line. The cumulative link model

$$F^{-1}[P(y_j \leq r)] = \alpha_r + \mathbf{x}_j' \boldsymbol{\beta}$$

links the cumulative probabilities to the real line using the link function  $F^{-1}$ . This model assumes that effects  $\mathbf{x}_j$  are the same for each cutpoint,  $r = 1, \dots, R-1$ .

For the segmentation model the Proportional Odds model is extended by considering a latent class model with  $K$  segments. Let  $\boldsymbol{\phi} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\pi})$  be the vector comprising the parameters of the latent class model with  $K$  segments. The  $n^{\text{th}}$  density function is of the form

$$P(\mathbf{Y}_n = \mathbf{y}_n | \boldsymbol{\phi}) = \sum_{k=1}^K \pi_k \cdot P(\mathbf{Y}_n = \mathbf{y}_n | \boldsymbol{\alpha}, \boldsymbol{\beta}_k)$$

$\pi_k$  are the proportion of respondents that are allocated to each segment such that  $\sum_{k=1}^K \pi_k = 1$  and  $P(y_{jn} = r | \boldsymbol{\alpha}, \boldsymbol{\beta}_k)$  is the Proportional Odds model. The log-likelihood function

$$l(\boldsymbol{\phi}) = \sum_{n=1}^N \ln \sum_{k=1}^K \pi_k \cdot P(\mathbf{Y}_n = \mathbf{y}_n | \boldsymbol{\alpha}, \boldsymbol{\beta}_k)$$

is maximized through the EM algorithm. The procedure is similar to the one described for latent class regression models where observed data is augmented by introducing unobserved data  $\lambda_{nk}$  in the complete likelihood function.

$$L(\boldsymbol{\phi} | \boldsymbol{\Lambda}) = \prod_{n=1}^N \prod_{k=1}^K [\pi_k \cdot P(\mathbf{Y}_n = \mathbf{y}_n | \boldsymbol{\alpha}, \boldsymbol{\beta}_k)]^{\lambda_{nk}}$$

In the E-step, the expected log-likelihood function is derived by replacing  $\lambda_{nk}$  by posterior probabilities  $\hat{p}_{nk}$ .

$$E[l(\boldsymbol{\phi} | \boldsymbol{\Lambda})] = \sum_{n=1}^N \sum_{k=1}^K [\hat{p}_{nk} \cdot \ln(\pi_k) + \hat{p}_{nk} \cdot \ln P(\mathbf{Y}_n = \mathbf{y}_n | \boldsymbol{\alpha}, \boldsymbol{\beta}_k)]$$

In the M-step, the two terms on the right hand side of  $E[l(\boldsymbol{\Phi}|\boldsymbol{\Lambda})]$  are maximized separately. The maximization of the first term with respect to  $\pi_k$ , is carried out by the method of Lagrange multipliers subject to the constraint  $\sum_1^K \pi_k = 1$ . The maximization of the second term with respect to  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}_k$  is carried out by transforming the polychotomous responses as a vector of 0-1 indicators, which allows the use of Poisson likelihood in the model fit. Hence each term of  $\sum_1^N \sum_1^K p_{nk} \cdot \ln P(\mathbf{Y}_n = \mathbf{y}_n | \boldsymbol{\alpha}, \boldsymbol{\beta}_k)$  is considered as a weighted Poisson log-likelihood function.

## SIMULATION STUDY

In order to assess the performance of the two proposed models, synthetic data sets were generated using a GLIM algorithm. The simulation study was devised to mimic the application of Camilleri and Green (2003) in which four car brands; four price values and two door features were generated to define the car attributes. In the application, a full profile design was employed in which 32 items were generated where each item had a unique attribute combination. This guaranteed a full factorial approach.

To simulate subjects' responses, parameter values were required for each of the  $K$  segments. A set of parameter values was generated for each model described in the preceding section using a data set illustrated in Camilleri and Green (2003). Model 1 is the latent class regression model using mixtures of multivariate conditional normal distributions. Model 2 is the latent class model, which accommodates a Proportional Odds model using a probit link function. The linear predictor used in both models included all main effects and all pairwise interactions of the three car attributes. The parameter vectors  $\boldsymbol{\beta}_k$  and  $\boldsymbol{\pi}$  were used to simulate rating responses and segment allocation for each of the  $N$  hypothetical subjects.

To allocate the  $N$  hypothetical subjects into  $K$  segments, the known proportions  $\pi_k$  were used to compute the cumulative probabilities,  $q_0, q_1, \dots, q_K$ , where  $q_0 = 0$ ,  $q_k = \sum_{i=1}^k \pi_i$  and  $q_K = 1$ . A set of uniformly distributed pseudo-random real values was then generated in the range  $[0, 1]$  to allocate the hypothetical subjects to one of the  $K$  segments. Subjects whose corresponding pseudo-random value was in the range  $[q_{k-1}, q_k]$  were allocated to the  $k^{th}$  segment. This classification gave each subject a random segment allocation.

To simulate the 32 synthetic data values for each subject, given his segment allocation, the known parameter values  $\boldsymbol{\beta}_k$  and the known item attribute values were substituted in the linear predictor. The linear predictors were then perturbed by adding an error terms  $\varepsilon_i$  to these utility values to have either Logistic, Normal or Extreme value distribution. These error terms were generated by

transforming pseudo-random real values  $u_i$  in the range  $[0, 1]$  from a uniform distribution.

$$\varepsilon_i = \log\left(\frac{u_i}{1-u_i}\right) \text{ if } \varepsilon_i \text{ has a Logistic distribution}$$

$$\varepsilon_i = \Phi^{-1}(u_i) \text{ if } \varepsilon_i \text{ have a Normal distribution}$$

$$\varepsilon_i = \log[-\log(1-u_i)] \text{ if } \varepsilon_i \text{ has an Extreme distribution}$$

A set of 6 specified cutpoint values  $\alpha_r$  was used to modify these perturbed linear predictors to rating scores ranging from 1 to 7. Utility values ranging from  $\alpha_{r-1}$  to  $\alpha_r$  were categorized as having a rating score  $r$ . This classification gave the rating responses for each hypothetical subject a random category allocation.

To investigate model performance of the two latent class models, a number of data sets were generated by using different sets of values  $u_i$ . Four factors that are listed in literature (Vriens, Wedel and Wilms 1996; Wedel and DeSarbo 1995) as having potential effect on model performance include:

- Number of simulated respondents
- Number of segments
- Size of perturbation
- Distribution of the error terms  $\varepsilon_i$

The size of the perturbation was varied by multiplying the error term  $\varepsilon_i$  by a specified scalar  $c$ . These four factors, which reflect a variation in conditions in applications, are examined as parameters of the experiment since they are expected to affect model performance.

The following six measures are normally used to assess computational effort, goodness of fit, predictive power, parameter recovery and segment membership recovery.

- The percentage of variance,  $R^2$  accounted for by the latent class model is a measure of the goodness of fit.
- The number of iterations required for convergence is a measure of the computational effort.
- The root-mean-squared error between the true and estimated parameters and the root-mean-squared error between the true and estimated segment membership probabilities are measures of parameter recovery.

$$RMS(\hat{\boldsymbol{\beta}}) = \left[ \sum_{p=1}^P \frac{(\beta_p - \hat{\beta}_p)^2}{P} \right]^{\frac{1}{2}}$$

$\hat{\beta}_p$  and  $\beta_p$  are respectively the estimated and true parameters; whereas  $P$  is the number of parameters.

$$RMS(\hat{\boldsymbol{\pi}}) = \left[ \sum_{k=1}^K \frac{(\pi_k - \hat{\pi}_k)^2}{K} \right]^{\frac{1}{2}}$$

$\hat{\pi}_k$  and  $\pi_k$  are respectively the estimated and true segment membership probabilities; whereas  $K$  is the number of segments.

- The root-mean-squared-error between the true and predicted responses is a measure of the predictive power.

$$RMS(\hat{y}) = \left[ \sum_{n=1}^N \sum_{j=1}^J \frac{(y_{nj} - \hat{y}_{nj})^2}{N \cdot J} \right]^{\frac{1}{2}}$$

$\hat{y}_{nj}$  and  $y_{nj}$  are respectively the estimated and true responses; whereas  $N$  and  $J$  are respectively the number of simulated respondents and the number of items assessed by each subject. A predicted response for model 2 was set to the rating category with the highest predicted probability.

- The percentage number of subjects that are correctly classified into their true segments is a measure of segment membership recovery. A subject is assigned to the segment with highest posterior probability.

## RESULTS OF THE SIMULATION STUDY

A major limitation of the EM algorithm is that it can converge on local stationary points and global maxima are not guaranteed. To overcome this limitation, five data sets were generated for each model varying the type of distribution for  $\varepsilon_i$ , number of subjects  $N$ , number of segments  $K$  and the perturbation constant  $c$ . These starting values were selected from a wide range of seed numbers.  $N$  and  $K$  were each varied at two levels (200 and 400 subjects; 2 and 4 segments). Two values were considered for the constant  $c$  (0.3 and 1). The Logistic, Normal and Extreme value distributions were considered for  $\varepsilon_i$ . Each simulated data set was re-fitted using either the latent class model 1 or model 2.  $RMS(\hat{\beta})$ ,  $RMS(\hat{\pi})$  and  $RMS(\hat{y})$  were computed after permuting the parameters and predicted responses to match estimated and true segments optimally. All the six measures were averaged over these five data sets.

| Measure              | N   | Model 1 | Model 2 |
|----------------------|-----|---------|---------|
| $R^2$                | 200 | 0.8216  | 0.8613  |
|                      | 400 | 0.8056  | 0.8473  |
| Number of iterations | 200 | 27.4    | 30.4    |
|                      | 400 | 28.9    | 28.8    |
| Segment recovery     | 200 | 95.1%   | 99.0%   |
|                      | 400 | 97.3%   | 99.4%   |
| $rms(\hat{\beta})$   | 200 | 0.2385  | 0.2109  |
|                      | 400 | 0.1679  | 0.1473  |
| $rms(\hat{\pi})$     | 200 | 0.0352  | 0.0461  |
|                      | 400 | 0.0309  | 0.0302  |
| $rms(\hat{y})$       | 200 | 2.9651  | 2.1538  |
|                      | 400 | 2.8945  | 2.0956  |

**Table 1:** Model performance by number of subjects

Table 1 displays model performance for the two models by considering 2 segments, a perturbation constant  $c = 0.3$ , and assuming a Normal distribution for  $\varepsilon_i$ . An increase in the number of subjects improves parameter recovery; however, goodness of fit deteriorates with an increase in the sample size. A change in the sample size has negligible effect on computational effort and segment membership recovery. A latent class model that accommodates the proportional odds model outperforms the latent class regression model in all the measures except computational effort.

| Measure              | K | Model 1 | Model 2 |
|----------------------|---|---------|---------|
| $R^2$                | 2 | 0.8512  | 0.8852  |
|                      | 4 | 0.9033  | 0.9361  |
| Number of iterations | 2 | 28.3    | 29.7    |
|                      | 4 | 35.6    | 37.4    |
| Segment recovery     | 2 | 97.6%   | 98.3%   |
|                      | 4 | 94.8%   | 97.1%   |
| $rms(\hat{\beta})$   | 2 | 0.1732  | 0.1231  |
|                      | 4 | 0.2536  | 0.2246  |
| $rms(\hat{\pi})$     | 2 | 0.0319  | 0.0322  |
|                      | 4 | 0.0168  | 0.0159  |
| $rms(\hat{y})$       | 2 | 2.1156  | 1.6897  |
|                      | 4 | 3.0598  | 2.6781  |

**Table 2:** Model performance by number of segments

Table 2 displays model performance for the two models by considering 400 subjects, a perturbation constant  $c = 0.3$ , and assuming a Logistic distribution for  $\varepsilon_i$ . Computational effort increases when fitting a latent class model with more segments. Parameter recovery, predictive accuracy and segment membership recovery deteriorate with an increase in the number of segments; however, goodness of fit and recovery of segment membership probabilities improve by an increase in the number of segments. Model 2 performed better than model 1 in almost all the measures. The number of iterations required and the mean  $RMS(\hat{\pi})$  are comparable for both models.

| Measure              | c   | Model 1 | Model 2 |
|----------------------|-----|---------|---------|
| $R^2$                | 0.3 | 0.8017  | 0.8223  |
|                      | 1   | 0.5452  | 0.5489  |
| Number of iterations | 0.3 | 30.4    | 30.9    |
|                      | 1   | 32.2    | 33.1    |
| Segment recovery     | 0.3 | 93.1%   | 95.3%   |
|                      | 1   | 73.4%   | 79.6%   |
| $rms(\hat{\beta})$   | 0.3 | 0.2481  | 0.2134  |
|                      | 1   | 0.3261  | 0.3144  |
| $rms(\hat{\pi})$     | 0.3 | 0.0316  | 0.0329  |
|                      | 1   | 0.1896  | 0.1529  |
| $rms(\hat{y})$       | 0.3 | 2.8614  | 2.5319  |
|                      | 1   | 3.3614  | 3.2189  |

**Table 3:** Model performance by perturbation constant

Table 3 displays model performance for the two models by considering 400 subjects and 2 segments and an extreme value distribution for  $\varepsilon_i$ . Inevitably, the amount of added

error decreases the performance of the algorithms, which is a well-known result in estimation theory. Furthermore, goodness of fit, parameter recovery, segment membership and predictive accuracy recovery deteriorate by increasing the error variance. Once more, model 2 is outperforming model 1 in most of the measures.

| Measure                   | Distribution | Model 1 | Model 2 |
|---------------------------|--------------|---------|---------|
| $R^2$                     | Logistic     | 0.9011  | 0.9216  |
|                           | Normal       | 0.9349  | 0.9356  |
|                           | Extreme      | 0.8544  | 0.8874  |
| Number of iterations      | Logistic     | 35.7    | 36.4    |
|                           | Normal       | 36.9    | 37.5    |
|                           | Extreme      | 35.9    | 35.8    |
| Segment recovery          | Logistic     | 97.8%   | 98.2%   |
|                           | Normal       | 98.3%   | 98.2%   |
|                           | Extreme      | 90.3%   | 94.1%   |
| $\text{rms}(\hat{\beta})$ | Logistic     | 0.2786  | 0.2586  |
|                           | Normal       | 0.2589  | 0.2572  |
|                           | Extreme      | 0.3325  | 0.3063  |
| $\text{rms}(\hat{\pi})$   | Logistic     | 0.0258  | 0.0227  |
|                           | Normal       | 0.0211  | 0.0231  |
|                           | Extreme      | 0.0312  | 0.0208  |
| $\text{rms}(\hat{y})$     | Logistic     | 3.0716  | 3.0511  |
|                           | Normal       | 3.0599  | 3.0435  |
|                           | Extreme      | 3.2566  | 3.1207  |

**Table 4:** Model performance by type of distribution

Table 4 displays model performance for the two models by considering 200 subjects, 4 segments and a perturbation constant  $c = 0.3$ . For each model, the six measures give comparable results when the choice of the distribution is Normal or Logistic. Goodness of fit, parameter recovery, segment membership and predictive accuracy recovery deteriorate when the extreme value distribution is used. The performance of model 1 is at best comparable to model 2 when the distribution of  $\varepsilon_i$  is Normal. This is due to the fact that model 1 assumes mixtures of multivariate conditional normal distributions.

## CONCLUSIONS

A latent class model that accommodates the proportional odds model outperforms the latent class regression model in segment membership and parameter recovery, goodness of fit and predictive accuracy. Goodness of fit improves for fewer data points and more segments; however, parameter recovery, segment membership recovery and predictive accuracy deteriorate with an increase in fitted segments. Computational effort increases with an increase in the number of clusters but is not affected by the choice of distribution, sample size and error variance. The choice of the error distribution has noticeable effect on the performance of the two models. The Normal and Logistic distributions outperform the Extreme value distribution. This may be attributed partly to the similarity of Normal and Logistic distributions and partly to the choice of the model to generate the parameter vectors  $\beta_k$  and  $\pi$ .

## REFERENCES

- Camilleri, L. and Green, M. (2004), Statistical Models for Market Segmentation, *Proceedings of the 19<sup>th</sup> International Workshop Statistical Modelling, Florence*. 120-124.
- Camilleri, L. and Portelli, M. (2007), Segmenting the heterogeneity of tourist preferences using a Latent Class model combined with the EM algorithm, *Proceedings of the 6<sup>th</sup> APLIMAT International Conference, Bratislava*. 343-356.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data via the EM algorithm, *Journal of the Royal Statistical Society, B*, 39, 1-38.
- DeSarbo, W., Wedel, M., Vriens, M. and Ramaswamy, V. (1992), Latent Class Metric Conjoint Analysis, *Marketing Letters*, 3,3, 273-288.
- Green, M. (2000), Statistical Models for Conjoint Analysis, *Proceedings of the 15<sup>th</sup> International Workshop on Statistical Modelling, Bilbao*. 216-222.
- Helsen, K., Jedidi, K. and DeSarbo, W.S. (1992): A new approach to country segmentation using multinational diffusion patterns. *Journal of Marketing*, 57: 60-71
- Nelder, J.A and Wedderburn, R.W.M. (1972), Generalized Linear Models, *Journal of the Royal Statistical Society, A*, 135, 370-384.
- Ramaswamy, V., DeSarbo, W.S., Reibstein, D., Robinson (1993), An Empirical Pooling Approach to Estimate Marketing Mix Elasticities with PIMS Data, *Marketing Science*, 12, 103-124.
- Vriens, M., Wedel, M. and Wilms, T. (1996), Conjoint Segmentation Methods A Monte Carlo Comparison, *Journal of Marketing Research*, 23, 73-85.
- Wedel, M. and DeSarbo, W.S. (1995), A Mixture Likelihood Approach for Generalized Linear Models, *Journal of Classification*, 12, 1-35.
- Wedel, M. and Kamakura, W.A. (2000), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers.

## AUTHOR BIOGRAPHY

**LIBERATO CAMILLERI** studied Mathematics and Statistics at the University of Malta. He received his PhD degree in Applied Statistics in 2005 from Lancaster University. His research specialization areas are related to statistical models, which include Generalized Linear models, Latent Class models, Multi-Level models and Random Coefficient models. He is presently a lecturer in the Statistics department at the University of Malta.

# PREDICTING POST MACHINING DISTORTIONS VIA FE SIMULATIONS AT DESIGN PHASE

Waqas Saleem  
Fan Yuqing  
37-Xueyuan Lu, Beihang University  
New Building 428-B, Beijing, China

## KEYWORDS

ANSYS, CATIA, Computer Aided Engineering (CAE),  
Computer-aided analysis, Manufacturing,

## ABSTRACT

Problems of post machining distortions and dimensional instability are frequently emerged in thin walled components. These are mainly instigated and developed due to presence of initially induced residual stresses that generate during the material processing operations. This research work deals with the prediction of post machining distortions of a thin walled aerospace component at design synthesis phase by exploiting the FE based manufacturing simulations. For this purpose, large scale cutting simulations in an optimized sequence are executed in ANSYS software. These simulations are accomplished by employing the element deactivation technique after developing nonsymmetrical initial residual stress field. Based on predicted distortions, certain modifications in the initially developed design configuration are incorporated to generate a manufacturable design configuration. Final CAD feasible design (3D model) is developed in CATIA software. Dimensional accuracy of actual machined component is determined via 3-axis coordinate measuring machine (CMM). Design viability, structural performance and cutting simulation results of novel design proposal are evaluated and compared with the statistics of actual operating machined component. Evaluations reveal that simulations and analysis result corroborate with the actual data and thus validate the approach adopted.

## REALIZATION OF MACHINING DISTORTIONS AT DESIGN STAGE

FEM (Finite element method) is a powerful tool that provides good approximate solutions to continuum problems. This can be used to simulate the metal removal action in cutting procedures. Presently, FE cutting and manufacturing simulations techniques have become imperative to validate the manufacturing aptness of developed design configurations. These simulations have now been realized as important aspect of today's contemporary design environment because these provide cogent means to replace the costly experimentation or prototyping. By executing these simulations and analysis, a designer can deduce the post machining inaccuracies and can induce appropriate measures at an early design stage to

avoid the distortions and recondite problems that generally deem to be inevitable in actual machining processes. With the aid of these simulations a designer can also predict the clamping effects, optimized cutting scheme, cutting tool impact and other important requirements early at the design stage. By imparting the results and data of these simulations to other design and manufacturing sections, important redesign decisions can be incorporated earlier to save time, efforts and development costs. Indeed FE based manufacturing and analysis simulations help to bridge the gap between design and computational engineers and result in more robust design proposals (Marcus et al. 2005).

Generally, dimensional distortion problems are observed less or more in any component that is thin in thickness, complex in structure and tight in machining tolerances. Specially, in case of low rigidity components (like thin walled multi pocket shaped machined parts) distortion potential increases (Ratchev 2004) greatly. In designing these types of components, analysis of post machining distortions via FE based cutting simulations become indispensable. For example, to reduce the weight of airplane, designers strive to develop integrally stiffened ribbed type components that increase the overall stiffness. However, during the machining of such thin walled aero parts approximately 90 % of the material is removed, that frequently result in severe distortions. Therefore prediction of post machining distortions, effect of material residual stresses and assessment of manufacturability constraints are necessary to develop an optimal manufacturable design proposal. This research deals with the large scale material removal simulations without taking into account some factors like variable cutting forces, tool deflections, heat dissipation effects etc.

## METHODOLOGY

Based on stipulated loads, constraints and other requirements a designer finalizes the initial design model that is sent to CAD department for detailed drawings. At this stage, the suggested initial design proposal can be analyzed to predict post machining distortions. In this research, effects of post machining distortions and optimal cutting sequence is determined via cutting simulation technique (Element Birth/Death feature) in ANSYS software (ANSYS V10 Help). The necessary steps are described below:

- 1) Transformation of initial design model to a suitable FE model for machining simulations

- 2) FEM Preprocessing
- 3) Meshing
- 4) Definition of loads and constraints
- 5) Incorporation of machining simulation restrictions
- 6) Coupled field thermomechanical analysis for developing the initial residual stresses
- 7) Execute element deactivation process step by step (for example selected pockets, material layers etc)
- 8) Analysis and evaluation of results

For machining simulations, the initial design proposal (3D model) is first modified to an acceptable FE model that is suitable for machining and manufacturing simulations. In this research, the initially developed model (in CATIA) of a thin walled aero part is transformed to a model apt for machining simulations. This is shown in figure 1.

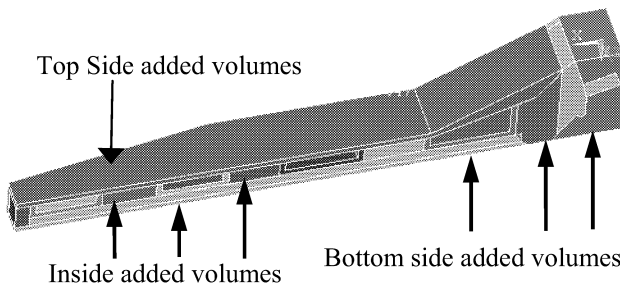


Figure 1: Model for Machining Simulations (Section view)

The initial emptied pockets are filled with equivalent volumes and then integrated to the main volume before meshing.

The underlying assumption behind all the element deactivation techniques is that any removal of material with certain stiffness (having a certain initial residual stress state) causes the majority of distortions. The virtual material removal from any position of component imparts certain distortions in the component when it attains a new equilibrium state. When an element is deactivated via Birth/Death feature in ANSYS software, it actually does not remove; rather its stiffness is multiplied by a severe reduction factor which means the presence of this element is almost ignored (ANSYS V10 Help). Similarly mass, damping and other properties associated with this element are also zeroed out. However, the effect of removal of this element is cumulatively transferred to all the live elements present in the model.

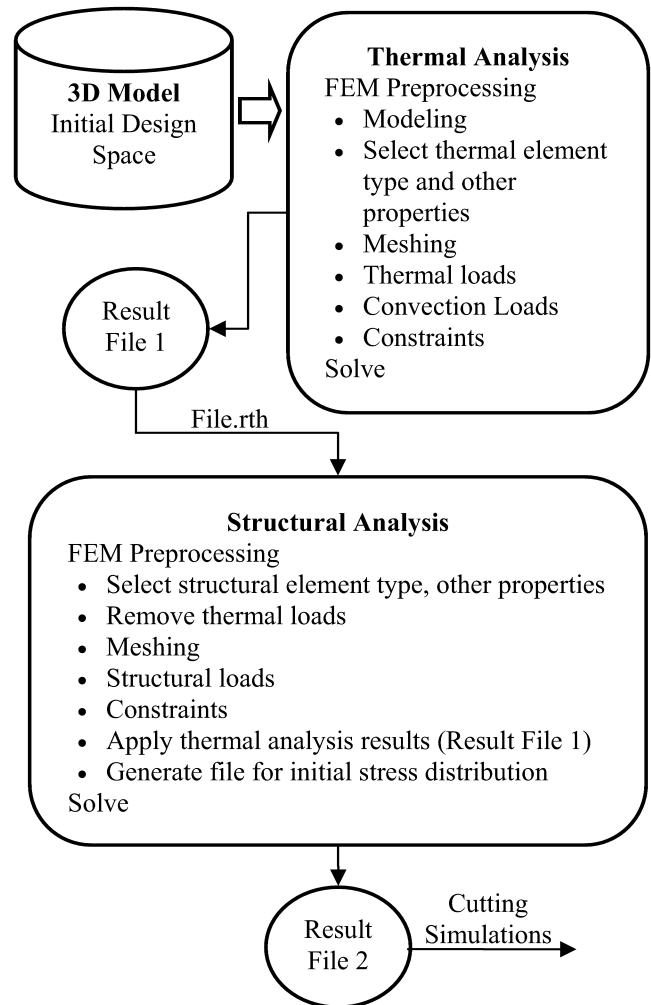
#### Development of Initial Residual Stress Field - Sequential Coupled Field Analysis

Before the cutting simulations, a sequential coupled field analysis is performed to develop an initial residual stress field in the model. Accounting the effects of initial residual stresses at the component design stage is essential because these are major source of post machining distortions (Wei and Wang 2007). Generally, their effect at the early design stage is not considered. However, through manufacturing simulations these can be studied and controlled by effective measures like adding an additional stress relieving processes among machining steps that cause maximum distortions.

In this research, following assumptions are made for developing an initial residual stress field:

- 1) Residual stresses distribute symmetrically along the mid plane.
- 2) Residual stress varies along the thickness direction, and distributes evenly paralleling with the mid plane.
- 3) Material is isotropic and perfectly elastic that deforms in a stable elastic manner.

The material of selected component is A17075-T6. Because the finite element deactivation is a nonlinear analysis, therefore model is meshed with suitable mesh size to cope with computational limitations. In this case, the model is meshed into 74122 elements by using element thermal Solid87. For the sequential coupled field analysis, a thermal analysis is performed first by applying temperatures and convection loads. The process is explained in figure 2.



Figures 2 : Sequential Coupled Field Thermo-Mechanical Analysis

Material heat processing temperatures, convection properties and other inputs are taken from the valid published source (ASM Handbook, Heat Treating, Vol. 04). A temperature of 60°C to all the nodes, an additional temperature of 15°C to all the surface areas and a convection load to all the outer areas are applied. The attained initial temperature distribution (result file) is then used for generating a nonsymmetrical residual stress field via structural analysis. This is performed by switching



thermal elements to structural and then by applying structural loads, symmetry and other constraints. The developed stress state is then evaluated by comparing with the actual published data. After developing a nonsymmetrical residual stress field, cutting simulations are performed.

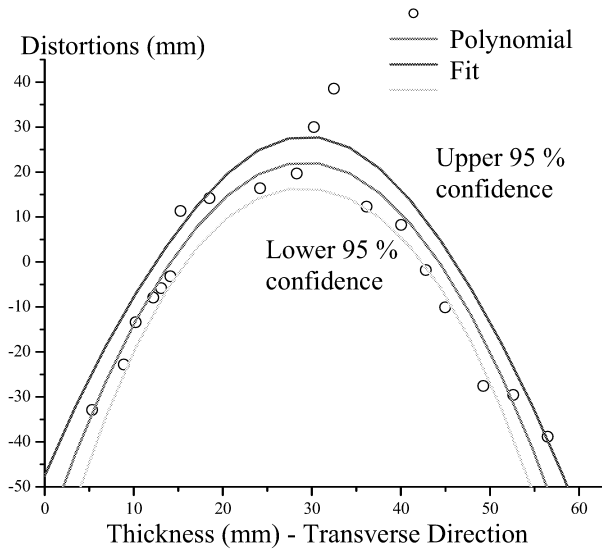


Figure 3: Initial Residual Stresses

Initial residual stress field developed in the component is measured in thickness direction and then plotted (by curve estimation), as shown in figure 3. Generally, the residual stresses in pre-stretched aluminum alloys after material heat processing lie within the range of  $\pm 30$  MPa (ASM Handbook, Heat Treating, Vol. 04). From figure 3, it is apparent that the methodology and analysis performed to develop the initial residual stresses in FE model are adequately suitable.

## Manufacturing Simulations

By applying the symmetry constraints, model is constrained from two side areas in transverse direction. Machining steps 1 ~ 5 follow the material removal (elements deactivation simultaneously one by one) entirely from top face then inside pockets, inside through pockets, side pockets and finally the bottom face. The model is solved for each step under the same loading and constraints. These steps are shown in figure 4.

Table 1: Distortions (mm) Measured via Cutting Simulations

|                  | Step 1  | Step 2 | Step 3 | Step 4 | Step 5       |
|------------------|---------|--------|--------|--------|--------------|
| <b>Resultant</b> | 0.59    | 0.343  | 0.348  | 0.391  | 2.824        |
| <b>X(Max)</b>    | 0.474   | 0.34   | 0.316  | 0.319  | 0.628        |
| <b>X(Min)</b>    | -0.188  | -0.215 | -0.204 | -0.202 | -0.137       |
| <b>Y(Max)</b>    | 0.37    | 0.081  | 0.0775 | 0.0731 | 0.422        |
| <b>Y(Min)</b>    | -0.0768 | -0.188 | -0.199 | -0.202 | -0.071       |
| <b>Z(Max)</b>    | 0.25    | 0.139  | 0.141  | 0.146  | <b>2.743</b> |
| <b>Z(Min)</b>    | -0.184  | -0.187 | -0.196 | -0.197 | -2.742       |

The distortions observed in each cutting simulation step are presented in table I. It is noted that maximum distortion occurred when the bottom side elements are deactivated

and model is solved. It is seen from the results that the deformation in Z direction amplifies step by step. While in X direction it decreases till step 4 and then increases considerably in step 5 when bottom side elements are deactivated. A similar deformation trend is also seen in Y direction. It can be anticipated that removal of more and more elements increases the distortions because stresses in the work piece redistribute to reach new equilibrium state when elements are removed step by step. The inner part of the piece is stressed compressively. That is, the stress changes from the top surface to the bottom in a tensile-compressive manner. Although the overall stresses are balanced, the concentration of the tensile stresses in the upper part of the specimen cause both ends of the specimen to deflect upwards. After removal of the fifth layer, the lower part tensile stress of the specimen becomes strong enough to balance the upper part; therefore deflection reduces as a result. This is shown in figure 4.

The stresses developed after the machining simulations that are considered as the post machining stresses are shown in figure 5. The maximum value of stress is 68.2 MPa and minimum 31 MPa. This range of stress is a little high as compared to general published data for post machining stresses on aluminum plate specimens. This may be due to assumptions made in the cutting simulations for computational restrictions and neglecting tools deflections, material removal heat dissipation effects. However, the results still are feasible.

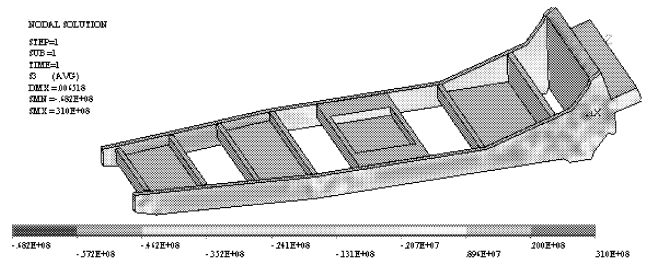


Figure 5: Post Machining Distortions (After Simulations)

## ACTUAL EXPERIMENTATION/MEASUREMENT

The proposed design is validated by comparing with the actual operating component of aircraft vertical tail fin. The machining clamping scheme actual component is shown in figure 6.

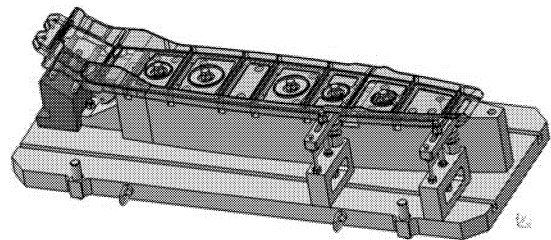


Figure 6: Fixture Assembly for Actual Part Machining

The machined component in finished form is shown in figure 7. The component is machined on CNC milling machining center with a cutting speed 18000 rpm, depth of

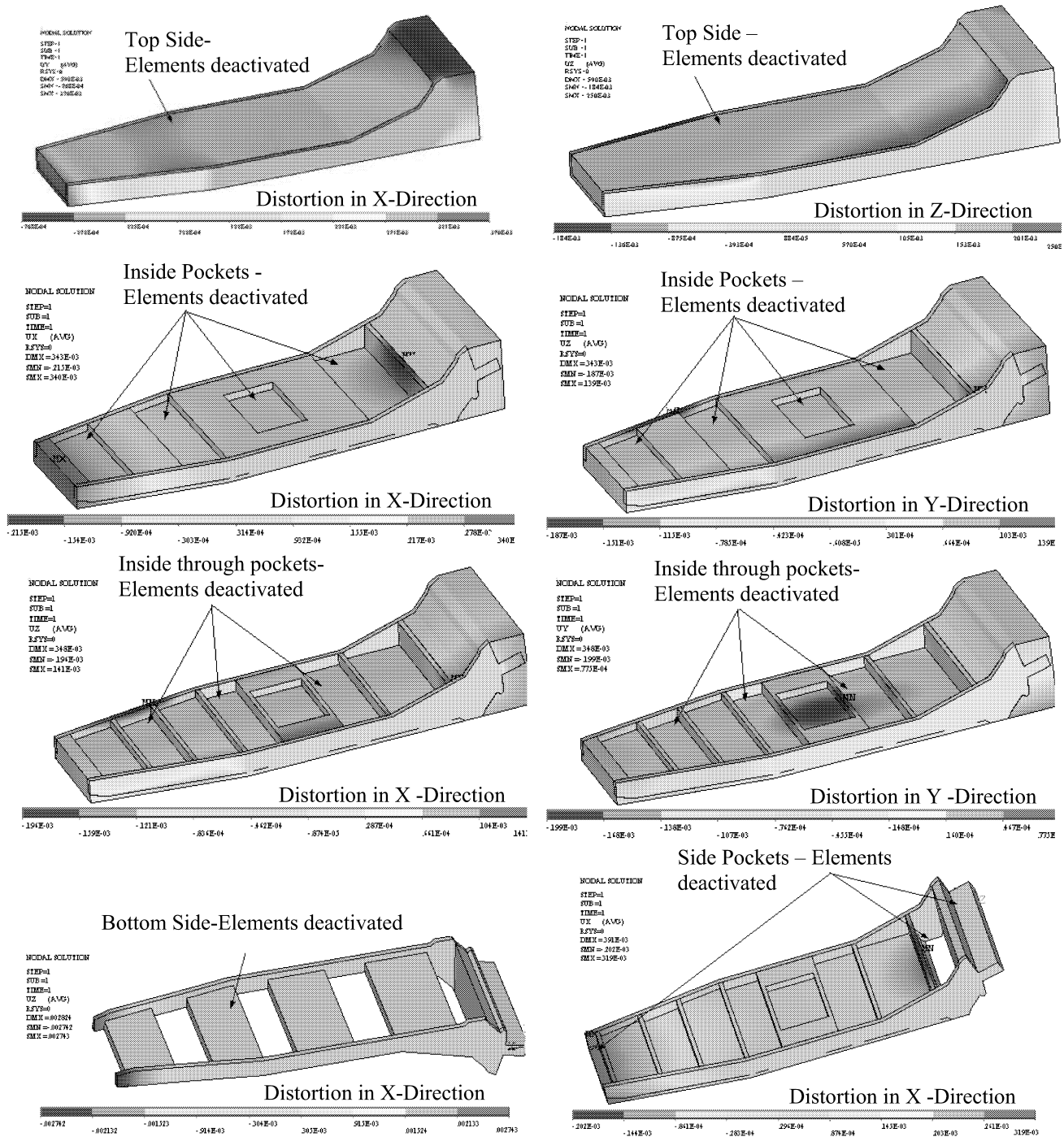


Figure 4: Machining Simulations

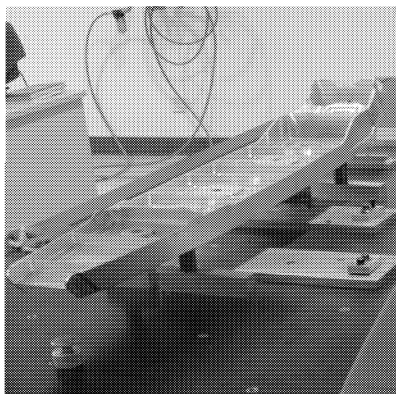


Figure 7: Actual Part Machining

cut 0.015 mm, and the feed rate as 4000 mm/min. The dimensional accuracy of finished component is measured by a three axis coordinate measuring machine (CMM) at different points. Points marked for measuring the dimensions in a defined order are shown in figure 8. Figure 9 gives the dimensional inaccuracy and deformation trend of finished machined component.

## DISCUSSION AND CONCLUSIONS

Machining simulation results are compared with the actual machining data. In first look it seems that machining simulation distortions of the new design are more than the actual component in finished form. The reason why the

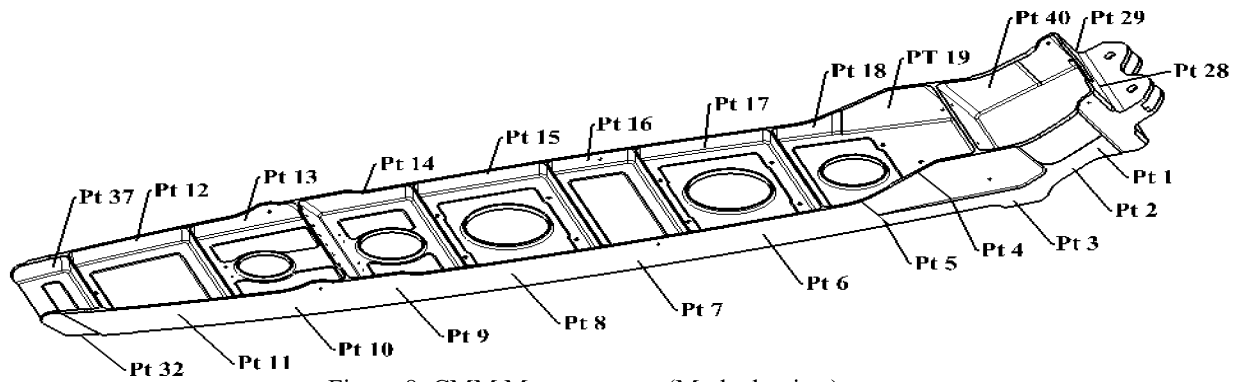


Figure 8: CMM Measurements (Marked points)

actual machining results are better is due to approximations made in the simulation process. However, when the new design proposal will actually be machined it results would be better because the most critical steps that produce the maximum distortions and the best cutting sequence is already simulated at the design phase. By comparing the simulation deviation values with the actual measurements, it seems that both the deflection trends are consistent. If the factors like variable cutting forces, tool deflections, finite element mesh size etc are taken into account, the percentage error between simulation and experiment will be small.

The clamping scheme of component also plays an important role. Optimized clamping scheme for the new design proposal is not studied in this research. With these inputs, a designer gets confidence about the design fidelity and alterations are minimized when design is validated by actual experimentation. The methodology envisaged in this research, can be used for predicting post-machining distortions for any component at the design stage for developing robust design proposals.

## REFERENCES

- ANSYS V10 Help, Nonlinear Analysis, Element birth/Death
- ASM Handbook, Heat Treating, Vol. 04, 1991, ISBN 0-87170-379-3, pp 1905-1949
- He Ning, Wang Zhigang, Jiang Chengyu, Zhang Bing, 2003, "Finite element method analysis and control stratagem for machining deformation of thin-walled components", Elsevier, Journal of Materials Processing Technology, 139 (2003) 332–336, 2003
- Marcus Sandberg, Tobias Larsson, Peter Astrom and Mats Nasstrom, 2005, "A design tool integrating cad and virtual manufacturing for distortion assessment", International conference on engineering design iced.
- Ratchev S, W. Huang, S. Liu, A.A. Becker, 2004, "Modeling and simulation environment for machining of low-rigidity components", Journal of Materials Processing Technology 153–154 (2004) 67–73
- Wei Yu, X. W. Wang, 2007, "Computer simulation and experimental study of machining deflection due to original residual stress of aerospace thin-walled parts", The international journal of advanced manufacturing technology, Vol 33, No 3-4, pp 260-265

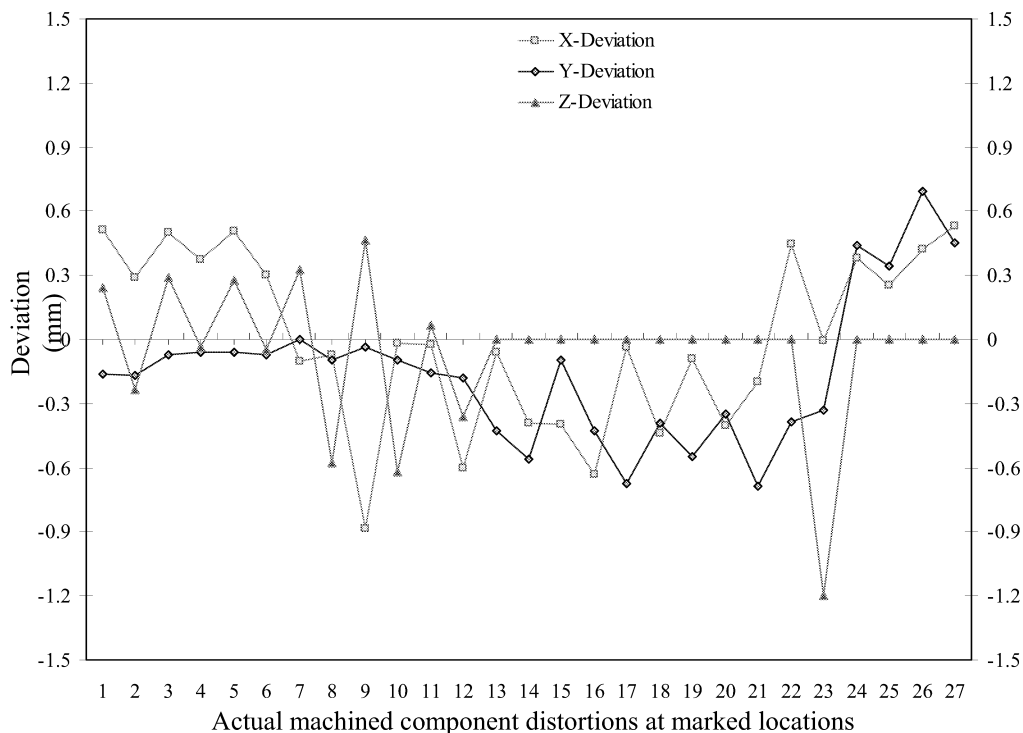


Figure 9: Dimensional variations measured by CMM

# ON-LINE FAULT DETECTION OF A CLOSE LOOP CIRCUIT HYDRAULIC SYSTEM WITH A VARIABLE DISPLACEMENT PUMP

C. Angeli

Department of Mathematics and Computer Science  
Technological Education Institute of Piraeus  
Konstantinoupoleos 38, N. Smirni  
GR-171 21 Athens, Greece  
E-mail: [c\\_angeli@otenet.gr](mailto:c_angeli@otenet.gr) or  
[c.angeli@computer.org](mailto:c.angeli@computer.org)

A. Chatzinikolaou

S. Patsi 62  
GR-118 55 Athens, Greece  
E-mail: [achatz@otenet.gr](mailto:achatz@otenet.gr)

## KEYWORDS

Modelling, Simulation, Fault Detection, Fault Compensation

## ABSTRACT

The purpose of development of on-line fault detection methods for drive and control systems is of main importance for the modern production technology in order to improve productivity, reduce operational and maintenance faults and increase safety. In this paper, sensor data are recorded and modelling information is used in comparison to these data for the on-line detection and compensation of faults in a close circuit hydraulic system with a variable displacement pump. Additionally, stored knowledge is used in cooperation with these data for the final diagnostic process.

## INTRODUCTION

Fluid power systems have been used in a wide variety of applications ranging from robotics and aerospace to heavy industrial systems and they are becoming more complex in design and function. Reliability of these systems must be supported by an efficient maintenance scheme.

Due to a component wear or failure, some system parameters may change causing abnormal behaviour in components or in the overall circuit. The development of a suitable system, that is able to detect on-line a defective component and potentially compensate the consequences of the fault on-line offers a higher degree of reliability and operational efficiency of the total system.

Faults in hydraulic components are often caused by a small incipient leakage and can be recognised by an increased temperature, a decreased performance or both. A lot of monitoring and fault detection techniques have been developed for the detailed cause of component malfunction as (Chen and Lu 1995, Muenchof et al. 2003, Tan and

Nariman 2001, Oehler et al. 1997, Kwan et al. 2003) that makes them complicated and expensive for the requirements of practice. On the other hand, an operator is mainly concerned with the performance of the system than to identify a fault within the component. In consequence potential on-line diagnostic approaches for hydraulic components are of great interest to the industrial maintenance systems.

On-line fault detection methods for hydraulic components use signal information depending on the nature of the component. In this paper, pressure signals information is used as a source for the fault detection purpose. It takes little effect of the background noise and offers sufficient information for the fault detection (Hehn 1995 and Gao et al. 2003).

The main components of the hydraulic system were modelled and on-line data referring to pressure signals and other kinematic signals, as the angular velocity signal of a hydraulic motor, are acquired from the actual system and compared with the relevant variable values of the simulation process.

The data acquisition process was performed using the DASyLab software. The diagnostic decisions are concluded by the knowledge base of the system. Recent research work in developing of diagnostic procedures for hydraulic systems or hydraulic components includes (Angeli and Chatzinikolaou 2005, Mundry and Stammen 2002, Straky et al. 2003, Atkinson et al. 1993, Murrenhoff et al. 2004, Angeli 2008).

## DESCRIPTION OF THE SYSTEM

A hydraulic close circuit system with a variable pump was used that drives a hydraulic motor through a cyclical routine. It is needed a high speed for a short time and then it returns to a low speed according to a periodically changed voltage  $U$  to a proportional valve that controls the pump flow.

Figure 1 shows the diagram of the hydraulic system that consists mainly of a variable displacement axial piston pump (1), a fixed displacement hydraulic motor (2) and a close circuit control block.

The hydraulic motor accelerates and rotates a mass with a given speed. The rotation speed of the hydraulic motor depends on the actual flow rate through the motor and on its displacement volume.

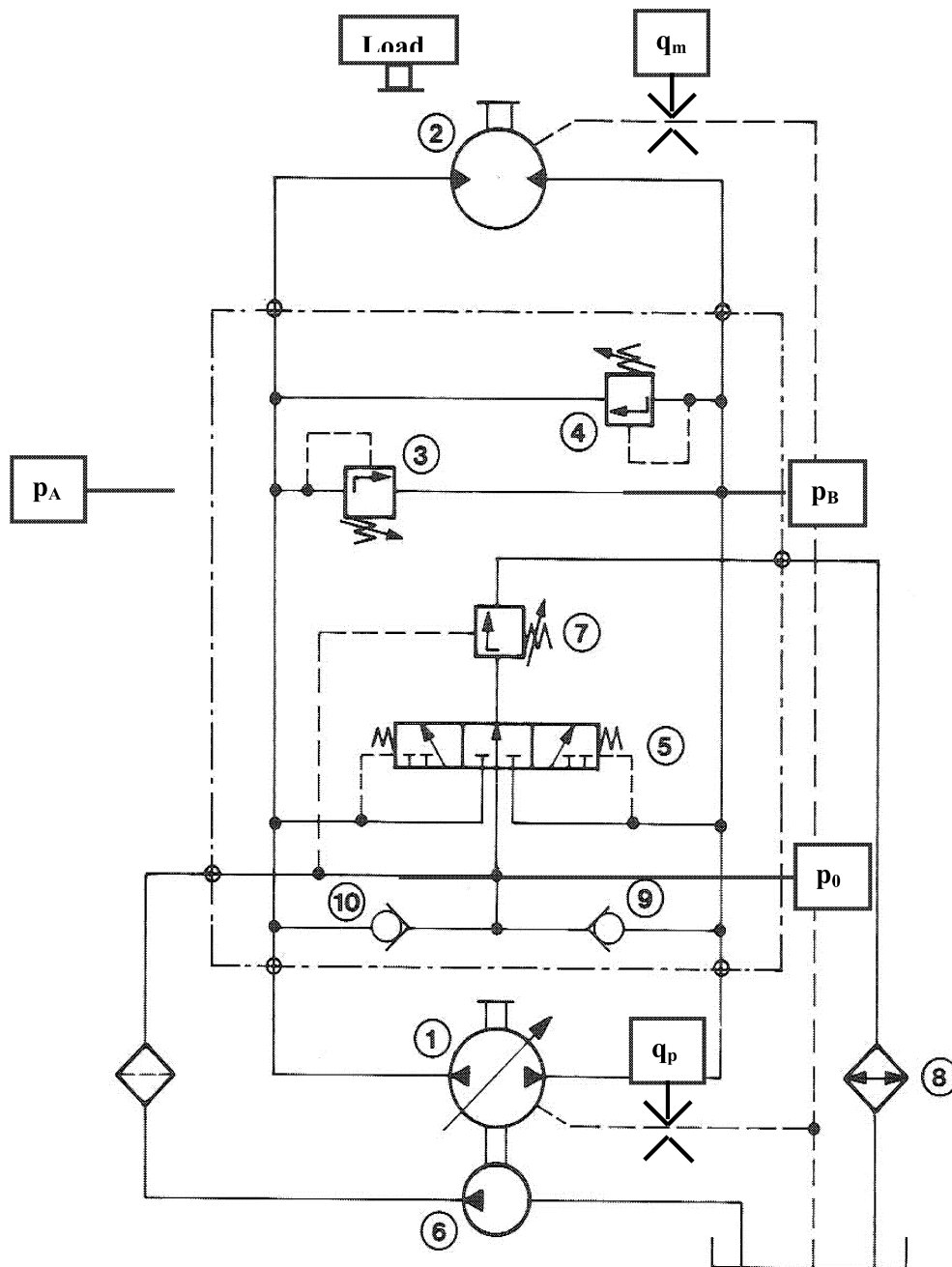


Figure 1: The actual system

In every hydraulic circuit, the pump is an indispensable component. Pumps convert mechanical power into hydraulic power by delivering a variable flow at the pump output.

The pump used is a variable displacement axial piston pump and is presented in Figure 2.

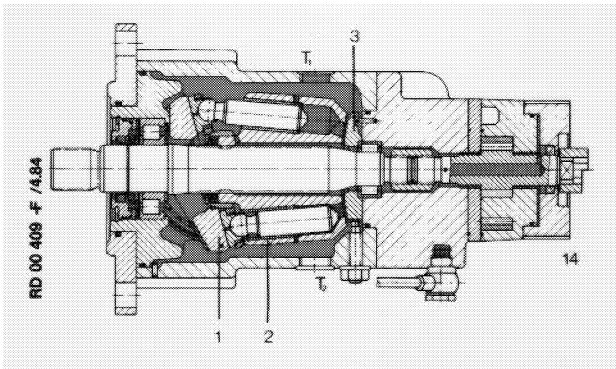


Figure 2: Variable displacement axial piston pump

Typical faults that affect the performance of the hydraulic system are following:

Typical pump faults:

1. Internal leakage (piston or swash plate wear) that has as effect overheating and flow and pressure losses.  
Extensive pump leakage causes low volumetric efficiency of the pump and hence increased temperatures locally and can affect the performance of the system. Piston wear affects the flow waveform from the pump output. Since pressure is related to flow, the pressure waveform will also be affected.
2. Wear of bearings that has as effect noise, overheating, vibration, increased input torque and hence input power.
3. Defective pump flow control system.

Typical motor faults :

1. Internal leakage that has as effect reduction of angular velocity and output torque.
2. Wear of bearings that has as effect noise, overheating and vibration.

Typical fault of the close circuit control block:

Reduced pressure  $p_0$  at the suction ports of the pump that has as effect strong cavitation noise and overheating due to insufficient or missing progressive replacement (regeneration) of the close circuit oil quantity.

## MODELLING AND SIMULATION PROCESS

A model of the hydraulic system was developed. This model was validated in comparison to the real process measurements. Physical relations as well as relations derived from the technical specifications data sheets were used for the development of this model.

For those components which contain large quantities of oil as pipes and hydraulic actuators it may be assumed that the

change in pressure in the dynamic state is proportional to the net inflow of oil, that is:

$$dp / dt = (E / V_o) \cdot \Sigma Q \quad \text{where}$$

$E$  is the elasticity module of the oil plus the included air,

$V_o$  is the Volume of the pipe plus a part of the volume of the attached actuator,

$$\Sigma Q = Q_{in} - Q_{out},$$

$Q_{in}$  is the incoming flow to the volume  $V_o$  of a connecting pipe and

$Q_{out}$  is the outgoing flow from  $V_o$ .

The variable displacement pump is controlled by a proportional valve according to the relation:

$$V = V_{max} \cdot U / U_{max} \quad \text{where}$$

$V$  is the displacement (flow per revolution) of the pump

$V_{max}$  is the maximal displacement

$U$  is the voltage to the amplifier of the proportional control valve and

$U_{max}$  is the maximum control voltage (10 V)

The pump flow to the hydraulic motor is given by:

$$Q = V \cdot n \cdot \eta_v \quad \text{where}$$

$n$  is the number of the revolutions per minute of the pump shaft and

$\eta_v$  is the volumetric efficiency of the pump

The modelling of the hydraulic elements with the coupled moving masses leads to a non-linear system of equations.

The model was validated in comparison to the real process data to determine the uncertain parameter values and define the acceptable limits of deviation between the measurements and the simulation results in order to establish the fault detection system

## MONITORING PROCESS

The connection of the electrical devices and the measurement instruments with the built-in I/O card was realized by an electronic interface, Figure 3.

The proportional valve for the control of the pump flow is controlled by a voltage of 0 to  $\pm 10$  V via the electronic amplifier of type VT5005. The analogue input of 0 to  $\pm 10$  V for the amplifier comes from a PCI-multi I/O card.

The input to the actual system is the voltage signal  $U$  from the control system and the outputs which are fed to the expert system are the angular velocity  $\omega$  of the load, the pressures  $p_0$ ,  $p_a$ ,  $p_b$ , the leakage flows  $q_m$  and  $q_p$  and the state signals from the devices of the power unit.

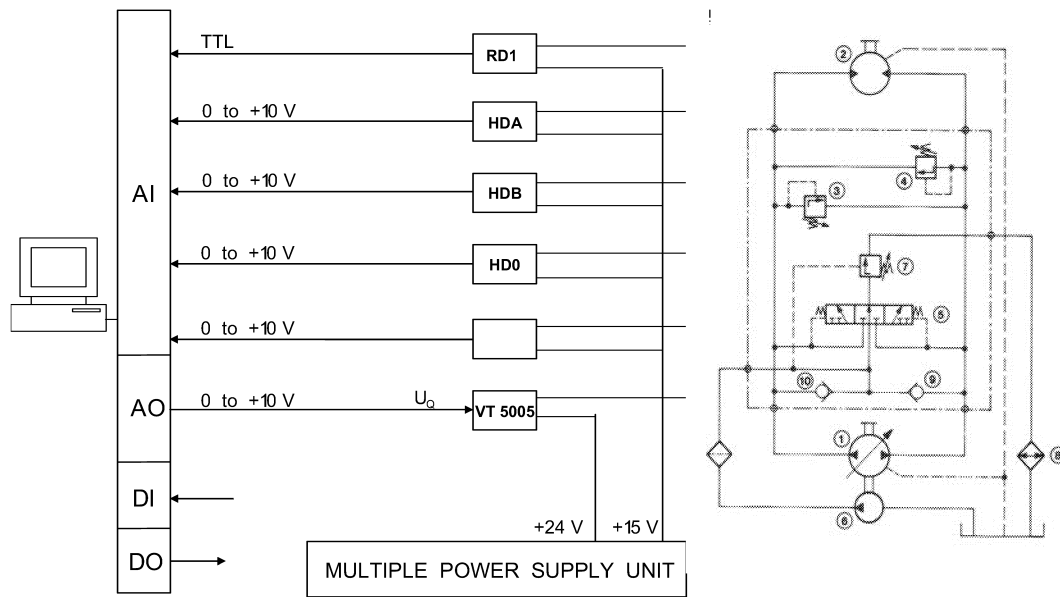


Figure 3: Diagram of data acquisition system

For the monitoring process a suitable environment which is capable to generate and evaluate the information provided was developed using suitably connected modules of the DASYLab software. The data acquisition system measures the pressures  $p_0$ ,  $p_a$ ,  $p_b$  and the angular velocity  $\omega$  as well as the monitoring of the digital signals that indicate the correct operation of the electric motor and other devices of the system.

The measured values of the pressures  $p_a$  and  $p_b$  as a response to a step signal  $\Delta U$  to the proportional valve for the pump flow and for a leakage free case are illustrated in Figure 4. It can be observed that pressure  $p_a$  at the A-port of the hydraulic motor increases at the beginning because of the inertia of the rotated motor load and then decreases and takes a lower constant value. Pressure  $p_b$  at the B-port of the hydraulic motor decreases at the beginning and then increases according to the increasing flow resistance.

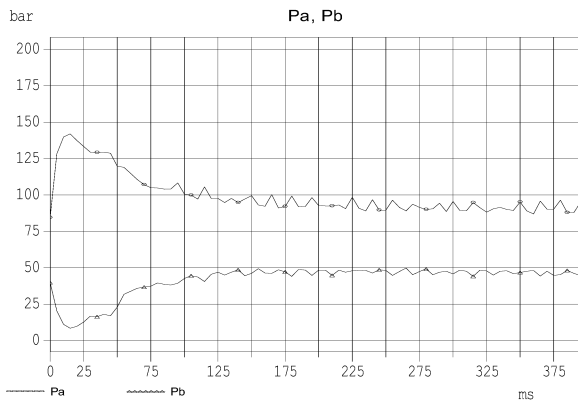


Figure 4: Measured pressures  $p_a$  and  $p_b$  in a leakage free case

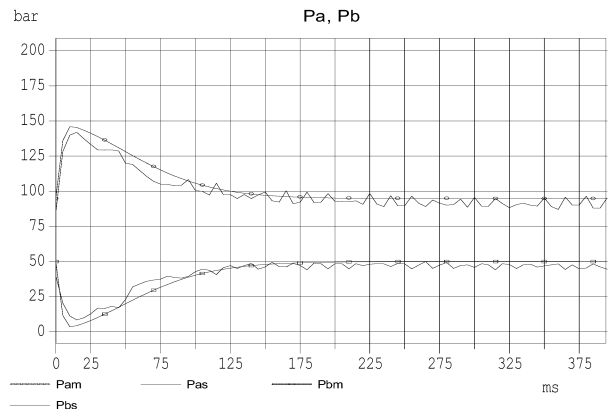


Figure 5: Comparison between measured and calculated pressure values  $p_a$  and  $p_b$  in a leakage free case

The comparison of the measured and the calculated values, Figure 5, was then performed through another DASYLab subsystem and the difference was written to an output file that is used for the fault diagnosis process.

### FAULT DETECTION PROCESS

The fault detection process is based on the model of the normal behaviour. The simulation program calculates the values of the angular velocity response to the input voltage and the pressures for the same period of time. Comparison of measured and calculated values is then performed through a DASYLab subsystem.

The differences are written to output files after their transformation to linguistic information and interact with the knowledge base of the system for the final diagnostic process. When a pump fault is detected the system should stop operating. In this case expert knowledge suitably

formatted is used for the diagnosis of the specific pump fault. A decision tree was used as technique to define the various logical paths that knowledge base must follow to reach conclusions. From the decision tree the relevant rules to each node were written and so the initial knowledge base was constructed. In the following example a rule is presented as it is needed to make a decision:

```

if      '?reduced pressure' is Yes and
        '?down' is No and
        '?motor' is No
then    '?electrical failure' is Yes.

```

The system searches for the topics 'reduced pressure', 'down' and 'motor' to satisfy the rule. Each of these topics may be further a set of rules, or simply a question asked to the user.

## FAULT COMPENSATION FOR HYDRAULIC PUMPS

The decreased speed drive due to a slightly worn hydraulic motor or a flow reduction of the defective pump is compensated by adding a correction voltage value  $\Delta U$  to the command value  $U$  of the proportional valve that controls the variable pump of the close circuit system and thus the motor speed according to the relation:

$$\Delta U = U \cdot (\Delta \omega / \omega_m)$$

## CONCLUSION

Faults in hydraulic components are often caused by a small incipient leakage and the development of a suitable system, that is able to detect and eventually compensate the consequences the leakage before the repair of the defective component is beneficial for the production technology.

In this paper, modelling information in comparison to on-line measurements referring to the pressure signals as well as to the angular velocity signal were used for the detection of abnormal behaviour in a close loop hydraulic system with a variable displacement pump.

Diagnosis of the cause of fault is performed by expert knowledge suitably formatted when the operation of the system is stopped. The developed system is also able to compensate leakage faults without the need of additional devices.

## REFERENCES

- Chen, Z. and Y.Lu. 1995. "State Monitoring and Fault Diagnosis for Hydraulic systems", *Pneumatics and Hydraulics*, 2, pp. 3-7.
- Muenchof, M.; H. Straky and R. Isermann. 2003. "Model-based fault detection and diagnosis for hydraulic braking systems", *Safeprocess 2003*, Washington, USA, pp.307-312.
- Tan, Hong-Zhou and S. Nariman. 2001. On Condition Monitoring of Pump pressure in a Servo-Driven System. *Proceedings of the 2001 American Control Conference*, pp. 4478-4483.
- Oehler, R.; A. Shoenhoff and M. Schreiber. 1997. "On-line model-based fault detection and diagnosis for a smart aircraft actuator". In *Proceedings Safeprocess 1997*, Kingston upon Hull, U.K. pp. 575-580.
- Kwan, C.; R. Xu and X. Zhang. 2003. "Fault detection and Identification in Aircraft Hydraulic Pumps Using MCA", In *Proceedings Safeprocess 03*, Washington, U.S.A. pp. 1137-1143.
- Hehn, A. H. 1995. Fluid power troubleshooting. Marcel Dekker, Inc. New York, USA.
- Gao, Y.; O. Zhang and X. Kong. 2003. Wavelet-based Pressure Analysis for Hydraulic Pump Health Diagnosis, Transaction of ASAE, Power and Machine Division, Vol. 46(4), pp. 969-976.
- Angeli, C. and A. Chatzinikolaou. 2005. Troubleshooting in Hydraulic Systems using knowledge-based Methods, *International Journal of Engineering Simulation*, Vol. 6, Nr.1, pp. 24-29.
- Mundry, S. and C. Stammen. 2002. "Condition Monitoring for fluid technology", *o+p Oelhydraulic and Pneumatic* Vol. 46, Nr.2.
- Straky, H.; M Muenchhof. and R. Isermann. 2003. "Model-based fault detection and diagnosis for hydraulic braking systems", *Safeprocess 2003*, Washington, USA, pp.307-312.
- Atkinson, R.M; M.R. Montakhab.; D.J. Woollons; P.A. Hogan; C.R. Burrows and K.A. Edge. 1993. "DESHC: a diagnostic expert system for hydraulic circuits" CERT-ONERA International Conference on fault diagnosis, Toulouse, France.
- Murrenhoff, H.; Th. Meindorf and C. Stammen. 2004. Condition Monitoring in Fluid Technology. *Proceedings of 4<sup>th</sup> International Fluid Power Conference*, Vol. 2, Dresden, pp. 219-244.
- Angeli, C. 2008. On-line Expert Systems for Fault Diagnosis in Technical Processes. *Expert Systems*, Vol. 25, No2, pp. 115-132.



# Comparison of Steel, Aluminum and Composite Bonnet Concerning Adult/Child Pedestrian Protection

Abolfazl Masoumi  
School of Mechanical Engineering  
College of Engineering  
University of Tehran  
Tehran 11155/4563, Iran

Amir Najibi  
Automotive Engineering Department  
Iran University of Science & Technology  
Tehran 16864/13114, Iran

E-mail: amasomi@ut.ac.ir

## KEYWORDS

Bonnet, Pedestrian, Composite Material, HIC.

## ABSTRACT

Material selection for automotive closure components is influenced by different factors like cost, weight and structural performance which include both safety and strength. The vehicle bonnet has significant effect on the collision between pedestrian head and automotive when it occurred among the bonnet regions; therefore, in this paper new finite element adult and child headform impactors are developed and validated according to European Commission standard. Then the behavior of three identical bonnets made of steel, aluminum and composite are measured by the adult and child headform impactors. It is shown that the energy absorption of an aluminum bonnet is smaller than of steel one and to keep the aluminum bonnet at the same level of stiffness, it is necessary to increase the thickness. Therefore, the aluminum bonnet needs a larger space between the bonnet and the parts in the engine compartment to reduce head injuries. Also by using composite bonnet the displacement of headform is more than that of steel or aluminum bonnet. Finally, the comparison of the general pedestrian friendliness of steel, aluminum and composite, used as bonnet material and measurement of HIC values in which headform impactor collide bonnet are studied.

## INTRODUCTION

Pedestrians are a high-risk group in vehicle impacts, especially in urban areas. About 7000 pedestrians per year are killed by traffic accidents in the European Union. For this reason, the European Enhanced Vehicle-safety Committee (EEVC) has proposed test procedures to evaluate vehicle aggressiveness against pedestrians (Davies et al. 1997). Proposals for such procedures have focused mainly on subsystem tests because they seem less expensive, simpler and more repeatable than application of full-scale pedestrian dummies (Harris 1989). However, it has not been determined yet to what extent these procedures represent real-world pedestrian accidents. For instance, in these procedures, one standard speed (40 Km/h) of headform impactor is used, although several mathematical modeling and experimental studies have indicated that the head-to-bonnet impact speed varies depending on a vehicle shape and can be appreciably below the initial vehicle speed (European Enhanced Vehicle-safety Committee 1998).

The proposed tests consist of three impact test procedures (Figure 1):

- The impact of headforms into the vehicle bonnet.
- The impact of a legform into the front bumper.
- The impact of an upper legform into the leading edge of the bonnet.

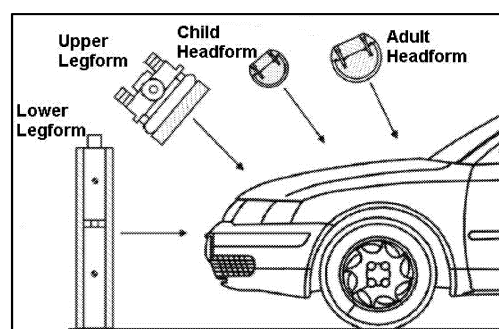


Figure 1: Pedestrian Test Made of 4 Impactors Propelled against the Car Front-End (European Enhanced Vehicle-safety Committee 1998).

Research into pedestrian protection has been carried out since 1960s (Yong and Young 2002). A computer simulation model for pedestrian subsystem impact tests was conducted through the cooperation of TNO and JARI in Japan (Konosu et al. 2000). Using aluminum bonnet structure is suitable for pedestrian protection and weight reduction. Therefore the finite element analysis were performed to prove that the new type of aluminum bonnet with a wave-shaped cross section is able to absorb more energy in a smaller space than the beam-type aluminum bonnet and more importantly, this new wave-shaped aluminum bonnet prevents bad impact with the parts in the engine compartment (Ikeda and Ishitobi 2003). Comparison between two bonnet materials was conducted on a car that is still available on the market with either steel or aluminum bonnet that both have the same design (Schwarz and Bachem 2004). Because the bonnet design was not developed to meet the pedestrian safety requirements, the results compare the application of both steel and aluminum to assess which bonnet material is favorable for pedestrian protection.

In order to provide protection to pedestrians in collision with passenger vehicles, in this research vehicle bonnet is simulated to meet EEVC pedestrian protection legislation. An explicit finite element code, ABAQUS, is used to implement the computation. In this study at first it is focused on developing of a new numerical adult and child headform impactors, to fulfill the requirements of EEVC/WG17 standards then these impactors are used for impact analyses. The results are compared to select the best material among

steel, aluminum and composite for bonnet for pedestrian protection.

### EEVC/WG17 ADULT/CHILD HEADFORMS IMPACTORS AND CERTIFICATION EQUIREMENT

The adult and child headform impactors should be rigid spheres fitted with a vinyl skin (Figure 2). The total impactor masses should be  $4.8 \pm 0.1$  kg for the adult and  $2.5 \pm 0.05$  kg for the child headform impactor, respectively. A triaxial, accelerometer should be mounted in the centre by which the acceleration of the complete adult and child headforms is read.

According to EEVC/WG17 legislation for validation of the headform impactor a dynamic test is performed. Specific impactors are used for each kind of headform. The velocity of impactor when it collides child headform impactor is different from that collides the adult one. Headform impactors should be suspended with the rear face at an angle between  $25^\circ$  and  $90^\circ$  with the horizontal, as shown in Figure 3. Three different points should be chosen between these angles, because each point on the headforms must comply with the regulation of standard.

The certification impactor should have a mass of  $1.0 \pm 0.01$  kg and propelled horizontally at a velocity of  $7.0 \pm 0.1$  m/s into the stationary child headform impactor and at a velocity of  $10.0 \pm 0.1$  m/s into the stationary adult headform impactor as shown in Figure 3. When the child headform impactor is impacted, the peak resultant acceleration which is measured by triaxial accelerometer in the headform should not be less than 405g and not more than 495g and peak resultant acceleration measured by triaxial accelerometer in the headform should not be less than 337.5g and not more than 412.5g for adult headform.

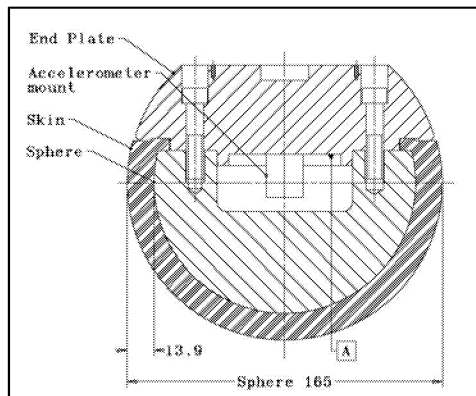


Figure 2: Adult Headform Configuration (European Enhanced Vehicle-safety Committee 1998).

### DEVELOPING NEW MODELES FOR ADULT/CHILD HEADFORM IMPACTOR

Starting with the geometrical information of pedestrian headform impactor, the finite element models were generated.

The aluminum headform core parts are extremely stiff compared to the headform vinyl skin; therefore, we can understand these parts of the numerical model are considered as a rigid body element based on an analytical surface. A reference point is located in the center of gravity of each

headform to measure the time acceleration history which is applied to the headform.

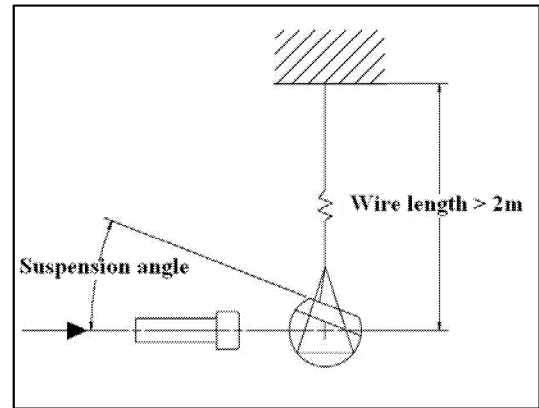


Figure 3: Headform Test Certification (European Enhanced Vehicle-safety Committee 1998).

The vinyl skin model is generated by means of solid elements connected to the aluminum core using a kinematic coupling element between the base of the skin and the reference node of the rigid body.

It is very important to use the accurate material properties for the thick skin that represent the real world properties. Also, applying acceleration on the headform which is caused by the rigid impactor in validation dynamic tests is affected strongly by the skin material properties. Therefore, the behaviour of the material approximated as a hyperelastic model. In ABAQUS all hyperelastic models are based on the assumption of isotropic behavior throughout the deformation history. Hence, the strain energy potential can be formulated as a function of the strain invariants. Hyperelastic materials are described in terms of a strain energy potential, which defines the strain energy stored in the material per unit of reference volume (volume in the initial configuration) as a function of the strain at any desired point in the material. There are several forms of strain energy potentials available in ABAQUS to model approximately incompressible isotropic elastomer which from Mooney-Rivlin form has been chosen. For a homogeneous material, homogeneous deformation modes suffice to characterize the material constants test data from the following deformation modes:

- Uniaxial tension and compression
- Equibiaxial tension and compression
- Planar tension and compression (also known as pure shear)
- Volumetric tension and compression.

The material properties that is used for child skin is different from that of adult one (the skin thickness and headform size are different), therefore for appropriate material properties some efforts are needed to enhance the material properties. Figure 4 shows the Equibiaxial nominal tension and compression of child/adult skins, which are compared with each other. It is obvious that the nominal stress in child skin is more than adult one. Having the headform models ready, the simulation of test procedure is the next step. Due to nonlinearity of the behaviour of material and contact between surfaces, convergence of analyses in software was a time consuming issue.

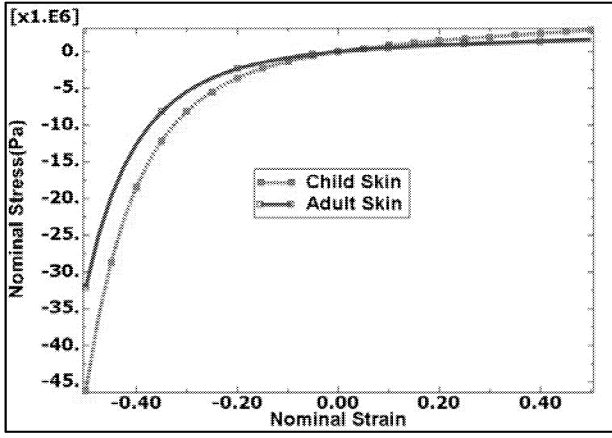


Figure 4: Equibiaxial Nominal Tension and Compression of Child/Adult skins

Consequently in some cases new mesh generation and contact property changes were needed.

In adult and child headforms Certification Tests, the headform collides with the rigid impactor in three different points. The result of analysis must show good agreement between numerical models and EEVC/WG17 standard in all points.

As it was mentioned above, before performing any analyses the adult/child headform impactor must be validated; therefore, three impact analyses that show the certification of impactor in three different points were performed. These points have been chosen in three different angles which show the new headform has the ability of acceptable response in any location of its surface. These suspension angles of headform were 90, 60 and 30 degree, respectively.

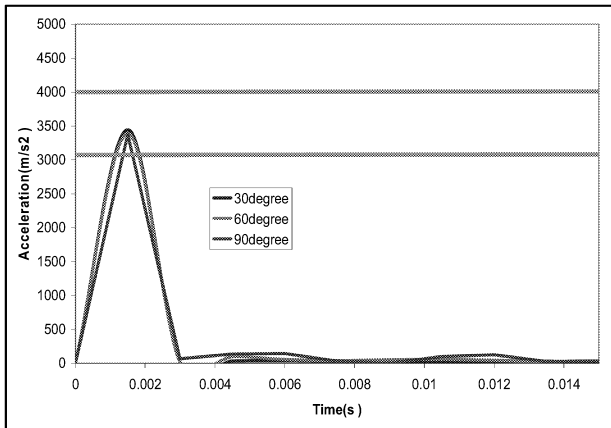


Figure 5: Validation of Numerical Adult Headform.

Figures 5 and 6 show the peak values of acceleration of three impact points are located between the specified boundaries which depicting lower and upper acceleration limitation in center of gravity. Therefore, the new adult/child headform impactor was validated to be used as pedestrian headform in numerical analysis.

As it can be seen in Figure 5, the results at the angles 30 and 60 degree points of impact are the same. Whereas in figure 6 this resemblance is seen at the angles 60 and 90 degree; therefore, the applied acceleration on the headform is depend on the location of impact. A headform impactor is acceptable when at any points of its surface the peak values of acceleration to be located within the acceptable boundary.

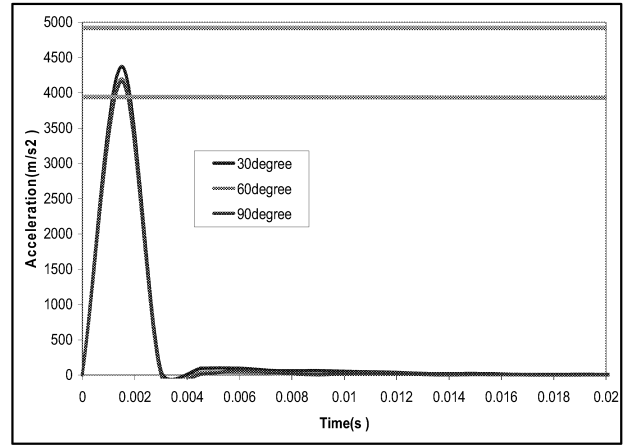


Figure 6: Validation of Numerical Child Headform.

## HEAD INJURY CRITERIA (HIC)

Head injuries are most likely to result in pedestrian fatality following a collision with a motor vehicle. The mechanisms of head injury are complex and are the focus of considerable research. Both linear and angular acceleration are debated as the major determinant of injury to the brain (King et al. 2003). Other factors include lateral versus frontal impact, duration of the acceleration/deceleration phase, brain contusion versus concussion and movement of the brain relative to the skull. The Wayne State University tolerance curve, determined from cadaver testing, describes the injury threshold with regard to the duration of a linear acceleration impulse based on the likelihood of skull fracture following an impact. Points above the curve are thought to indicate a high likelihood of brain injury or death (Gurdjian et al. 1966). The search for a common criterion for evaluating potential head injury during car crash testing led to the Gadd Severity Index (GSI), which is an integration of the Wayne State tolerance curve, with the acceleration component weighted by applying a power of 2.5 (Gadd 1966). This value represents the slope of the Wayne State tolerance curve when plotted logarithmically between 2.5 and 50 milliseconds. Thus the expression for the GSI is:

$$GSI = \int_{t_1}^{t_2} a^{2.5} dt \quad (1)$$

Where  $a$  is the average linear acceleration and  $t_1$  and

$t_2$  are the beginning and end of the time interval, respectively. Gadd proposed a threshold GSI value of 1000 for concussion resulting from frontal impact.

The Head Injury Criterion (HIC) was subsequently developed by (Versace 1971). This focuses the integration time interval on the most injurious part of the impulse. By

defining  $t_1$  and  $t_2$  as the time at which equal levels of acceleration occur either side of an instant of maximum acceleration; HIC can be expressed as (Versace 1971):

$$HIC = (T_2 - T_1) \left\{ \int_{T_1}^{T_2} A_v dt / (T_2 - T_1) \right\}^{2.5} \quad (2)$$

Where  $A_v$  is the resultant head acceleration in g's and  $T_1$  and  $T_2$  are two time instants, in seconds. During the impact  $T_1$  and  $T_2$  are selected so as to provide a maximum HIC value for a given time interval. For contact with hard surfaces a maximum of 15 milliseconds time interval has been employed by the author. Calculation of HIC's for getting the maximum values of HIC during one collision needs some arithmetic iterations which are performed by using the prepared computer programming code.

## FINITE ELEMENT MODELING AND MATERIAL PROPERTIES

When the impactors are used for the simulation of vehicle and pedestrian collision, some predefined zones on the bonnet must be identified. Therefore, based on the simulation of standard, the zones were defined as follows.

The top surface of bonnet was divided into 6 zones for collision of adult headform and 6 other zones for collision of child headform. These zones are between wrap around line 1500 mm to 2100 mm for adult headform collision and between wrap around line 1000 mm to 1500 mm for child headform collision which these are the geometric traces described on the top of bonnet by the end of flexible tape or wire with 1500 and 2100 mm long and with 1000 and 1500 mm long, respectively. Then, the surface between these each two wrap around lines are divided into 6 zones which are the locations that probably the adult and child headform impact there.

The headform impactor for the bonnet top test should be in "free flight" at the moment of impact. The impactor should be released to free flight at such a distance from the vehicle that the test results not influenced by contact of the impactor with the propulsion system during rebound of the impactor. The headform is launched against the bonnet at a speed of 40 km/h in such a way that the direction of impact to be kept 65 degree from centerline of vehicle for adult and 50 degree for child. The rear part of the bonnet, wrap around lines 1500-2100 mm and 1000-1500 mm, is impacted by an adult and child headform, respectively. Figure 7 shows the configurations of adult and child headform in the time of impact to the bonnet.

One of the most important points in this study is to apply the correct material properties which are close to the real world. The material properties of skin on the rigid core of headform impactor have a significant rule when the impactor collide the headform for validation test; therefore, in finite element code, the properties of vinyl skin approximated with Mooney-Rivlin. Aluminum bonnet is simulated such as steel bonnet with an increase in thickness to maintain the same stiffness, but the inner structure for composite bonnet is changed.

The mechanical properties that are used for steel was obtained from sheet metal which used for bonnet production; therefore, the results for steel bonnet also can be used as criteria for the current vehicle regarding to pedestrian friendliness and the thickness of the steel bonnet is 0.7 mm.

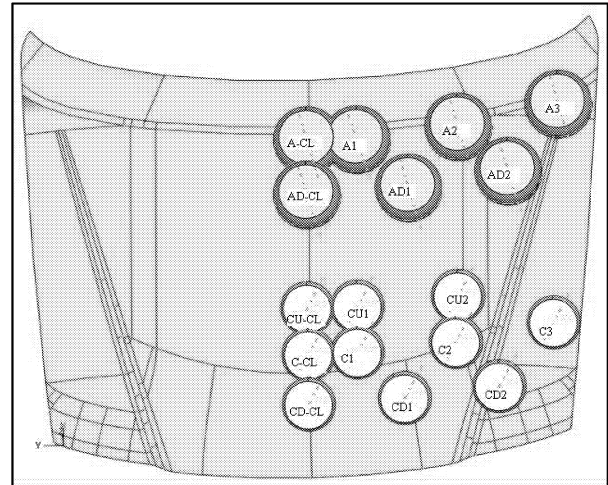


Figure 7: The Configuration of Impacted Headforms.  
(C: child. A: adult)

The mechanical properties of aluminum were modeled as an isotropic elastic-plastic material, characterized by elastoplastic behavior with strain hardening.

The composite bonnet including 5-ply Glass/Epoxy [0/0/0/45/-45] and a steel frame to keep the strength of bonnet was simulated for this study. The elastic properties of the lamina of the Glass/Epoxy were  $\rho=2100\text{Kg/m}^3$ ,  $E_1=55000\text{MPa}$ ,  $E_2=9500\text{MPa}$ ,  $G_{12}=5500\text{MPa}$ ,  $G_{13}=3000\text{MPa}$ ,  $G_{23}=3000\text{MPa}$  and  $\nu_{12}=0.33$ . For the aim of energy absorption in composite bonnet the damage initiation and evaluation properties of composite were applied according to Hashin theory (Hashin 1980).

## RESULTS AND DISCUSSION

Metallic bonnets are usually made of two components, upper and inner; that inner body is used for structural strength and upper body is almost used for getting homogenous style and maintains the aerodynamic of vehicle body. Although, the upper body is less important in the case of strength; it has a vital role in pedestrian bonnet friendliness, because the first contact surface is the upper body of the bonnet. Consequently, the strength, the shape, the curvature and geometric shapes that applied for style of vehicle affects the acceleration of the head when it collides bonnet. The inner body is different for any kind of vehicles according to the shape and the size of the bonnet.

The influence of the bonnet inner body in pedestrian safety is not such as its effect in case of structure strength, but it also affects on intense of impact during collision. The engine and its parts and the other segments that are located below the bonnet are rigid and in many cases the distance between them and bonnet is less than 70 mm. Therefore, after the intense of collision the displacement of head is also important; because the critical values at first but the second impact or rebound because of impact to the rigid parts leads to extreme acceleration. It means that this structure not only must be strengthened in front of static and dynamic forces such as aerodynamic, slam and dent, it also should be able to reduce the intense of impact and avoid extra deformation of bonnet.

In order to calculate the HIC in each zone which adult and child headforms impact, 4 points are chosen for the zones that headform impactors collide bonnet according to Figure 7. The A-CL, A1, A2 and A3 and C-CL, C1, C2 and C3 are the most likely points that adult and child head collide in this specific vehicle. These points have been chosen in a manner that A-CL and C-CL locate on the center line of vehicle and the other points place with equal distances on the bonnet along the specified wrap around line. For example, the distance between A1 and A2 is equal to distance between A2 and A3 along the wrap around line 1800 mm and it also applicable for the points which are labeled for child headform collision. Figure 8 shows the acceleration vs. time of adult headform collision zones in steel bonnet. The acceleration and HIC in point A3 is the highest among the points. Point A3 is located above the hinge and near the fender; therefore, the head faces to a strong structure which means increase in acceleration.

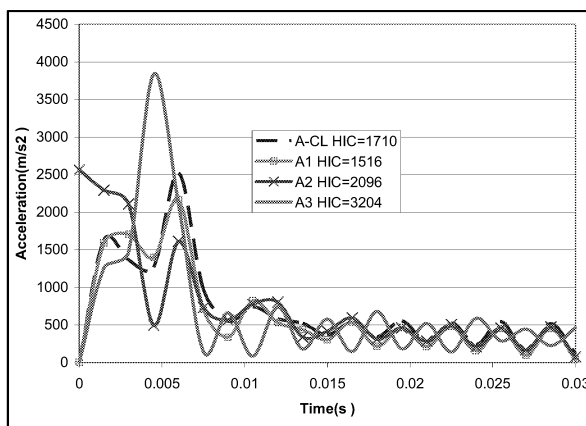


Figure 8: Acceleration in the Adult Collision Points of Steel Bonnet.

The acceleration curves have the peak values which are common in crash analysis so that the peak value and the time that it occurs, depends on the intense of impact. Any increase in the intense leads to the higher peak value. Figure 9 shows the highest peak value of acceleration in the point C3 where child headform collides with the edge of the bonnet near the fender and also on the beam-type inner structure of the bonnet.

When a simplified finite element model is performed as a simulation model that stands for evaluating the real world, HIC values lonely can not be enough for evaluation; therefore,

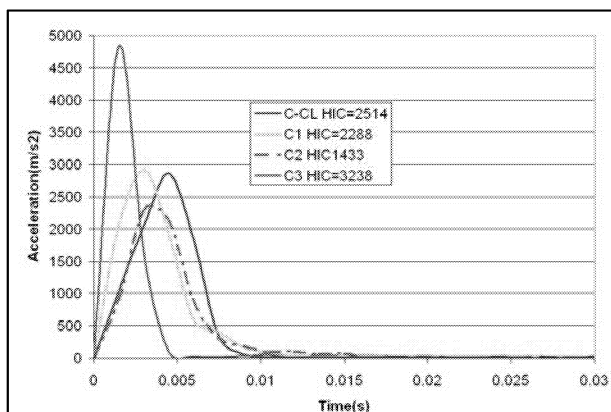


Figure 9: Acceleration in the Child Collision Points of Steel Bonnet.

HIC values and displacement of headform simultaneously should be considered. As, the maximum distance between bonnet and the engine compartment parts could be considered less than 70 mm; therefore, displacements more than 70 mm means an extreme HIC; however, the displacement between 40 and 70 mm is considered as critical zone, too.

Figure 10 illustrates the comparison between all three bonnets in one graph considering acceleration and displacement. It is obvious that the composite bonnet has less acceleration that leads to HIC=1100 in point A-CL. whereas, an enormous bonnet deflection is occurred. It is also shown that the displacement of aluminum bonnet is even exceeds than the critical value of 40 mm.

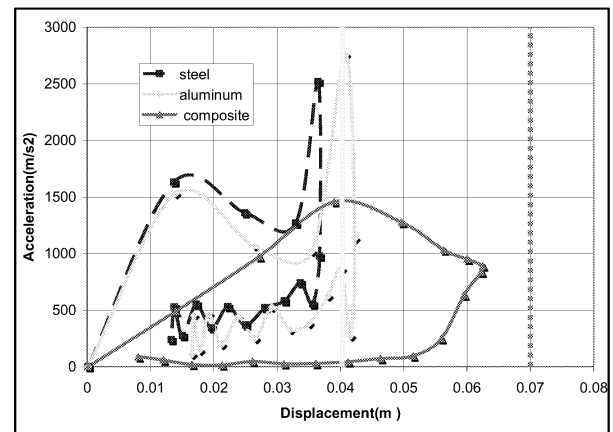


Figure 10: Comparison between Steel, Aluminum and Composite Bonnet in Terms of Acceleration and Displacement at Point A-CL.

For the child impact zone on the center line of the vehicle, the acceleration vs. displacement curves for three kinds of the materials show the similar results which obtained in the adult case. Consequently the composite bonnet has the lowest value of HIC=337 for which displacement of the child headform is more than 40 mm (Figure 11).

According to EEVC/WG17 standard, the maximum value of HIC must not exceed than 1000. Rigid components within the engine compartment will increase the HIC, if the displacement increases more than certain amount. Therefore, the deflection must kept minimum while having maximum absorbed energy.

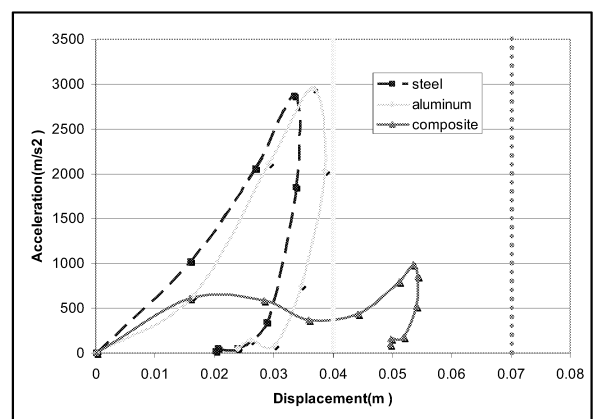


Figure 11: Comparison between Steel, Aluminum and Composite Bonnet in Terms of Acceleration and Displacement at Point C-CL.

Results obtained from the analysis show that the current steel bonnet is not able to fulfill the pedestrian safety regulations. By using the aluminum bonnet with the same stiffness of the steel one, the values of the HIC are improved but they are not below the 1000, while the displacements of the headform were increased. Although the HIC values calculated in composite bonnet fulfill the regulations except in the end points of wrap around lines, the displacements of headform impactors exceed than 40 mm.

The outcome of this study on different bonnet materials can be demonstrated in two plots which cover the displacement and HIC for all of points together.

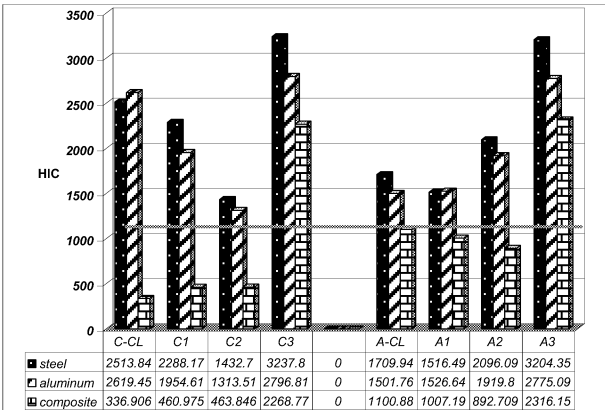


Figure 12: Comparison of HIC for Different Type of Bonnets.

Figure 12 shows the comparison of HIC values for different type of bonnets. Almost in all points, HIC for aluminum bonnet is lower than steel bonnet, but HICs exceed than 1000. HIC for composite bonnet is lower than 1000 except in points A3 and C3. Since the bonnet hinge structure are located under this points, sudden increase is inevitable.

Figure 13 illustrates the maximum displacement value of headform. In all points this value in the aluminum and composite bonnet is more than steel bonnet, however in some points in composite bonnet, it exceeds from 70 mm.

The benefit of simplified finite element model of the bonnet is that it makes possible to investigate the influence of bonnet itself on head injury which is caused by impact into bonnet and it helps the automotive designer to study the relation between structural strength and reliability of bonnet in order to absorb energy absorption and keeping the head injury less as possible.

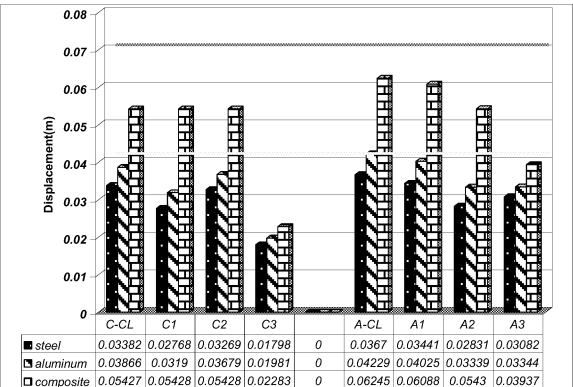


Figure 13: Maximum Displacement Value of Headform.

## CONCLUSION

The head injury risk for adult and child pedestrians who impacted to bonnet was examined by numerical simulation. Because of using new material in automotive industry for light weighting to decrease fuel consumption, different bonnet materials were investigated in case of pedestrian safety. Impact characteristic and HIC value and deformation were the main criteria for assessing of bonnet pedestrian friendliness and the results were compared in each point of impact and each material lonely or with each other. Two bonnet materials and original one were examined by adult and child headforms impacting the bonnet at eight different locations under conditions described in a proposed directive. The simulations showed that softer bonnet structure causes reduction in bonnet HIC and increase of deformation. Using aluminum bonnet causes reduction of HIC values and increase of head displacement in comparison to steel bonnet. The lowest HIC value and the highest head displacement are derived in the case of the composite bonnet.

## REFERENCES

Davies, R. G.; L. C. Clemo; and D. G. C. Bacon. 1997. *Study of Research into Pedestrian Protection MIRA, European Commission.*

European Enhanced Vehicle-safety Committee. *Improved Test Methods to Evaluate Pedestrian Protection Afforded by Passenger Cars.* 1998, *EEVC Working Group 17 Report.*

Gadd, C. M. 1966. "Use of a Weighted Impulse Criterion for Estimating Injury Hazard." Proc. In 10th Stapp Car Crash Conference, Society of Automotive Engineers, New York, pp 164-174.

Gurdjian, E. S.; V. L. Roberts; and L. M. 1966. "Thomas Tolerance Curves of Acceleration and Intracranial Pressure and Protective Index in Experimental Head Injury." *Journal of Trauma*, 6(5): 600-604.

Harris, J. 1989. "A Study of Test Methods to Evaluate Pedestrian Protection for Cars." Proc. 12th International Conference on the Enhanced Safety of Vehicles, pp. 1217-1225.

Hashin, Z.1980. "Failure Criteria for Unidirectional Fiber Composites." *Journal of Applied Mechanics*, vol. 47, pp. 329–334.

Ikeda, K. and H. Ishitobi. 2003. "Development of Aluminum Hood Structure for Pedestrian Protection." Proc. SAE International, Document number: 2003-01-2793(Oct).

King, A. I.; K. H. Yang; L. Zhang; W. Hardy; and D. C.Viano. 2003. "Is Head Injury Caused by Linear or Angular Acceleration?" IRCOBI Conference, Lisbon (Portugal) (Sep).

Konosu, A.; H. Ishikawa; and R. Kant. 2000. "Development of Computer Simulation Models for Pedestrian Subsystem Impact Test." *JSAE Review* 21.

Schwarz, D. and H. Bachem. 2004. "Comparison of Steel and Aluminum Hood with Same Design in View of Pedestrian Head Impact." Proc. SAE 2004 World Congress & Exhibition, Detroit, MI, USA, Session: Pedestrian Protection (Part 1 of 2) (Mar).

Versace, J. 1971. "A Review of the Severity Index." Proc. In 15th Stapp Car Crash Conference, Society of Automotive Engineers, New York, pp 771-796.

Yong, H. and w.Young. 2002. "Optimization of Bumper Structure for Pedestrian Lower Leg Impact." Proc. SAE 2002 World Congress & Exhibition, Detroit, MI, USA, Session: Biomechanics (Part A&B) (Mar).

# **DATAMINING**





# Information Extraction from Solution Set of Simulation-based Multi-objective Optimisation using Data Mining

Catarina Dudas and Amos Ng  
Centre for Intelligent Automation  
University of Skövde  
Box 408, 541 28 Skövde  
Sweden

Henrik Boström  
Informatics Research Centre  
University of Skövde  
Box 408, 541 28 Skövde  
Sweden

## KEYWORDS

Output analysis, Data mining, Information extraction.

## ABSTRACT

In this work, we investigate ways of extracting information from simulations, in particular from simulation-based multi-objective optimisation, in order to acquire information that can support human decision makers that aim for optimising manufacturing processes.

Applying data mining for analyzing data generated using simulation is a fairly unexplored area. With the observation that the obtained solutions from a simulation-based multi-objective optimisation are all optimal (or close to the optimal Pareto front) so that they are bound to follow and exhibit certain relationships among variables vis-à-vis objectives, it is argued that using data mining to discover these relationships could be a promising procedure. The aim of this paper is to provide the empirical results from two simulation case studies to support such a hypothesis.

## RELATED WORK

Data mining is a technique which has been used in both private and public sectors and clearly with different objectives. Companies within banking, insurance and retailing use data mining to reduce cost, detect frauds and to advertise in more effective ways. Homeland security is yet another application area of growing interest, in which data mining also has been used.

One of the first uses of artificial intelligence in manufacturing applications was accomplished in the 1980's according to (Kusiak 2006). In the beginning of the 1990's, the use of data mining techniques was introduced for production, something which has been growing since then. A comprehensive review of papers considering data mining applications within manufacturing is presented in (Kusiak 2006). Manufacturing operations, fault detection, design engineering and decision support systems have been in focus as

research topics, but there is still an enormous potential for further research in other application areas, such as maintenance, layout design, resource planning and shop floor control.

The combination of multi-objective optimisation solutions and data mining techniques is a fairly unexplored area. Therefore the literature reveals quite few reports. (Chiba et al. 2006) and (Jeong et al. 2005) apply the use of analysis of variance (ANOVA) and Self-Organizing Maps (SOM) in the design process for aerodynamic optimisations problems. It is found that the ANOVA obtains the quantitative correlation between objective function and design variable. The result from SOM is qualitative and subjective and can be used for understanding of the design variable influence. Furthermore, SOM explains the trade-off between the competing objectives.

## DATA MINING

Data mining is an automated or semi-automated technique used to discover and interpret hidden relationships, patterns or trends in large data sources. A blend of concepts and algorithms from machine learning, statistics, artificial intelligence, and data management are borrowed to the field of data mining.

Figure 1 shows the data mining process as an iterative procedure (Fayyad et al. 1996). The process can be divided into three parts: selection/pre-processing, mining and presentation.

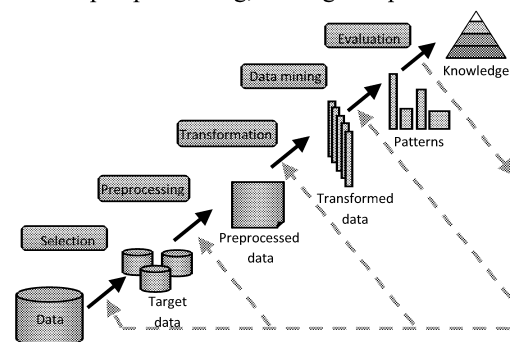


Figure 1: Data Mining Process

The first step involves gathering, organising and cleaning data before it can be used. The mining process involves choosing appropriate method(s) to be used for searching patterns in data. The final step is about how to present the results of the prior processes in a suitable way. After evaluation of the presented results, the entire process, or parts of it, may be re-iterated.

Data mining is a rapidly expanding field with growing interests and importance. Simulation based optimisation is certainly an application area where the use of this technology can provide a significant advantage.

### **Predictive and Descriptive Data Mining**

Data mining techniques have become standard tools to develop predictive and descriptive models in situations where one wants to exploit data collected from earlier observations in order to optimise future decision making (Witten and Miller 2005). In the case of predictive modelling, one typically tries to estimate the expected value of a particular variable (called the dependent variable), given the values of a set of other (independent) variables. In the case of a nominal dependent variable (i.e., the possible values are not given any particular order), the prediction task is usually referred to as classification, while the corresponding task when having a numerical dependent variable is referred to as regression. One usually wants the model to be as correct as possible when evaluated on independent test data, and several suggestions for how to measure this have been proposed. For classification, such measures include accuracy, i.e., the percentage of correctly classified test examples, and the area under the ROC curve (AUC), i.e., the probability that a test example belonging to a class is ranked as being more likely belonging to the class than a test example not belonging to the class (Provost et al. 1998). Besides the ability to make correct predictions, one is also often interested in obtaining a comprehensible (descriptive) model, so that the reasons behind a particular classification can be understood, and also that one may gain insights into what factors are important for the classification in general. Examples of such comprehensible models are decision trees and rules, e.g. (Quinlan 1993), while examples of models not belonging to this group, often called black-box, or opaque, models, include artificial neural networks and support vector machines; see e.g. (Hastie et al. 2001).

### **Decision Trees and Ensembles**

Techniques for generating decision trees are perhaps among the most well-known methods for predictive data mining. Early systems for generating decision trees include CART (Breiman et al. 1984) and ID3 (Quinlan 1986), the latter

being followed by the later versions C4.5 (Quinlan 1993) and C5.0 (Quinlan 1997). The basic strategy that is employed when generating decision trees is called recursive partitioning, or divide-and-conquer. It works by partitioning the examples by choosing a set of conditions on an independent variable (e.g., the variable has a value less than a particular threshold, or a value greater or equal to this threshold), and the choice is usually made such that the error on the dependent variable is minimised within each group. The process continues recursively with each subgroup until certain conditions are met, such as that the error cannot be further reduced (e.g., all examples in a group belong to the same class). The resulting decision tree is a graph that contains one node for each subgroup considered, where the node corresponding to the initial set of examples is called the root, and for all nodes there is an edge to each subgroup generated from it, labelled with the chosen condition for that subgroup.

Decision trees have many attractive features, such as allowing for human interpretation and hence making it possible for a decision maker to gain insights into what factors are important for particular classifications. However, recent research has shown that significant improvements in predictive performance can be achieved by generating large sets of models, or ensembles, which are used to form a collective vote on the value for the dependent variable (Bauer and Kohavi 1999). It can be shown that as long as each single model performs better than random, and the models make independent errors, the resulting error can in theory be made arbitrarily small by increasing the size of the ensemble. However, in practice it is not possible to completely fulfil these conditions, but several methods have been proposed that try to approximate independence, and still maintain sufficient accuracy of each model, by introducing randomness in the process of selecting examples and conditions when building each individual model. One popular method of introducing randomness in the selection of training examples is bootstrap aggregating, or bagging, as introduced by (Breiman 1996). It works by randomly selecting  $n$  examples with replacement from the initial set of  $n$  examples, leading to that some examples are duplicated while others are excluded. Typically a large number (at least 25-50) of such sets are sampled from which each individual model is generated. Yet another popular method of introducing randomness when generating decision trees is to consider only a small subset of all available independent variables at each node when forming the tree. When combined with bagging, the resulting models are referred to as random forests (Breiman 2001), and these are widely considered to be among the most competitive and robust of current methods for predictive data mining. The

drawback of ensemble models are however that they can no longer be easily interpreted and hence provide less guidance into how classifications are made.

The Rule Discovery System™ (RDS) addresses this problem by providing some insight into what factors are of importance in an ensemble of decision trees by presenting the variable importance of each independent variable, i.e., how much the variable, relative to all other variables, contributes to reducing the squared error of the dependent variable.

## TWO DATA MINING APPLICATIONS

The use of data mining in manufacturing applications can have different aims and purposes. In this paper a specific approach is presented: data mining for identifying patterns in data sets generated by multi-objective optimisation. The first data set handles a buffer allocation problem and the second is for identifying dispatching rules setting in a production line. The data mining software used is RDS for both studies.

### Buffer Allocation

The simulation model used for this study was developed by a simulation optimisation system called FACTS Analyser (Ng et al. 2007). The model of the production line consists of 5 stations and 5 buffers and is controlled by a Critical Work-In-Process (CWIP) strategy. A Pareto Front was found after 40 generations of multi-objective optimisation (MOO) with MA-NSGA-II.

There is a constraint for each individual buffer size ( $0 \leq \text{Buffer Size} \leq 50$ ) and no constraint on the total buffer size. The CWIP level varies between [0-100] in terms of percent. For the MOO the objectives are to minimise lead time (LT) and maximise throughput (TP).

The data set in RDS™ consists of 6 input variables (Buffer capacity 1-5 and CWIP) and two output variables, LT and TP. For each generation there are 200 observations and in this study the set for generation 39 (G39) is explored. As validation method 10-fold cross validation is used.

Since the data mining software make predictions with one output variable at time, LT and TP have to be studied separately. The algorithms used in these experiments are trees and ensembles of trees. Trees are useful for interpretation of the important input variables and the benefits of an ensemble of trees are that you receive a model with higher correlation and lower error rate.

### Results for the buffer allocation case

The optimal solution data set for G39 is as close as possible to an optimal Pareto front. In order to find the key information in the data set the importance

score and the decision tree has to be examined. The importance score plot enlightens the variables that are most informative, i.e. contribute to the model primarily. On the other hand, the decision tree can be used to illuminate more detailed information about the settings of the variables.

The importance score for G39 can be found in Figure 2. It reveals that for both LT and TP the most informative variables are buffer capacity 1 and 2 (B1, B2), where B1 is dominating.

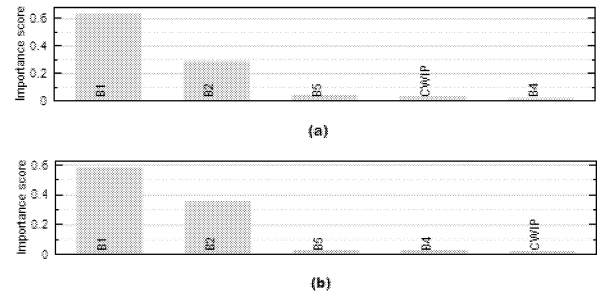


Figure 2: Importance Score for G39 (a) LT, (b) TP

Figure 3 and 4 show the decision tree with TP and LT as output variable where the detailed information can be found.

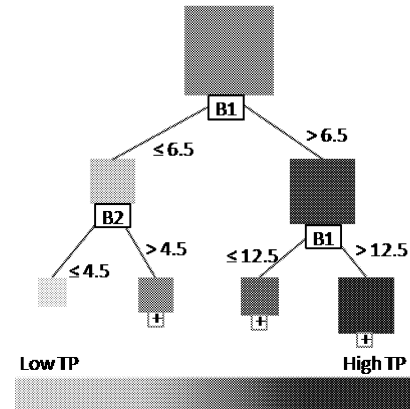


Figure 3: Decision Tree with TP as Output Variable

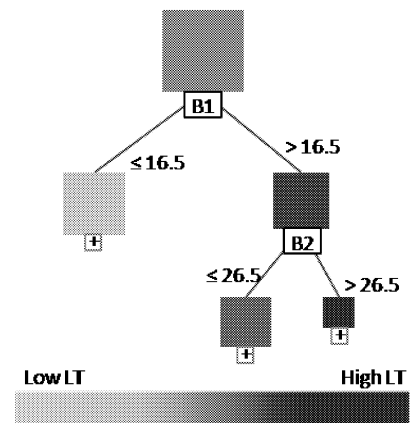


Figure 4: Decision Tree with LT as Output Variable

To get an acceptable TP in this production line the buffer capacity for B1 must be greater than 6. In order to receive as high TP as possible the capacity in this buffer should be greater than 12.

It is desirable to have as low LT as possible and Figure 4 reveals the settings in order to accomplish this. Notable is that the buffer capacity for buffer 1 is the main divider and if the capacity is less than 16 the lead time will be satisfying.

The capacity of the first buffer has the most impact on the throughput and the lead time in this study. The conclusion of this is that the station after the first buffer is the bottleneck in this production line. Letting the capacity for buffer 1 vary between 12 and 16 will result in low lead time and simultaneously provide the highest throughput.

### Dispatching rules in a production line

The aim of this experiment is to understand how dispatching rules affect the outcome of a production line. The result of a multi-objective optimisation study is used to discover patterns in dispatching rule settings in order to maximise throughput (TP) and minimise all delayed products, i.e. the total tardiness (TT).

Input to the data mining experiment is output from a Discrete Event Simulation (DES) model. This simulation model is a representation of the H-factory at Volvo Cars in Skövde. The H-factory is committed to camshaft processing and 15 variants are handled on the production line. The H-factory consists of thirteen different groups of operations with one to seven machines in each group. All machines within a group of operations have the same capability and in front of every group of machines is a buffer.

When a product enters to a buffer it checks if a machine is free, if so the product is directly moved there. But, if there is not a machine available then the product is placed on a free spot in the buffer. The buffer is checked every time a machine has finished a product. If there is only one product there; move that one to the machine. The dispatching rules are considered each time there is more than one product in the buffer. The product to pick is dependent on the current dispatching rule assigned to that specific buffer.

There are eight different dispatching rules: shortest processing time (SPT), longest processing time (LPT), earliest due date (EDD), total working remaining (TWR), least work remaining (LWKR), most work remaining (MWKR), minimum slack time (MST) and operation due date (OPNDD).

There are 13 groups of operations and 8 dispatching rules. Due to the great number of different dispatching rule settings, simulation based optimisation (SBO) is used to generate an optimal configuration of the production line. The optimisation objective parameters are TP and TT.

The output of the simulation based optimisation is the dispatching rules used for each operation with its resulting TP and TT. The number of different settings is  $8^{13}$ , approximately  $5.5 \cdot 10^{11}$ . The optimisation generates about 400 solutions which are all non-dominated and on the Pareto front. These are used in the post processing step for investigation of better understanding of the solutions.

### Results for the dispatching rules case

The data set with 13 input variables (applied dispatching rule for each buffer) and the output variables are TP and TT was used to generate decision trees. These are simple to interpret and they have therefore high usefulness.

It can easily be seen in Figure 5 that the bottlenecks, i.e. the most important variables, are op20 and op90. This is true for both TP and TT as output variables. In order to generate a small and more interpretable model all other input variables are excluded and new models are built.

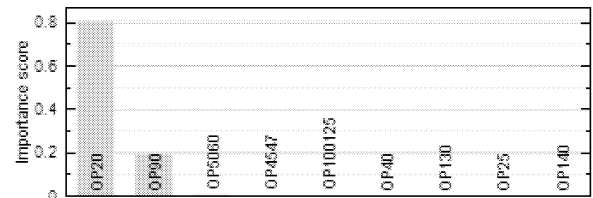


Figure 5: Important Variables in the H-factory.

The decision tree from which detailed information can be found is shown in Figure 6.

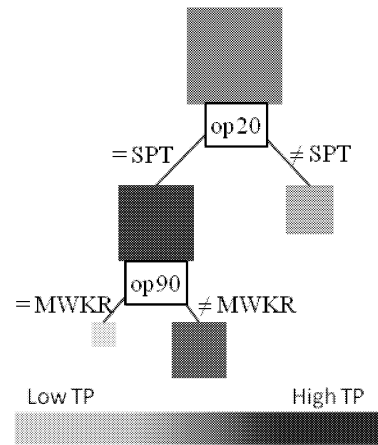


Figure 6: The Tree Structure for Maximising TP.

The information which can be extracted from the tree structure is that for a higher average TP use dispatching rule SPT in op20 and use any of the others but MWKR for the buffer before op90.

The experiment for the TT is performed in a similar way. As an initial study all input variables are used to identify the variables with most importance and the most important variables were identified to be

op20 and op90. Its tree structure can be seen in Figure 7.

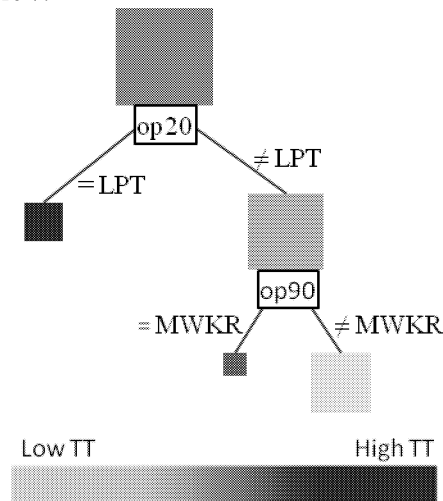


Figure 7: The Tree Structure for Minimising TT.

In contrast to the TP experiment the TT case focus on discovering the settings for a low prediction values. Letting the most severe bottleneck have a dispatching rule that is not LPT and the second one having any but MWKR will result in a low TT. Combining the results will lead the decision maker to draw the conclusions that the dispatching rule for the most influencing buffer before op20 should be shortest processing time (SPT) and for op90 most work remaining (MWKR) should be chosen. The 11 other buffers are not significantly influencing the total tardiness or throughput.

## SUMMARY AND FUTURE WORK

In this work, we have shown how information can be extracted from simulation data by means of data mining, providing support for a human operator aiming for optimising manufacturing processes. For example, the operator may learn how various process parameters affect different optimisation criteria.

One main question for future research concerns how to most effectively exploit the information that has been acquired by analysing simulation data, with other sources of information in actual decision situations. Another question for future work is to determine whether data mining can outperform various experimental design methods. While these methods, ranging from orthogonal arrays to stratified Latin Hypercube design, can be used to explore the input variables space uniformly and effectively for generating the required data sets, it is questioned whether these techniques are sufficient enough to unravel the relationships between input decision variables and the output parameters, which is an important purpose of many data mining processes.

## ACKNOWLEDGEMENTS

This work was supported by the Information Fusion Research Program ([www.infofusion.se](http://www.infofusion.se)) at the University of Skövde, Sweden, in partnership with the Swedish Knowledge Foundation under grant 2003/0104.

## REFERENCES

- Bauer E. and Kohavi R., 1999. An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants, *Machine Learning*, Vol. 36, Issue 1-2, 105-139
- Breiman L., 1996. Bagging Predictors, *Machine Learning*, Vol. 24, Issue 2, 123-140
- Breiman L., 2001. Random Forests, *Machine Learning*, Vol. 45, Issue 1, 5-32
- Breiman L., Friedman J.H., Olshen R.A. and Stone C.J., 1984. *Classification and Regression Trees*, Wadsworth, Belmont
- Chiba, K., Jeong S., and Yamamoto K., 2006. "Efficient Data Mining for Multi-Objective Design Exploration regarding Aerospace Vehicle", *ICNPAA-2006: Mathematical Problems in Engineering and Aerospace Sciences*
- Fayyad, U. M., Piatetsky-Shapiro, G. and Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases, *AI Magazine*, Vol. 17, Issue 3, 37-54.
- Hastie T., Tibshirani R. and Friedman J., 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, London.
- Jeong, S., Chiba, K., and Obayashi, S., 2005. "Data Mining for Aerodynamic Design Space," *Journal of Aerospace Computing, Information, and Communication*, Vol. 2, No. 11, 452-469.
- Kusiak, A. 2006. Data Mining in Manufacturing: A Review, *Journal of Manufacturing Science and Engineering*, Vol. 128, Issue 4, 969-976.
- Ng, A., 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.
- Provost F., Fawcett T. and Kohavi R., 1998. *The case against accuracy estimation for comparing induction algorithms*, Proc. Fifteenth Intl. Conf. Machine Learning, 445-553
- Quinlan J.R., 1986. Induction of decision trees, *Machine Learning*, Vol. 1, Issue 1, 81-106
- Quinlan J.R., 1993. *C4.5: Programs for Machine Learning*, Morgan Kaufman, San Francisco
- Quinlan J.R., 1997. Data Mining Tools See5 and C5.0, <http://www.rulequest.com/see5-info.html> (accessed March. 28, 2009)
- Witten I. and Frank E., 2005. *Data Mining: Practical Machine Learning Tools and Techniques* (Second Edition), Morgan Kaufmann Publisher, San Francisco.

# USING ANT COLONY IN OPTIMIZING ASSOCIATIVE CLASSIFICATION

Kifaya Qaddoum  
Philadelphia University  
Amman, Jordan  
E-mail: [sweetkq@hotmail.com](mailto:sweetkq@hotmail.com).

## KEYWORDS

Ant Colony, Associative Classification, Data Mining, Classification Rules

## ABSTRACT

As we can see from research, ant algorithms have sparked enormous interest within the search/optimization community. In the data mining search space, associative classification [3] (AC) approaches have proven to be strongly able to extract more accurate classifiers than traditional classification techniques. In this paper the two are brought together, namely an investigation of a variant of ant colony algorithms linked to associative classification mining. In the proposed algorithm, each class is labeled and assigned to represent the consequential part of the rules within a processor. The ants groups are attached to the main processor to search for the antecedent part of the rules. The ants, depending on the pheromone, and heuristic information, select the values of the attributes according to the importance of an attribute to the class. The proposed algorithm is expected to discover classification rules with magnificent enhancement on accuracy and less redundancy than comparing methods.

## INTRODUCTION

Classification is considered to be an important task of data mining, aimed at finding a small set of rules from training data sets with predetermined targets [6]. Many classification algorithms are being used to extract relevant relationships in data, such as decision trees [15], neural networks [13], statistical and rough sets [17] etc. Although these classification techniques are algorithmically strong, they require significant expertise to work effectively and normally do not provide lucid rules [6].

The problem of classification becomes NP-hard when the number of possible different combinations of parameters is so high and as such, many researchers have devoted their attention to nature-based approaches to find an intelligible solution for that. Especially, there have been numerous attempts to apply Genetic Algorithms (GAs) [7, 8] in Data

Mining to accomplish classification tasks and achieve better results [5]. In addition, Ant Colony Optimization (ACO) techniques [4], which have emerged recently as the new meta-heuristic derived from nature have attracted

NP-hard combinatorial optimization problems, on the other, however, the use of the algorithm for mining classification rule in the context of data-mining, is a research area still in its infancy.

Building effective classification systems is one of the main tasks of data mining and machine learning. Past researches have produced many techniques (e.g. Decision Tree [12], Naive-Bayesian [10], Support Vector Machines [5], Neural Networks [13], and Statistical approaches [12], [8]. In recent years a new classification approach that integrates classification with association rule mining called associative classification arises, which is considered a promising technique that produces highly accurate classifiers. AC in many experiments proved to give more accurate rules, and less error rates related to the produced rules, compared with other leading techniques in data mining classification. [please be consistent with how you represent data mining, see, e.g., high lighted above and here]

The ACO algorithm [4] has been inspired by colonies of real ants, which deposit a chemical substance called pheromone on the ground. This substance influences the choice they make. The larger the amount of pheromone is on a particular path, the larger the probability is that an ant selects the path. Artificial ants in ACO algorithms behave in similar way. The ACO algorithm is basically a multi-agent system where low level interactions between single agents (i.e., artificial ants) result in a complex behavior of the whole ant colony. ACO algorithms have been successfully applied to several NP-hard combinatorial optimization problems, thereby achieving better results [3].

Referring [18] are the first to propose the Ant Colony Optimization (ACO) [14] algorithm for discovering classification rules, using the system Ant-Miner system. An entropy-based heuristic function is used in their method. In [12], Liu et al, a modified version of Ant-Miner (i.e., Ant-Miner2) was presented, where the core heuristic value was based on a simple density estimation. In addition, Liu et al [11] have further introduced another ant-based algorithm, which uses a different pheromone updating strategy and state transition rule. Recently Ani, has investigated the Ant Colony Miner for the problem of

feature subset selection [1]. According to [17] have applied a new rule pruning procedure for the ant colony optimization classification algorithm.[2]. The objective of this paper is to investigate the capability of ACO algorithms to discover classification rules with higher predictive accuracy and much smaller rule lists by introducing threshold value during the rule construction. The advantage of the introduced threshold value is to reduce the rule length and enhance the simplicity by controlling the addition of the terms with less information gain.

## AN EXACT DEFINITION OF ASSOCIATIVE CLASSIFICATION

Classification rule mining and association rule mining are two important data mining techniques [4]. The target of mining is not pre-determined for association rule mining, while for classification rule mining there is one and only one pre-determined target, i.e., the class.

Associative classification [3] proposes the integration of association rule mining and classification. Association rules are unsupervised learning describing the co-occurrence among data items in a large amount of collected data, [1]. whereas, associative classification is a supervised task that predicts the class label of test cases.

This section introduces the associative classification technique. A training dataset  $R$  is characterized by a schema  $(A_1, \dots, A_k, c)$ , where  $(A_1, \dots, A_k)$  are  $k$  distinct attributes and  $c$  is a class attribute. Each tuple in  $R$  is a data object, which is associated to a unique identifier called *tid*. The attributes may have either a categorical or a continuous domain. For categorical attributes, all values in the domain are mapped to consecutive positive integers. For continuous attributes, the value range is discretized into intervals, which are mapped onto consecutive positive integers. In this way, all attributes are treated uniformly.

Each data object in  $R$  can be described as a collection of pairs (attribute, integer value), with a class label (a value belonging to the domain of the class attribute  $c$ ). Each pair (attribute, integer value) is called an *item* in the remaining part of the paper.

In associative classification, classification rules are used to model the most relevant properties characterizing classes of data, and to predict the class label for unknown (unlabeled) data. A classifier is a function from  $2^I$  to  $C$  that allows the assignment of a class label to a data object. A classifier, able to predict the class label for data objects with high accuracy, is generated from a collection of transactions (i.e., data objects with a label, also called training dataset). In associative classification, the classifier is generated by selecting the most appropriate set of association rules. A rule  $r : X \Rightarrow c$  classifies or matches a data object  $d$  when

$X \subseteq d$ . In this case, rule  $r$  assigns class label  $c$  to data object  $d$ .

Several measures have been proposed to quantify the “interestingness” or the quality of an association (and a classification) rule. Frequently used quality indices are support and confidence [1]. In our setting, these measures correspond to the support and confidence of the *l-itemset* encoding the classification rule. For an arbitrary classification rule  $Xc$ , the support  $\text{sup}(Xc)$  is the fraction of transactions in  $D$  which contain  $X$  and are labeled by class label  $c$ . For the rule antecedent  $X$ , the support  $\text{sup}(X)$  is the fraction of transactions in  $D$  which contain  $X$ .

## PROBLEM DEFINITION

The main objective of data mining is to find interesting/useful knowledge for the user, as Rules are an important form of knowledge; some existing research has produced many algorithms for rule mining. These techniques use the whole dataset to mine rules and then filter and/or rank the discovered rules in various ways to help the user identify useful ones.

There are many potential application areas for association rule technology which include catalog design, store layout, customer segmentation, telecommunication alarm diagnosis, and so on.

The data mining task is to generate all association rules in the database, which have a support greater than *min sup*, i.e., the rules are frequent, and also have a confidence greater than *min conf*, i.e., the rules are strong. Here we are interested in rules with a specific item, called the *class*, as a consequent, i.e., we mine rules of the form  $A \Rightarrow c_i$  where  $c_i$  is a class attribute ( $1 \leq i \leq k$ ).

This task can be broken into two steps:

1. Find all frequent *itemsets* having minimum support for at least one class  $c_i$ . The search space for enumeration of all frequent *itemsets* is  $2^m$ , which is exponential in  $m$ , the number of items.

2. Generate strong rules having minimum confidence, from the frequent *itemsets*. We generate and test the confidence of all rules of the form  $X \Rightarrow c_i$ , where  $X$  is frequent. For example, consider the sales database of a bookstore [22] shown in Figure 1, where the objects represent customers and the attributes represent books. The discovered patterns are the set of books most frequently bought together by the customers. An example could be that, "40 percent of the people who buy Jane Austen's *Pride and Prejudice* also buy *Sense and Sensibility*". The store can use this knowledge for promotions, shelf placement, etc.

There are five different items (names of authors the bookstore carries), i.e.,  $I = \{A, C, D, T, W\}$ , and the database consists of six customers who bought books by these authors. Figure1 [4] shows all the frequent *itemsets* that are contained in at least three customer transactions, i.e., *min sup* = 50 percent.

Table 1

Figure 1 Mining Frequent Itemset

| TID | Items |
|-----|-------|
| 1   | ACTW  |
| 2   | CDW   |
| 3   | ACTW  |
| 4   | ACDW  |
| 5   | ACDTW |
| 6   | CDT   |

| Support | Itemset                           |
|---------|-----------------------------------|
| 100 %   | C                                 |
| 83 %    | W, CW                             |
| 67 %    | A, D, T, AC, AW<br>CD, CT, ACW    |
| 50 %    | AT, DW, TW, ACT,<br>ATW, CDW, CTW |

There is one main difference between classification and ARM which is the outcome of the rules generated. In case of classification, the outcome is pre-determined, i.e. the class attribute. Classification also tends to discover only a small set of rules in order to build a model (classifier), which is then used to forecast the class labels of previously unseen data sets as accurately as possible. On the other hand, the main goal of ARM is to discover correlations between items in a transactional data set. In other words, the search for rules in the classification is directed to the class attribute, whereas, the search for association rules are not directed to any specific attribute.

Associative Classification (AC) is a branch in data mining that combines classification and association rule mining. In other words, it utilizes association rule discovery methods in classification data sets. Many AC algorithms have been proposed in the last few years, i.e. [4], [14], which have produced highly competitive results with respect to classification accuracy if compared with that of traditional classification approaches such as decision trees [12], and probabilistic ones. [7].

## ASSOCIATIVE CLASSIFICATION PROBLEM AND RELATED WORKS

According to [14] the AC problem was defined as: “Let a training data set  $T$  have  $m$  distinct attributes  $A_1, A_2, \dots, A_m$  and  $C$  be a list of class labels”. The number of rows in  $T$  is denoted  $|T|$ . Attributes could be categorical (meaning they take a value from a finite set of possible values) or continuous (where they are real or integer). In the case of categorical attributes, all possible values are mapped onto a set of positive integers. For continuous attributes, a discretisation method is first used to transform these attributes into categorical ones.

**Definition 1:** An *item* can be described as an attribute name  $A_i$  and its value  $a_i$ , denoted  $(A_i, a_i)$ .

**Definition 2:** The  $j$ th row or a *training object* in  $T$  can be described as a list of items  $(A_{j1}, a_{j1}), \dots, (A_{jk}, a_{jk})$ , plus a class denoted by  $c_j$ .

**Definition 3:** An *itemset* can be described as a set of disjoint attribute values contained in a training object, denoted  $\langle (A_{i1}, a_{i1}), \dots, (A_{ik}, a_{ik}) \rangle$ .

**Definition 4:** A *ruleitem*  $r$  is of the form  $\langle \text{cond}, c \rangle$ , where condition  $\text{cond}$  is an itemset and  $c \in C$  is a class.

**Definition 5:** The actual occurrence (*actoccr*) of a *ruleitem*  $r$  in  $T$  is the number of rows in  $T$  that match  $r$ 's itemset.

**Definition 6:** The support count (*suppcount*) of *ruleitem*  $r = \langle \text{cond}, c \rangle$  is the number of rows in  $T$  that matches  $r$ 's itemset, and belongs to a class  $c$ .

**Definition 7:** The occurrence (*occitm*) of an itemset  $I$  in  $T$  is the number of rows in  $T$  that match  $I$ .

**Definition 8:** An itemset  $i$  passes the minimum support (*minsupp*) threshold if  $(\text{occitm}(i)/|T|) \geq \text{minsupp}$ . Such an itemset is called *frequent* itemset.

**Definition 9:** A *ruleitem*  $r$  passes the *minsupp* threshold if,  $\text{suppcount}(r)/|T| \geq \text{minsupp}$ . Such a *ruleitem* is said to be a *frequent ruleitem*.

**Definition 10:** A *ruleitem*  $r$  passes the minimum confidence (*minconf*) threshold if  $\text{suppcount}(r) / \text{actoccr}(r) \geq \text{minconf}$ .

**Definition 11:** An associative rule is represented in the form:  $\text{cond} \rightarrow c$ , where the antecedent is an itemset and the consequent is a class.

The problem of AC is to discover a subset of rules with significant support and high confidence. This subset is then used to build an automated classifier that could be used to predict the classes of previously unseen data. It should be noted that MinSupp and MinConf terms in ARM are different than those defined in AC since classes are not considered in ARM. Only item sets occurrences are used for the computation of support and confidence.

Classification Based on Associations (CBA) was presented by [4] and it uses the A priori candidate generation method [1] for the rule discovery step. CBA operates in three steps, where in step 1, it discretises continuous attributes before mining starts. In step 2, all frequent ruleitems which pass the MinSupp threshold are found. Finally, subsets of these that have high confidence are chosen to form the classifier in step3. Due to a problem of generating many rules for the dominant classes or few and sometimes no rules for the minority classes, CBA (2) was introduced by [4], which uses multiple support thresholds for each class based on class frequency in the training data set. Experiment results have shown that CBA (2) outperforms CBA and C4.5 in terms of accuracy.

Classification based on Multiple Association Rules (CMAR) adopts the FP-growth ARM algorithm for discovering the rules and constructs an FP-tree needs to mine large databases efficiently. It consists of two phases, rule generation and classification. It adopts a FP-growth algorithm to scan the training data to find the complete set of rules that meet certain support and confidence thresholds. The frequent attributes found in the first scan are sorted in a descending order, i.e. F-list. Then it scans the training data set again to construct an FP-tree.



For each tuple in the training data set, attribute values appearing in the F-list are extracted and sorted according to their ordering in the F-list. Experimental results have shown that CMAR is more accurate than CBA and C4.5 algorithms. The main drawback documented in CMAR is the need of large memory resources for its training phase.

Classification based on Predictive Association Rules (CPAR) is a greedy method proposed by [20]. The algorithm inherits the basic idea of FOIL in rule generation [21] and integrates it with the features of AC. Multi-class Classification based on the Association Rule (MCAR) is the first AC algorithm that has used a vertical mining layout approach [22] for finding rules. As it uses vertical layout, the rule discovery method is achieved through simple intersections of the itemsets Tid-lists, where a Tid-list contains the item's transaction identification numbers rather than their actual values.

The MCAR algorithm consists of two main phases: rules generation and a classifier builder. In the first phase, the training data set is scanned once to discover the potential rules of size one, and then MCAR intersects the potential rules Tid-lists of size one to find potential rules of size two and so forth. In the second phase, the rules created are used to build a classifier by considering their effectiveness on the training data set. Potential rules that cover a certain number of training objects will be kept in the final classifier. Experimental results have shown that MCAR achieves 2-4% higher accuracy than C4.5, and CBA.

The Multi-class, Multi-label Associative Classification (MMAC) algorithm [14] consists of three steps: rules generation, recursive learning and classification. It passes over the training data set in the first step to discover and generate a complete set of rules. Training instances that are associated with the produced rules are discarded. In the second step, MMAC proceeds to discover more rules that pass MinSupp and MinConf from the remaining unclassified instances, until no further potential rules can be found. Finally, rule sets derived, during each iteration, are merged to form a multi-label classifier that is then evaluated against the test data. The distinguishing feature of MMAC is its ability to generate rules with multiple classes from data sets, where each data objects is associated with just a single class. This provides decision makers with useful knowledge discarded by other current AC algorithms.

Where  $\gamma_u$  is all the state reachable from  $u$ ,  $\beta$  and  $\alpha$  are arbitrary parameters and  $c(u,v)$  is the pheromone rate deposited by the ant colony on the edge  $(u,v)$ . When an ant crosses an edge, it deposits a pheromone value on it. After a complete iteration, the whole pheromone rate on each edge  $(u,v)$  is computed by

$$c^{t+1}(u,v) = \sum_{k \in A} c^{t+1}(u,v,k) + \rho \cdot c^t(u,v) \quad (2)$$

where  $c^{t+1}(u,v,k)$  is the pheromone value deposited by the ant  $k$  from the set  $A$ , at the iteration  $t+1$ ,  $\rho$  is an evaporation factor which allows to decrease the previous pheromone rate. This negative feed-back allows for forgetting some initial and non optimal solution to make new ones emerge in a dynamic way.

To the best of the author's knowledge and during the learning step, most of the above AC algorithms join frequent itemsets of size  $K$  regardless of their class values to derive candidate itemsets of size  $K+1$ . Whereas, our proposed training algorithm only joins frequent itemsets with common class values of size  $K$  to produce candidate itemsets of size  $K+1$ . This significantly reduces costs associated with memory usage and training time as discussed in details in Section 4.

## ANT COLONY ALGORITHM FOR CLASSIFICATION RULES DISCOVERING

Ant colony techniques have been successfully applied to many problems including the Traveling Salesman Problem [8], Quadratic Assignment Problem [8], Job-Shop Scheduling Problems [9], generic constraint satisfaction problems [10], University Course Timetabling Problems [11], cutting and packing [2] graph coloring problems [9] and load balancing in simulation distribution [15]. The standard technique is a constructive one: a colony of ants moves in a solutions space and evolves in collaborative way to find the best solution. Each ant contributes to this collaborative construction by making decisions stochastically, using existing problem constraints and heuristics combined with experience (which is analogous to a substance called pheromone). The colony then reinforces decisions in the construction process according to their successes by adding pheromone, which also decays, to mitigate against poorer decisions. Our purpose in this paper is to investigate the ant algorithm technique as a means of constructing effective class labels associated with an attribute value in the training data set rather than just the single most obvious class.

For our purpose, we define a weighted graph  $G = (V,E)$ , such that  $V$  is the set of vertices corresponding to the class labels and  $E$  is the set of edge  $(u,v)$  representing the transactions between two nodes  $u$  and  $v$ . The edge weight value corresponds to  $w(u,v)$  is the value of  $\sup(u)$ , the fraction of transactions which contain  $u$ . At each step, an ant will choose to move, crossing an edge  $(u,v)$ , according to a probabilistic value  $p(u,v)$ :

$$p(u,v) = (w(u,v))^\alpha (c(u,v))^\beta / (\sum_{w \in \gamma_u} (w(u,w))^\alpha (c(u,w))^\beta) \quad (1)$$

This distributed computation contributes in finding an appropriate transactions path for emergent association classification.

Even though Ant-Miner is working well, the system has the following drawbacks: 1.) If the rule quality measure  $Q$  is very small, the evolutionary process will become stagnant. 2) The  $H(W_j A_i = V_{ij})$  of term  $ij$  is always the same when computing heuristic function  $\eta_{ij}$ , regardless of the contents of the rule in which the term occurs. This is impossible in real applications. 3) During the rule construction the unfit terms have unnecessarily been added and the same is pruned later during the pruning phase. This increases the computational

complexity and cost of rule construction considerably. In order to overcome the above drawbacks, we propose a new classification rule mining algorithm.

In the proposed Miner, a new procedure is introduced in the rule construction. The information gain is mainly used to check the credibility of the term that is being selected to be present in the antecedent part of the rule. If the information gain of the selected term is below the threshold value of 0.6, the term is rejected from inclusion, otherwise it is added. The study shows that 0.6 is the best value for the selection criteria. However, it still requires deeper study to obtain the exact value. Thus only the fittest terms are ensured in the antecedent part of the rule and hence the rules with variable lengths are obtained. Even if we fix the number of terms in the antecedent part as in ACO-miner, the unfit terms may still come in to play. The selection is done, based on probability instead of roulette wheel selection as in the Ant miner. The Algorithm REPEAT Choose the first term IF (Information Gain  $\geq 0.6$ ) THEN add the term in the empty rule set ELSE Reject the term from inclusion END IF UNTIL No more term is left

The performance of the proposed Miner algorithm has been tested on five datasets obtained from the UCI repository [9]. In Table 1, the selected data sets are summarized in terms of the number of cases, the number of attributes (including categorical and continuous attributes), and the number of the classes of the data set. As Miner mines rules refer only to categorical attributes, continuous attributes are discretised using the C4.5-Disc Discretisation method [10] in a preprocessing step. All the results of the comparison were obtained on a 64 bit AMD turion processor, 798 MHZ 512MB RAM. In the following experiments, Miner has the following user-defined parameters:

Number of Ants = 200, Minimum number of cases per rule = 8, Maximum number of uncovered cases in the training set = 15. iv) Number of rules used to test convergence of the ants= 8. v) In order to control the relative importance of pheromone and heuristic function in state transition rule, we set parameter

$\beta=0.4$ . vi). In order to provide a direct way to balance between exploration of new terms and exploitation of a priori and accumulated knowledge about the classification, we set parameter  $q_0=0.5$ .

## COMPARISON RESULTS

The performance of MACO-Miner is evaluated by comparing it with Ant-Miner, ACO-Miner and C4.5. The first experiment was carried out to compare predictive accuracy of discovered rule lists by well-known ten-fold cross-validation procedures. The predictive accuracies of the ten runs are averaged as the predictive accuracy of the discovered rule list. Table 2 shows the results comparing the predictive accuracy of MACO-Miner, Ant-Miner and ACO-Miner and C4.5, where the “ $\pm$ ” symbol denotes the standard deviation of the corresponding predictive accuracy. It shows that the predictive accuracies of TACO-Miner are higher than those of Ant-Miner, ACO-Miner and C4.5. In additional, we compared the simplicity of the discovered rule list by the number of discovered rules and the average number of literals (conditions) per rule. The results comparing the simplicity of the rule lists discovered by MACO-Miner, Ant-Miner, ACO-Miner and C4.5 are shown in Table 3 and Table 4. As shown in those tables, taking into account the number of rules discovered and number of literals per rule, TACO-Miner mined rule lists are much simpler (smaller) than the rule lists mined by Ant-Miner, ACO-Miner and C4.5.

Table 2. Comparison of Predictive Accuracy: MACO, nt,ACO-Miner and C4.5

| C4.5(%)  | ACOMiner(%)  | Ant-Miner (%) | MACO-Miner (%) | Data Set                |
|----------|--------------|---------------|----------------|-------------------------|
| 73±0.32  | 77.14 ± 0.53 | 75.28 ±2.24   | 77.76 ±0.43    | Ljubljana Breast Cancer |
| 95±0.25  | 97.15 ± 0.89 | 96.04 ±0.93   | 97.98 ±0.65    | Wisconsin Breast cancer |
| 83± 0.22 | 98.43 ± 0.36 | 73.04 ±2.53   | 71.87 ±0.23    | Tic-Tac-Toe             |
| 89±0.17  | 96.57 ± 0.74 | 94.29 ±1.20   | 97.26 ±0.45    | Dermatology             |
| 86±0.56  | 94.63 ± 0.58 | 90.00 ±3.11   | 95.50 ±0.37    | Hepatitis               |

Table. 3. Comparison of Number of Discovered Rules: MACO, Ant, ACO-Miner and C4.5

| C4.5(%) | ACO-Miner(%) | Ant-Miner(%) | MACO-Miner(%) | Data Set                |
|---------|--------------|--------------|---------------|-------------------------|
| 06 ± 45 | 6.25 ±0.23   | 7.10 ±0.31   | 5.73 ±0.23    | Ljubljana Breast cancer |
| 11 ± 65 | 4.63 ±0.12   | 6.20 ±0.25   | 4.52 ±0.32    | Wisconsin Breast cancer |
| 83 ± 23 | 6.47 ±0.43   | 8.50 ±0.62   | 5.57 ±0.56    | Tic-Tac-Toe             |
| 23 ± 72 | 6.87 ±0.56   | 7.30 ±0.47   | 7.87 ±0.54    | Dermatology             |
| 4.0± 45 | 3.07 ±0.24   | 3.40 ±0.16   | 3.31 ±0.43    | Hepatitis               |

Table 4. Comparison of Number of Literals per Rule: TACO, Ant, ACO-Miner and C4.5

| C4.5(%) | ACO-Miner(%) | Ant-Miner(%) | MACO-Miner(%) | Data Set                |
|---------|--------------|--------------|---------------|-------------------------|
| 2.23    | 1.13         | 1.28         | 1.10          | Ljubljana Breast cancer |
| 2.51    | 1.58         | 1.97         | 1.52          | Wisconsin Breast cancer |
| 2.95    | 1.06         | 1.18         | 1.03          | Tic-Tac-Toe             |
| 2.48    | 2.85         | 3.16         | 3.55          | Dermatology             |
| 2.11    | 1.98         | 2.41         | 1.84          | Hepatitis               |

Table 5

| C4.5(%) | Miner(%) | Ant-miner(%) | A-Miner(%) | Data Set                |
|---------|----------|--------------|------------|-------------------------|
| 51.32   | 46.63    | 55.28        | 40.21      | Ljubljana Breast Cancer |
| 55.76   | 52.39    | 58.74        | 51.25      | Wisconsin Breast Cancer |
| 84.69   | 57.53    | 61.18        | 55.23      | Tic-Tac-Toe             |
| 49.34   | 37.68    | 49.58        | 38.16      | Dermatology             |
| 50.62   | 41.84    | 56.57        | 35.56      | Hepatitis               |

Finally we compared the running time of the MACO-Miner with Ant-Miner, ACO-Miner and C4.5. The experimental results are reported in Table 5, and as expected we can see that MACO-Miner's running time is shorter than Ant Miner's, ACO-Miner's and C4.5's for all data sets.

The main reason was that we used threshold to limit the entry of the unfit terms and hence the pruning effort required. Table 5. Comparison of Running Time between MACO, Ant, ACO-Miner and C4.5

## REFERENCES

- [1] Agrawal, R. and Srikant, R. 1994. "Fast algorithms for mining association rules." VLDB-94, 1994.
- [2] E. Burke and G. Kendall, Applying Ant Algorithms and the No Fit Polygon to the Nesting Problem, Proceedings of 12th Australian Joint Conference on Artificial Intelligence, Sydney, Australia, December 6-10, 1999, Lecture Notes in Artificial Intelligence vol. 1747, Foo, N. (Ed), pp. 453-464, 1999.
- [3] Baralis, E. And Garza, P. 2002. A lazy approach to pruning classification rules. In Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM'02). IEEE Computer Society Press, Los Alamitos, CA, 35-42.
- [4] .B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In Proc. of 4th Intl. Conf. on Knowledge Discovery and Data Mining (KDD), Aug. 1998.
- [5] Cristianini, N. And Shawe-Taylor, J. 2000. An Introduction to Support Vector Machines. Cambridge University Press, Cambridge, U.K.
- [6] D. Costa and A. Hertz, Ants can colour graphs, Journal of the Operations Research Society 48, pp. 295-305, 1997.
- [7] Duda, R. And Hart, P. 1973. Pattern Classification and Scene Analysis. John Wiley and Sons, New York, NY.
- [8] M. Dorigo, V. Maniezzo and A. Colomi, The Ant System: Optimization by a colony of cooperating agents, IEEE Transactions on Systems, Man and Cybernetics Part B, vol. 26, no. 1, 1996, pp. 1-13.
- [9] M. Dorigo, V. Maniezzo and A. Colomi. Distributed optimization by ant colonies Proceedings of the first European Conference on Artificial Life Paris, ed. By FJ Varela and P Bourguine, MIT/Press/Bradford Books, Cambridge, Massachusetts: pp. 134-142, 1992.
- [10] S. Pimont and C. Solnon, A Generic Ant Algorithm for Solving Constraint Satisfaction Problems, Abstract Proceedings of ANTS' 2000 From Ant Colonies to Artificial Ants: Second International Workshop on Ant Algorithms, pp. 100-108, 2000.
- [11] K. Socha, J. Knowles and M. Sampels, A MAX-MIN Ant System for the University Course Timetabling Problem, Proceedings of the Third International Workshop on Ant Algorithms (ANTS'02), Springer LNCS vol. 2463, pp. 1-13, 2002.
- [12] Quinlan. J. C4.5: Programs for machine learning. Morgan Kaufmann, 1993.
- [13] R. Lippmann. An introduction to computing with neural nets. IEEE ASSP Magazine, 4(22), 1987.
- [14] Thabtah, F, Cowling, P and Peng, Y, 2004, MMAC: A new multi-class, multi-label associative classification approach. In Proceedings of the 4th IEEE International Conference on Data Mining (ICDM'04), Brighton, UK, pp. 217-224.
- [15] Bertelle C, Dutot A, Guinand F and Olivier D, 2007, "Organization detection for dynamic load balancing in individual-based simulation", Multiagent and Grid System, pp 141-163, Volume 3, Number 1, 2007.
- [16] Ani, A.A., "Feature Subset Selection Using Ant Colony Optimization", International Journal Of Computational Intelligence Volume 2 Number 1 2005 ISSN 1304-4508.
- [17] Chan, A., and Freitas, A., "A New Classification Rule Pruning Procedure for an Ant Colony Algorithm", Artificial Evolution 2005: 25-36.
- [18] Parpinelli, R.S., Lopes, H.S., and Freitas, A.A. "Data mining with an ant colony optimization algorithm", IEEE Transactions on Evolutionary Computing 6(2002), 321-332.
- [19] Goldberg, D.E. "Genetic Algorithms in Search, Optimization, and Machine Learning", Addison-Wesley (1989).
- [20] Yin, X., and Han, J. (2003) CPAR: Classification based on predictive association rule. Proceedings of the SDM (pp. 369-376). San Francisco, CA.
- [21] Cohen, W. (1995) Fast effective rule induction. Proceedings of the 12th International Conference on Machine Learning, (pp. 115-123). CA, USA.
- [22] Zaki, M., Parthasarathy, S., Ogihara, M., and Li, W. (1997) New algorithms for fast discovery of association rules. Proceedings of the 3rd KDD Conference (pp. 283-286). Menlo Park, CA.

# **AI APPLIED TO MANUFACTURING**



# A HIERARCHICAL OBJECT-ORIENTED SOFTWARE STRUCTURE FOR MANUFACTURING SIMULATION

Lina A. M. Huertas Quintero\*  
Andrew A. West  
Diana M. Segura Velandia  
Paul P. Conway  
Radmehr Monfared.

Wolfson School of Mechanical and Manufacturing Engineering  
Loughborough University  
Loughborough, Leicestershire, LE11 3TU  
UK  
E-mail: L.A.Huertas@lboro.ac.uk

## KEYWORDS

Object-oriented, industrial application, manufacturing, simulation, software design.

## ABSTRACT

Modern manufacturing companies have been forced to face a demanding environment due to global pressures like dynamic markets, globalisation and competition. Simulation has become an essential tool to respond to these challenges by supporting continuous improvement of manufacturing systems for different applications. Different kinds of simulation tools have been proposed by the academic and industrial sector providing valuable contribution to this research topic. However, these approaches usually fail to provide software that copes with the specific requirements of an industrial application (e.g. low development times, no specific skills required and changeability). The aim of this paper is to provide a description of an innovative software structure developed at Loughborough University. The structure tackles the simulation requirements identified for industrial environments. A comparative example between the proposed simulation software and a commercial package proves the benefits provided by the new structure. The advantages include short model development times, low specific skills requirements, and enhanced changeability. These features not only offer improved usability, but also flexible software capability that can easily migrate to other organisations or even different domains.

## INTRODUCTION

The globalised economy coupled with economic and political pressures has resulted in the offshore relocation of the UK's high volume, low variety manufacture of electronics products, in particular, printed circuit assemblies (PCA). Apart from the drain of manufacturing knowledge and design expertise, the loss of high volume capability has put tighter constraints on the remaining UK electronics companies since the costs of the early lifecycle phases are not recovered via the sale of large volumes of product (e.g. specification, design, analysis and setup). In the face of these pressures UK PCA manufacturing has migrated towards supplying low volume solutions for demanding applications

in terms of time, cost, quality and reliability (e.g. aerospace, defence). This case illustrated the challenging conditions that many companies in this and other manufacturing sectors have to face.

With manufacturing companies driven to reduce lead times and costs and improve quality and reliability, evaluation and decision support tools have become essential. The most powerful and popular tool to cover these needs, according to numerous authors in the research field, is simulation (Hollocks 1995; Terzi and Cavalieri 2004). Various simulation approaches have been proposed, including a very promising trend towards utilising object-oriented paradigm. Nonetheless, these attempts have, generally, failed to cover important non-functional requirements that would enable a feasible adaptation in an industrial context (e.g. low building times and no specific skills requirement).

A piece of research aimed to design a simulation structure that satisfies these industrial requirements is presented in this paper. The application domain for this specific study is PCA manufacturing in the UK. This document is organised as follows. In the first place, a brief review of previous contributions to this specific research topic is outlined. Afterwards, the method followed for the simulation software design and development is described. The next section illustrates the conceptual framework employed, together with a recounting of the design activity. Finally, an example that proves the benefits gained from the new design is shown prior to a summary of the main conclusions of the research and some future prospects.

## SIMULATION TO SUPPORT MANUFACTURING

Simulation has become an important tool to analyse the performance of manufacturing systems (Wang 2005). Its prediction capability provides support for continuous improvement and increasing competitiveness (Hibino 1999; Murphy and Perera 2002). The utilisation of the object oriented paradigm to develop simulation tools has proved to facilitate the software abstraction, implementation, application, verification and validation, and provide a more intuitive mapping between the models and the real world (Narayanan 1998).

Various commercial tools have been developed in the last decades (e.g. Simul8, WITNESS, Prosim, Arena). These software tools offer great flexibility since they offer low level constructs as well as pre-built blocks. Furthermore, the graphical user interfaces of these software are usually well advanced and user friendly. Valuable experience related to this kind of tools was gained during the research as some of them were intensively used in early phases and various end users were interviewed. In spite of the advantages aforementioned, this experience unveiled a downside of commercial simulation packages. The main drawbacks identified are: i) non-integrated approach (product models are not considered), ii) customised templates are not provided and iii) it is difficult or impossible to dynamically control the simulation during time execution.

The academic community has also provided massive contributions to the field. Important research effort has been dedicated to the development of object-oriented simulation tools for the manufacturing domain (Karcal 1998; Narayanan 1998; Hibino 1999; Borenstein 2000; Anglani 2002). In general, these attempts share the common aim of reducing the effort required from the end user of the software to develop a model. Some of these research works have made valuable contributions to this research topic. For instance, (Hibino 1999) identifies key non-functional requirements of an appropriate manufacturing simulation software like the encapsulation of complex behaviour. Another example is the approach presented by (Martínez 2002) where the concept of identifying and formalising common objects in a specific domain is employed.

Even though important advances have been done regarding object-oriented simulation, existing designs fail to cover important non-functional requirements. These requirements were identified from the literature review and are summarised in Table 1.

Table 1: Non-functional requirements

| Non-functional requirements |
|-----------------------------|
| Low building times          |
| Low expertise requirements  |
| Modelling simplicity        |
| Intuitive                   |
| Configurability             |
| Scalability                 |
| Customisation               |
| Reusable                    |

None of the current tools cover all the requirements in Table 1 in an effective manner. This means that the industrial application of existing object-oriented simulation software is difficult, and usually unfeasible. These pitfalls are the focus of the research exposed in this paper. To overcome them, an innovative simulations structure is proposed.

## METHOD

In the previous section of this paper, the requirements for simulation software to design or improve manufacturing systems in the PCA manufacturing in the UK were identified and its functionality specified. During the research reported,

simulation software to cover these requirements was developed. The method developed to design and implement this software is presented in this section.

The object-oriented paradigm (OOP) was selected to program the simulation software. Along with this paradigm, a hierarchical approach was adopted to design and develop the software structure. The layers of the hierarchy designed were inspired on the instantiation proposed by CIMOSA framework (Vernadat 1996). The method is illustrated in Figure 1.

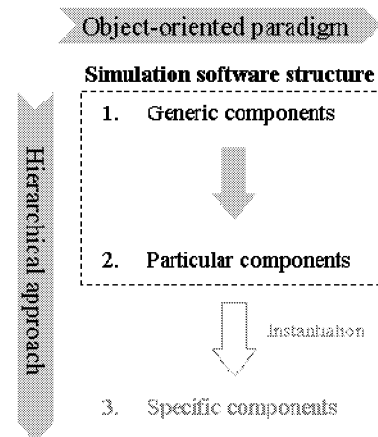


Figure 1: Simulation structure generation method

Accordingly to Figure 1, the first components generated were generic and based on standard manufacturing systems. These components were defined from selected modelling techniques that cover the most relevant aspects of a manufacturing system (e.g. enterprise modelling, cost estimation and models of physical phenomena within processes and activities). More specific components (i.e. the particular components) are derived from the generic ones by identifying attributes and behaviour characteristic and repeated in the application domain. The final step in the hierarchical approach is the development of the most specific components. These components are generated during the deployment of the software when specific instances of the particular objects are created to execute the simulation. This method leads to the generation of the simulation structure that tackles the research problem.

## A NOVEL SIMULATION STRUCTURE

A simulation software structure is proposed in this paper to tackle identified non-functional requirements in an industrial context. The main features of the approach are the exploitation of object oriented principles (i.e. mainly encapsulation and inheritance) and a hierarchical configuration. Although, UK PCA manufacturing was the application domain for this study, one of the advantages of the structure is the facility of migration across domains. The structure developed is referred to as HOOSS (hierarchical object oriented software structure) in this document for simplicity.

The HOOSS design consists of three levels: i) generic, ii) particular and iii) specific. This three layer hierarchy is inspired on the instantiation rank proposed in the CIMOSA modelling framework (Vernadat 1996). The idea behind the



hierarchical approach is to improve the performance of the software in terms of non-functional requirements. The top level layer (i.e. generic layer) contains standard elements or classes shared by manufacturing systems regardless the application. The following layer (i.e. particular layer) is derived from the generic layer by specialising its classes according to characteristics and behaviour, and the repeated elements particular to a specific application. In this case, the particular layer is based on characteristics of PCA manufacturing in the UK. Finally, the specific layer consists of instances or objects of the classes in the particular layer. The top two layers of the HOOSS represent a modelling framework proposed. Sub-hierarchies are also possible within each layer. From this framework, cases studies can be derived and represented in the bottom layer. The nature of the elements in different layers is explained better with the examples illustrated in Figure 2.

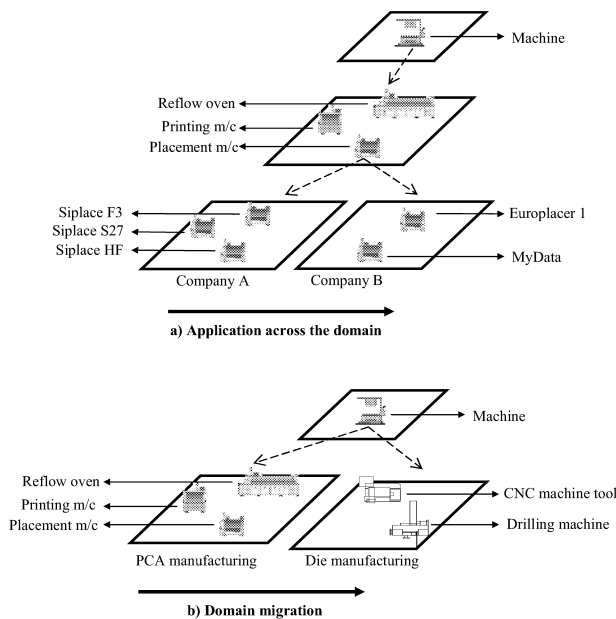


Figure 2: HOOSS conceptual view

In the example illustrated in Figures 2.a. and 2.b, there is one element in the generic layer that is a standard element of manufacturing systems (i.e. a machine). When the particular case of manufacturing systems in the UK PCA manufacturing domain is considered, several elements can be derived from these categories (see Figure 2.b.). The example shows particular machines that are common across the domain i.e. printers, placers and reflow ovens. The particular layer should include the most important elements in a manufacturing system in the application domain and encapsulate their characteristics, behaviour and relationships. Finally, to construct a case study elements are created in the specific layer as derivations of the particular and generic elements. In the example, instances of placement machines, in a specific company, are shown in Figure 2.b.

(i.e. Siplace F3, Siplace S27 and Siplace HF). These objects are instances for a specific system of a company that is to be analysed.

Situations where the advantages of HOOSS can be exploited include deployment of the simulation software across the application domain or, even, the migration of the software to other domains (see Figures 2.a and 2.b). The case illustrated in Figure 2.a is the application of the simulation software in different companies across the domain. The figure shows an example where two implementations of the software are made, one in Company A and the other in Company B. The companies in the example belong to the same application domain i.e. they manufacture PCAs in the UK. However, their choices of equipment are different. Using a flat simulation structure, the elements generated for a case study in Company A would be the three machines of the example i.e. Siplace F3, Siplace S27 and Siplace HF. To implement a case study in Company B, the new placement machines (i.e. two Europlacer machines) would have to be created from scratch including the specification of their characteristics and behaviour.

On the other hand, the HOOSS allows reusing elements generated in the particular layer during the case study in Company A for the implementation in Company B. Following the example, when creating the elements for Company A, the behaviour of machines, in a manufacturing environment in general, is embedded in the class “Machine” in the generic layer. Then, particular behaviour is encapsulated in the class “Placement machine” and in the end the specific characteristics of each Siplace machine are generated. To implement a new case study in Company B with HOOSS, the classes “Machine” and “Placement Machine” can be reused to instantiate the Europlacer machines. A considerable reduction on the implementation time and effort can be achieved by using the structure proposed for repeated implementations. Figure 2.b illustrates equivalent benefits in the case of migration of the simulation software to other domains.

## A HIERARCHICAL OBJECT-ORIENTED SIMULATION STRUCTURE

The generic layer of the HOOSS designed in the study herein reported is based on selected modelling techniques including enterprise modelling, cost estimation, and models of the physical phenomena within a process or activity (i.e. physical models). These modelling techniques were selected and integrated into the simulation structure to cover the most relevant aspects of a manufacturing system. Three root classes are derived from the models i.e. processes, products and production. These classes are the top level elements in the HOOSS design presented. From these classes, other generic elements are derived. For instance, it is common to have resources associated to processes, materials associated to products and orders associated to the production of a company.

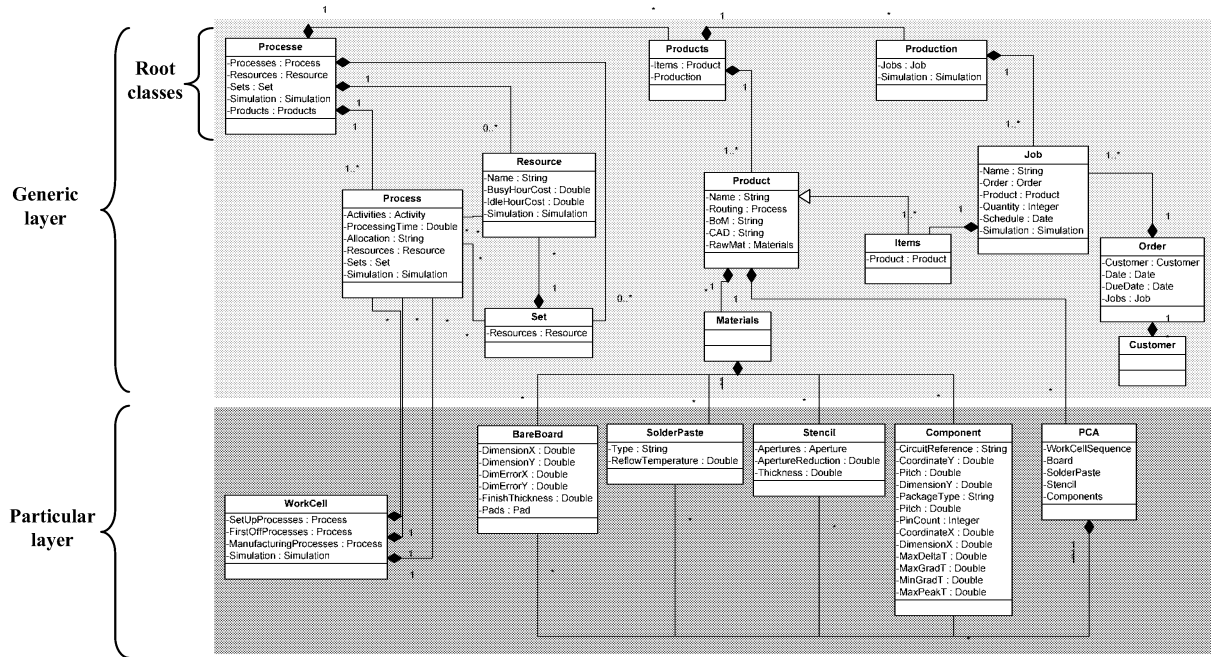


Figure 3: Main classes of the HOOSS

The particular layer encapsulates knowledge specific to the application domain, such that the instantiation process (i.e. simulation scenarios implementation) is as straightforward as possible. This layer also separates generic constructs of manufacturing systems, from constructs specific to the domain. As illustrated in Figure 2, this feature facilitates the migration of the tool not only across the domain but also to other domains. The main classes in this layer are workcells (i.e. composed structures made of processes to represent production lines) and classes related to the specific products of the application domain.

The generic and particular layers of the HOOSS from the framework of available constructs that the user of the simulation can instantiate to model a particular manufacturing system or sub-system. On the other hand, the specific layer is a temporal layer that consists of the instances created from the generic and particular layers. The elements in this layer are not complementary part of the HOOSS and are eliminated every time the simulation software is closed unless saved. The main classes in the HOOSS developed during this research are illustrated in Figure 3.

The HOOSS described in this chapter was implemented using object oriented programming environment Microsoft Visual Basic 2005 Version 8.0.50727.42 together with the software Microsoft Visual Studio 2005 (RTM.050727.42) and Microsoft .NET Framework Version 2.0.50727. Additionally, a discrete event simulation (DES) commercial engine called Delsi™ was integrated to the software to drive the simulation. These software was used given the benefits provided in terms of interface with other software and dynamic control. However, any other object oriented software could have been used to implement the HOOSS, including object oriented commercial simulation packages. The behaviour of SMT production lines, or what we call in

this study work cells, is not as simple as a sequential series of processes. In the first place, before a batch of products goes through a work cell a number of set up of the processes should be performed to assure that their parameters comply with the specific parameters for the product. For instance, the placement machine has to be set up according to a specific CAD file to assure that the electronic components are placed in the right place on the bare board. It should be noted that the product is not transformed during the set up times and just the processes and resources may be altered. A second characteristic of product lines in the domain is that, due to the demanding quality and reliability standards required in the domain, in particular in the UK, an extra number of process named “first off processes” are performed. First off processes are simply some or all the manufacturing process in the production line performed just for the first item in the batch to be produced. After first off processes are finished, the first off item is inspected to reassure that the processes are generating the expected results. Once this procedure is over, the whole batch is send through the work cell, following a normal flow from one manufacturing process to the next one.

## APPLICATION OF THE HOOSS

One of the main process structures in electronics manufacturing companies are surface mounted technology (SMT) production lines. The basic processes in a SMT line are the printing process where the solder paste which joins the board and the components is disposed on the board, the placement processes where the electronic components are placed over the wet deposits of solder paste and the reflow processes where the solder paste is melted to form the joints between the component pins and the board.

Production Line 1

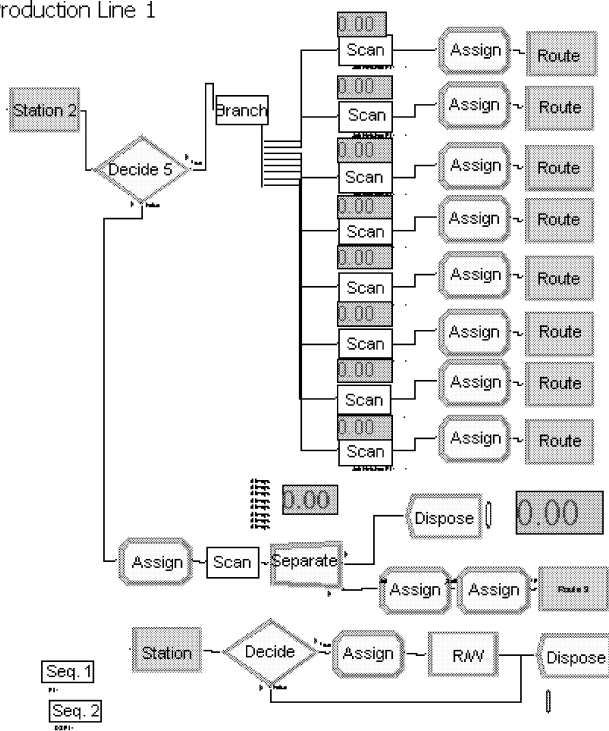


Figure 4: SMT line implemented in Arena

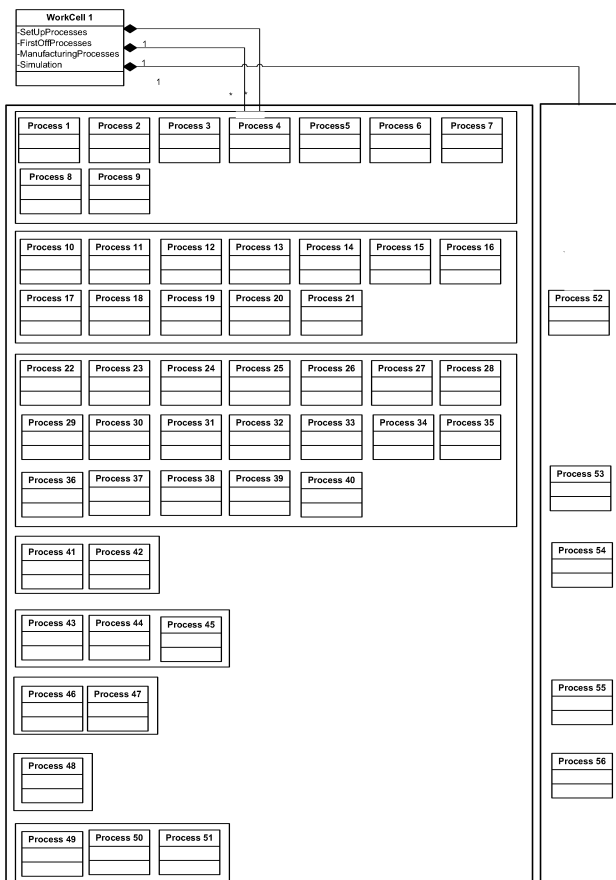


Figure 5: SMT line implemented with HOOSS

Usually, to implement a work cell in a commercial package for manufacturing simulation, the whole control flow has to be designed and implemented in the program using logical modules. One example of a traditional SMT work cell implemented in Arena (Rockwell Automation Technologies 2006) is shown in Figure 4.

As illustrated in Figure 4, the implementation consists of two parts. On the left hand side of the figure the logic blocks that build up the control structure of the simulation is illustrated. The processes within the production line, including set up processes, first off processes and manufacturing processes, are presented on the opposite side. The implementation process of the simulation model in Figure 4 took approximately 5 hours (including validation and verification) for a developer with previous experience on the development of Arena models and on algorithms design. Most of the time invested on the model implementation (i.e. circa 4 hours) was spent designing the control structure that simulates the internal process flow of the work cell. Thus, the development time is expected to rise considerably for a developer without previous knowledge regarding Arena constructs and production lines work flow.

In the case of HOOSS, the internal process flow is already embedded in a single construct. This construct is represented by the “WorkCell” class in Figure 3. Instances of this class were used to implement the production line in Figure 4 using HOOSS. The class diagram of the implementation is shown in Figure 5.

In the case of the implementation illustrated in Figure 5, the number of processes to be instantiated is exactly the same as in the implementation in Figure 4 (56). Hence, the effort input required to implement the process structure is the same with both tools. However, in the HOOSS

implementation flow control is already encapsulated in the workcell object and, therefore, the control blocks become redundant and design or implementation is needed. As a result, to model the work cell, the only task that the user should perform is the instantiation of each process and its assignation to the WorkCell attributes. It should be noted that in a HOOSS workcell, the processes are associated to and not packed in the object. Hence, the process are still available to be reused for different process flows of workcells.

Given the nature of the HOOSS, the expertise and the time required to implement a model is significantly reduced and its usability in an industrial context is enhanced. In the case presented as example, the implementation took little more than 1 hour (including validation and verification) for the same developer. The implementation time with HOOSS is then reduced to approximately 20% of the time taken with a commercial software using a traditional approach. This reduction is a proof of the major simplification achieved on the implementation of simulation models. Furthermore, the time is not expected to increase significantly for developers without previous experience, as it was the case of the model development with commercial tools. It is important to note from the example that the particular layer of the HOOSS is the essential piece in the solution of the problem to be solved in this chapter.

## DISCUSSION

The time savings achieved by using a simulation software based on the HOOSS presented in this paper are noteworthy and prove the usefulness of the approach. However, their meaning in terms of net benefits and their scope ought to be further discussed. In the first place, to be able to refer to net savings, it should be considered that the development and implementation of the HOOSS itself require significant time and effort. The real benefits of the approach are hence not associated to the development of one-off models using the approach, as this would result in time losses rather than gains. The real contribution of the approach lays in providing a tool that can be repeatedly used in an industrial environment. In this context, constant utilisation of the tool by users that are not experts on software development or simulation modelling is the typical scenario. This situation leads to the following benefits: (i) repeated use of the simulation approach result in time savings higher than the time invested on the software development and (ii) simulation constructs with complex behaviour embedded eliminates the need to program dense logic, not only reducing the modelling time, but also reducing the risk of incorrect models.

In terms of the scope of the results presented, a few observations should also be made. As the case study description indicates, the time savings presented are the result of a one-off comparative experiment. In the comparison, just one model for each approach (i.e. commercial simulation and HOOSS-based simulation) were taken into account. Therefore the results obtained are valid purely as a way to illustrate the approach proposed and the benefits that could be potentially achieved. Nonetheless, the result do not have statistical significance and further experimentation is required. The performance of complete usability experiments are part of the work to be performed

by the authors in the near future and the results will be published in upcoming publications.

## CONCLUSIONS AND FUTURE WORK

The paper presents the development of an innovative simulation structure that complies with the requirements of an industrial environment. The main needs addressed with the new structure are reductions in the modelling times, lower specific skills requirement, and improved changeability. This structure was achieved by exploiting the basic principles of object-oriented and using a hierarchical structure that categorise the software objects from generic to specific. The most generic elements correspond to standard elements in manufacturing systems. The next layer, a particular layer, consists of elements specific to the application domain; in this case it is electronics manufacturing. These two layers are static and form the main part of the structure. The elements in the most specific layer are temporal instances generated for each simulation model implemented in the software.

A comparative example that illustrates the benefits provided by the proposed structure is described. The example shows benefits in the process of implementation of a simulation model. In the first place, the process requires significantly less time than commercial software packages (i.e. the reduction in time is approximately 80%). Additionally, the implementation requires less or no software skills from the user. Finally, different parts of the software structure can be changed to migrate to other case studies or domains, with reduced impact on the rest of the structure. Future prospects of the research include comprehensive usability experiments to characterise and quantify the contribution of the approach.

## ACKNOWLEDGEMENTS

The authors wish to express their gratitude to the industrial collaborators. This work was supported financially by the IeMRC DISCOVER grant, SP/05/01/02 and DTI CLOVES grant, TP/3/DSM/6/1/16333.

## REFERENCES

- Anglani, A. (2002). "Object-oriented modeling and simulation of flexible manufacturing systems: a rule-based procedure." *SIMULATION MODELLING PRACTICE AND THEORY* 10(3-4): 209-234.
- Booch, G. (1994). *Object-Oriented Analysis and Design with Applications*.
- Borenstein, D. (2000). "Implementation of an object-oriented tool for the simulation of manufacturing systems and its application to study the effects of flexibility." *International Journal of Production Research* 38(9): 2125-2142.
- Hibino, H. (1999). "The development of an object-oriented simulation system based on the thought process of the manufacturing system design." *International journal of production economics*. 60: 343.
- Hollocks, B. W. (1995). "The impact of simulation in manufacturing decision making." *Control Engineering Practice* 3(1): 105-112.

- Jacobson, I. (1994). Object-Oriented Software Engineering: A Use Case Driven Approach.
- Karcal, S. C. (1998). "A formal structure for discrete event simulation. Part II: Object-oriented software implementation for manufacturing systems." IIE Transactions 30(3): 217.
- Martínez, F. M. (2002). "Modular simulation tool for modelling JIT manufacturing." International Journal of Production Research 40(7): 1529-1547.
- Murphy, S. P. and T. Perera (2002). "Successes and failures in UK/US development of simulation." SIMULATION MODELLING PRACTICE AND THEORY 9(6-8): 333-348.
- Narayanan, S. (1998). "Research in object-oriented manufacturing simulations: an assessment of the state of the art." IIE TRANSACTIONS 30(9): 795-810.
- Rambaugh, J., M. Blaha, et al. (1991). Object-Oriented Modeling and Design.
- Rockwell Automation Technologies, I. (2006). Arena: Template Developer's Guide. USA, Supersedes Publication ARENDG-RM001B-EN-P.
- Terzi, S. and S. Cavalieri (2004). "Simulation in the supply chain context: a survey." Computers in Industry 53(1): 3-16.
- Vernadat, F. B. (1996). Enterprise Modeling and Integration: Principles and Applications. Chapman & Hall.
- Wang, Q. (2005). "Key issues and developments in modelling and simulation-based methodologies for manufacturing systems analysis, design and performance evaluation." The International journal, advanced manufacturing technology. 25(11-12): 1254-12.

# KNOWLEDGE MANAGEMENT OF SHEET HYDROFORMING PROCESSES

Antonio Del Prete, Alfredo Elia and Gabriella Romano

Department of Engineering of Innovation

University of Salento

Via per Arnesano, 73100 - Lecce

Italy

E-mail: antonio.delprete@unisalento.it; alfredo.elia@unisalento.it; gabriella.romano@unisalento.it

## KEYWORDS

Decision support systems, Industrial engineering, Model design, Computer-aided analysis, Performance analysis.

## ABSTRACT

Sheet Hydroforming is considered a good opportunity when it is necessary to deal with complex shapes.

For a non conventional forming process, metal forming simulation has been extensively used in order to design an appropriate experimental tooling and, at the same time, to understand the influence of each parameter on the process performances. Large scale computation corresponding to a very high number of simulated process set up configurations has led authors to consider the option to develop a customized software application.

The primary aim of this work is to test the applicability of Business Intelligence tools in the technological field of the modelling of non conventional sheet metal forming production processes.

In order to achieve this objective, a traditional Decision Support System based on OLAP database has been applied in the Engineering Intelligence context.

The analyzed test cases are three hydroformed components characterized by different geometrical shapes, whose production processes were studied in I.T.Idro project by using both numerical simulation and experimental analysis.

From the whole simulation data, by the use of DSS methodology and the adopted EI tool, a small set of Key Performance Indicators has been identified, in order to monitor the process output and highlight the optimum set of parameters per each studied case.

## INTRODUCTION

Nowadays hydroforming techniques have been largely accepted by industries for the production of components characterized by good surface quality, high-dimensional accuracy together with high drawing ratio and complex shapes (Siegert et al. 2000). The main distinction among the large variety of hydroforming techniques is usually based on the used raw material: tubular or sheet material. Today the tubular material is predominantly used for the mass production; actually sheet hydroforming is used for small batch production because both its cycle time and clamping forces (large tooling investments) are higher than the tube hydroforming ones. However, sheet hydroforming process offers the possibility of no restrictions in the final shape of the part (on the contrary tube hydroforming is simply used

for the production of space frame structures or longitudinally extended parts) together with a large number of variants. (Del Prete et al. 2007a,b).

The work presented illustrates the analysis of these non conventional deformation process. It is generally used for medium-low volume productions as an alternative to conventional metal forming processes due to its highest efficiency in terms of cost reductions. The technological opportunity and potential of the process is often not fully exploited due to the difficulty in managing the high number of process variables: fluid chamber pressure, blankholder force distribution (Shulkin et al. 2000), pre-bulging height, etc (Rimkus et al. 2000).

This study refers to the results obtained in I.T.Idro (a research project developed by the Department of Engineering Innovation – University of Salento), whose target is the individuation of the relationship among variables (geometrical and process parameters) with a direct influence and the process itself.

A numerical campaign has been conducted (fractionary Design of Experiment – DOE –  $2^k$ ) followed by experimental analysis, based on well known parameters relating to a test-case composed by three hydroformed components characterized by different geometrical shapes. The primary aim of the project is to demonstrate how to drive the process designer choices in the n-dimensions process space, where n is the number of its variables. The investigation can be exploited by the usage of numerical and experimental campaigns, to define quality functions and to analyze new strategies for sheet hydroforming process tryout (Del Prete et al. 2008a,b).

## DSS AND OLAP

Decision Support Systems (DSS) disciplines deal with the use of information technology to support human decision-making processes. Michael Scott-Morton, who virtually invented the discipline in the early 1970s, offered this definition of DSS:

“Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision-makers who deal with semi-structured problems”.

The key elements of a decision-making model are fairly common, and include (Demarest 2005):

- a *decision-maker*: an individual or group charged with making a particular decision;
- a *set of inputs* to the decision-making process: data, numerical or qualitative models for interpreting that data,

historical experience with similar data sets or similar decision-making situations, and various kinds of cultural and psychological norms and constraints associated with decision-making;

- *the decision-making process itself*: a set of steps, more or less well-understood, for transforming the inputs into outputs in the form of decisions;
- *a set of outputs* from the decision-making process, including the decisions themselves and (ideally) a set of criteria for evaluating decisions produced by the process against the set of needs, problems or objectives that occasioned the decision-making activity in the first place.

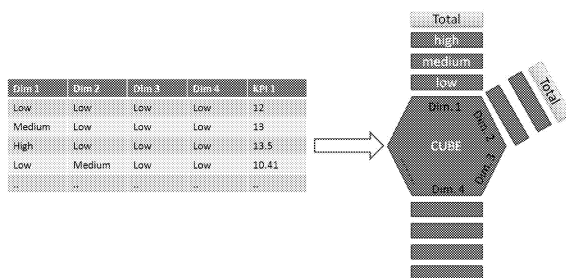
There is a particular kind of DSS called: Data-driven DSS; it emphasizes access to manipulation of a time-series of internal company data and/or external data. Simple file systems accessed by query and retrieval tools provide the most elementary level of functionality. Data warehouse systems, that allow the manipulation of data by computerized tools tailored to a specific task and setting or by more general tools and operators, provide additional functionality.

The OLAP (On-Line Analytical Processing) term is always linked to data warehouse; it refers to a series of technics used mainly for business reporting. Using OLAP, it is possible to analyze data in all manner of different ways, including budgeting, planning, simulation, data warehouse reporting, and trend analysis.

A main component of OLAP is its ability to organize data into multidimensional structures (called hypercubes), in which it is possible to make multidimensional calculations, allowing a wide and lightning-fast array of possibilities. Each hypercube is characterized by:

- *a fact*: that is the analysis field;
- *measures*: a set of KPI related to the fact;
- *dimensions*: hierarchical categories of information. Each dimension has a defined cardinality;
- *cells*: they are unambiguously identified by a set of parameter values and contain KPI values.

Using this structure, the user's first view of the data in an OLAP report is a "top-level" one that reveals patterns and trends at a glance. If users have identified issues or questions in summary-level information, OLAP reports enable them to interactively explore data from any angle, to any level of detail (Grotevant and Foth 1999). This is impossible to obtain using a traditional "Facts table", that represent a flat and static representation of data.



Figures 1: From "Fact Tables" to "Hypercubes"

The previous picture summarizes what it's happening: many companies, which were using fact tables for their data

analysis, now move on dynamical data access, based on OLAP hypercubes, to perform multidimensional analysis.

Multidimensional calculations enable a large business to complete in seconds what it otherwise would have waited a handful of minutes to receive (Pérez-Martínez et al. 2007).

Current data warehouse and OLAP technologies can be efficiently applied to analyze the huge amounts of structured data that companies produce.

Now, an increasing number of DSS have adopted data warehousing (Shi et al. 2007). Data-driven DSS with On-line Analytical Processing (OLAP) provides the highest level of functionality and decision support that is linked to analysis of large collections of historical data.

Moreover, data warehouses, DSS and OLAP technologies are in full swing in industry as major Business Intelligence technologies (Trujillo and Song 2007).

## MODEL DESIGN

### Methodology

The primary aim of this work, as indicated above, is the automation of some phases of the decisional process connected to a non conventional forming process. It has been realized the application of a DSS in a context where a set of parameters at the same time takes part in realization of a complex model (like Engineering Intelligence field). As immediate result it is possible to assert that the developed methodology has given a substantial advantage to simplify the process modelling: the test case is analyzed and evaluated considering a reduced set of Key Performance Indicators.

In this way, the analyst can take decisions to several levels based on a particular scenario (real or presumed) that becomes more and more clear and simple.

Moreover, exploiting the developed model it is possible: to save hystorical data of a forming process, to evaluate relationships among two or more process parameters and their influence on feasibility, to evaluate cause and effect relationships among controllable process parameters and KPI and, therefore, to foresee effects that could occur when corrective actions on controllable process parameters are undertaken.

For the referenced test case (described below), seven fundamental KPI have been utilized:

- Maximum Thickness reduction (called "Tmax%");
- Maximum drawing depth (called "Himb×Tmax");
- Maximum feasible drawing depth (called "Himbfatt");
- Ratio between Maximum drawing depth and Maximum feasible drawing depth;
- FLD Plot;
- Thickness Plot Contour;
- Drawing ratio for each class of models analyzed (called "Rimb").

The post processing of the great amount of obtained results needs the involvement of an expert engineer. This is a key step in order to extract useful data for the OLAP DB. Specific automation programs have been implemented for a semi-automatic extraction of synthetic information by post processing simulated data. The application of the customized routines allows to have a set of metadata for each post

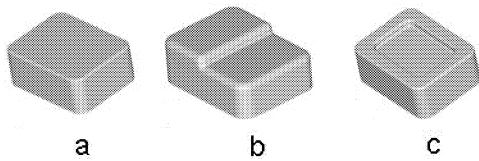
processed simulation (characterized by specific geometrical and process parameters).

Once data are correctly prepared, they can be organized into a multidimensional space where, any considered process parameter will be a dimension of the OLAP cube with cardinality equal to the number of investigated levels.

HiQube is the chosen DSS: it is a decision support and data analytics platform that empowers users to explore engineering and business KPI, easily create and share visually rich reports and dashboards.

### Test cases

To improve the knowledge in sheet hydroforming, three different geometrical shapes have been selected as test-cases:



Figures 1: Different analyzed shapes: (a) Mod1; (b) Mod2; (c) Mod3

For the given models a numerical investigation is developed in order to evaluate the influence of some parameters on the process performance.

The main factors of models are described in the following tables:

Table 1: Mod1 main parameters

| n° Factors | Factors           | Levels           |                  |
|------------|-------------------|------------------|------------------|
|            |                   | Lower Level (LL) | Upper Level (UL) |
| 1          | Hpre-bulging [mm] | 15               | 45               |
| 2          | Thickness [mm]    | 0,7              | 1                |
| 3          | A1 [ton]          | 10               | 22,5             |
| 4          | A2 [ton]          | 10               | 40               |
| 5          | A3 [ton]          | 10               | 22,5             |
| 6          | Himb [mm]         | 100              | 150              |
| 7          | Rp [mm]           | 10               | 50               |
| 8          | Rm [mm]           | 10               | 20               |

Table 2: Mod2 main parameters

| n° Factors | Factors           | Levels           |                  |
|------------|-------------------|------------------|------------------|
|            |                   | Lower Level (LL) | Upper Level (UL) |
| 1          | Hpre-bulging [mm] | 15               | 45               |
| 2          | Thickness [mm]    | 0,7              | 1                |
| 3          | A1 [ton]          | 12               | 25               |
| 4          | A2 [ton]          | 10               | 30               |
| 5          | A3 [ton]          | 10               | 20               |
| 6          | A4 [ton]          | 10               | 25               |
| 7          | A5 [ton]          | 15               | 35               |

|    |           |     |     |
|----|-----------|-----|-----|
| 8  | A6 [ton]  | 10  | 25  |
| 9  | Himb [mm] | 100 | 150 |
| 10 | H2 [mm]   | 20  | 35  |
| 11 | Rm [mm]   | 10  | 20  |
| 12 | Rp [mm]   | 10  | 50  |

Table 3: MOD3 main parameters

| n° Factors | Factors                     | Levels           |                  |
|------------|-----------------------------|------------------|------------------|
|            |                             | Lower Level (LL) | Upper Level (UL) |
| 1          | Hpre-bulging [mm]           | 15               | 45               |
| 2          | Thickness [mm]              | 0.7              | 1                |
| 3          | A1 [ton]                    | 10               | 18               |
| 4          | A2 [ton]                    | 8                | 20               |
| 5          | A3 [ton]                    | 12               | 18               |
| 6          | Himb [mm]                   | 100              | 150              |
| 7          | H2 (H reverse drawing) [mm] | 20               | 30               |
| 8          | L [mm]                      | 65               | 130              |
| 9          | Rp [mm]                     | 10               | 25               |
| 10         | Rm [mm]                     | 10               | 20               |

Where (Vermeulen 2001):

- “Ai” values are the blankholder forces applied by each actuator;
- “Hpre-bulging” is the value of the preforming height;
- “Himb” is the maximum drawing depth;
- “Rp” and “Rm” are punch and die radius, respectively;
- “H2” and “L” are geometric parameters necessarily defined in order to fully describe the geometric profile of the formed part.

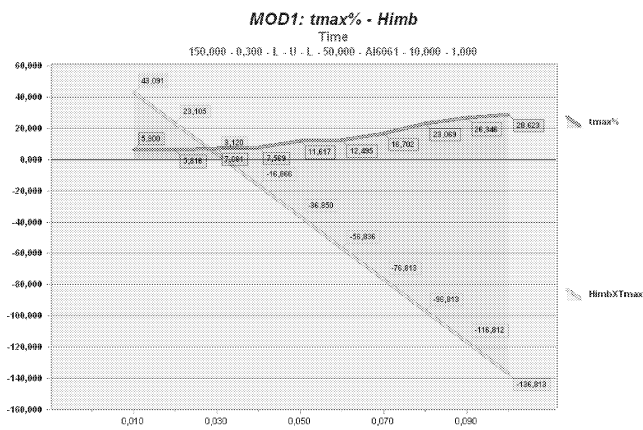
### Results analysis

Using the chosen DSS, for each considered model, it has been possible to develop four kinds of analysis: a) simulations analysis based on their temporal lapse; b) simulations analysis focalized on special issues of technological interest; c) Prediction analysis; d) Cross-model analysis.

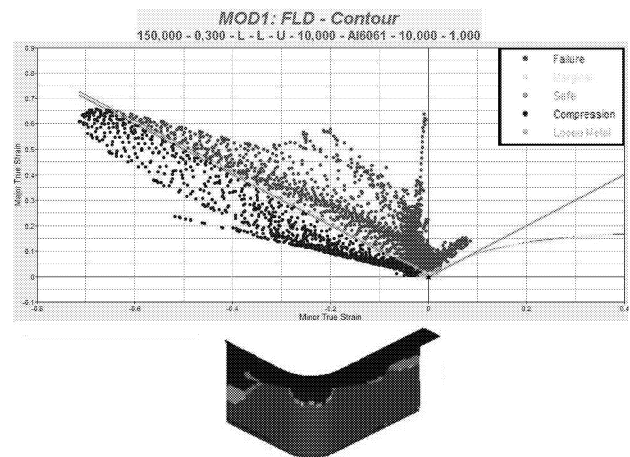
In particular, for each class of models, after having opportunely constructed the database, a set of diagrams have been realized:

- Maximum thickness reduction and maximum drawing depth as a function of the process time (e.g Fig. 2);
- Evaluation of KPI using the “bulb diagram” format (e.g Fig. 3);
- Maximum feasible drawing depth variation as function of geometrical parameters (like punch and die radius) (e.g Fig. 4);
- Visualization of FLD and Thickness Contour Plots (e.g Fig. 5);
- Maximum thickness reduction and drawing ratio as function of pre-bulging height and blank thickness (e.g Fig. 6).

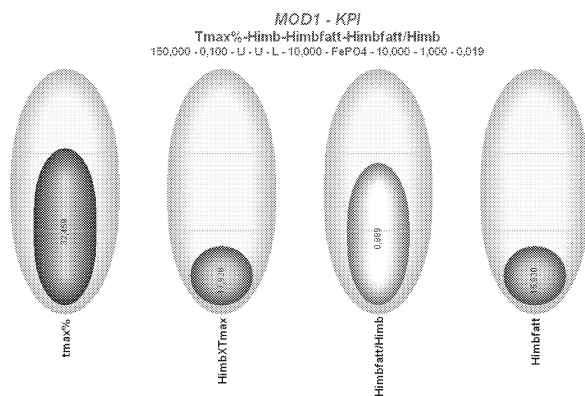




Figures 2: Maximum thickness reduction plotted versus maximum drawing depth for Mod1

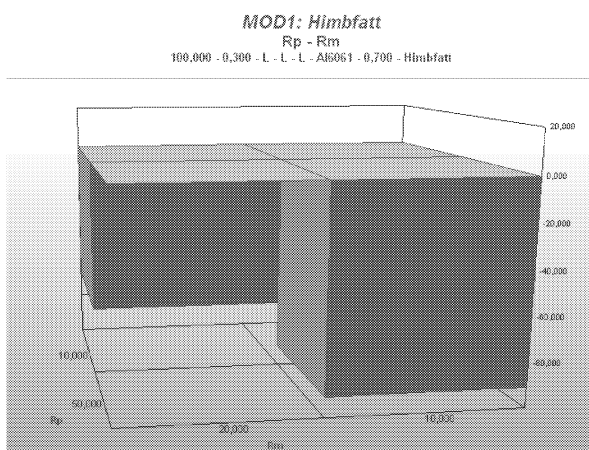


Figures 5: FLD and Thickness Contour Plots for Mod1

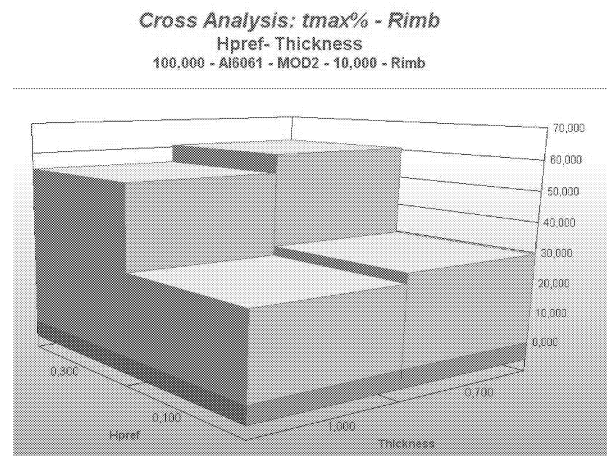


Figures 3: Bulb diagram for Mod1

In the previous diagram (Fig.3) it is possible to define upper and lower bounds for each considered indicator: it allows to easily monitor KPI's even if the user is not a hydroforming expert adopting a "semaphoric representation".



Figures 4: maximum feasible drawing depth plotted versus geometrical parameters for Mod1



Figures 6: Maximum thickness reduction and drawing ratio plotted versus performing height and blank thickness

The innovative character of these representations is due to the extreme easy way to inquire the relationship among the factors that determine the variation of a particular analysis response, without implementing some complex queries in a programming language (like SQL or others).

Moreover, the adopted Decision Support System makes available a powerful method of data navigation through a valid visualization of them that, varying the metadata set, keeps the results publication up-to-date (in a graphical or tabular format).

Furthermore, response surfaces, previously calculated in the I.T.Idro research project, have been inserted in developed OLAP model, giving the right answer to the requirement to obtain informations about not calculated experimental points.

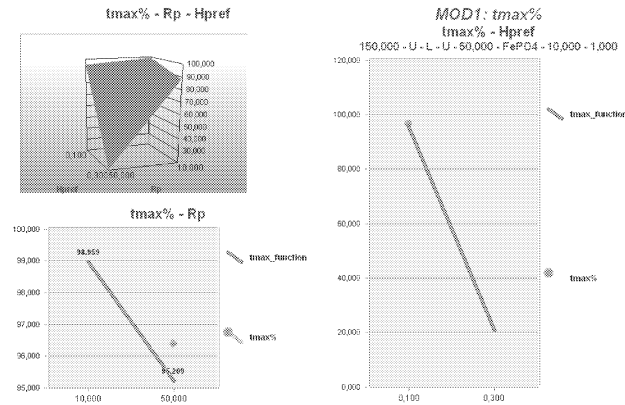
It is important to remind that, according to the developed DOE the response surface for each material and model has been built by using an additive model as follows:

$$Y = \beta_0 + \sum_i \beta_i X_i + \sum_{i \neq j} \beta_{ij} X_i X_j \quad (1)$$

Where Y is the maximum thickness reduction.

After implementing this response surface in HiQube, an appropriate graphic has been created: it shows the maximum thickness reduction as a function of geometrical parameters

(punch radius and preforming height). The following figure shows the response surface and its projection on “tmax% - Rp” and “tmax% - Hpref” plans for Mod1 in the case of low carbon steel.



Figures 7: Response surface for Mod1 in the case of low carbon steel (FeP04)

The yellow points show how the calculated response approximates the real results obtained thanks to the simulations.

During the development of the presented activity a limit on the applicability of some functionalities available in the chosen DSS emerged: the impossibility to have hierarchy data aggregation for the selected test case (and generally speaking for engineering information) doesn't give the opportunity to apply the “Delta Drill” operation, that allows to determine which factors contribute to the variation (positive or negative) of a KPI.

## CONCLUSIONS

Thanks to the developed activity it has been possible to demonstrate how a traditional DSS can be used in a process automation flow, in order to perform a DOE analysis of a non conventional forming process.

The cartesian character of the adopted OLAP database was successfully used to analyze the numerical experimentation campaign, focused on a specific class of models and aimed to infer on process performances even for the not simulated factors combinations.

The developed model could constitute a valid “decision support tool” for the examined forming process: using these systems it is possible to easily increase the process know-how. As soon as the database increases it gradually supplies more and more detailed information about components classes feasibility. This procedure gives the advantage to reduce the analysis field when it has to be defined the optimal set of process parameters for:

1. the production of a new component;
2. the adoption of a different material;
3. the adoption of alternative thicknesses.

It is also noteworthy the possible advantage coming from a structured data management which could be helpful in order to evaluate which is the corporate know how for each specific application.

## ACKNOWLEDGEMENTS

Authors are very grateful to “MUR: Ministero dell'Università e della Ricerca” for funding this work recognized as I.T.Idro: innovative solutions for sheet hydroforming.

Special thanks go to Stamec srl which has the role of industrial partner of the project.

## REFERENCES

- Grotevant, S.; and D. Foth. 1999. "The Power of Multidimensional Analysis (OLAP) in Higher Education Enterprise Reporting Strategies." CUMREC '99, College and University Information Services Conference.
- Rimkus, W.; Bauer H.; and Mihsein M.J.A.. 2000. "Design of load-curves for hydroforming applications", Journal of Materials Processing Technology 108, 2000, pp. 97–105.
- Shulkin L.; R. Posteraro; M. Ahmetoglu; G. Kinzel; and T. Altan. 2000. "Blankholder force (BHF) control in viscous pressure forming (VPF) of sheet metal." Journal of Materials Processing Technology 98 (2000), p. 7-16.
- Siegert, K.; M. Häussermann; B. Lösch; and R. Rieger. 2000 "Recent developments in hydroforming technology." Journal of Materials Processing Technology 98 (2000), p. 251-258.
- Demarest, M.. 2005 "Technology and Policy in Decision Support Systems", DSSResources.COM, 07/08/2005.
- Vermeulen, M.; K. Siegert; and A. Schwager. 2001. "Sheet Metal Formability Test for Hydroforming." Papers of the International Conference on Hydroforming, Fellback (Germany), 6/7 November 2001.
- Del Prete, A.; T. Primo; G. Papadia; and B. Manisi. 2007. "Process Rules for Sheet Metal Hydroforming." ISC 5th International Simulation Conference, Delft, The Netherlands, EUROSIS publication, pp. 109-113.
- Del Prete, A.; A. Elia; T. Primo; and B. Manisi. 2007. "Process automation tools development for sheet metal hydroforming simulation." In ISC'07, 5th International Simulation Conference, Delft, The Netherlands.
- Del Prete, A; T. Primo; and A. Anglani. 2007. "Improvement of Sheet Metal Hydroforming Simulation Reliability." Enhancing the Science of Manufacturing, Convegno AITeM 2007, Montecatini Terme.
- Shi, Z.; Y. Huang; Q. He; L. Xu; S. Liu; L. Qin; Z. Jia; J. Li; H. Huang; and L. Zhao. 2007 "MSMiner—a developing platform for OLAP." Decision Support Systems 42.
- Pérez-Martínez, J.; R. Berlanga-Llavori; M. Aramburu-Cabo; and T. Pedersen. 2007 "Contextualizing data warehouses with documents." Decision Support Systems 45.
- Trujillo, J.; and I. Song. 2007 "New Trends in Data Warehousing and OLAP." Decision Support Systems 45.
- Del Prete, A.; B. Manisi; and M. Strano. 2008. "Sheet metal hydromechanical deep drawing process optimization." Conference NUMISHEET 2008, September 1-15, Interlaken, Switzerland.
- Del Prete, A; A. Anglani; T. Primo; and B. Manisi. 2008. "Numerical and experimental validation for sheet metal hydroforming process rules." 12th International Conference METALFORMING08, September 21-24 2008, Krakow, Poland.
- Del Prete, A; A. Anglani; T. Primo; and A. Spagnolo. 2008. "Computer Aided Simulation as valid tool for sheet hydroforming process development." International Journal of Material Forming, Springer Paris, ISSN 1960-6206 (Print) 1960-6214 (Online), Symposium MS05: Hydroforming, DOI 10.1007/s12289-008-0340-5.

# MEMORY-BASED IMMUNE NETWORK FOR MULTI-ROBOT COOPERATION

Sazalinsyah Razali  
Qinggang Meng  
Shuang-Hua Yang

Department of Computer Science  
Loughborough University  
Loughborough, Leicestershire  
LE11 3TU United Kingdom

E-mail: {S.Razali|Q.Meng|S.H.Yang}@lboro.ac.uk

## KEYWORDS

Multi-robot, Immune Systems, Cooperation, Memory, Simulation

## ABSTRACT

In this paper, basic biological immune systems and their responses to external elements to maintain an organism's health state are fully described. The relationship between immune systems and multi-robot systems are also discussed. Our proposed algorithm is based on immune network theories that have many similarities with the multi-robot systems domain. The paper describes a memory-based immune network that enhance a robot's action-selection process and can obtain an overall a quick group response. The algorithm which is named as *Immune Network T-cell-regulated—with Memory* (INT-M) is applied to the dog and sheep scenario. Simulation experiments are being conducting on the Player/Stage platform and experimental data are being evaluated.

## INTRODUCTION

Usually mobile robots need to interact and engage with one another in order to achieve assigned tasks more efficiently. These autonomous multi-robot systems would be highly beneficial in assisting humans to complete suitable tasks. In such systems, distributed intelligence is highly needed in the team whereby decisions are processed in each individual robots (Parker 1998). Furthermore, these robots would need to have the mechanism to cooperate so that they would achieve the assigned task (Cao et al. 1997).

Biological systems are examples of distributed information processing that are capable of solving problems in living organisms in a distributed manner. Some of these biological systems have neural networks in the brain that is capable of processing information through impulses at the synapses, genetic systems in constructing the organism genes and immune systems which protect and

maintain the homeostatic state of the living organism. Biological immune systems are particularly interesting, not only because they have no central processing but also exhibit cooperative capability among the antibodies in maintaining the internal stable environment of the body.

This leads to the advances in research on Artificial Immune Systems (AIS) and the application of AIS in engineering fields particularly in Multi-Robot Systems (MRS) domain (Cao et al. 1997, Li et al. 2007, Parker 1998). Situations faced by multi-robot systems require real-time processing and response. Furthermore, such situations would also require these systems to be robust to changes in the environment and some unexpected events, such as failure of robots in the team. Thus, mimicking the biological immune system is appropriate.

This paper proposes a memory-enhanced immune system algorithm to achieve cooperative behavior in a team of robots. Using the algorithm inspired by the immune network theory, the robots have the capability of performing their mission in a dynamically changing environment. The proposed algorithm is applied to the dog and sheep scenario (Li et al. 2007, Schultz et al. 1996). Simulation experiments are arranged to investigate the proposed algorithm using the above scenario.

The proposed approach would be suitable in various application domains such as military, disaster rescue operation and even service robotics in domestic environment. These example applications usually require several robots and have continuously changing environments. The test scenario can be extended for the chosen application domain.

## BACKGROUND

This section explains the principle of the biological immune response and the Idiotypic Network Hypothesis which describe the cooperative behaviour achieved by

immune systems in vertebrate organisms. This is followed by the generic relation between immune systems and multi-robot systems.

## Biological Immune Systems

Immune system is a system that eliminates foreign substances from an organism's body. These foreign substances such as bacteria, fungi or virus cells that can harm the host are called pathogens. When such substance activates an immune response it is called *antigen*, which stimulates the system's antibody generation. Each antigen has a unique set of identification on its surface called *epitope*. These antigenic determinants are where the host's antibodies would attach to by using their paratope, as shown in Figure 1. *Antibodies* are cells in the immune system that kill antigens in order to maintain the host homeostatic state—i.e. balancing the body's health status.

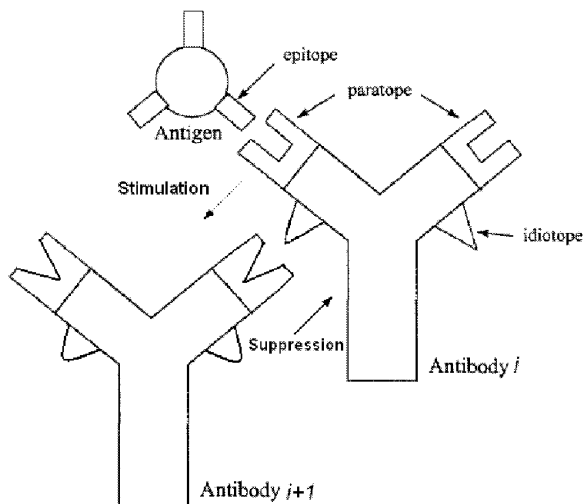


Figure 1: Antigen-antibody binding and Jerne's Idiotypic Network Theory

The immune system can be divided into two general categories, innate immunity and adaptive immunity. Innate immunity is the first line of defense of the immune system. Generic pathogens that can be recognized and killed by the innate immunity cells would not be able to harm the host further. However, certain disease carrying antigens would bypass this defense mechanism because the innate immunity does not adapt to antigens that originate from various types of illnesses. The adaptive immunity would then play its role through the use of *lymphocytes* which are generally known as white blood cells. Lymphocytes have two main types, *T-cells* that mainly help in recognizing antigen cells and *B-cells* that mainly produce antibodies to fight specific antigens. In humans, T-cells are primarily produced in the thymus while B-cells are produced in bone marrows. These innate and adaptive immune responses make up effective

and important defense mechanism for living organisms.

## Immune Response

The immune response basically can be viewed in six phases of recognition and activation, as seen in Figure 2. Pathogen is digested by Antigen Presenting Cells (*APCs*) where it is broken down into *peptides* (de Castro and Timmis 2002). These peptides will then bind to Major Histocompatibility Complex (*MHC*) molecules, then presented on the APC surface. T-cells recognize these different APC receptors and thus become activated. They divide and release *lymphokines* that transmit chemical signals to stimulate other immune system components to take action. B-cells would then travel to the affected area and be able to recognize the antigen. This would activate the B-cells which then mature into *plasma cells*. Plasma cells are the ones which release specific antibody molecules that neutralize the particular pathogens.

This immune response cycle results in the host's immunity against the antigen which triggers it, thus having protection in future attacks (de Castro and Timmis 2002). Prominent characteristics of the immune system is that there is no central control of the lymphocytes in fighting antigens that invade the host and the system's adaptability in responding to various kind of antigens. The B-cells cooperatively merge at the affected area and produce appropriate antibodies for that particular situation. This phase of immune response exhibits cooperative behavior of the related cells.

## Idiotypic Network Hypothesis

Studies in immunology have suggested that antibodies are not isolated but they 'communicate' with each other. Each type of antibody has its specific *idiotope*, an antigen determinant as shown in Figure 1. Jerne (1984) who is an immunologist proposed the Idiotypic Network Hypothesis (also known as Idiotypic Network Theory) which views the immune system as a large-scale closed system consisting of interaction of various lymphocytes (i.e. B-cells). Referring to Figure 1, idiotope of antibody *i* stimulates antibody *i+1* through its paratope. Antibody *i+1* views that idiotope (belonging to antibody *i*) simultaneously as an antigen. Thus, antibody *i* is suppressed by antibody *i+1*. These mutual stimulation and suppression chains between antibodies form a controlling mechanism for the immune response (de Castro and Timmis 2002).

Farmer et al. (1987) proposed differential equations of Jerne's idiotypic network theory. These equations consist of antibodies' stimulus and suppression terms, antigen-antibody affinity, and cell's natural mortality rate (Farmer et al. 1987). This large-scale closed system

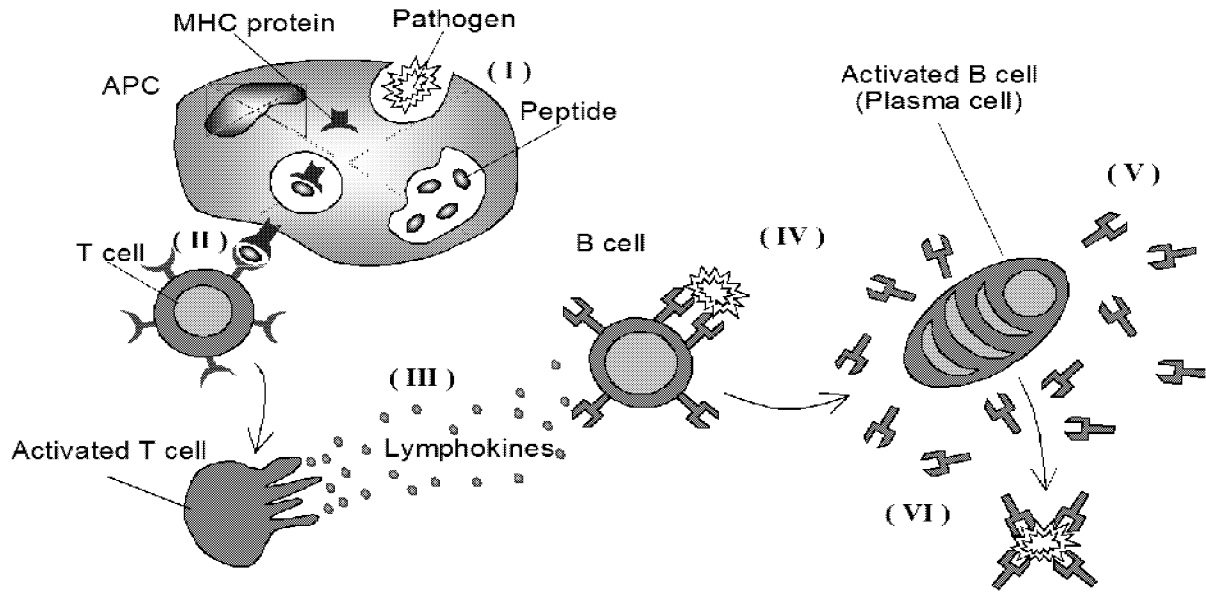


Figure 2: Basic biological immune systems response (de Castro and Timmis 2002)

interaction is the main mechanism that can be used for cooperation of multi-robot systems.

### Immune Systems and MRS

The relationship of the immune systems with multi-robot systems is evident where obstacles, robots and their responses are antigens, B-cells and antibodies respectively. Table 1 lists the obvious parallel of MRS and immune systems terminologies.

Table 1: Immune Systems and MRS relationship

| Immune Systems   | Multi-Robot Systems          |
|------------------|------------------------------|
| B-cell           | Robot                        |
| Antigen          | Robot's Environment          |
| Antibody         | Robot's action               |
| T-cell           | Control parameter            |
| Plasma cell      | Excellent robot              |
| Inactivated cell | Inferior robot               |
| Immune network   | Robots interaction           |
| Stimulus         | Adequate robot stimulation   |
| Suppression      | Inadequate robot stimulation |

Immune network theory as previously described is suitable as a basis for emulating cooperative behavior in a multi-robot environment. This is because the immune network uses affinity measures that are dependent on other cells concentration and location in determining the next action. Other than that, multi-robot systems require recognition ability of obstacles and other robots, which is parallel to the immune system recognition and activation phase of an immune response. Obviously, in immune network the processing of information is done

in real-time and in a distributed manner—as what a multi-robot system requires.

### IMMUNE NETWORK BASED MULTI-ROBOT COOPERATION

Sun et al. (2001) have proposed a model based on Farmer's immune network equation that involves T-cells as control parameter which provides adaptation ability in group behaviour.

The group control or coordination phase is done in a distributed manner via local communication between nearby robots. When a robot encounters other robot and both have the same or similar strategy, this strategy is stimulated; if not, the strategy is suppressed. This facilitates the group to self-organize towards a common action which is optimal for the local environment. If a robot is stimulated beyond a certain threshold—which makes it an excellent robot, its behaviour is regarded as adequate in the system such that it can transmit its strategy to other inferior robots. This is a metaphor of the plasma cell in the biological immune systems.

The advantage of adding the T-cell model is that the system adapts quickly to the environment by recovery of antibody concentration to the initial state, when antigens have successfully been removed. Thus, the system is more adaptable to environmental changes.

Our proposed approach is based on Sun et al. (2001) work, with the extension of Memory ability so that quick responses can be achieved in certain relevant situation.

## IMMUNE NETWORK WITH MEMORY

In biological immune response, there is a Clonal Selection process, whereby various B-cells try to identify the antigen. Once the appropriate B-cell is selected, it is activated and multiplied (i.e. proliferate) so that adequate immune response could be mounted later. The activated B-cells will proliferate and differentiate into Plasma cells that will secrete specific antibodies and memory cells which will be in the host body for quite a long time (de Castro and Timmis 2002). These memory cells will act as catalysts in mounting a quick immune response to the same antigen in the future.

### The INT-M Model

In order to improve the approach by Sun et al. (2001), a specific memory mechanism is proposed in order to retain the appropriate action for relevant environment condition. This mechanism is introduced when the newly sensed environment is similar to the previous environment. Thus, a quick action-selection process can be executed without the need of re-evaluating the new situation.

The approach is aptly named as *Immune Network T-cell-regulated—with Memory* (INT-M) as it involves modelling the memory part of the biological immune systems. The general algorithm is shown in algorithm 1 which is an extension of Sun et al. (2001). The algorithm being displayed is for each robot in the group, and uses Equations (2), (3) and (4).

$$A(t) = \sum_{j=0}^{N-1} (m_{ij} - m_{ji}) s_j(t-1) \quad (1)$$

$$S_i(t) = S_i(t-1) + \left( \alpha \frac{A(t)}{N} + \beta g_i - c_i(t-1) - k_i \right) s_i(t-1) \quad (2)$$

$$s_i(t) = \frac{1}{1 + \exp(0.5 - S_i(t))} \quad (3)$$

$$c_i(t) = \eta (1 - g_i(t)) S_i(t) \quad (4)$$

In equations (2) and (3),  $S_i(t)$  is the stimulus value of antibody  $i$  where  $i, j = 0 \dots N$ ,  $N$  is the number of antibody types.  $m_{ij}$  is the mutual stimulus of antibody  $i$  and  $j$ , which is referred to in Table 2.  $g_i$  is the affinity of antibody  $i$  and antigen, which can arbitrarily be assigned using a function.  $s_i(t)$  is the concentration of antibody  $i$ . The difference with Farmer et al. (1987) immune network equation is that  $s_j(t)$  is not the concentration of self-antibody, but that of other robot's antibody obtained by communication.

Equation (4) is the T-cell model whereby  $c_i(t)$  is the concentration of T-cell which controls the concentration

Table 2: Mutual stimulus coefficient,  $m_{ij}$

| robot $i \setminus$ robot $j$      | <b>Ab<sub>0</sub></b> | <b>Ab<sub>1</sub></b> | <b>Ab<sub>2</sub></b> | <b>Ab<sub>3</sub></b> |
|------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Aggregation, <b>Ab<sub>0</sub></b> | 1                     | -0.4                  | -0.2                  | -0.4                  |
| Search, <b>Ab<sub>1</sub></b>      | -0.4                  | 1                     | -0.4                  | -0.2                  |
| Dispersion, <b>Ab<sub>2</sub></b>  | -0.2                  | -0.4                  | 1                     | -0.4                  |
| Homing, <b>Ab<sub>3</sub></b>      | -0.4                  | -0.2                  | -0.4                  | 1                     |

of antibody  $i$ .  $\alpha$ ,  $\beta$ , and  $\eta$  are constants, whereby  $\alpha$  and  $\beta$  are parameters of response rate of other robot and the environment (antigen) respectively. In biological immune systems, helper T-cells activate B-cells when antigen invades, and suppressor T-cells prevent the activation of B-cells when the antigen has been eliminated thus ensuring that the system adapts quickly to the environment by recovery of antibody concentration to the initial state.

Equations (5) and (6) are the functions and its corresponding values for the upper ( $\bar{\tau}$ ) and lower ( $\underline{\tau}$ ) thresholds in determining whether a robot becomes an excellent (i.e. plasma cell) or an inferior (i.e. inactivated cell) robot.

$$\bar{\tau} = \frac{1}{1 + e^{-0.5}} = 0.622 \quad (5)$$

$$\underline{\tau} = \frac{1}{1 + e^{0.5}} = 0.378 \quad (6)$$

## SIMULATION

In this research we investigate shepherding behavior of robots. Shepherding behavior is similar to a flocking behavior but having agents/robots outside of the flock guiding or controlling the members (Lien et al. 2004). Figure 3 shows the screenshot of the dog and sheep scenario.

In a dog and sheep problem, a few dogs try to guide a few sheep to the grazing site (also called the safety zone) without going beyond the borders (Schultz et al. 1996). Dogs are required to cooperate in shepherding the sheep which are moving away from the dogs or wandering randomly inside the area. The objective is to prevent the sheep from going out of the grazing site while having partial information of what is happening in the area.

This problem is highly dynamic and obviously requires the robots to have real-time processing of partial information of the environment. The robot dogs use the proposed immune-inspired approach in cooperating with one another while the robot sheep have basic avoidance and flocking behaviours.

The proposed approach as described in algorithm 1 is applied to the dog and sheep problem and adjusted

---

**Algorithm 1** Immune Network T-cell-regulated—with Memory (INT-M)

---

**Require:**  $t = 0$ ,  $S_i(0) = s_i(0) = 0.5$  for  $i = 0 \dots N - 1$ ,  
 $N$  is number of actions

**Ensure:** retain previous  $Ab$  if robot is not inferior within similar environment, execute  $Ab_{max}$

$Ab_{max} \leftarrow Ab_1$   
robot  $\leftarrow$  inferior  
environment  $\leftarrow$  similar

**loop**

Execute  $Ab_{max}$

{robot is activated (normal) or excellent}

**if** robot  $\neq$  inferior **then**

{environment sensed is *similar* to previous}

**if**  $g_i(t) \approx g_i(t - 1)$  **then**

$S_i(t) \leftarrow S_i(t - 1)$

$s_i(t) \leftarrow s_i(t - 1)$

$c_i(t) \leftarrow c_i(t - 1)$

**else**

environment  $\leftarrow$  changed

**end if**

**end if**

{robot is inferior or environment has changed}

**if** (robot=inferior) || (environment=changed) **then**

**for**  $i \leftarrow 0$  to  $N - 1$  **do**

Calculate  $S_i(t)$

Calculate  $s_i(t)$

Calculate  $c_i(t)$

**end for**

**if**  $S_i(t) > \bar{\tau}$  **then**

robot  $\leftarrow$  excellent

**else if**  $S_i(t) < \underline{\tau}$  **then**

robot  $\leftarrow$  inferior

**if** robot encounter  $robot_{excellent}$  **then**

**for all**  $i$  **do**

receive  $Ab_i$

renew  $s_i(t)$

**end for**

**end if**

**end if**

**end if**

**if**  $Ab_i$  has  $\max(s_i(t))$  **then**

$Ab_{max} \leftarrow Ab_i$

**end if**

$t \leftarrow t + 1$

**end loop**

---

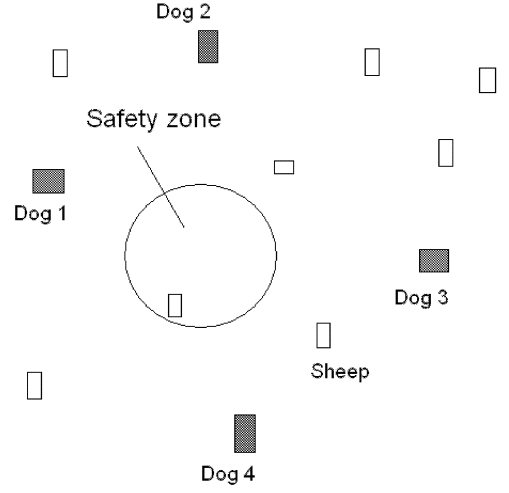


Figure 3: The Dog and Sheep problem environment

where necessary. The Player/Stage simulation platform (Gerkey et al. 2003) on a Fedora Core 6 Linux operating system is being used to test the proposed algorithm. Figure 4 shows a sample screenshot of the simulation platform. Experimental data is currently being collected to analyze the behaviours of the simulated robots.

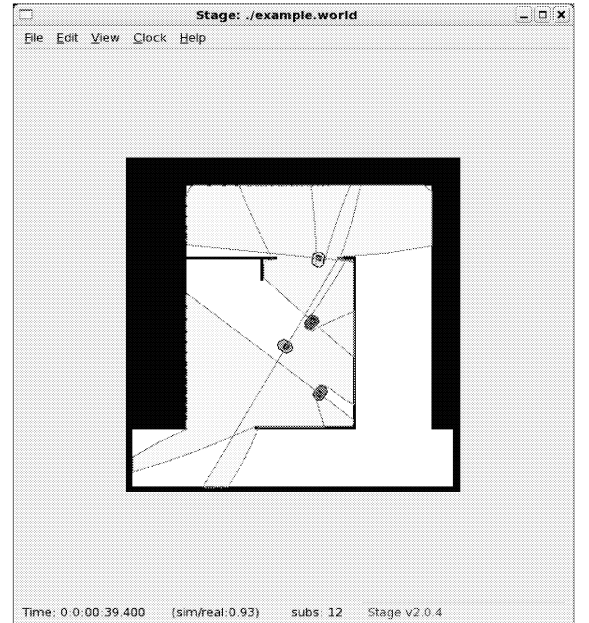


Figure 4: The Player/Stage simulation platform

A distinct part of this study is that we are looking into the memory-based immune network cooperation approach by the robots (i.e. dogs) in maintaining the herd (i.e. sheep). This utilizes the advantage of memory in the action-selection phase and affects the resulting dynamic behaviour of both the robot dogs and the

robot sheep. The dog-sheep scenario partly contains the pursuit-evasion problem and can be further applied to other robot coordination problems, such as robot formation and the likes.

## CONCLUSIONS

In this paper a memory-based immune system inspired approach for cooperation in multi-robot systems has been proposed. We have described the basic concepts and mechanisms of biological immune systems, and argued that the immune network is a suitable analogy for multi-robot cooperation problem. We have also proposed a multi-robot cooperation algorithm—the INT-M model, and applied to the dog-sheep test scenario. An experimental simulation environment has been setup to evaluate the proposed approach and algorithm. The approach can be easily extended to other application domains which require several agents (robots) to work cooperatively in a distributed way in a dynamic environment.

## ACKNOWLEDGEMENTS

Mr. Razali gratefully acknowledges Malaysian Ministry of Higher Education and Universiti Teknikal Malaysia Melaka (UTeM) for sponsoring his PhD study. The authors thank the reviewers for their constructive comments.

## REFERENCES

- Cao Y.U.; Fukunaga A.S.; and Kahng A., 1997. *Cooperative Mobile Robotics: Antecedents and Directions. Autonomous Robots*, 4, no. 1, 7–27.
- de Castro L.N. and Timmis J., 2002. *Artificial Immune Systems: A New Computational Intelligence Approach*. Springer.
- Farmer J.D.; Kauffman S.A.; Packard N.H.; and Perelson A.S., 1987. *Adaptive Dynamic Networks as Models for the Immune System and Autocatalytic Sets. Annals of the New York Academy of Sciences*, 504, no. 1, 118–131.
- Gerkey B.; Vaughan R.T.; and Howard A., 2003. *The player/stage project: Tools for multi-robot and distributed sensor systems. In Proceedings of the 11th International Conference on Advanced Robotics*. 317–323.
- Jerne N.K., 1984. *Idiotypic Networks and Other Preconceived Ideas. Immunological Reviews*, 79, no. 1, 5–24.
- Li J.; Xu H.; Wang S.; and Bai L., 2007. *An Immunology-based Cooperation Approach for Autonomous Robots. In Proceedings of the 2007 International Conference on Intelligent Systems and Knowledge Engineering (ISKE 2007)*. vol. 4. doi: 10.2991/iske.2007.42.
- Lien J.M.; Bayazit O.B.; Sowell R.T.; Rodriguez S.; and Amato N.M., 2004. *Shepherding behaviors. In Proceedings of the 2004 IEEE International Conference on Robotics and Automation (ICRA 2004)*. vol. 4, 4159–4164. doi:10.1109/ROBOT.2004.1308924.
- Parker L.E., 1998. *ALLIANCE: an architecture for fault tolerant multirobot cooperation. IEEE Transactions on Robotics and Automation*, 14, no. 2, 220–240.
- Schultz A.; Grefenstette J.J.; and Adams W., 1996. *Roboshepherd: Learning a complex behavior. Robotics and Manufacturing: Recent Trends in Research and Applications*, 6, 763–768.
- Sun S.J.; Lee D.W.; and Sim K.B., 2001. *Artificial immune-based swarm behaviors of distributed autonomous robotic systems. In Proceedings of the 2001 IEEE International Conference on Robotics and Automation (ICRA 2001)*. vol. 4, 3993–3998. doi: 10.1109/ROBOT.2001.933241.



# A fuzzy randomness approach for preventive and corrective maintenance

F. Aguirre  
LGIPM, ENIM, Ile de Saulcy  
57045 Metz Cedex 01, France.  
Universidad EAFIT  
Medellin, Colombia.  
e-mail: felipeam86@gmail.com

S. Hennequin  
LGIPM, ENIM, Ile de Saulcy  
57045 Metz Cedex 01, France  
e-mail: hennequin@enim.fr

N. Rezg  
LGIPM, Université de Metz,  
Ile de Saulcy-57045  
Metz Cedex 01, France  
e-mail: rezg@univ-metz.fr

## KEYWORDS

Fuzzy weibull distribution, Reliability, Fuzzy numbers, Corrective maintenance and Preventive maintenance

## Abstract

Uncertainty in a machine reliability is commonly present, but not always completely modeled. In general, statistical laws like Weibull, normal or exponential probability distribution are used, but they only model randomness of the reliability and not the fuzziness. Additionally, the systems parameters like the costs and durations of maintenance, are used to be taken like real constants without taking into account the unsharpness and imprecision of these parameters. In this paper, the concept of fuzzy random probability distribution is introduced to the maintenance model to cover in a wider level the uncertainty of the reliability of the machines. This method gives a better representation of the randomness as well as the fuzziness of the reliability. The systems parameters are modeled not as real constants but as fuzzy constant numbers to introduce their unsharpness. Finally the goal is to minimize the total cost of maintenance per unit time, by finding the optimal age at which preventive maintenance must be performed in a fuzzy environment.

## INTRODUCTION

The determination of the time between preventive maintenance interventions usually concerns the engineer as to minimize the long term cost of maintenance (Al-Najjar & Alsayouf, 2002). When interventions are too close together the cost can be fairly high because there may be more interventions than necessary; or when interventions are too separated, the cost can also increase because the machine can fail more often, causing more reparations to be made which are considerably more expensive than preventive interventions. To overpass this situation, appropriate maintenance models and suitable matched computational models must be applied to try to describe the behavior of a system as close to reality as possible (Nakagawa, 1942). These models are then used

to optimize the system usually by finding the correct time between preventive interventions that minimizes the long term cost of maintenance per time unit.

These models and their parameters are usually established on the basis of observations, experiences, expert knowledge, codes, standards, use of statistics, etc. All of this is surrounded by uncertainty that is normally created by human mistakes (Hennequin and al.), hypothesis made, errors in the manufacture, interventions not well done, lack of information, etc. In general there would never be certain and precise information (Möller and al., 2003). When talking about uncertainty we have to understand that there are two types or sources of uncertainty as described by A. F. Shapiro (2008): randomness and fuzziness. The first one is usually modeled with statistical laws (like the Weibull probability distribution for the maintenance model presented here) but the second one is rarely taken into account in the literature. Randomness models the stochastic variability of all possible outcomes of a situation, and fuzziness relates to the unsharp boundaries of the parameters of the model.

The present study focuses on a maintenance policy for a failure prone manufacturing system composed of one machine. The production unit is submitted to a maintenance action as soon as it reaches a certain age  $m$  or at failure, whichever comes first. A mathematical model of the cost function is established where the decision variable is the critical preventive maintenance age  $m$ . The objective being to minimize the total expected cost per time unit. Additionally, this paper presents a method for appropriately taking into account uncertainty in the definition of this model as well as of their parameters. The method is adapted from the work done by Möller & Beer (2004).

To better understand the approach presented in this paper, the concept of fuzzy number, fuzzy random variable (FRV) and fuzzy random probability distribution must be introduced.

As described by Dubois & Prade (1978), a fuzzy number is a fuzzy subset (Zadeh, 1965) of the real line  $\mathbb{R}$  whose highest membership values are clustered around a given real number called the mean value; the member-

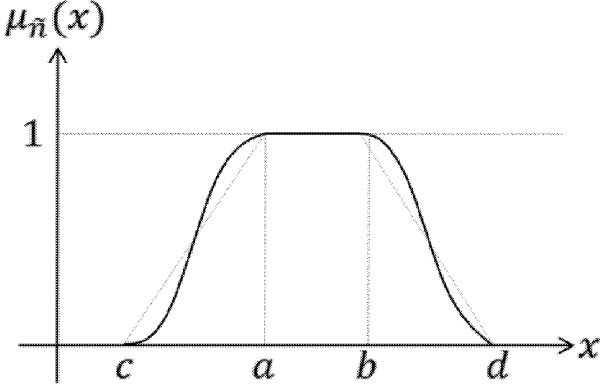


Figure 1: Fuzzy number.

ship function is monotonic on both sides of this mean value and is:

- (i) A continuous mapping from  $\mathbb{R}$  to the closed interval  $[0, 1]$ .
- (ii) Constant on:  
 $(-\infty, c) : \mu_{\tilde{n}}(x) = 0 \forall x \in (-\infty, c)$
- (iii) Strictly increasing on  $[c, a]$
- (iv) Constant on:  
 $[a, b] : \mu_{\tilde{n}}(x) = 1 \forall x \in (a, b)$
- (v) Strictly decreasing on  $[b, d]$
- (vi) Constant on:  
 $(d, +\infty) : \mu_{\tilde{n}}(x) = 0 \forall x \in (d, +\infty)$

$a, b, c$  and  $d$  are real numbers. Eventually we can have:

$$c = -\infty, \text{ or } a = b, \text{ or } c = a, \text{ or } b = d, \text{ or } d = +\infty$$

The mean value  $m$  of the fuzzy number would be an element of  $[a, b]$ , often  $m = (a + b)/2$ . The following cases must be remarked:

- If  $a = b = c = d$ , is an ordinary real number.
- If  $a = b$ , is the representation of a fuzzy number, the value of which is “approximately  $a$ ”.
- $\mu_{\tilde{n}}$  is the truth value of the assertion “the value of  $\tilde{n}$  is  $x$ ”.
- Two fuzzy numbers are equal if and only if they have the same membership functions.

Fuzzy numbers are used to be represented as “ $L$ - $R$  fuzzy numbers” (Dubois & Prade, 1978). Let  $L$  be the restriction of on  $[c, a]$ ,  $L$  is continuous and strictly increasing in that interval. Let  $R$  be the restriction of on  $[b, d]$ ,  $R$  is continuous and strictly decreasing in that interval. Now, an  $L$ - $R$  fuzzy number can be defined as follows:

$$\mu_{\tilde{n}} = \begin{cases} L\left(\frac{a-x}{\beta}\right) & c < x < a, \beta > 0 \\ 1 & a \leq x \leq b \\ R\left(\frac{x-b}{\delta}\right) & b < x < d, \delta > 0 \end{cases} \quad (1)$$

$\beta$  and  $\delta$  are called respectively left and right spreads of and will be called an  $L$ - $R$  fuzzy number. Our interest for this type of fuzzy numbers is that the elemental operations like addition and multiplication does not depend on the analytical expressions of  $L$  and  $R$  when we work with the same type of functions for all our fuzzy numbers. We are only interested in the values of  $a, b, c$  and  $d$  of each fuzzy number, and at the end of the calculation we use those expressions to represent the results.

The first definition of FRVs is due to Kwakernaak (1978, 1979) who introduced them as “random variables whose values are not real but fuzzy numbers”. Later on Kruse & Meyer (1987) worked on an expanded version of a similar model. To better illustrate the concept of FRVs we present the example given by Kwakernaak (1978). Consider an opinion poll, during which a number of individuals are questioned on their opinion concerning the weather in Europe in a particular summer. The responses are classified into three categories, respectively characterized as “very warm”, “warm”, and “no opinion”. Randomness occurs because it is not known which response may be expected from any given individual. Once the response is available, there still is uncertainty about the precise meaning of the response. The latter uncertainty will be characterized by fuzziness, since each of the responses very warm, warm, and no opinion will be represented by a fuzzy set, in particular by a fuzzy number.

In our case, the randomness occurs because the likeliness of a machine failing at a given time is unknown. On the other hand, fuzziness occurs because the expected type of failure is unknown, thus the cost of maintenance and the time to repair are fuzzy.

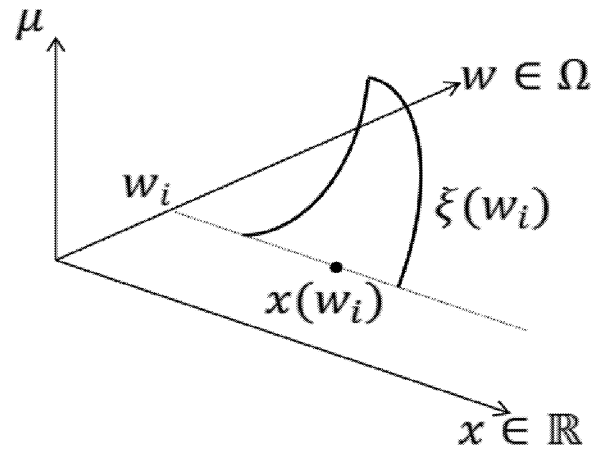


Figure 2: Model of a fuzzy random variable (FRV)

As described by Shapiro (2008) operationally speaking, a FRV is a RV taking fuzzy values, so we start with a probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  and let  $u_1, u_2, \dots, u_n$  be fuzzy variables. Then, for each event  $w_i$  in  $\Omega$ ,  $\xi(w)$  is a FRV, where  $\xi(w) = u_i$  if  $w = w_i$ ,  $i = 1, 2, \dots, n$ . In Figure 2, the real-valued realization is represented by  $x(w_i)$ , while the fuzzy realization is represented by  $\xi(w_i)$ .

As presented by Möller & Beer (2004) the potential for FRVs becomes more apparent when we explore probability distributions. Here we won't get a crisp value of the probability for a given event but a fuzzy value of the probability. In our case the probability distribution is the reliability of the machine and it follows a weibull distribution. The fuzziness is introduced in the shape and scale fuzzy parameters, which are defined by triangular fuzzy numbers (TFN).

As fuzzy parameters are considered instead of real crisp parameters, the reliability, that will also be fuzzy, takes the same shape for the membership function as the parameters, in this case a triangular distribution. The fuzzy weibull probability distribution is then defined by the following equation:

$$\tilde{R}(x) = e^{-\left(\frac{x}{\tilde{a}}\right)^{\tilde{b}}} \quad (2)$$

An example is presented in Figure 3 for a fuzzy scale parameter  $\tilde{a} = (90, 100, 110)$  and a fuzzy shape parameter  $\tilde{b} = (1.8, 2, 2.2)$ .

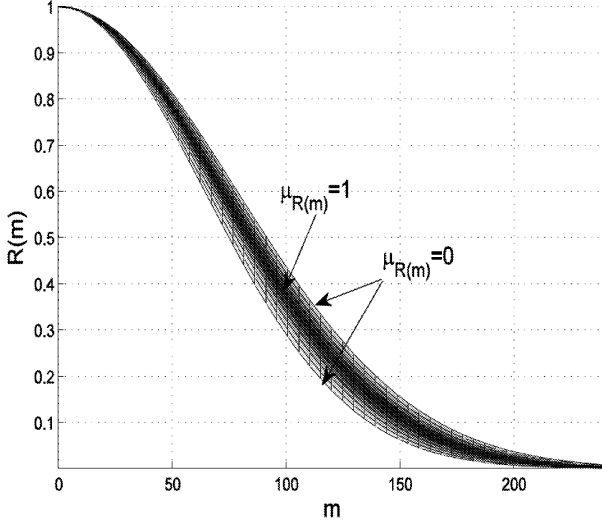


Figure 3: Fuzzy weibull probability distribution

## STUDIED SYSTEM AND MATHEMATICAL MODEL

The manufacturing system is a failure-prone system composed of one machine. The maintenance strategy under consideration is the age-based preventive maintenance policy (Barlow & Proschan, 1996). It consists

of submitting the machine to a preventive maintenance action as soon as it reaches a certain age  $m$  or at failure, whichever comes first. In this paper the following hypothesis are considered:

- The production cycle is divided in two: working periods TBM (Time Between Maintenance), before the start of a maintenance action (preventive or corrective), and a maintenance period, TTR (Time To Repair).
- The costs of preventive and corrective maintenance are known and would be treated as constant fuzzy numbers.
- Each maintenance action leaves the machine as new.
- Failures are detected instantaneously.
- All of the resources needed to perform the maintenance actions are available at the right time and place.
- The maintained system has an increasing failure rate.

The following notation will be used in this paper:

|              |   |
|--------------|---|
| $C_{cm}$     | Corrective maintenance cost   |
| $C_{pm}$     | Preventive maintenance cost   |
| $\mu_c$      | Average duration of corrective maintenance action                           |
| $\mu_p$      | Average duration of preventive maintenance action                           |
| $N_c(t)$     | Quantity of corrective maintenance actions during $(0, t)$                  |
| $N_p(t)$     | Quantity of preventive maintenance actions during $(0, t)$                  |
| $F(t)$       | Probability distribution function associated to the machine time to failure |
| $R(t)$       | Reliability function associated to the machine ( $R(t) = 1 - F(t)$ )        |
| $\varphi(t)$ | Cost of maintenance per time unit   |
| $m$          | Age at which preventive maintenance must be performed                       |

The corresponding total cost of maintenance over the interval of time  $(0, m)$  is given by:

$$CT_m(m) = C_{cm} \cdot N_c(m) + C_{cp} \cdot N_p(m)$$

And the total expected cost per time unit  $\varphi$ , is given by:

$$\varphi(m) = \lim_{t \rightarrow \infty} \left\{ \frac{E[CT_m]}{t} \right\}$$

By applying the renewal theory (Cox 1962) we obtain the explicit form of  $\varphi(m)$ :

$$\varphi(m) = \frac{C_{cm} \cdot F(m) + C_{pm} \cdot R(m)}{\int_0^m R(u)du + \mu_p \cdot R(m) + \mu_c \cdot F(m)} \quad (3)$$

Normally, the next step to follow after we have the model is to find the optimal value for  $m$  that we will call from now on  $m^*$ . As the machine has an increasing failure rate, there is a finite unique optimal strategy  $m^*$  that minimizes  $\varphi(m)$ .

In this paper we will use this model as a basis to represent the system in a fuzzy environment by introducing the concept of fuzzy random probability distribution and fuzzy parameters.

Now that we introduced the concept of fuzzy random probability distribution, we can introduce it to the cost of maintenance from equation (3) as well as the fuzzy maintenance parameters, to complete the representation of fuzziness and randomness. So we obtain the new fuzzy cost of maintenance:

$$\varphi(m) = \frac{\tilde{C}_{cm} \cdot \tilde{F}(m) + \tilde{C}_{pm} \cdot \tilde{R}(m)}{\int_0^m \tilde{R}(u)du + \tilde{\mu}_p \cdot \tilde{R}(m) + \tilde{\mu}_c \cdot \tilde{F}(m)} \quad (4)$$

To resume we have a fuzzy cost of maintenance that for a given set of fuzzy parameters ( $\tilde{C}_{cm}$ ,  $\tilde{C}_{pm}$ ,  $\tilde{\mu}_p$ ,  $\tilde{\mu}_c$ ) and a given fuzzy random reliability distribution varies depending on the age at which maintenance interventions are performed. The parameters are triangular L-R fuzzy numbers. The reliability follows a fuzzy random weibull distribution with triangular L-R fuzzy numbers for the scale and shape factors.

## APPLICATION

To solve the problem of finding the optimal age  $m^*$  we have developed a MATLAB code. The algorithm uses the concepts of fuzzy operations found in Dubois & Prade (1978) to resolve the fuzzy cost function. What we obtain is a fuzzy function like the one shown in Figure 4.

As the values of cost are fuzzy numbers, our optimal solution will also be one. Thereby we would get a triangular fuzzy number with the mean value having a grade of possibility of 1 and the other values of the fuzzy set having grades of possibilities defined by the L-R fuzzy membership functions.

For the convenience of our optimization procedure, we divide our fuzzy cost functions in three different functions that later on would be optimized:

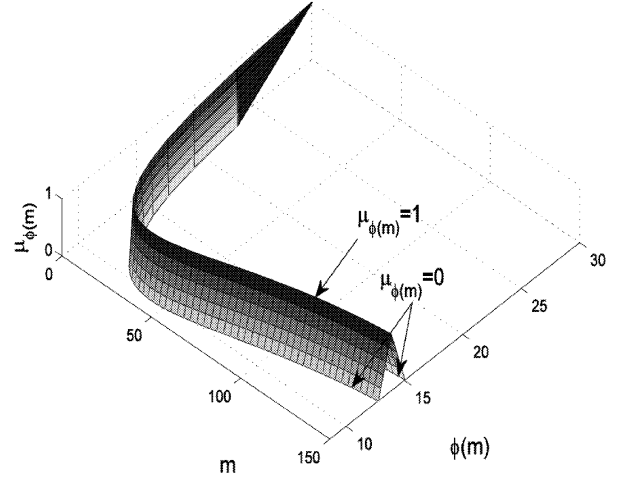


Figure 4: Fuzzy cost of maintenance per time unit.

$$\tilde{\varphi}(m) = [\tilde{\varphi}_L(m), \tilde{\varphi}(m), \tilde{\varphi}_R(m)]$$

Where:

- $\tilde{\varphi}_L(m)$  Is the function that describes the values of cost for the left spread of each fuzzy cost realization with  $\mu_{\tilde{\varphi}(m)} = 0$
- $\tilde{\varphi}(m)$  Is the function that describes the values of cost for each fuzzy cost realization with  $\mu_{\tilde{\varphi}(m)} = 1$
- $\tilde{\varphi}_R(m)$  Is the function that describes the values of cost for the right spread of each fuzzy cost realization with  $\mu_{\tilde{\varphi}(m)} = 0$ .

Now that we have divided the fuzzy cost function, we minimize each of the three functions and we obtain the optimal fuzzy cost  $\tilde{\varphi}^*$ :

$$\tilde{\varphi}^*(m) = [\min(\varphi_L(m)), \min(\varphi(m)), \min(\varphi_R(m))]$$

For this optimal fuzzy cost there would also be an associated optimal fuzzy age of intervention  $m^*$ :

$$m^* = (m_L, \bar{m}, m_R)$$

Two cases were analyzed to study the influence of the fuzzy parameters in the model.

1. First, we introduced only the fuzzy random weibull distribution (FRWD) and left the maintenance parameters as crisp real numbers (Figure 5 and 6).
2. Then we introduced the FRWD as well as the fuzzy maintenance parameters (Figure 7 and 8).

For both cases we used the following input data:

$$\tilde{a} = (90, 100, 110)$$

$$\tilde{b} = (1.8, 2, 2.2)$$

$$\mu_c = (45, 50, 55)$$

$$\mu_p = (9, 10, 11)$$

$$C_{cm} = (1800, 2000, 2200)$$

$$C_{pm} = (270, 300, 330)$$

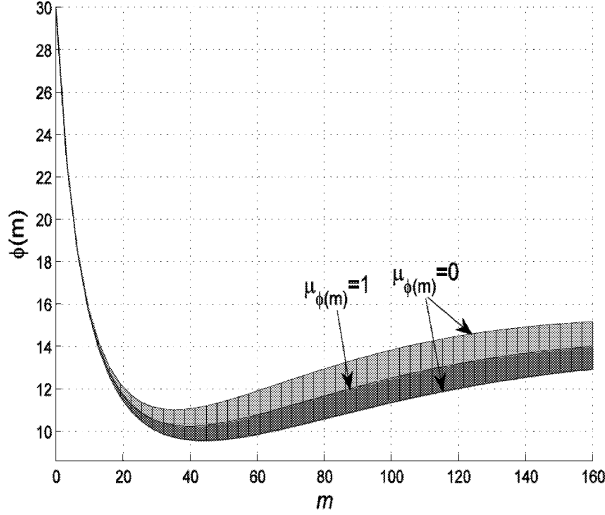


Figure 5: Fuzzy cost function for the first case.

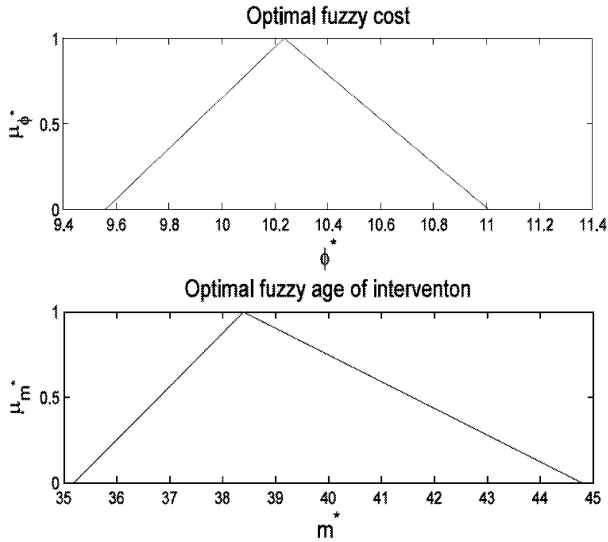


Figure 6: Optimal fuzzy cost and age of intervention for the first case.

For the first case it could be noticed that the fuzzy cost becomes fuzzier when we increase  $m$ . In the same way that we introduced uncertainty into our input data, our output data has as well some uncertainty, so we can interpret our optimal cost as been approximately 10.23 (as we can see in Figure 6) for an age of intervention of approximately 38.3.

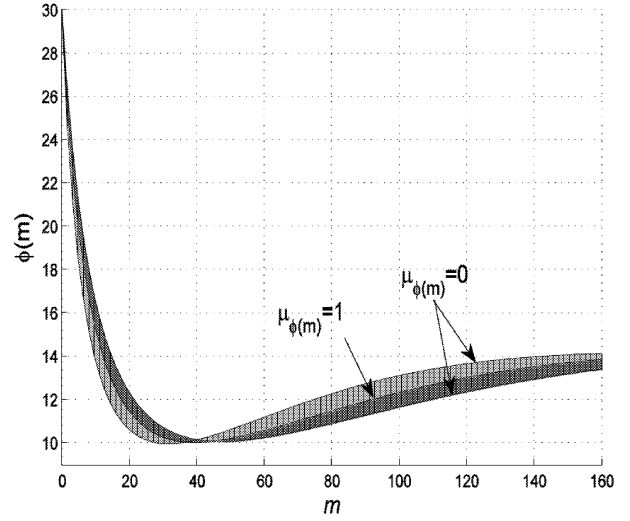


Figure 7: Fuzzy cost function for the second case.

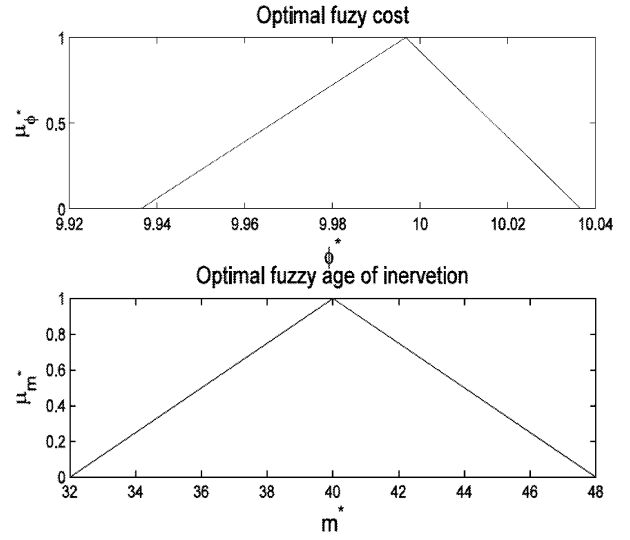


Figure 8: Optimal fuzzy cost and age of intervention for the second case.

For our second case we can see that the fuzzy cost narrows down around the optimal cost. From Figure 8 we can see that the optimal fuzzy cost is less fuzzier than the first case, but, the age of intervention is even fuzzier than the first case. What we can say here is that we have a wider interval for  $m^*$  for which we would obtain approximately a fuzzy cost of 10.

## CONCLUSIONS

The model represented provides an advanced algorithm capable of introducing uncertainty from its two types of sources: Randomness and fuzziness. Hereby representing the unsharpness of the model and its parameters contrary to the classical method with real crisp parameters and non fuzzy probability distributions.

With this model we get a closer look to reality as we take into account the human error, the lack of information and the non perfection of the actual model by using the concept of fuzziness.

Although the algorithm used gives us a good solution to the problem, it is not the best suited procedure as there are more advanced techniques like the  $\alpha$ -level optimization and the extension principle (Möller & Beer 2004).

The continuing research will be focused in implementing the  $\alpha$ -level optimization to resolve the problem. Later on we will be focusing in developing a joint policy of maintenance and inventory control, thereby taking as decision variables not only the age  $m$  but also the stock level  $h$  and the inventory build-up start time  $A$ . The model will be based on the different scenarios that could arise depending on the fact that there is or not a machine failure and the fact that there is or not a backlog. For the latter we will use the theory of possibilities to weigh the occurrence of each scenario.

Eventually we may also impose an availability constraint to the model. We can try to find the optimal age  $m^*$  that minimizes the cost, satisfying a minimal availability level.

## References

- [Al-Najjar and Alsyounf, 2002] Al-Najjar, B. & Alsyounf, I., 2002. Selection the most efficient approach using fuzzy multiple criteria decision making. *Int. Journal of Production Economics*.
- [Barlow and Proschan, 1996] Barlow, R.E. & Proschan, F., 1996. *Mathematical theory of reliability*, Society for Industrial Mathematics.
- [Cox, 1962] Cox, D.R., 1962. *Renewal theory*, Chapman & Hall.
- [Dubois and Prade, 1978] Dubois, D. & Prade, H., 1978. Operations on fuzzy numbers. *International Journal of Systems Science*, 9(6), 613- 626.
- [Hennequin and al.] Hennequin, S., Arango, G. & Rezg, N., Optimization of imperfect maintenance based on fuzzy logic. To be published in *Journal of Quality in Maintenance Engineering*.
- [Kruse and Meyer, 1978] Kruse, R. & Meyer, K.D., 1987. *Statistics with vague data*, D Reidel Pub Co.
- [Kwakernaak, 1978] Kwakernaak, H., 1978. Fuzzy random variables-I. Definitions and theorems. *Information Sciences*, 15(1), 1-29.
- [Kwakernaak, 1979] Kwakernaak, H., 1979. Fuzzy random variables-II. Algorithms and examples for the discrete case. *Information Sciences*, 17(3), 253-278.
- [Möller Beer, 2004] Möller, B. & Beer, M., 2004. *Fuzzy randomness: uncertainty in civil engineering and computational mechanics*, Springer.
- [Möller and al, 2003] Möller, B., Graf, W. & Beer, M., 2003. Safety assessment of structures in view of fuzzy randomness. *Computers and Structures*, 81(15), 1567-1582.
- [Nakagawa, 1942] Nakagawa, T., 1942. Maintenance theory of reliability. In *Springer series in reliability Engineering*. pp. 171- 197.
- [Shapiro, 2008] Shapiro, A.F., 2008. *Fuzzy random variables*. Insurance Mathematics and Economics.
- [Zadeh, 1965] Zadeh, L.A., 1965. Fuzzy sets. *Information and Control*, 8(3), 338-353.

# A threaded Java concurrent implementation of the Monte-Carlo Metropolis Ising model

Carlos Castañeda-Marroquín  
Alfonso Ortega de la Puente  
Manuel Alfonsaca  
Departamento de Ingeniería  
Informática  
Escuela Politécnica Superior  
Universidad Autónoma de Madrid  
Campus de Cantoblanco  
28049 - Madrid (España)  
[Carlos.Castanneda@uam.es](mailto:Carlos.Castanneda@uam.es)  
[Alfonso.Ortega@uam.es](mailto:Alfonso.Ortega@uam.es)  
[Manuel.Alfonseca@uam.es](mailto:Manuel.Alfonseca@uam.es)

James A. Glazier  
Maciej Swat  
Department of Physics  
Swain Hall West 159  
  
Indiana University  
727 East Third Street  
Bloomington, IN 47405-7105  
U.S.A  
[glazier@indiana.edu](mailto:glazier@indiana.edu)  
[mawat@indiana.edu](mailto:mawat@indiana.edu)

## KEYWORDS

Ising Model, Metropolis – Monte-Carlo Algorithm, Parallelization, Java

## ABSTRACT

This paper describes a concurrent Java implementation of the Metropolis Monte-Carlo algorithm that is used in 2D Ising model simulations. The presented method uses threads, monitors, shared variables and high level concurrent constructs that hide the low level details. In our algorithm we assign one thread to handle one spin flip attempt at a time.

We use special lattice site selection algorithm to avoid two or more threads working concurrently in the region of the lattice that "belongs" to two or more different spins undergoing spin-flip transformation. Our approach does not depend on the current platform and maximizes concurrent use of the available resources.

## MOTIVATION

Ising model (Ising 1925) in its original form was used to simulate statistical behavior of ferromagnetic (or anti-ferromagnetic) domains. The Ising model, defined by Ernst Ising in the twenties, has been extended to model systems composed of many particles each having more than two spin states (Potts model). In 1991 Glazier and Graner have realized that biological cells can be represented as adjacent domains having same value of spin. With this assumption they were able to successfully simulate sorting process in the system composed of cells with different adhesivities. Subsequent extensions and modifications of the Potts model led to creation of the Glazier-Graner-Hogeweg model (GGH) also known as Cellular Potts Model (CPM) (Glazier et al. 2007) y (Graner and Glazier 1992). The GGH uses as an input phenomenological cell behavior description and produces evolution of complex cell-based biological systems, for example, living tissues.

Mathematical simplicity of the GGH model and its ability to simulate systems composed of up to  $10^7$  cells (on a single processor) made it one of the most promising tools in the

field of biomedical modeling of tissues, organs or even entire organisms.

The common feature of Ising, Potts and GGH model is their connection to statistical mechanics. The real systems they usually simulate are built of multiple copies of microscopic elements (spins, or cellular domains). The behavior of the overall system is driven by the local changes in the microscopic element. For example, in the cell sorting simulation, cells attempt to spontaneously rearrange themselves to reach the most energetically favorable configuration. Every microscopic change in the configuration of a single cell depends only on the local change of the energy due to minimal change of system configuration. The complex behavior of the entire system is thus a derivative of the local rules. Computationally, the simulation is implemented using Monte-Carlo techniques and relies on Metropolis Monte-Carlo Algorithm.

The snapshots of the simulation taken at regular intervals, measured by the number of system perturbation attempts in the Metropolis algorithm, represent temporal evolution of the studied system. Typically, Metropolis algorithm is implemented as a serial algorithm and suffers from computational inefficiency, especially when multi-processor architectures and computer clusters are available. Even if scientists implement a parallel version of the algorithm, their codes are specific to a given application, un-portable and architecture dependent. The goal of our approach is to propose a general technique to implement parallelism in the simulation using maximum number of system available resources, in a way that does not depend on the details of the used platform. Our approach seems natural and simple, and adds minimal amount of computational overhead.

## INTRODUCTION TO THE ISING MODEL AND ITS VARIANTS

The Ising Model originally was a discrete model of ferromagnetism, based on the magnetic moments or spins ( $\sigma$ ) of individual atoms and their interaction energies ( $J$ ) that contribute to the global energetic behavior via a function called Hamiltonian ( $H$ ). The Ising model considers the material as a lattice where spins are regularly distributed.

Spins are represented by two discrete values ( $\sigma \in \{1, -1\}$ ) and are supposed to minimize the following energy expression when there is no external magnetic field:

$$H = -J \sum_{i,j} \sigma_i \sigma_j$$

The summation adds the interaction energy of a given spin with that of its four nearest neighbors. If  $J > 0$  (systems of this kind are called *ferromagnetic*), couplings with the same spin projection lower system energy, while couplings with different spin projections make this energy higher. Thus homogeneous configurations of the system are more frequent. On the other hand, in *non-ferromagnetic* materials ( $J < 0$ ), dissimilar couples are less energetic, and disordered configurations are more frequent.

The Ising model and its derivatives are frequently used in simulations where an analytical or numerical approach is difficult. The Metropolis Monte-Carlo (Metropolis et al. 1953) (or Metropolis-Boltzman) method is one of the most popular algorithms used to simulate, not only the Ising model, but also its variants (for example, Potts and GGH). This method can be described by means of the following schema:

1. Choose a lattice site at random.
2. Compute the energy change ( $\Delta H$ ) when the current configuration energy is replaced by a configuration where the selected spin is flipped (changed into its opposite value).
3. Flip the spin at the current site according to the following probability function,

$$P = \begin{cases} 1 & \text{if } \Delta H \leq 0 \\ e^{-\Delta H / kT} & \text{if } \Delta H > 0 \end{cases}$$

where  $k$  is the Boltzman constant and  $T$  the temperature. These first steps are called a *spin-copy attempt*.

4. Return to step 1.

In a 2D lattice of dimensions  $x_{max}, y_{max}$ , a *Monte-Carlo step (MCS)* involves  $x_{max} \times y_{max}$  spin-copy attempts.

The main variants of the Ising model are the Potts model, and GGH. The Potts model was proposed in the fifties by Renfre Potts (Potts 1952). This extension allows the use of more than two kinds of spins. Domains with the same *spin* value are named *cells*. The Potts model successfully simulates different physical processes, such as the growth of grain, or foam coarsening. There are similarities between the behavior of these systems and that of the Ising model: in both, there is a critical temperature, below which the cells dissociate, and above which they disappear; similarly to the Ising model, the Potts model is usually simulated by means of Metropolis-Boltzman or Monte-Carlo methods.

Glazier et al. have extended these models from Potts model to GGH. They realized that biological cells can be represented as spatially extended domains characterized by the same value of the "spin", by analogy to the ferromagnetic domains. In addition to simple spin-spin interactions that describe ferromagnetic elements in the Ising model, GGH model encompasses many cell behaviors actually observed in nature: volume and shape constraint, cell-cell adhesivity, response of cells to chemical signals, secretion of signals by a cell, cell division, cell death, cell type differentiation etc...

## INTRODUCTION TO PARALLEL IMPLEMENTATIONS OF THE ISING MODEL

Computer simulations either complement analytical models or are used where analytical models do not exist or when analytical solutions are so complex that they are of limited use. Mathematical simplicity of numerical models combined powerful computers can, and in many cases is a substitute for the lack of the elegant and compact analytical solution to a given problem. Physical and biological phenomena simulated using Ising or GGH models fall, in most cases, into category of models for which no analytical solution is available.

The attractiveness of GGH (or Ising) models could be greatly increased if the simulations could be done in a relatively short time so that scientists can for example study dependence of the simulations on various parameters. Running parameter sweeps is a trivially parallel problem. However, in many cases in addition to knowing parameter dependence of the model we want to build models whose complexity prevents them to be executed in a timely fashion on a single processor. For example realistic modeling of extended, 3D fragment of tissue in its all complexity on a single computer is very often impractical if not impossible. Thus it is desirable to build parallel version of the simulation. The main principles of parallelizing Ising-types models are the same Ising, Potts or GGH models.

Several methods have been proposed, using both homogeneous and heterogeneous clusters of computers. One of the main problems is the communication and synchronization of processes for their concurrent execution in different processors. This is a very hard, non-deterministic optimization problem whose optimal solution is not always feasible. In many cases, only an approximation can be calculated.

We are interested in an abstract approach that hides the low level details, delegating them to the network modules of the operating system. We are also interested in offering versions of our simulators accessible via web. From the viewpoint of a software developer, an object oriented language like Java, which supports concurrent threads, is a reasonable choice, despite the fact that Java Virtual Machine (JVM) cannot be run on clusters. It is possible, however, to run Java programs on a cluster of computers by means of a special Distributed Java Virtual Machine (DJVM), which supports the parallel execution of Java threads. In this way, a multithreaded Java application runs on a cluster just as if it were running on a



single machine. There are several different kinds of DJVM, for example: Java-Enabled Single-System-Image Computing Architecture 2 - JESSICA2 (Wenzhang Zhu et al. 2002) the cluster virtual machine for Java developed by IBM (Aridor et al. 1999), Proactive (Baude F et al.), DO! (Launay and Pazat 1997), JavaParty (Philippsen M and Zenger M. 1997), Jcluster (Zhang et al. 2006), MPJ Express (Baker et al. 2006), and Terracotta (Terracotta 2008).

To create a parallel algorithm to run Metropolis Algorithm for the Ising model, we had to choose among

three kinds of techniques (Newman and Barkema 1999): *a trivial parallel algorithm*, where the problems can be divided into completely different tasks without the need for inter-process communication; *functional decomposition* that breaks up the task into a variety of different jobs executed by different processors, whose results have to be put together; and *domain decomposition*. Most of the approaches to parallelize the Ising model are of the last kind. Barkema and MacFarland (Barkema and MacFarland 1994) proposes a parallel version in which the lattice is decomposed into different parts (domains) that are then assigned to different processors. Korniss G (Korniss et al. 1999) describes a parallel version of the  $n$ -fold way to be run at low temperatures. Hermann DW and Burkitt AN (Hermann and Burkitt 1991) introduces a checkerboard pattern decomposition method, while other approaches (Shim and Amar 2006) are useful for a variety of parallel Kinetic Monte-Carlo simulations and have recently been applied to simulations of epitaxial copper growth.

## A THREADED JAVA CONCURRENT IMPLEMENTATION OF THE MONTE-CARLO METROPOLIS METHOD TO SIMULATE THE ISING MODEL

In the description bellow we will use the following names:

- $x_{\max}, y_{\max}$  is the dimension of the lattice.
- *main* is the name of the main thread of the algorithm.
- Threads associated with each spin location are named *spin threads*.

### Pseudo-code of the main thread

The main thread monitors the execution of each MCS: it creates the *spin threads*, starts them to be concurrently run, waits until all of them finish, and properly updates the lattice for the next MCS. It is easy to realize that this is the only synchronization condition: the family of *spin threads* of the next MCS cannot start before the family of the current step finishes.

1. Initialize randomly the initial configuration.
2. The next configuration is initially a copy of the first configuration.
3. For each MCS
  - a. For each family of spin threads and while the system supports a new spin thread

- I. Select a position in the lattice at random.
- II. While the position is the neighbor of a previously selected position.
  - Try another random position (the maximum number of trials is a parameter of the algorithm).
- III. Create a new thread for the selected position and start its execution.
- b. Select the next configuration.
- c. Copy the previous configuration into the next one.

### Pseudo-code of the *spin thread*

All the *spin threads* are created by the main thread. They share two configurations: the current and the next. They all have to get the values of their neighbors from the same copy of the current configuration, and, if needed, they have to change the value of their spins in the same copy of the next configuration. The code below summarizes this:

1. Compute  $\Delta H$  at this position in the lattice.
2. Change the spin value according to equation 2. Notice that all the *spin threads* change the same copy of the next configuration. A Java monitor controls the configuration by means of a *synchronized* method, to ensure that only one *spin thread* updates the configuration at the same time.

This approach fulfils the following important characteristics:

- The design of the algorithm makes it impossible to simultaneously try to change two close neighbor spins.
- The way in which these positions are selected does not introduce asymmetric factors, i.e. all the spins in the lattice can be visited, possibly in a different order, in each MCS, for the following reasons:
  - Let  $n$  be the average number of spins simultaneously studied. We need, therefore,  $(x_{\max} \times y_{\max}) \div n$  families of concurrently studied spins for each MCS. Avoiding neighbor spins ensures that  $n \leq (x_{\max} \times y_{\max}) \div 2$ . Some algorithms, for example Creutz's cellular automaton (Creutz 1986) or Q2R (Vichniac 1984), are able to reach this maximum. We thus need more than one *family* to complete each MCS.
  - These *families* do not depend on one another.
- Our approach accepts spin changes with the same criteria than Metropolis Monte-Carlo method, which ensures that both approaches share the same characteristics.

We can conclude from the above that our algorithm is an alternative parallel implementation of the Metropolis Monte-Carlo with the same properties.

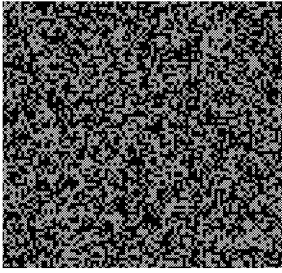
Figures 1 to 9 show different configurations generated by our algorithm. The typical behavior shown by other simulations of the Ising model appears here.

- Figure 1 shows an initial  $100 \times 100$  lattice.
- Figures 2 to 6 show the configurations reached after 1000 MCS for different values of  $T$  (respectively 0.2, 0.5, 1, 1.5 and 2.5). The presence of clear domains with these low  $T$  values can be observed.
- Figures 7 to 9 show the configurations reached for higher values of  $T$  (respectively 3.5, 4.5 and 6). They show a rather random distribution of spins.

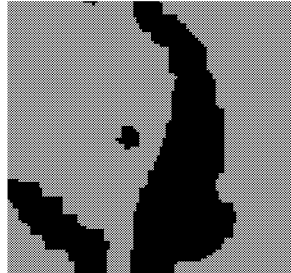
One of the main features of our approach is the fact that we can tune the *amount of parallelism* (the number of threads or spins simultaneously studied) to reach the maximum supported by the hardware. There are two critical factors: memory and the number of processors. Our algorithm can use all the available memory to define, at most,  $(x_{max} \times y_{max}) \div 2$  threads. The operating system is responsible of running the threads. In this way, our code does not depend on any specific platform.

Our mechanism to select the spins is more time expensive than Metropolis Monte-Carlo, because we have to check that each position does not overlap the neighborhood of any other spin, rather than just randomly select a location. Once the  $n$  threads of the current family have been created, it is possible (depending on the specific platform) that all of them can be simultaneously analyzed. In the best case, we spend the same time to analyze the whole family as an individual spin. To find the next position, we have to repeat several times the following procedure: select a random position in the lattice; check if it is available, i.e. it does not overlap the neighborhood of any previously selected spin. For each spin, this process is faster than the algorithm which changes a given spin. In this way, in most cases, our approach improves the performance of the original Metropolis Monte-Carlo method but runs on the multiple processors as opposed to serial Metropolis algorithm.

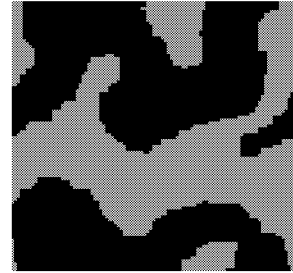
Our system takes into account this performance criterion, because we can adjust the number of attempts to find the position of the next spin.



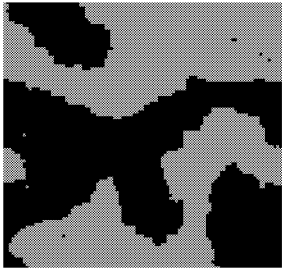
**Figure 1. Initial  $100 \times 100$  random configuration.**



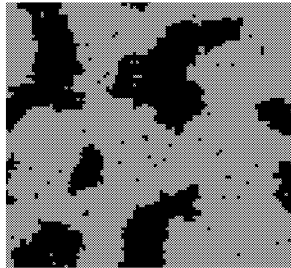
**Figure 2. Configuration after 1000 MCS,  $T = 0.2$ .**



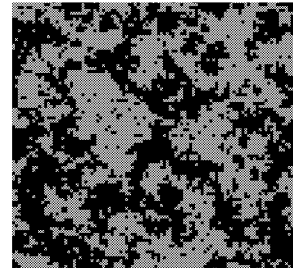
**Figure 3. Configuration after 1000 MCS,  $T = 0.5$ .**



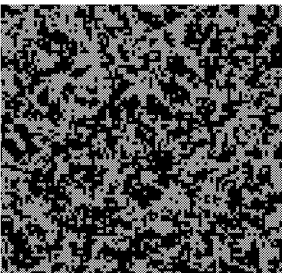
**Figure 4. Configuration after 1000 MCS,  $T = 1$ .**



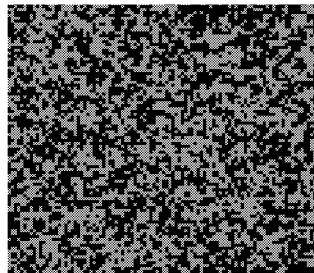
**Figure 5. Configuration after 1000 MCS,  $T = 1.5$ .**



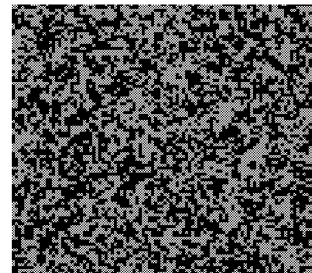
**Figure 6. Configuration after 1000 MCS,  $T = 2.5$ .**



**Figure 7. Configuration after 1000 MCS,  $T = 3.5$ .**



**Figure 8. Configuration after 1000 MCS,  $T = 4.5$ .**



**Figure 9. Configuration after 1000 MCS,  $T = 6$ .**

## CONCLUSIONS AND FURTHER RESEARCH LINES

Some of the authors of this paper have developed COMPUCELL3D (Cickovski et al. 2005; Izaguirre et al. 2004), a multi-tiered, flexible and scalable problem-solving environment for morphogenesis simulations, that is written in C++ and Python using object-oriented design patterns. Current version of COMPUCELL3D does not currently support parallel execution. The parallelism in the presented algorithm is achieved by careful spin selection and spawning of the spin threads. One can fairly easily adapt serial code to be re-implemented using presented approach. We will use presented method to implement parallel version of COMPUCELL3D.

## ACKNOWLEDGMENTS

This work has been partially sponsored by MICINN, project TIN2008-02081/TIN and by DGUI CAM/UAM, project CCG08-UAM/TIC-4425

## REFERENCES

- Aridor, Y., Factor, M. and Teperman, A., 1999. "cJVM: a single system image of a JVM on a cluster". Parallel Processing, 1999. Proceedings. 1999 International Conference on}. P 4-11.
- Baker Mark, Carpenter Bryan, Shafi Aamir. 2006 "An Approach to Buffer Management in Java HPC Messaging". In V. Alexandrov, D. van Albada, P. Sloot, and J. Dongarra, editors, International Conference on Computational Science (ICCS 2006), LNCS. Springer, 2006.
- Barkema G.T., MacFarland T. 1994. "Parallel simulation of the Ising model". Physical review E. Vol. 50. 1623-1628.
- Baude F., Baduel L., Caromel D., Contes A., Huet F., Morel M. and Quilici R. 2006 "Programming, Composing, Deploying for the Grid" in "GRID COMPUTING: Software Environments and Tools", Jose C. Cunha and Omer F. Rana (Eds), Springer Verlag, January 2006.
- Cickovski T.M, Chengbang Huang Chaturvedi, R. Glimm T., Hentschel H.G.E., Alber M.S, Glazier J.A., Newman S.A., Izaguirre J.A., 2005. "A framework for three-dimensional simulation of morphogenesis." Computational Biology and Bioinformatics, IEEE/ACM Transactions on. Vol 2. P. 273-288.
- Creutz, Michael. 1986. "Deterministic Ising Dynamics". Annals of Physics 167, 62-76.
- Glazier J.A., Balter A, Poplawski NJ. 2007 "Magnetization to morphogenesis: a brief history of the Glazier-Graner-Hogeweg model". Single-Cell-Based Models in Biology and Medicine A. R. A. Anderson, M. A. J. Chaplain and K. A. Rejniak, editors (Birkhäuser, Basel, Boston and Berlin), 79-106.
- Graner F., Glazier J.A., 1992."Simulation of biological cell sorting using a 2-dimensional extended Potts Model". Phys. Rev Lett. V 69. P.2013-2016.
- Heermann DW, Burkitt AN. 1991. "Parallel algorithms in computational science". Springer Series in Information Sciences. Springer-Verlag, Berlin, Heidelberg.
- Ising, Ernst. 1925. "The theory of ferromagnetism". Zeitschrift fuer Physik.31: 253-258.
- Izaguirre JA, Chaturvedi R, Huang C, Cickovski T, Coffland J, Thomas G, Forgacs G, Alber M, Newman S, Glazier JA. 2004 "COMPUCELL, a multi-model framework for simulation of morphogenesis". Bioinformatics V. 20 No. 7. 1129-1137.
- Korniss G., Brown G, Novotny M. A., Rikvold P.A. 1999. "Hard Simulation Problems in the Modeling of Magnetic Materials: Parallelization and Langevin Micromagnetics". Computer Simulation Studies in Condensed Matter Physics XI, edited by D.P. Landau and H.-B. Schuttler, Springer Proceedings in Physics Vol. 84 (Springer, Berlin).
- Launay P., Pazat J. 1997 "A Framework for Parallel Programming in Java". INRIA Rapport de Recherche Publication Internet. 1154 decembre 1997.
- Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., and Teller E. 1953. "Equation of State Calculation by Fast Computing Machines". Journal of Chemical Physics, vol. 21, pp. 1087 – 1092.
- Newman M., Barkema G.T. 1999. "Monte Carlo Methods in Statistical. Physics", Oxford University Press. Oxford, 1999. 356-362.
- Philippsen M., Zenger M. 1997. "JavaParty - transparent remote objects in Java". Concurrency: Practice and Experience 1997; 9(11): 1225-1242.
- Potts. R. B. Proc. Camb. Phil. So c. 48, 106 (1952).
- Vichniac Grard Y. 1984. "Simulating physics with cellular automata". Physica D: Nonlinear Phenomena Volume 10, Issues 1-2, January 1984, P. 96-116
- Shim Y. and Amar J.G.. 2006. "Growth Instability in Cu Multilayer Films Due to Fast Edge/Corner Difusion". Physical Review B. 73. V. 73. PP
- Terracotta, Inc, 2008. The Denitive Guide to Terracotta: Cluster the JVM for Spring, Hibernate and POJO Scalability (The Denitive Guide), Apress Publishing. Expert's Voice in Open Source Series. P. 41-66.
- Wenzhang Zhu, Cho-Li Wang, and Francis C. M. Lau.2002 "JESSICA2: A Distributed Java Virtual Machine with Transparent Thread Migration Support". In IEEE Fourth International Conference on Cluster Computing, Chicago, USA, September 2002.
- Zhang B., Yang G., Zheng W. 2006. "Jcluster: an efficient Java parallel environment on a large-scale heterogeneous cluster". Concurrency and Computation: Practice and Experience. V 18. N. 12. P 1541-1557.



# **SIMULATION IN CIM**



# NEW SYSML BASED APPROACH FOR INTEGRATED SYSTEM DESIGN

G. Auriol, C. Baron

<sup>1</sup>LATTIS Université de Toulouse

INSA 135 avenue de Rangueil,

31077 Toulouse, France

E-mail: {guillaume.auriol, claud.baron}@insa-toulouse.fr

## KEYWORDS

System design, system engineering standards, EIA 632, system modeling, SysML, requirements engineering

## ABSTRACT

The objective of the work presented in this paper is to demonstrate the interest to follow the methodological recommendation standards to develop complex systems. After having pointed out the principles of system engineering, we present a well-spread electronic and aeronautic standard and illustrate its application for the development of a small scale sailing ship. We demonstrate how this standard offers an effective methodological support, and allows an early detection of possible weaknesses from the very first steps of design.

## HOW TO BENEFIT FROM SYSTEM ENGINEERING STANDARDS?

Design tasks have naturally evolved towards the design of more and more complex systems. This evolution progressively led to study not only the “how to make” but also the “how to organize itself” to do it. If, separately, each stakeholder involved in the system development has very specific competencies, none has the sufficient visibility to consider the problem in its entirety. System engineering thus appeared as an emerging discipline which helps to solve the problems arising from complex system development.

### *Definition And Objectives*

According to (Thomé, 1993), *System Engineering (SE)* is an interdisciplinary approach allowing the realization and the deployment of successful systems. It can be seen as the application of engineering techniques, as well as the application of a systematized engineering approach. SE is thus an interdisciplinary approach whose objective is the design and the realization of complex systems, such as defined in (Shisko and Al, 1995). It consists of two principal activities (DAUP, 2001): technical knowledge of the

domains implemented in the system and the management of their implementation. It does not only consider the realization of a physical product, but also guides and coordinates all the processes involved in the design and the realization of the product.

The design and the development of a system indeed require the contribution of various processes, stakeholders and techniques. Each one often concentrates on a specific aspect of the system without any visibility on the other aspects: designers have an architectural and a functional vision of the system, the industrialization team considers it as an element to be integrated in the assembly line, the procurement like a list of supplies... SE must then bring a global point of view on the system to guaranty the consistency of specific contributions on the system; it considers the whole life cycle of the system: definition of the customer requirements, design, realization, marketing, etc. It defines methods and means allowing satisfying best the technical and organisational needs which, in practice, encounter constraints of costs, times and of productivity.

### *Methods, Tools And Models*

The SE approach consists in coordinating useful interdisciplinary actions during the whole system life cycle. This coordination is based on the definition and the use of the concept of *process* and on the implementation of *methods* defined by (Estefan, 2007). These methods are supported by *tools* based on *models*.

A *process* is a logical sequence of tasks performed to achieve a particular goal. A process defines “what is” to be made without specifying “how to” do it. The structure of a process can be represented under different levels of aggregation to allow analyzes and to support the various needs for decision-making. A *method* defines “how to” achieve each task. In addition to these processes and methods, *tools* represent the system throughout its development: expression of requirements, logical then physical solution ... One of the stakes of SE is to ensure the consistency and the interoperability of these representations during the project. For that, new tools based on languages

like UML (*Unified Modeling Language*) or SysML (*System Modeling Language*) try to unify the representations of the system in a single model throughout its life cycle. These languages rely on a decomposition of the system into diagrams. Each diagram is dedicated to the representation of a particular aspect of the system: requirements, structure, behaviour... Developers do not manipulate any more a multitude of models written in different formalisms but a single one offering various views linked with a unique language. They thus guaranty consistency between views of the model.

#### *Standards Of System Engineering, The EIA 632*

The objective of SE is to master and control the development and organizational processes in order to guarantee the quality of a product or a service, and thus the satisfaction of the customer. Within this framework, the SE standards define processes on which developers can rely. They are used as reference to manage a system from its initial concept to its delivery, through its design and realization.

In SE, three standards: EIA 632 “Processes for Engineering a System”, IEEE 1220 “Standard for application and Management of the Systems Engineering Process” and ISO 15288 “Systems Engineering - System Life-Cycle Processes” are particularly well adapted to the development of complex systems. The EIA 632 (Martin, 1998) is the one that widely covers the system life cycle. Its role is to provide a set of fundamental processes to help a developer in the design or the improvement of a system. The respect of the processes and the recommendations of this standard allows the developers: to establish and make evolve a complete and consistent set of requirements in order to obtain feasible and cost-effective systems, to meet these requirements according to cost, time and risk constraints, to provide a system that satisfies each stakeholder, etc.

As illustrated on Figure 2, the system is not only constituted by *end products* but also by *enabling products*, such as products of test, training... End products are the products which will perform the operational functions of the system; they will be delivered to and used by the customer. Enabling products do not perform any operational functions but “allow” the end products to be developed, tested, manufactured, deployed, or are used for the staff training.

It can happen that some end products and enabling products are available “on shelf”. However, in the other cases, end products and enabling products must be decomposed and developed applying the same processes and recommendations of the EIA 632. For that, the EIA 632 introduced the concept of *building block* (Figure 2).

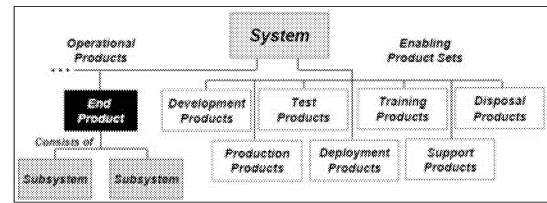


Figure 2. Definitions of the blocks of a system according to the standard EIA 632

A *building block* is a concept combining all the elements useful to the system development. It includes the decomposition into end products and enabling products of the system, the requirements, and any other information useful to the system realization. Therefore, it represents much more than a simple architectural decomposition of the system. It leads to a representation of the conceptual framework of the system and is used to organize the requirements, the work and other information associated with SE. The use of building blocks thus allows the developer to be sure that all end products through the system life cycle were taken into account.

In a building block, end products and enabling products can be decomposed into subsystems. By associating a building block of a lower level to subsystems, the decomposition of the system (Figure 3) is carried on either until all the subsystems are defined, specified and developed, or until they are available “on shelf”, or until they can be acquired from suppliers. We then speak of decomposition into layers. In the same way, an enabling product which requires a further development will be associated to a building block.

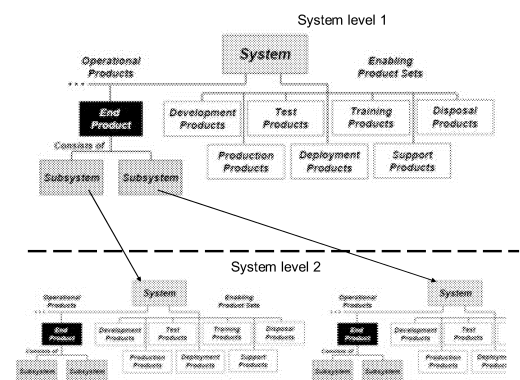


Figure 3. Decomposition in layers of a system

The processes of design, verification and validation can be applied to any building block. In accordance with a “V cycle” development, the EIA 632 recommends to follow a top down approach from specifications to prototyping together with a bottom up physical realization. In the top



down stage, the needs identified at a level will be used as requirements for building blocks of lower levels. Once all the levels specified, the bottom up realization is performed by checking each level. Thus, the standard ensures the conformity of the global developed system with regard to the initial requirements, whether it is for its end products or its enabling products. Once all building blocks specified, the corresponding products are acquired or manufactured. Assembling these elements leads to the construction of the system. Then, this one must be validated with regard to its specifications.

## APPLICATION OF A SE STANDARD TO THE DESIGN OF A SMALL SCALE SAILING SHIP

This section illustrates the application of the EIA 632 recommendations for the development of a small scale sailing ship. The objective of this application is to illustrate: the definitions and concepts of a general standard in a precise context and the contribution of the standard compared to a traditional approach of design.

### *Small Scale Sailing Ship*

The project presented has been led with 4 teams of 6 engineers; we made them compete in order to build four small scale of remote-controlled sailing ships of International One Meter (IOM) class in fibreglass. The project is multidisciplinary (mechanic and electronic, composite materials, hydrodynamics), easily decomposable into independent sub-projects however requiring a precise integration. Each team has to deliver a boat at the end of the project, composed by a hull, hull appendages (rudder, keel) and of one rig (Figure 6).

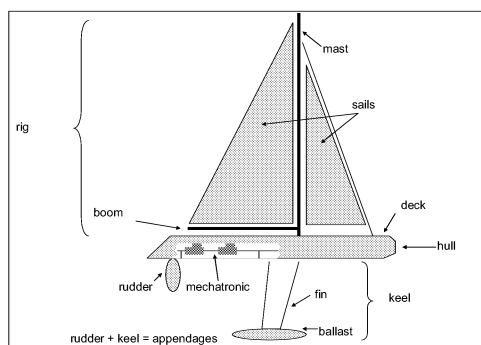


Figure 4. Architecture of a sailing ship

To obtain such a result, engineers have to cope with many problems: needs and requirement analysis, functional analysis, multi-technologies integration, collaborations in a context of multiple partners. The interest to develop a small

scale model in accordance with the IOM class is that the regulation authority fixes constraints which have to be taken into account during design and realization. Some parameters are strictly defined (dimensions of the mast, the sails ...), others can have free values (height of the deck, form of hull...).

### *Application Of The EIA632 Standard*

The realisation of an IOM class sailing ship is mainly constrained by technical limitations fixed by the class rules: dimensions, weight and materials of components are strongly regulated. On the one hand, these rules specify some requirements at the structural level (limitation of physical parameters...) as well as at the functional level (maximum 2 channels of piloting, short use in time...). On the other hand, manufacturers are free to design their small scale model as they want if the ship, once built, satisfies the requirements. Thus multiple design options are offered.

In this sub-section, we first illustrate the concept of stakeholders then we show how to proceed to the capture of needs from stakeholders and to the requirements specification; finally, we detail the decomposition of the system into end products and enabling products and how requirements can be allocated to building blocks.

In this project appear the following stakeholders:

- the referees, considered as customers in an industrial context,
- the technical staff to access hazardous manufacturing resources (milling machines, turning machines, furnaces)
- the suppliers,
- the financial department,
- the team members, having 2 different roles: (1) subcontractors, (2) customers (end user).

We can define as other nonhuman stakeholders which will impact the development of the system:

- the rules of international IOM class,
- the authority of certification which will verify the conformity of the product with the rules of class,
- the conditions of environment (strength and type of wind, condition of temperatures, type of race).

These stakeholders will intervene during one or several processes of system development: the design, the supply, the realization, the tests, and the validation.

To collect the requirements from the different stakeholders, many methods can be considered, for instance the very widespread method "Voice of Customer" (VOC) (Griffin and Al, 1993) and (Yusop and Al, 2007) or the KAOS method, developed by the University of Louvain (Nakatani, 2007). We adopted the method APTE for his graphic character; it is based on the European standard IN 1325-1

and on the French standard NF X50-151. This method represents the system, the stakeholders and all external constraints that impact the preliminary design, on an environment diagram in form of octopus (Figure 5).

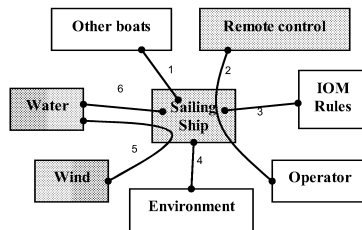


Figure 5. Representation of the development context of the system "sailing ship" (APTE method)

After the stage of identification of the stakeholders, it is necessary to define their respective constraints: from customer needs until technical requirements. In a preliminary study, the set of collected constraints is listed and classified in a table. A classification by type of requirement: functional, operational, performance, maintenance or due to an enabling system is used.

The formalization of the system decomposition supported by the iterative design process helps structuring the system. Through the evolution of the system representation, several layers of building blocks appear for the *hulls*, *rig* and *appendage* (Figure 6); the evolution of the enabling products (here the *cradle* and *sail box*) is not detailed. It shows the state of the sailing ship structure at the steps of predefinition of the system, definition of the system, design of the subsystems and detailed design.

In parallel with this step of decomposition, requirements are allocated to the different building blocks of the system in subsystems. This allocation is represented using a SysML requirements diagram.

The application of this approach allows structuring the design. By the systematization of the technical verification, the detection of design inconsistencies is facilitated. Because of the iterative character of processes, the specifications of the subsystems or components are then modified in order to correct the early detected errors.

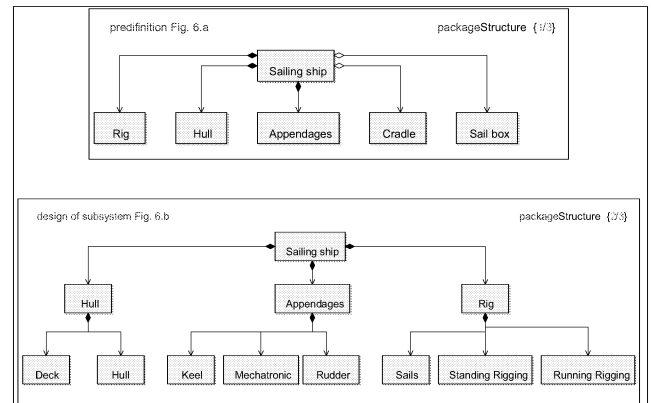


Figure 6. Preliminary decomposition

### Main Benefits From The Standard

Formalizing these elements allowed 24 engineers, completely neophytes in the construction of small scale sailing ships and knowing nothing about the maritime environment, to design, produce and test 4 sailing ships in assigned time.

Following the recommendations of the standard led to an improvement of the development process of the end product. We extracted from this case study two examples testifying of the contribution of the standard for the development of the system:

- the 1<sup>st</sup> example illustrates how two processes (see Figure 1) are coupled: the process of *requirements definition* in "system design" and the process of *assessment* in "technical management",
- the 2<sup>nd</sup> example illustrates the possibility of taking into account an error during the realization process.

### Example of the through deck block

This example illustrates the case in which the definition of a subsystem leads to re-examine not only the specifications of the system of which it depends but also the specifications of the surrounding systems.

In this case, the definition of the system led to decompose the sailing ship into a hull, a rig, the appendage and the mechatronic subsystem (mechanic and electronic elements). At this level of decomposition, the specifications of interface requirements between mechatronic subsystem and the hull are limited to the mechanical fixing part of the electronic elements to the hull. However, while designing the mechatronic subsystem appears the circuit of mainsail and headsail sheet control line; it is connected on a side to the actuator (under the desk) and other, to the running rig (on the desk).

The process of “requirements specification” of the design of the circuit of sheet control line then forces the designers to modify the specification of the deck so as to allow the passage of sheet control line. The implementation of the technical evaluation process makes it possible to transfer this need on the deck by modifying its specifications. The mechatronic subsystem, through the component “sheet control line” thus will have an impact on an element of the hull whereas they belong to different building blocks.

#### *Example of the mast strut*

The example of the mast strut illustrates the case of the modification of a building block consecutively to the detection of a defect at the integration stage.

It can happen that the moulded mast strut produced by the engineers is too short for reasons of defect of stratification and turn out from the mould. In this case, the deck, which must level the top of mast strut to avoid risks of infiltration, must be lowered compared to what was envisaged. However, the class rules impose a height of boom on the level of the gooseneck compared to the reference point marked on the deck. To respect this height, the position of the boom must be lowered. The height of mast being fixed by the class rule, the simplest solution to compensate for the difference in height is to remove a section of the bottom of the mast and to take again the length of the shrouds.

The integration process allows modifying the specifications of a subsystem, in our case the rig, to catch up with the failures of another subsystem (the mast strut). Of course, this operation is not systematic and sometimes the failing element is replaced. However, in certain cases, it is more interesting to allow modifying specifications than to modify or rebuild an element.

In both examples, following the recommendations of the standard concerning the need to connect processes (the process of requirements definition in system design and the process of assessment of technical management in the first example, or the process of integration in the second one) allows to detect and correct as soon as possible the errors due to a wrong requirements specification and those induced by the partly defective realization of a subsystem.

## CONCLUSION

The work presented in this article addressed system engineering standards and the interests of their application in the complex system development. After having pointed out the principles of system engineering, we presented a standard, the EIA 632, on which the case study is based. We

then illustrated the application of this standard for the development of a small scale sailing ship and demonstrated how a system engineering standard offered an effective methodological support. The suggested iterative approach is very structured: definition of needs and requirements, successive decomposition of the systems into building blocks, test and systematic validation of the enabling products... Thus, we show that relying on a standard guides the stakeholders of a project and allows an early detection of possible weaknesses from the very first steps of design.

The obtained results thus encourage to consider the methodological recommendations of standards for the design, the realization and the verification of systems. However, the framework of standard EIA 632 goes beyond the technical aspects: as we showed, it also considers the organisational aspects. It thus satisfies the general objectives of system engineering, which addresses the resolution of all issues arising during complex systems development, proceeding according to a global approach.

## REFERENCES

- Bernhard Thomé, "Systems engineering : principles and practice of computer-based systems engineering", John Wiley and Sons Ltd. editor, Chichester, UK, 1993
- R. Shishko and R.G. Chamberlain, "Nasa systems engineering handbook". Technical Report SP-610S, NASA Center for AeroSpace Information, June 1995
- J.A. Estefan, "Survey of model-based systems engineering (MBSE) methodologies", Technical report, INCOSE MBSE Focus Group, May 2007
- G. Auriol, C. Baron, J-Y. Fourniols, "Teaching requirements skills within the context of a physical engineering project", The Third International Workshop on Requirements Engineering Education and Training (REET'08), September 9th, 2008, Barcelona, Spain
- Griffin, Abbie and John R. Hauser, "The Voice of the Customer," Marketing Science, vol. 12, No. 1, 1-27., Sciences Literature, Winter 1993
- Norazlin Yusop, Zafar Mehboob and Didar Zowghi, "The Role of Conducting Stakeholder Meetings in Requirements Engineering Training", 2nd International Workshop on Requirements Engineering Education and Training, 15 october 2007, New Delhi, India
- Takako Nakatani, "Improving The Engineering Mind In Eliciting Requirements", 2nd International Workshop on Requirements Engineering Education and Training, 15 october 2007, New Delhi, India

# A Virtual Enterprise Application for the Textile and Apparel Industry

Antonela Curteza and Daniela Farima  
"Gherghe Asachi" Technical University of Iasi  
B-dul. D. Mangeron, Nr. 53bis,  
700050, Iasi, Romania  
e-mail: [acurteza@yahoo.com](mailto:acurteza@yahoo.com)

Ovidiu Stoica and Liviu Lupu  
European Center for Information  
and System Security - CESIS  
Str. Otilia Cazimir nr. 9,  
700400, Iasi, Romania  
e-mail: [cesis.ro@cesis.ro](mailto:cesis.ro@cesis.ro)

## INTRODUCTION

The nowadays global economy, developed as a consequence of new technological changes of all kinds, puts pressure on the companies to increase the efficiency of all their own economic processes. Competition forces all firms to concentrate on their “*core-competencies*” during the transformation processes in which they take part in order to achieve new “inter-enterprises” types of systems by following the “virtual enterprise” paradigm.

Like the majority of the “vertically” organised sectors, the textile and apparel industry has recorded a major “boom” in B2B (business-to-business) type of advertisements during the last few years. Retailers and manufacturers from all levels reacted to these technological innovations and many of them embraced enthusiastically the vision of the new “business model”. Others answered with more caution, taking “careful” steps and waiting for a new solution to be settled instead of taking the opportunity to define one by themselves. In both cases it has become clear that an efficient communicational system and the re-engineering of the business model are paramount elements in developing an extended textile and apparel company by using *e-business* type of technologies.

The study started from and shows the stringent need to adapt the company’s policies inside the textile and apparel industry in order to survive the harsh competition on the profile market and to satisfy the client’s needs, which become more and more sophisticated; and all this, by sticking to norms of environment protection, which grow harsher day by day, and by saving material resources, some of them being nearly finished. The increase in the performances of the communication and informational technology allows companies to carry out their activities everywhere in the world, thus having access to raw materials and labour from low-cost regions, to meet the needs of clients from all cultures and geographical areas (through *eCommerce*, etc.), **to establish a different kind of relationship than the traditional one of business partners (through Virtual Enterprise)**, to simulate the entire manufacturing process of a product before releasing it for production (through Virtual Factory), to benefit from modern instruments when designing the products (through Virtual Reality), to communicate, with heterogeneous informational systems in terms of: operational systems, data basis servers and so on (through XML or YAML types of data structures). In the textile and apparel

industry, modern instruments such as expert systems, virtual factory and so on are missing. The most frequent problem raised regarding the *virtual enterprise* concept is “how can an entire business system be organised in a way that combines the advantages of vertical integration with the advantages of subcontracting?”. The apparel industry is the archetype of a “virtual enterprise”. The Benetton story is an example of success that can be attributed to the concept of “virtual enterprise” organisation. Benetton, like many other companies from the apparel industry, is a “vertically de-integrated company” in all the activities of the business system: “styling, design, manufacturing, logistics, distribution, and sales”. The majority of these crucial activities is based on people and companies from the outside. The globalisation trend of manufacturing led to the increase of competition among the manufacturers in the apparel field, both in highly and not so highly developed countries. The new ICT technologies (Information and Communication Technologies), together with the globalisation processes, have already changed what we call “manufacturing”. Nowadays, the *manufacturing* processes meet the following characteristics:

- *Globalisation*, meaning that an array of functions starting with “Research and Development”, continuing with “marketing” and ending with distribution, is approached on a globally integrated basis
- *Networking* – the coordination of these functions is made by means of an intensive usage of electronic networks and virtual and geographical expertise clusters.
- *Mass customisation*, meaning that the production methods must allow a detailed customisation of the products in order to meet the demands of private markets and special clients.
- *Digitization*, meaning that many of these processes are controlled by “advanced computer systems”, which limit the need of human intervention.

The most eloquent example of virtual enterprise that uses *e-business* and ICT is Cisco, which projected, developed, and delivered a management system for suppliers, business partners, and customers (Cisco Connection Online, Tapscott, Ticoll, and Lowy 2000). In the apparel industry, the most successful companies that operate as virtual enterprises are the “high fashion apparel” ones in the U.S.A. They kept a small number of “in house” processes, namely the ones identified as being strategic, and *core competencies* that require a coherent supervision of the *brand*. Companies such as Ralph Lauren, Calvin Klein or Donna Karan decided to create international networks

because they could not put together a local network of potential partners and subcontractors. This way, they accomplished a flexible framework with the purpose of meeting the organisational needs and the requirement of market transformation.

## THE CONTEXT OF THE NEW DEVELOPMENT

Among the essential imperatives that are characteristic to the new paradigm, which is, the supply of products/services in the context of turbulent markets, there were identified important aspects, such as:

- the accomplishment process needs to be quick,
- the need of personalisation is more and more stringent,
- higher quality,
- being more careful about the environment and its protection,
- a more pressing demand to include the already existing enterprises or to build new ones, real or virtual, in order to respond more efficiently to the market opportunities that become more and more unpredictable.

These requests make the European textile and apparel industry to focus on:

- the relocation of the processes with a small added value, as well as those with high costs in the countries or regions with low costs;
- the modernising of the business processes and the use of *outsourcing* or *subcontracting* for *non-core* activities;
- the concentration on high added value and *knowledge-intensive* activities such as: design, product development, high quality, client relations management, brand construction;
- maintaining tight relationships with suppliers, clients and service providers;
- exploiting new market opportunities;
- hiring highly qualified personnel and a continuous training of employees in order to cope with the business demands.

A concept that seems to sustain these imperatives in the near future is the “virtual factory” one (**VF**). VF is defined as an integrated pattern that simulates major subsystems belonging to a certain *factory* (**F**), be it real or virtual. This *factory* is considered as a whole (*holistic approach*) and is capable of delivering an advanced decisional support. It is a new concept that goes beyond the typical modelling of a subsystem at a specific moment in time like: *manufacturing model*, *the business process model*, *communication network model* and so on, taken separately. In the context of accentuated turbulences on the profile markets (textile and apparel markets), a concept more refined than Virtual Factory is the Agile Virtual Enterprise (AVE). AVE is considered the most dynamic version of the VE model and its implementation supposes complying with some requirements in order to maintain the business partners affiliated to the market. These requirements include:

- diminishing the effort and the reconfiguration costs;

- the capability to preserve the firms’ private knowledge regarding the products or the processes.

The formation, integration and operation of an AVE are based on the existence of an adequate informational platform and on the communication technology. The environment for creating, integrating, operating, reconfiguring, and dissolving an AVE can be implemented as a *Market of Resources* (MR), a concept that covers the entire cycle of life of an AVE. Certain tendencies identified in the business competitive environment will lead, according to some experts, to dramatic changes in the present and future productivity and approach. Among the present tendencies one can identify:

- the life expectation – short – of the new products that undergo numerous *design* reconsiderations in order to maintain themselves in the competitive environment;
- the increase of product diversity, in time;
- rapid technological development;
- increased technological complexity;
- market globalization;
- frequent changes in demand;
- uncertainty;
- a stronger competition.

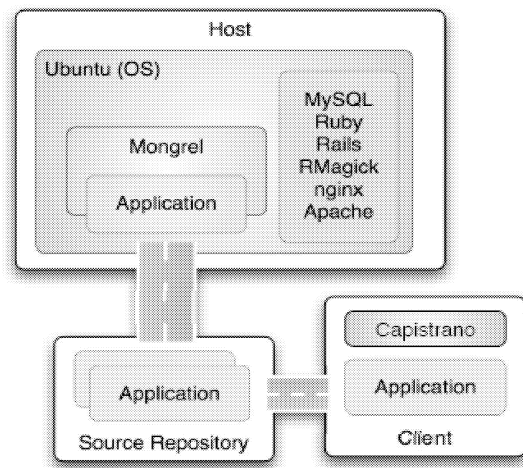
These tendencies demand that the companies have the capacity of incorporating in their products the best resources available on the market and dynamically adjust their inter-organisational structure in order to keep to the maximum their adaptability to the business opportunities.

## THE VIRTUAL ENTERPRISE APPLICATION

The changes that occur in the business environment require a permanent adaptation of the organisation partners in the virtual enterprise (VE), in other words the aligning with the business opportunities. By alignment, in this context, one understands the actions that must be taken in order to maintain the cohesion between the opportunity of the market and the delivery of the required product, which must have the required specifications, must be done on time, at the best cost and bringing the best profit. The reconfigurability, which means the ability to bring certain modifications in the structure of a VE because of unpredictable changes occurred on the market, is a major requirement of a VE in order to maintain its partners aligned to the market requirements. It is a consequence of the dynamics of the product life cycle. This requirement implies the ability to be:

- flexible and to be able to access almost instantaneously the optimal resources in order to integrate them in the company;
- apt for design, negotiation, management and production, independent from the distances at which the partners find themselves;
- apt for reducing to the minimum the reconfiguration or the integration time.

The virtual enterprise (VE) application is recommended to be used as a “**hosted application**”, benefitting at the same time from an administrator and a moderator. The application has the primary objective of creating an Internet portal that will offer information about “manufacturing skills and capacities of potential partners” as well as about potential AVEs (Agile Virtual Enterprises), incorporated on the site (figure 1).

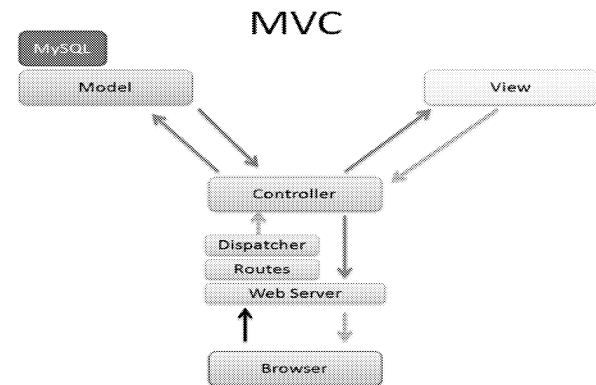


**Figure 1** The scheme of the site of the application and its exploitation

It is well known that the apparel industry is characterised by networks of companies that cooperate among them, each of them specialised in different stages of production. Generally, a first contractor is required, to collect the clients’ requests and to be responsible of their transformation into production requirements. Later, they are distributed to the subcontractors who supply finite or semi finite products to the first contractor. The selection of the subcontractors is a time consuming process that also implies costs, and is meant to lead to the best partner selection of a VE from a group of candidate firms. By using this portal and the calculation of the candidates’ agility, the process becomes shorter and it uses less resources. Once the partners are found, a simulation of the entire process of production of the goods required by the clients can be made by using **ARENA**. This is a simulation product linked to the home page and that can be purchased in advantageous conditions. The VE existing systems that do not use the new technologies are characterised by a long period of time between the receiving of the order and its processing, between finding and selecting the subcontractors, as well as by a high cost of the resources. Another negative aspect is the fact that the high and medium level employees allocate a great deal of their working time to activities that do not add valour to the final product. The new paradigm tries to eliminate these shortcomings.

### The structure of the application

The **Agility** application, which is Virtual Factory type, achieved by using the latest versions of the Ruby products 1.8.6, Ruby on Rails 2.0.2, MySQL 5.0, Appache 2.2, FireFox, Mongrel, etc., based on the MVC paradigm (Model Controller View), having the type of structure presented in figure 2, is meant to offer functions like CRM – Customer Relationship Management, CMS – Content Management System, ERP – Enterprise Resource Planning, to the *protagonists* from the textile and apparel industry.



**Figure 2** The structure of the application

Some explanations regarding figure 2:

- **The Browser** sends a request such as: <http://mysite.ro/gallery/show/3>;
- **The Web Server** (Mongrel, etc.) receives the request. It uses **Routes** to find the required *controller*. In the present example, the *controller* is *gallery*, the method is *show* and the *id* is 3. The web server uses the **dispatch** to create a new control command, calls the action (method) and transfers the parameter to it;
- **The Controller** verifies the request it receives from the user, the associated data, *cookies*, *sessions*, etc. and asks the **model** to provide the data from the base;
- **The Models** are Ruby categories. They “speak” to the data basis, memorise and validate the data, handle the logistics of the business, etc;
- **Views** are what the users see on the screen: HTML, CSS, XML, Javascript, JSON;
- **The Controller** sends the *corpus* of the answer to the server (HTML, XML, etc.). The server combines the raw data into an adequate HTTP answer and sends it to the user.

The application will provide a system – *full cycle* – of business management from the textile and apparel industry to the clients, suppliers, the manufacturers, etc.; more precisely, to the entire *Supply Chain*.

### The functions of the application

The functions of the application can be classified according to their specificity as following:

- The management of the contact between the *protagonists* in that field: an information source with the purpose of contacting the *protagonists*; the administration of the access accounts and assigning the access privileges to the application resources; the construction of *templates* such as: email, newsletter, etc., in order to exchange messages;
- The management of communications between *protagonists* , done by means of: email; forum; blog; news; newsletter; “wiki” pages; “user” WebPages made by the partners;
- The management of the commerce by generating online invoices, catalogues, etc., having the format: HTML; PDF;
- Project management: by monitoring the starting time, finishing time, the time needed to accomplish different stages as well as the entire project, to finish tickets with various actions that need to be taken care of as soon as possible; monitoring the tasks assigned to the partners; planning; obtaining graphic representations of the simulated/real results.

**Figure 3 The first page and the specific links**

The application displays the dialog interfaces either in the language in which the browser is set, or in another language, depending on what language the user requires. Anyone who connects to the address of the site that hosts the application (figure 3) will have access to public information, without having to connect via login. The access to the protected area of the application is granted if the users authenticate themselves by using the *login* page, like in figure 4. Depending on the settings, each project partner can be kept informed with the evolution of their project by email messages that are sent automatically by the system, or by accessing the *newsletter* pages, the *forum* pages attached to the project (only for that specific project they are involved in), the *News* or *Wiki* pages. Each project partner can *upload* graphic images (pictures) that can form a *Gallery*, and even documents (specifications, etc.) that can contribute to the well functioning of the project. The *Grantt* diagrams and the calendars can offer useful information to the project manager in order to achieve a better coordination of the activities linked to the project. The activity map can also be a useful monitoring instrument. The news, automatically conveyed via e-mail, can speed up the communication process between partners. *Chart* type of diagrams can offer useful information about various parameters that characterise the project. In case there is a need to insert new additional information attached to projects, tickets, etc., there is the possibility of creating new additional fields by authorised users. The intervention of the application's programmer is not required. Tickets can be used to assign, during the process, urgent tasks to some partners or auxiliary parties (translators, for example). The responsibility of opening a ticket belongs to the one who initiates the action, and its closure, usually, to the one who carries it out. The *workflow management* can be done by the manager of the project.

Each project partner can *upload* graphic images (pictures) that can form a *Gallery*, and even documents (specifications, etc.) that can contribute to the well functioning of the project. The *Grantt* diagrams and the calendars can offer useful information to the project manager in order to achieve a better coordination of the activities linked to the project. The activity map can also be a useful monitoring instrument. The news, automatically conveyed via e-mail, can speed up the communication process between partners. *Chart* type of diagrams can offer useful information about various parameters that characterise the project. In case there is a need to insert new additional information attached to projects, tickets, etc., there is the possibility of creating new additional fields by authorised users. The intervention of the application's programmer is not required. Tickets can be used to assign, during the process, urgent tasks to some partners or auxiliary parties (translators, for example). The responsibility of opening a ticket belongs to the one who initiates the action, and its closure, usually, to the one who carries it out. *The workflow management* can be done by the manager of the project.

The application has two main components: **the administration of the application** and **the proper application**.

- *General settings:*
  - the title and subtitle of the application;

- themes;
- the language in which the interface elements are displayed; the interface elements will be automatically displayed in the language in which the server is set to function or the language set by the user when creating the account;
- the format of the calendar date, depending on language or any other criteria;
- the name of the computer that hosts the application, or its IP;
- the used protocol in order to generate the links from the e-mail notifications (*http* or *https*);
- the formatting of the texts applied to the “description” of the fields from tickets, news, documents and so on;
- the limit of the Feed contents (the maximum number of registrations from a RSS feed);
- the coding of the folder data (UTF-8, ISO, 8859-15, CP1252, etc.);
- logging in (if wanted or not);
- auto-registration (Yes/No);
- whether the lost password can be retrieved or not;
- if one can or cannot auto-register.

- *The setting of tasks and permissions:*

Tasks allow establishing what the members of a project are allowed to do regarding the access to the functions of the associated website.

- *The user fields* allow to the user to bring additional information to the projects, tickets, users, etc. Such a field can be designed as:

- **an integer:** a positive or negative number;
- **a string:** a single row string;
- **a text:** displayed on multiple rows;
- **date:** calendar date;
- **boolean:** checkbox;
- **list:** drop down list.

At these parameters one can also associate some validations, such as: minimum and maximum length, regular expressions, etc.

- *The management of the projects* (figures 5-9):

- the title of the project;
- a short description of the project;
- the user states the public/non-public character of the project. If the project is public, then it can be viewed by anyone. If the project is private, then it can be seen only by the user who has the right to access it;
- **subprojects** can be mentioned if the project is divided into smaller subprojects that are connected to it;
- the calendar date when the project was created;
- the project can be erased, if the context demands it, in case it is not archived;
- the members and their tasks can be defined during the process;
- a **Wiki** page can be created for each project;
- **forums** can be added to each project that allows “online” debates about its contents.

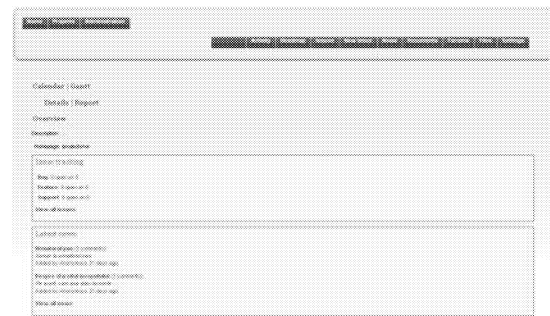


Figure 5 The general information about a project

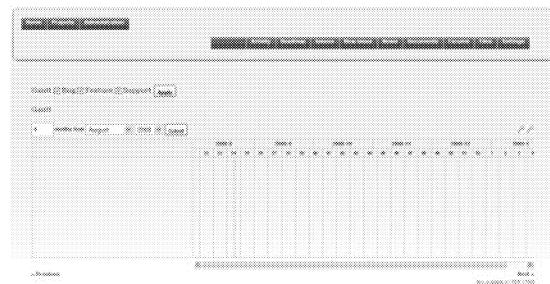


Figure 6 The GANTT diagram of the project

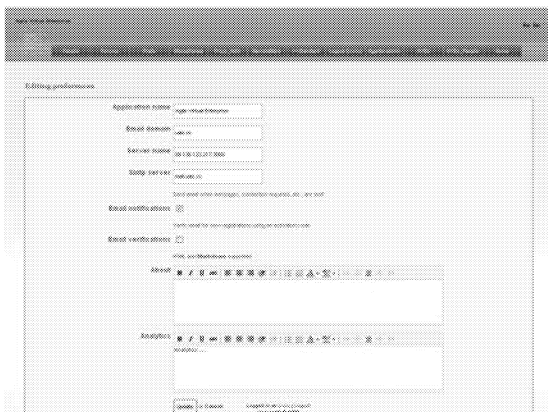
| Name                | Size    | Date                | Type    | P.C.    | M.P.    |
|---------------------|---------|---------------------|---------|---------|---------|
| 10-10-2008-10-20-08 | 1000000 | 10-10-2008-10-20-08 | 1000000 | 1000000 | 1000000 |
| 10-10-2008-10-20-08 | 1000000 | 10-10-2008-10-20-08 | 1000000 | 1000000 | 1000000 |
| 10-10-2008-10-20-08 | 1000000 | 10-10-2008-10-20-08 | 1000000 | 1000000 | 1000000 |

Figure 7 The list of the associated files

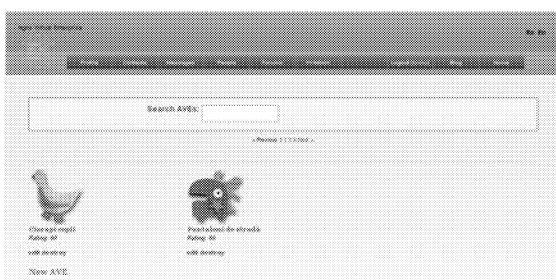
## CONCLUSIONS

Companies are focusing more and more on a collaboration based on computer networks, on Internet, etc., as a reaction to the accentuated changes that occur on a dynamic turbulent market. The concepts of Virtual factory, Virtual Enterprise, seem to offer solutions in order to surpass the periods of crisis by guaranteeing the survival of VE partners, the profit of such organisations, etc. In a nutshell, they can offer a prompt answer to clients' demands in a turbulent competitive context, they can contribute a great deal to the protection of the environment, they can guarantee the quality of products according to the customers' demands; therefore, they become agile.





**Figure 8 Setting the parameters**



**Figure 9 List of virtual enterprises created on the site**

By its two essential coordinates:

1. strategic market, linked to the management of the activities carried out in multiple locations by the suppliers and the business partners united in a strategic alliance, and
2. implementation techniques, linked to the sophisticated computer simulations of the distributed and integrated models belonging to the manufacturing environments,

the VF can significantly contribute to an agile behaviour of the business organisations. There is a tight connection between the two coordinates: the former is based on the latter.

The development, manufacturing and the marketing of new products, as well as the ability of reacting fast and flexible to the constant changing of the market demands, are decisive preconditions for assuring and developing the competitiveness on a global scale. Apart from flexible production, an adequate support for creating new products by using special computer-assisted design and fabrication methods represents an essential challenge in the context of general competitiveness. This is why the main objective of the 'Virtual Factory' module is to improve the general technical conditions in order to obtain products with a deep impact on the nowadays and future markets. The development of ICT in various enterprises has followed different coordinates from one to another, using different tools and standards, and the agility obtained by accepting the concepts of VF and VE seems the successful solution; this is why, in order to survive the harsh market competition, the elaboration of interfaces and new software products that allow interconnecting and

information exchange among partners as well as transparency towards the final user is imposed. There are more and more attempts to elaborate such software tools so as to reorient the current state of the European sector in the field of textile and apparel towards a competitive future of this industry. For simulations, the ARENA programme is recommended, which is 'the world's most effective simulation technology for modelling systems in manufacturing, transportation, logistics, warehousing, and business processing'.

## REFERENCES

- Luck, M., Marik, V., Stpankova, O., Trappl, R., Lecture Notes in: *Artificial Intelligence* LNAI 2086, pages 335-364, Springer, ISBN 3-540-42312-5, July, 2001
- Camarinha-Matos, L. M., Afsarmanesh, H., 2001, "Virtual Enterprise Modeling and Support Infrastructures: Applying Multi-Agent Systems Approaches", in: *Multi-Agent Systems and Applications*, 2001
- Yhai, W., Fan, X., Zan, J., Pengsheng, Y., "An Integrated Simulation Method to Support Virtual Factory Engineering", in: *International Journal of CAD/CAM*, Vol. 2, No. 1, pages 39-44, 2002
- Camarinha-Matos, L. M., Afsarmanesh, H., "Collaborative networks: A new scientific discipline", in: *Journal of Intelligent Manufacturing*, vol. 16, No. 4-5, pages 439-452, 2005
- Camarinha-Matos, L. M., Abreu, A., "Performance indicators based on collaboration benefits", in: *Proceedings of PRO-VE'05, Collaborative Networks and their Breeding Environments*, Springer, pages 273-282, Valencia, Spain, 26-28 September, 2005
- Camarinha-Matos, L. M., Afsarmanesh, H., Ollus, M., *Virtual Organizations – Systems and Practices*, Springer, 2005
- Camarinha-Matos, L. M., Afsarmanesh, H., "Collaborative Networks, Value creation in a knowledge society", In: *Proceedings of PROLAMAT 06*, Springer, Shanghai, China, 14-16 June, 2006

## BIBLIOGRAPHIE

**ANTONELA CURTEZA** is the Design and Product development course leader in The Knitting and Garment Technology Department, an authorised trainer for Sympatex Technologies GmbH Germany and director of research projects. Since 2008 she is full academic professor and starting with 2009 she is PhD coordinator also. She has a number of trainings in research and design/product development management. She is mainly interested in all areas of research regarding garments and fashion business.

# EMBEDDING SIMULATION IN LEAN PROJECTS

Yan Jia and Terrence Perera  
Faculty of Arts, Computing, Engineering and Sciences  
Sheffield Hallam University  
S1 1WB, Sheffield,  
UK  
E-mail: [Jayne.Jia@student.shu.ac.uk](mailto:Jayne.Jia@student.shu.ac.uk)

## KEYWORDS

VSM template, parameter optimization.

## ABSTRACT

This paper identifies some limitations of lean technologies and describes a simulation based approach to overcome some of these limitations. To enable the embedding of simulation in lean projects, simulation template has been developed for both push and pull (Kanban) systems. A case study is presented to demonstrate the use of template and optimization techniques within lean projects.

## 1 INTRODUCTION

Process improvement techniques such as Lean and Six Sigma methodologies have become mainstream management tools in many industrial sectors. Lean aims to continuously minimizing waste (non-value-added activities) to maximize flow whilst Six Sigma focuses on measuring “defects” in a process and eliminate them to get as close to “zero defects” as possible. Although these two methodologies are well known and have been used in a variety of industrial settings, they both have some limitations.

Referring to Lean Methods, Standridge and Marvel (2006) describes some limitations which include lack of modeling and assessing the effects of variation, non-use of all available data, inability to validate effects of proposed changes before implementation and inability to identify other possible improvements, and the interaction effects between system components.

Although Value Stream Mapping (VSM) has become an integral and essential technique in lean projects, it cannot capture complex interactions that exist in real world systems. For example, predicting inventory levels throughout the production process is usually impossible with only a future state map, because with a static model one cannot observe how inventory levels will vary for different scenarios (McDonald et al. 2002). In general, we need a complementary tool with VSM which is capable of generating resource requirements and performance statistics whilst remaining flexible to specific organizational details (Abdulmalek and Rajgopal. 2007). An obvious tool is simulation.

## 2 LITERATURE REVIEW

Many articles have studied the use of simulation in Lean manufacturing. For instance, Abdulmalek and Rajgopal (2007) develops a simulation model (using Arena 5) to analyze the benefits of Lean manufacturing and value stream mapping in a large integrated steel mill. Schroer (2004) used simulation as a tool in understanding the concepts of Lean manufacturing, including line balancing against Takt time, pull versus push manufacturing, Kanban inventory control, and process variability reduction. Lian and Van (2007) summarizes the modeling procedure and describes the formal value stream mapping paradigm (VSMP) in view of simulation and define the framework and elements of an enhanced VSM approach, we call simulation-based VSM (SimVSM). Via the use of SimVSM managers can see the impact of Lean transformation before the actual implementation. Many papers give specific case studies of how simulation could be used within the Lean manufacturing strategy, for example, Hild (2001) and Casanovas (2006) developed simulation models for the economic optimization of clinical trial designs and management. An economic optimization function integrating the statistical power, cost and duration variables of clinical trials was formulated, and the use of simulation to assess this optimization is described. Adams et al. (1999) presents a description of the tools of Lean manufacturing, the steps in the continuous improvement process and two case studies where simulation was used in the continuous improvement process. Wang et al. (2005) developed simulation models to compare the system's performance based on the same production line operated with fixed workers or with walking workers. Bodner and Rouse (2007) studies R & D management issues from the perspective of their effect on value creation and presented an application of the model to R & D in the forest products industry. Huang and Liu (2005) built simulation models to identify focused stages in which to start Lean control.

## 3 SIMULATION MODEL AND TEMPLATE BUILDING

From the literature review we can see the combination of Lean methodology and simulation has achieved greater success than merely using Lean method. Presently simulation method embedment is only in its infancy stage, and it does not play a full role in Lean projects. Since it is very important and useful to embed simulation into Lean Projects, it is urgent and necessary to develop the overall framework to embed simulation into practical engineering methods and applications. As a first step of the framework, a template which contains seven modules is developed with

Arena 10.0 Professional to simulation the Push and Pull scenarios of VSM.

In this section, we explain modules in the template we built by using Arena 10.0 Professional. Table 1 lists the templates we built for Push and Pull systems.

### 3.1 Simulation Templates

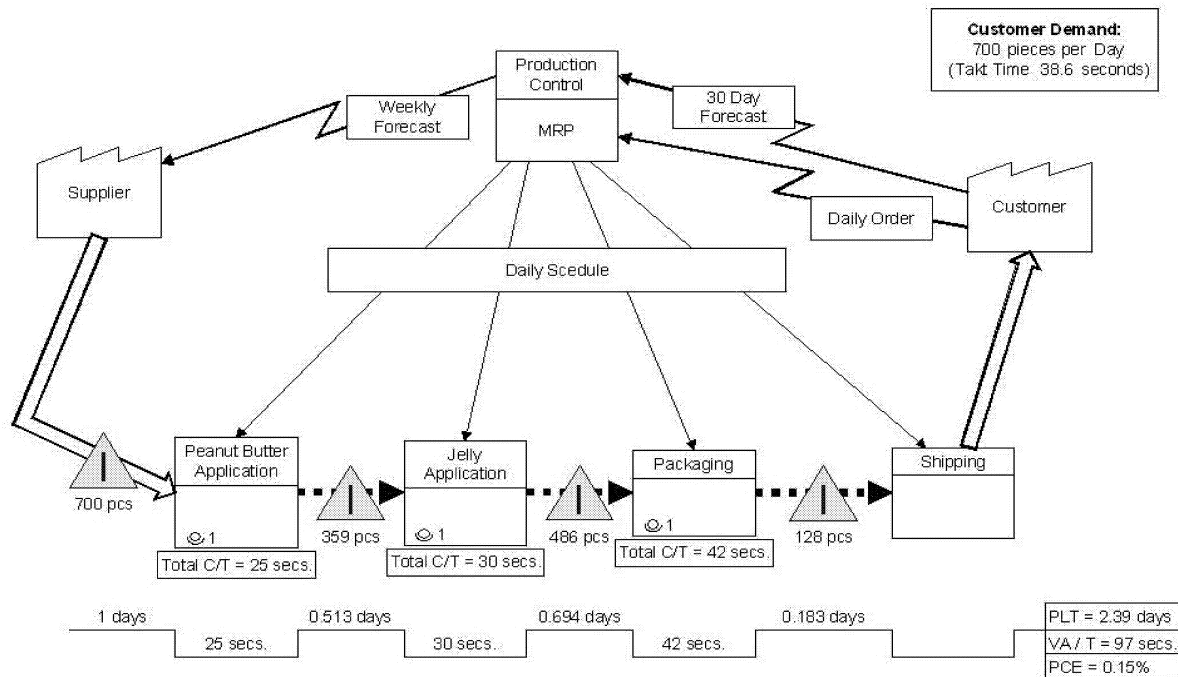


Figure 1: Push System

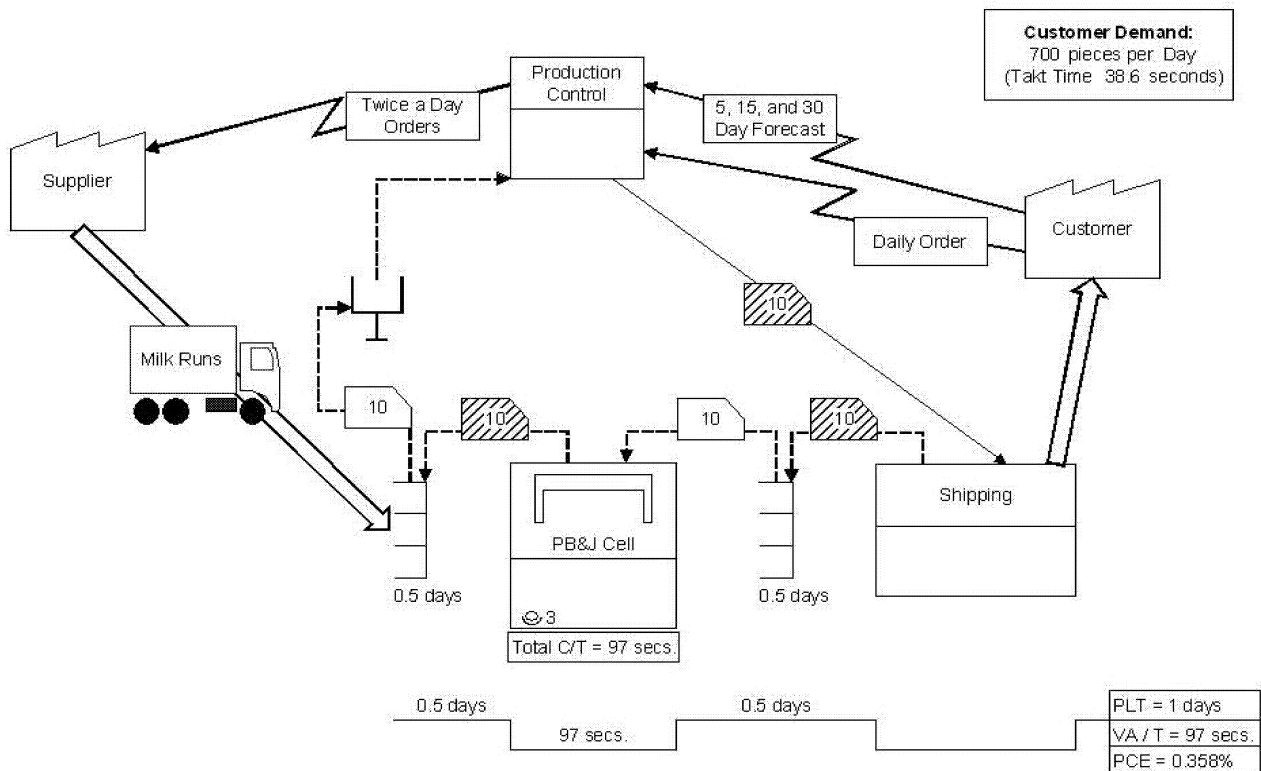


Figure 2: Pull System

### 3.2 Template Mapping from Visio 2007 to Arena 10.0

A mapping is built from Visio 2007 VSM Template to Arena 10.0 VSM Template, which means that by building a VSM in Visio 2007 and filling in the data, then exporting into Arena, a simulation model of dynamic VSM is built automatically and most data is collected.

Table 1: Simulation Templates for Push and Pull Systems

| Template Name  | Push | Pull | Descriptions   |
|----------------|------|------|--|
| Supplier       | ✓    | ✓    | Create entities  |
| Customer       | ✓    | ✓    | Start and end of the supply chain; Send kanban out in pull system  |
| Shipment Truck | ✓    | ✓    | Transport entities from Supplier or to the Customer  |
| Process VSM    | ✓    | ✓    | Manufacturing steps in the supply chain; Also send out withdraw kanban and receive production kanban in pull system  |
| Inventory      | ✓    |      | An interface / buffer between two processes ready to ship the parts when downstream process has capacity.  |
| Supermarket    |      | ✓    | An interface / buffer between two processes ready to ship the parts when receiving a withdraw kanban from downstream; a production kanban would then be released to upstream process to replenish the supermarket. |
| Kaizen Burst   | ✓    | ✓    | A symbol which important information can be marked on  |

### 4 CASE STUDY

PB&J manufacturing produces sandwiches (Pereira 2008), production data is listed below:

- Daily demand: 700 pieces
- Hours per shift: 8
- Break minutes per shift: 30
- Shifts per day: 1
- Days per week: 5
- Takt Time: 38.6

Two scenarios are simulated, push and pull system. The manufacturing processes in Figure 1 (push system) consist of material purchasing from supplier, Peanut Butter Application, Jelly Application, Packaging and finally Shipping to Customer.

The "future state" is a pull system (Figure 2) which contains supermarket and kanban. It is based on the logic that nothing will be produced until it is needed. The U-shaped flow line is a combination of Peanut Butter Application, Jelly Application and Packaging operations.

#### 4.1 Key Measurements

The definitions of our key measurements are given as follows:

- Production Lead Time (PLT)
- Value Added Time (VA/T)
- Process Cycle Efficiency (PCE)

#### 4.2 The Simulation Scenes

In push production system (Figure 3), there are inventory, scrap rate is set at 6%, we can see long lead times and low value-added ratios. In pull system (Figure 4), the inventory is replaced by supermarkets, and there is no scrap rate due to U-shape flow line design, and the lead times are shorter, the PCE is increased a lot.

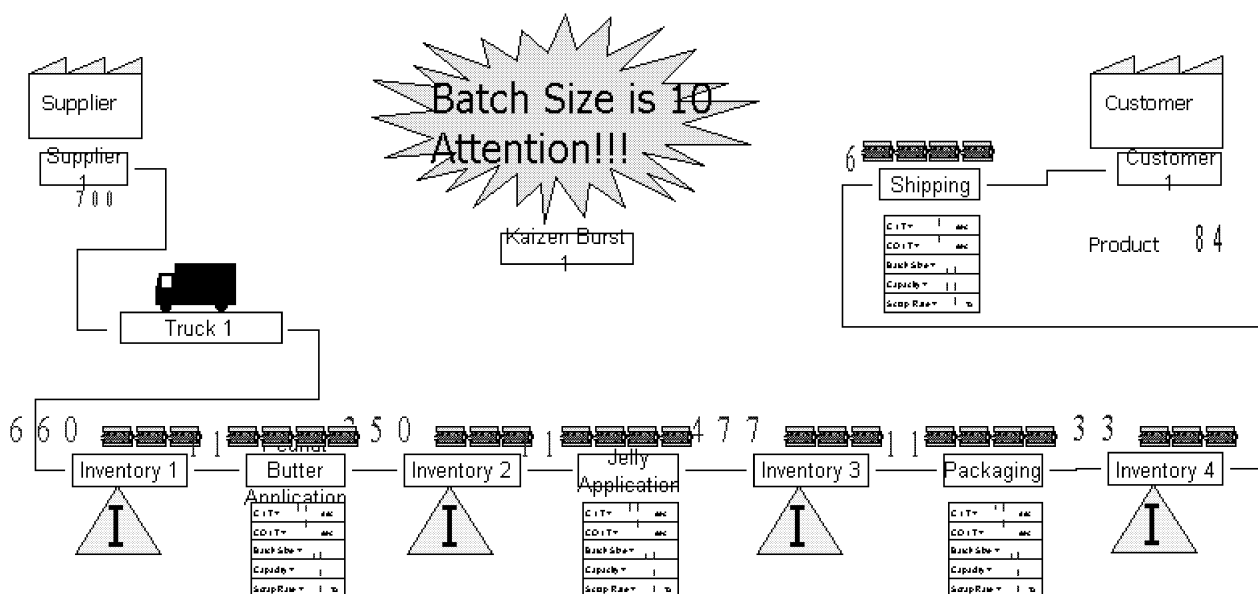


Figure 3: Push System in Simulation

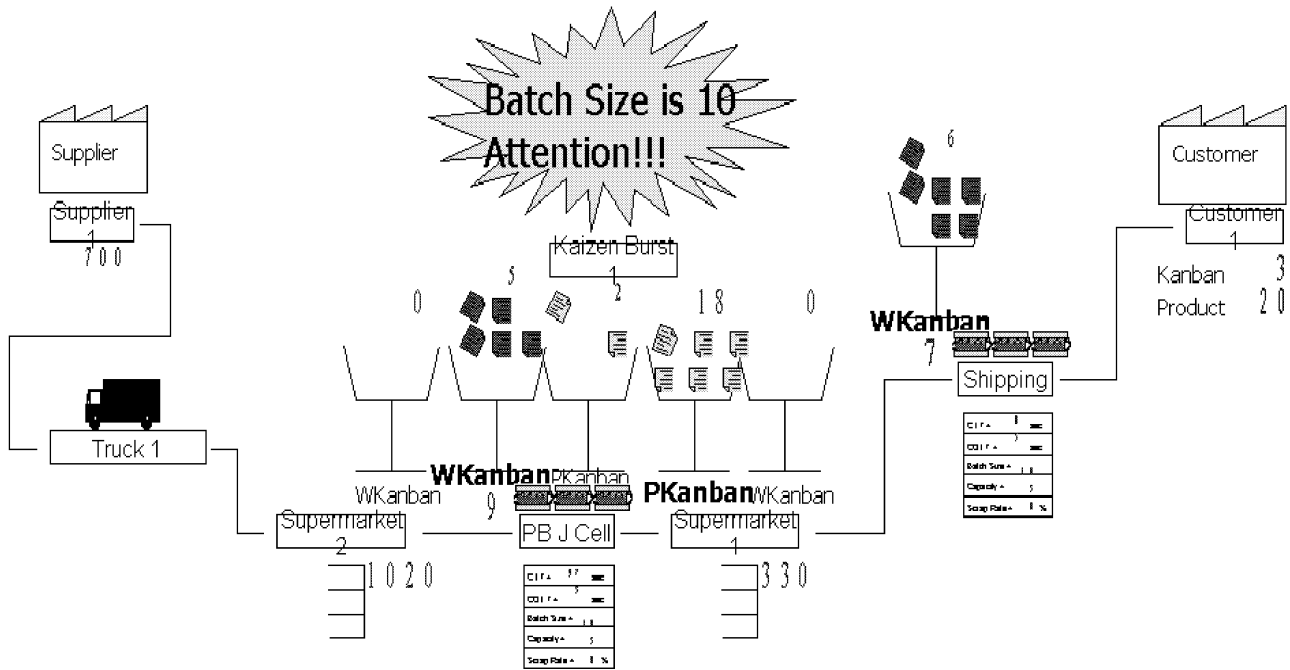


Figure 4: Pull System in Simulation

## 5 SIMULATION RESULTS

Table 2: Simulation Output Data

| Key Measurements | PUSH System | PULL System |
|------------------|-------------|-------------|
| PLT              | 2.14 days   | 0.91 days   |
| VA/T             | 93.6 sec    | 94.9 sec    |
| PCE              | 0.15%       | 0.36%       |
| Utilization (%)  |             |             |
| Peanut Operator  | 64%         | N/A         |
| Jelly Operator   | 74%         | N/A         |
| Package Operator | 93%*        | N/A         |
| U-Cell Operator  | N/A         | 81%         |

\*High Utilization – Bottleneck

The output data is the values of 1 replication, 8-hours per day, and 22 days per replication (22 working days per month).

Data in Table 2 shows that the Lean pull production system solves the bottleneck problem, shortens the lead time 58%, and increases PCE from 0.15% to 0.36%.

## 6 PARAMETER OPTIMIZATION IN SIMULATION

The biggest advantage of simulation is its ability to answer “what if” questions, as well as deal with complex situations, which are more appeal to realistic systems.

We use *OptQuest* software to do the parameter optimization process. For example in push system, if we change batch size in a range of [1, 10], change workers' capacity from 1 to 3, with the limit of 5 workers in total, the optimal parameter set for the object of minimum lead time is shown in Table 3. After optimization, the Lead Time reduced from 2.14 days to 0.51 days, PCE increased from 0.15% to 0.64%, and no bottlenecks in the system.

Table 3: The Optimal Parameter set for the Minimum Lead Time in Push System

| Optimal Parameters and Key Measurements | PUSH System |
|---|-------------|
| Batch Size                              | 6           |
| Capacity In Peanut Process              | 1           |
| Capacity in Jelly Operator              | 2           |
| Capacity in Package Operator            | 2           |
| PLT                                     | 0.51 days   |
| VA/T                                    | 94 sec      |
| PCE                                     | 0.64%       |
| Utilization (%)                         |             |
| Peanut Operator                         | 64%         |
| Jelly Operator                          | 37%         |
| Package Operator                        | 50%         |

In Pull system, if we change batch size in a range of [1, 10], change workers' capacity from 2 to 5, change inventory amount in supermarket 1 and supermarket 2 in a range of [10, 700], the optimal parameters and results are listed in Table 4. The data in Table 4 illustrate that when batch size reduces to 1, with workers' capacity of four, inventory amount of ten in each supermarket, lead time is reduced from 0.91 days to 0.43 days, PCE increases from 0.36% to 0.78%, operators' utilization decreases from 81% to 59%.

Through the case example, we can see the use of simulation templates achieves greater success than merely using static VSM:

- Firstly, more accurate data can be got by using simulation. Since simulation results shows the average output value of the system in a period (such as one day, one month, or even one year) while the static VSM is just a snapshot of the system, which is not reliable at all;
- Secondly, users can easily modify the parameters in the

simulation model and immediately run the model to watch the system results under different parameters;

- Finally, through the use of *OptQuest*, it is convenient to do the data analysis and automatic parameter optimization, which could not be completed or even imagined by using static VSM or rely on people's experience.

Table 4: The Optimal Parameter set for the Minimum Lead Time in Pull System

| Optimal Parameters and Key Measurements | PULL System |
|---|-------------|
| Batch Size                              | 1           |
| Capacity In U-Cell Operators            | 4           |
| PLT                                     | 0.43 days   |
| VA/T                                    | 97 sec      |
| PCE                                     | 0.78%       |
| Utilization (%)                         |             |
| U-Cell Operators                        | 59%         |

## 7 SUMMARIES AND FUTURE RESEARCH

This paper described simulation templates we have built for VSM tool. The case study illustrated that simulation can be used to get a better analysis of the system, calculate resource usage, discover bottlenecks, deal with more complex systems, change parameters to see the alternative results (change scrap rate in process, change capacity in processes, change batch size, etc.), and optimize parameters automatically to minimize lead time within seconds. This can not be done by any static methods or tools. The mapping from Visio 2007 to Arena 10.0 is built, which makes the model building much easier by drawing a VSM in Visio and exporting to Arena.

Future research:

Template building for VSM (Value Stream Map) is just one step of the research work. Next step is to study embedment of simulation in Six Sigma methodology, and finally propose a guidance framework about embedding simulation in Lean and Six Sigma projects.

## REFERENCE

- Bodner, D.A. and W.B. Rouse. 2007. "Understanding R and D value creation with organizational simulation" [online]. *Systems engineering*, 10 (1), 64.
- Adams, M.; P. Componation; H. Czamecki and B.J. Schroer. 1999. "Simulation as a tool for continuous process improvement" [online]. *Winter Simulation Conference Proceedings*, 1 766.
- Huang, C.C. and S.H. Liu. 2005. "A novel approach to lean control for Taiwan-funded enterprises in mainland china" [online]. *International journal of production research*, 43 (12), 2553.
- Standridge, C.R. and J.H. Marvel. 2006. "Why lean needs simulation". *Winter Simulation Conference Proceedings*, IEEE, Piscataway, NJ, USA. 1907-1913
- Mcdonald, T.; E.M. Van Aken and A.F. Rentes. 2002. "Utilizing simulation to enhance value stream mapping: a manufacturing case application". *International Journal of Logistics: Research and Applications* 5(2), 213-232.
- Wang, Q.; G.W. Owen and A.R. Mileham. 2005. "Comparison between fixed- and walking-worker assembly lines" [online]. *Proceedings of the institution of mechanical engineers. part B, journal of engineering manufacture*, 219 (11), 845.
- Abdulmalek, F.A. and J. Rajgopal. 2007. "Analyzing the benefits of Lean manufacturing and value stream mapping via simulation: A process sector case study" [online]. *International journal of production economics*, 107 (1), 223.
- Pereira, R. 2008. *The LSS Academy Guide to Lean Manufacturing* [e-book].
- Schroer, B.J. 2004. "Simulation as a tool in understanding the concepts of Lean manufacturing" [online]. *Simulation*, 80 (3), 171.
- Lian, Y.H. and L.H. Van. 2007. "Analyzing the effects of Lean manufacturing using a value stream mapping-based simulation generator" [online]. *International journal of production research*, 45 (13), 3037.

## AUTHOR BIOGRAPHY

**YAN JIA** is a Ph.D. student in Materials and Engineering Research Institute at Sheffield Hallam University. She finished her master's degree in Automatic Control Research Institute at North China University of Technology, China. Her research area includes simulation integration in lean and six sigma projects. Email to <Jayne.Jia@student.shu.ac.uk>

**TERRENCE PERERA** is a Professor of Manufacturing Systems and the Assistant Dean (Academic Resources) at the Faculty of Arts, Computing, Engineering & Sciences, Sheffield Hallam University in UK. He also leads the Systems, Modeling and Integration Research Group. His research interests include practices, integration and implementation of simulation modeling within the industrial sectors. Email to <T.D.Perera@shu.ac.uk>

# SOFTWARE TOOLS INTEGRATION FOR THE DESIGN OF MANUFACTURING SYSTEMS

Pavel Vik

Technical University of Liberec  
Faculty of mechanical engineering  
Department of production systems  
461 17 Liberec 1  
Studentská 2  
Czech Republic  
vikpavel@seznam.cz

Guilherme Pereira  
Luís Dias

University of Minho  
School of Engineering  
Production and Systems Dpt  
Campus de Gualtar  
4710 057 Braga  
Portugal  
gui/lsd@dps.uminho.pt

## KEYWORDS

Integration, Design of production system, Simulation, Witness, Autocad

MAYA, modules of virtual reality included in simulation tools like VR WITNESS or 3D ARENA PLAYER)

## ABSTRACT

Nowadays there is great pressure on production systems design to be developed rapidly and efficiently. Moreover, designs are faced with numerous requirements for system flexibility, modularity and robustness. For this purpose several software tools have been used – mainly for project analysis, design and validation. Nevertheless, these tools have been used with low integration levels, including absence of data coherence.

This paper deals with production systems design, and more specifically, with the integration of particular software tools into unified system architectures. The paper identifies a variety of tools, describing their functions and principles of integration. It also addresses the way this integration enables automatic generation of simulation programs. Finally, we discuss ways in which integration can contribute to the automatic generation of different patterns of project layouts.

## INTRODUCTION

The production system's design is a process of managing technical and organization variants in order to configure system elements with regard to optimal usage of all sources (i.e. material, energy, areas), productive and non-productive facilities (production machines, handling and checking facilities) and workers. These variants deal with spatial layout and also with dynamic time changes. (Zelinka 1995, Muther 1970)

Generally it is possible to divide the software tools used in production system's design, into four groups:

- **Databases** for data analysis, data storage, data mining (product-quantity analysis, cluster analysis etc.) – e.g. MS Access, MS Excel, SAP etc.)
- **CAD systems** for graphic designs, verification of spatial layout variants (for general work there are AUTOCAD, SOLID WORKS, MICROSTATION or specific software tools focused into layout design – FACTORYCAD, MATPLAN or FASTDESIGN)
- **Discrete computer simulation systems** for verification of variants with time changes (WITNESS, ARENA, SIMUL8, SIMPRO, etc.) (Dias, et.al, 2007)
- **Visual systems** for 2D, 3D animation or virtual reality for improved understanding and presentation of suggested principles (3DS MAX, BLENDER,

The idea of integration is not new. In fact the literature show us several types of application supporting some level of –integration (Moorthy 1999, Mecklenburg 2001, Havlik 2005). In the next section the aforementioned level of integration is discussed.

## Description of some software examples

The software tools mentioned above are usually used separately or with minimum integration possibilities. For example application MatPlan supports disposition of facilities based on material flows (triangular method - a simple constructive heuristic method) and it can generate a simple simulation model into Witness simulation tool, data-loading from data-source (excel file and text file) and it is necessary to manual re-write data obtained from simulation into these files. It is an example of data-based integration without direct feedback information flows. (Markt 1997)

Application FASTDESIGN consists of two parts – FASTPLAN (based on MS Access database, for analysis like Product-quantity or cluster analysis helping in design of manufacturing cells or production lines) and FASTGRAF (based on AutoCad for generation of layouts according to data from database).

Simulation software tools are generally focused into time-studies (layout behaviour –accordingly to dynamic time changes) and there are not direct relationship with CAD tool, There is just a one-way information flow into simulation tool, e.g through SDX (Simulation Data Exchange, file generated by CAD tools), containing all information (production times, capacities, costs etc.). for creation of a full functional simulation model There is not data connection from simulation software into CAD. (Mecklenburg 2001, Moorthy 1999, [www.ugs.com](http://www.ugs.com))

So-called "Digital factories" have become available, which deploy the before-mentioned tools into one common solution, connecting these by means of information flows. The principals drawbacks relate to the cost of these software packages; the fact that they focus mainly on mass production (automotive industry) in large-scale production systems; and finally, their excessive level of complexity. This level of complexity is reinforced and exacerbated by the fact that only two *digital factories* are currently available – TECNOMATIX ([www.ugs.com](http://www.ugs.com)) and DELMIA ([www.delmia.com](http://www.delmia.com)) (for example software package TECNOMATIX consists of: Simulation software PLANT

SIMULATION; Applications for design layouts – FACTORYCAD and FACTORYFLOW; And another logistics software tools). Both mentioned systems of *digital factory* largely overlook the particular issues and requirements of small and middle production systems, for instance according to ease of use, configurability and control. (Bureš et al. 2007)

We have identified the following disadvantages and limitations associated with using software tools:

- Absence of data-flow between programs
- Existence of several independent data-sources
- Absence of feedbacks which are necessary for registering of changes suggested during the design process
- Management of project variants

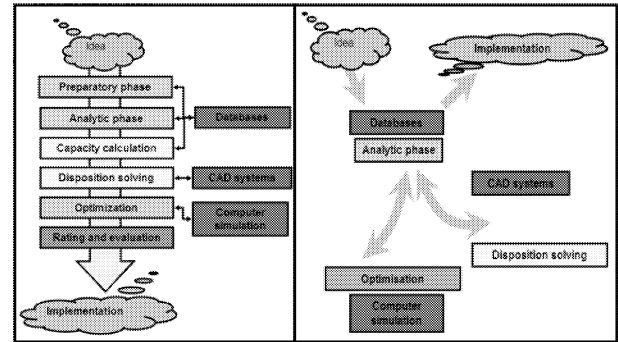
Manual project data-connection causes delays in the design process, duplication of work and can even be significant error sources.

From this point of departure, efforts are focused on conceptualization and elaboration of a compact system, which not only inter-connects particular systems, but simplifies and streamlines the entire design of production systems.

In this context, the following, most frequent barriers and challenges of the design process area are addressed: (Taylor 2008)

- Dimensioning of storages and buffers, establishment of storage limitation
- Optimal usage of sources
- Bottleneck analysis
- Design of handling systems (AGV, milk-run system, conveyor, manual handling etc.)
- Adaptability and issues related to long-term robustness against changes in – for instance - product mix, volume, costs or other parameters (Benjaafar 2002)
- Plant layout – taking into account real and specific constraints (current facilities, walls, pillars, ergonomics, etc.)
- The issues of related to layout or redesign – thus addressing scenarios more realistic and frequent than green-field design.

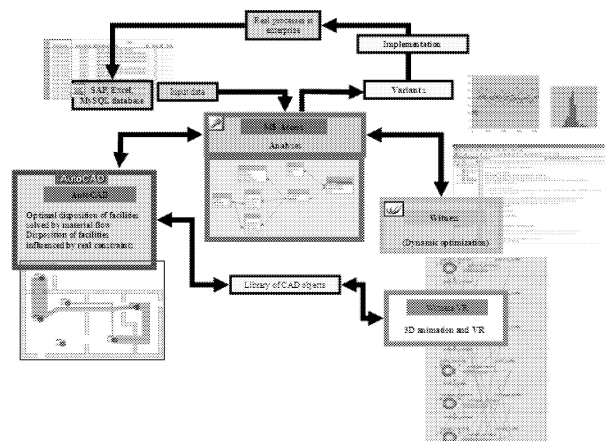
Figure 1 depicts a comparison between the currently most used principle (on the left side) and the new, suggested approach with relative interconnection (on the right side). Both approaches are based on “Systematic Layout Planning” developed by Muther (Muther 1973), the differences are in ways of application’s integration (Vik 2008).



**Figure 1** – Comparing of current and new approach of software interconnection

## DESCRIPTION OF THE NEW APPROACH

Below, a system is suggested on the basis of possibilities and given requirements. Three programs were chosen and integrated, **MS Access 2003** (database), **Witness Simulation tool v. 2008** (discrete simulation tool) and **AutoCAD v. 2004** (CAD system for drawing of layouts). These programs were chosen for their usage expansion in this area of design. It could be possible to use other software tools – e.g. Arena simulation tool and Solid Works CAD system due to similar properties as Witness and AutoCAD. Nevertheless authors experience with Witness and AutoCAD also contributed to explore its use in the present context.. In the diagram 2, overall integration with information flows is illustrated with -snapshots from individual programs.



**Figure 2** – Overall integration

## INTERCONNECTION

For integration and control of programs, Visual Basic for Application (VBA) was chosen as programming language. The main reasons behind this choice were:

- VBA’s presence in almost Microsoft, AutoCAD and other products (common language simplifies integration);
- easy language syntax and usage (especially interactive development, entertainment and visual tools);

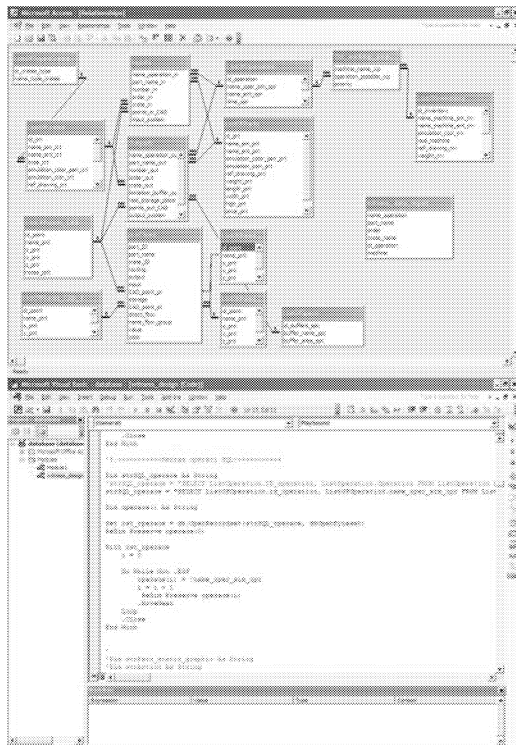


- ease of connection to databases (e.g. with help of OLE DB Microsoft Jet);
- support of **OLE Automation** (Object Linking and Embedding) enables control of other programs and use of their objects
- **DDE (Dynamic Data Exchange)** for data exchange between applications.

## DETAILED DESCRIPTION OF PARTICULAR TOOLS FUNCTION

### Database system

The common database (MS Access) serves as the source and deposit for project data (i.e. also for input and output data from simulation and CAD), and it builds a basic interconnection element. The database includes VBA macros for automatic generation simulation of elements and their properties with minimum user intervention. Figure 3 displays the relational database, its structure and VBA macros.



**Figure 3** – Snapshots from relational database and VBA macros

The needed data for the simulation model are:

- Definition of **operation** (input and output parts, size batches, assembly information, process times etc.);
- production **sequences** of parts (i.e. their routing thought production facilities);
- type of crates and **transported** quantity;
- productive and non-productive **facilities** (machines and handling equipment) and their properties (setup times, transport speeds);

- **scheduling** of production;
- **workers** and their properties.

The above data was used –in several examples and proved to be adequate to specify medium complexity productive systems.

The database, moreover, contains VBA macros for automatic loading of CAD objects into AutoCAD - also with minimum intervention - and elements of fundamental analysis (mainly PQ analysis). Throughout, emphasis is given to ease of data entry and modification – e.g. with respect to production plan or volumes. As an illustrative example, in the automotive industry, such flexibility is needed for multiple scenarios related to product facelifts, expansion/scaling, or the initiation of new production. On this platform, one can conduct experiments and test systems – and in turn - analyse properties and consequences into designed systems, test— the influence of uncertainty and determinate the general robustness and flexibility of the system. (Kulturel 2007)

### Simulation System

The generated simulation model consists of the before-mentioned elements and additional elements such as lists (operation, facilities, parts etc.), modules for project database establishment and modules involving SQL queries for data loading. These modules feed simulation elements by data and re, furthermore, automatically configured and updated. It is possible to configure the model on the basis actual data from databases (model's initialization in the beginning of simulation run). As a consequence, users only change data in the database though user-friendly forms in order to manage the simulation model.

After the generation and configuration of the model, subsequent simulation runs and experimentation is executed, and the obtained data is recorded back to the database. Output data are mainly stochastic material flows, average and maximal occupation of buffers (this information is used in design of the CAD layout), usages of sources (machines, operators and handling system), bottlenecks, system throughputs etc.

Figure 4 contains an illustration of a generated simulation model. The illustrative example – a simulation of unpacking and assembly of parts from transport crates - depicts the actual occupation of buffers (queues), usages of facilities (pie graphs) and a series of additional information (list of facilities, parts, and operation).

(Mujber 2005, Vilarinho 2003)

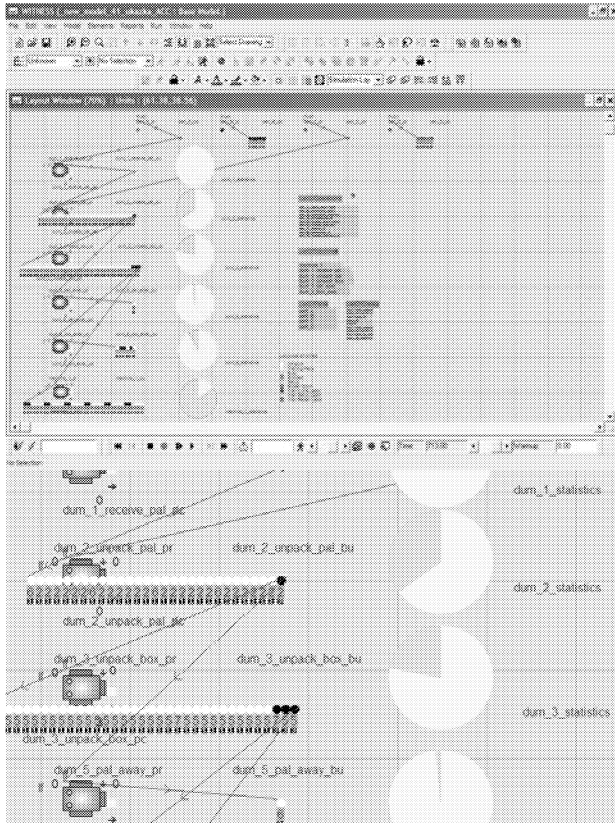


Figure 4 – Illustration of simulation model

### CAD System

The CAD system (exactly Autodesk AutoCAD v. 2004) is mainly used for displaying facilities in layout and their system settlement, i.e. finding the best location for each facility in the layout. For this purpose, VBA macros enable the identification of the optimal positions according to a triangular algorithm based on values of material flows. The detailed layout is influenced by several factors and limitations (e.g. current production facilities or building constraints). Graphic objects (drawings of sites/facilities) are automatically feed from data libraries into the layout according to the following sequence:

- 1) Material flows are displayed – deterministic or stochastic (obtained by simulation and involves stochastic influences).
- 2) Generated flows are defined as a direct two-point line, without respect to real deformation caused by real constraints (e.g. building). In order to display real material flows, it is necessary to change these lines into poly-lines (lines defined by several points) and modify their shapes with respect to real constraints.
- 3) The definition points of poly-lines are recorded into the database.

Figure 5 gives a simple exhibit of a layout with the material flows (ideal and real) of particular parts. For comparison of project variants based on material flows, it is possible to count a sum of material flows (real length multiplied by its width) and evaluate these through the numbers.

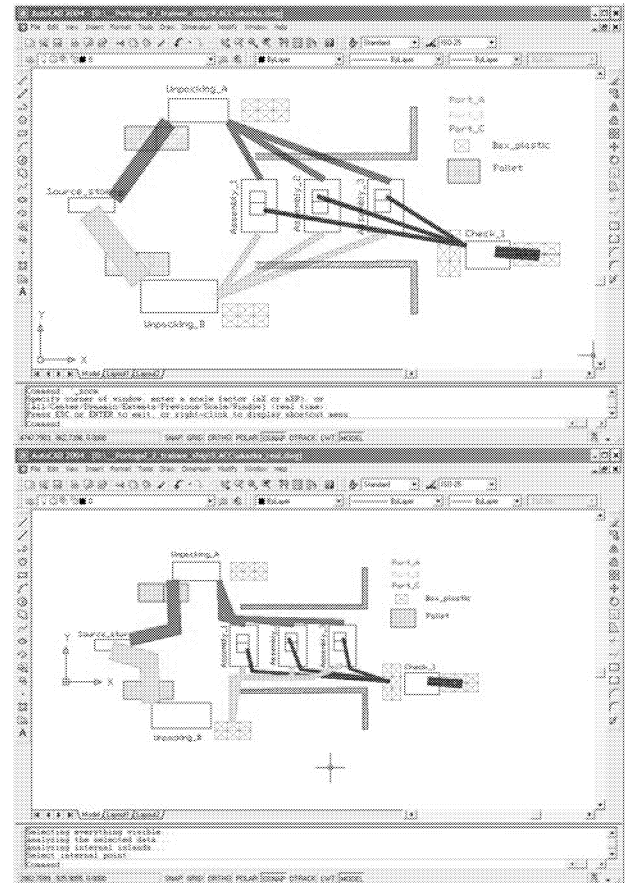


Figure 5 – Snapshots of layout in AutoCAD with material flows (ideal and real)

Flows can have various units (transported weight after a time interval or frequency of transport movement). Flows can be displayed according various parameters – flows of specific parts, flows of one transport mode or flows of crate type (e.g. pallets).

There are used data such as buffer sizes (necessary areas for storages) which helps to design production batches and logistic components like size of crates, size or shelves, etc.

Output data recorded into project database are coordinates of facilities, real material flows (coordinates of definition points) and also changes raised during layout design – e.g. changes of pallets number obtained by simulation (original task) is necessary to reduce because of lack of space (finds during layout design). In this case, there arises a question how this limitation influences designed production system.

### OTHER PROPERTIES OF INTEGRATION

Among the advantages from the discussed approach is the fact that it enables several ways of designing manufacturing systems, namely:

- Design of layout and subsequent verification through computer simulation

- Design of a simulation model – and subsequently, a plant layout
- Iterations with combinations of the two

Generally, for minimization of handling costs, the first option is the more suitable. In the event that the principal objective is processes optimisation (throughput, bottlenecks, batch sizes, usage of handling system etc.), the reverse sequence (computer simulation, layout design in CAD) represents the most effective strategy. (Aleisa 2005)

## CONCLUSION

The described solution integrates software tools used in production systems design into a unique system where all tools are connected by a database. The general aim of this work was to improve the process of production systems design, removing redundant activities and replacing “manual” work by automatic design phases. The work integrates static and dynamic layout optimization, increasing factory competitiveness while helping the design of flexible and robust production systems. Additional benefits include the intelligible illustration, e.g. rendering of material flows, 2D animation of processes within the simulation tool or the possibility to create virtual reality entertainment with Witness Virtual Reality Toolbox.

The key drawback of the approach is data-unification between companies and project databases, involving real constraints into layout definitions and some restriction as far as automatic generation of simulation models are concerned (models are generated just for solving specific and common tasks). Future work will be focused on using this integration for solving real production problems, improving and adding functions – such as solving routing problems, identifying the best routes for transport vehicles (e.g. milk-run systems) or creating a tool for designing storage systems.

## REFERENCES

- Aleisa E., Lin L. (2005) *For effective facilities planning: layout optimization then simulation, or vice versa?* Proceedings of the Winter Simulation Conference, p. 1381-1385
- Benjaafar, S., Heragu, S. S., Irani S.: (2002) *Next generation Factory layouts: Research challenges and recent progress*, Interface.
- Bureš M., Ulrych Z., Leeder E. (2007) *Digitální fabrika – softwarové produkty pro oblast dílčích fabriky*, In Mopp – Modelování a optimalizace podnikový procesů, Plzeň 8.-9. 2. 2007 ISBN 978-80-7043-535-9
- Dias, L.S.; G.B. Pereira and A .G. Rodrigues. (2007.) *A Shortlist of the Most Popular Discrete Simulation Tools*, Simulation News Europe, April 2007, vol 17, n° 1, 33-36. ISSN 0929-2268
- Francis R.L., McGinnis L.F., White J.A. (1992) *Facility Layout and Location: An Analytical Approach*, 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, USA.
- Havlík, R.; Manlig, F. (2004) *Vom 2D Layout bis zur virtuellen Realität. In: Priemyselne inženierstvo. Sborník příspěvků mezinárodní konference*, Dolný Smokovec 07.-08.10.04. Košice: TU v Košiciach, s. 149 – 154.
- Kulturel S. (2007) *Approaches to uncertainties in facility layout problems: Perspectives at the beginning of the 21st Century*, In Springer Science and Business Media, Springer Netherlands, ISSN 0956-5515
- Markt L., Mayer M. (1997) *Witness simulation software a flexible suite of simulation tools*, In Proceedings of the 1997 Winter Simulation Conference, s. 711-717.
- MECKLENBURG K. (2001) *Seamless integration of layout and simulation*, In Proceedings of the 2001 Winter Simulation Conference, s. 1487- 1491.
- Moorthy S. (1999) *Integrating the cad model with dynamic simulation: simulation data exchange*, Proceedings of the 1999 Winter Simulation Conference, s. 276-281
- Mujber T.S., Szecsi T., Hashmi M. (2005) *A new hybrid dynamic modelling approach for process planning*, Journal of Materials Processing Technology 167, 2005
- Muther R. (1973) *Systematické projektování (SLP)*, SNTL Praha.
- Taylor G. (2008) *Introduction to logistics engineering*, New York, Taylor & Francis Group, ISBN 978-1-4200-8851-9
- Vilarinho P., Guimaraes R. (2003) *A Facility Layout Design Support System*, Investigacao Operacional 23, p. 145–161, ISSN 0874-5161
- Vik P. (2008) *Integration of CAD systems and computer simulation and its usage in designing of manufacturing systems*, 3. mezinárodní konference Výrobní systémy dnes a zítra, 27.–28.11.2008, ISBN 978-80-7372-416-0
- Vik P. (2008) *Propojení nástrojů používaných pro navrhování výrobních systémů*. Proceedings of 4th annual International conference for Ph.D studentes and young researches, Zlín 10. Apr, 2008, ISBN 978-80-7318-663-0
- Wu Y., Appleton E. (2002) *The optimization of block layout and aisle structure by a genetic algorithm*, Computers and Industrial Engineering, v41, no 4, 371–387
- Zelinka A., Král M. (1995) *Projektování výrobních systémů*, Praha: ČVUT, 1995, ISBN 80-01-013302-2
- www.ugs.com; www.lanner.com; www.autodesk.com; www.delmia.com

## BIOGRAPHY

**GUILHERME PEREIRA** was born in 1961 in Porto, Portugal. He graduated in Industrial Engineering and Management in the University of Minho, Portugal. He holds an MSc degree in Operational Research and a PhD degree in Manufacturing and Mechanical Engineering from the University of Birmingham, UK. His main research interests are Operational Research and Simulation.

**LUÍS DIAS** was born in Vila Nova de Foz Coa, Portugal. He graduated in Computer Science and Systems Engineering in the University of Minho, Portugal. He holds a PhD degree in Production and Systems Engineering from the University of Minho, Portugal. His main research interests are Operational Research and Simulation.

**PAVEL VIK** was born in 1982 in Mladá Boleslav, Czech Republic. He studied at the Technical University of Liberec where he obtained his MSc degree in Manufacturing Systems, in 2005. He is making doctoral studies in

Manufacturing Systems and Processes and he is attending a ERASMUS-SOCRATES program at the University of Minho. His main research interests are 3D animation and virtual reality in design of production systems.

# OPTIMIZATION OF INDUSTRIAL ROBOT WORK CELLS USING THE AUTOMATED VARIANT SIMULATION

Prof. Dr.-Ing. Jürgen Roßmann  
Institute of Man-Machine-Interaction  
RWTH Aachen University  
Ahornstraße 55  
52074 Aachen, Germany  
rossmann@mimi.rwth-aachen.de

Dr.-Ing. Roland Wischnewski  
RIF e.V.  
Department Robot Technology  
Joseph-von-Fraunhofer-Straße 20  
44227 Dortmund, Germany  
roland.wischnewski@rt.rif-ev.de

Dipl.-Ing. Patrick Lenk  
FRIMO Technology GmbH  
System Automation  
Jacobsenweg 3-5  
22525 Hamburg, Germany  
lenk.p@frimo.com

## KEYWORDS

Industrial Engineering, Robotics, Optimization, 3-D-Models, Simulation.

## ABSTRACT

During the planning and optimization stages of industrial manufacturing work cells with robots, the engineer often faces an unmanageable amount of possible process alternatives. This is especially true for layout planning and robot path programming with regard to reachability, collision behavior and minimum cycle times. In this paper, we present the automated variant simulation as a means to support the engineer in his decision process. The automated variant simulation allows the definition of variable parameters of a process, varies these automatically, simulates the modified process and evaluates the simulation results to find optimum parameters and thus the “best” process variant. We show the general idea of the automated variant simulation and describe its integration into a 3-D simulation system. Finally, we present a case study of a successful application in industry.

## INTRODUCTION

Today, many industrial companies use means of the Virtual Engineering to cover the lifecycle of manufacturing plants. Many aspects of these plants are gathered in digital models enabling engineers to look at details without the need to access or even have the real plant. The typical procedure of the Virtual Engineering very roughly divides into the four sequential stages tender preparation, planning, optimization, and operation. Figure 1 shows these stages on the left side. Usually, companies seek to cover all stages with only one software tool to prevent problems with data formats and data conversions. Once created, the digital model of a real plant even allows detailed technical analysis in consideration of optimizations.

In the following, we will focus on manufacturing plants containing robots. During the planning of the work cell layout and the robot trajectories, the reachability of every single robot position and the absence of collisions for every path must be ensured. Moreover, a minimum cycle time for the whole process is requested. The problem is that this optimization task may have very many parameters and their combination may lead to an infinite number of process variants. Even today, the “felt optimum” is still picked by an experienced engineer who relies on his know-how. This approach may lead to good results, but there can be no prove that the true optimum is picked. The “felt optimum” even may turn out to be a wrong guess after the work cell has been set up in reality and the process does not work at all. The latter case usually produces high costs and demands a complete redesign. Our aim is to provide technically funded means to find the true optimum. This also includes an exact documentation of the original problem, the possibility to describe the process and a way to choose the variable parameters. In the following, we will present the automated variant simulation as a means to achieve these goals.

## THE AUTOMATED VARIANT SIMULATION

The automated variant simulation is based on ideas and methodologies first described in detail in (Lüdemann-Ravit 2005; Freund and Lüdemann-Ravit 2002). An industrial application has already been presented in (Roßmann et al. 2009). Unfortunately, most of these publications are in German language. For this reason, we will summarize some aspects of the mentioned publications in this paper. The main areas of application of the automated variant simulation are the optimization of the layout of manufacturing plants and the calculation of optimal robot trajectories for work piece handling and treatment. Generally, the methods presented are also suitable for any kind of optimization tasks encountered in manufacturing.

The simplified procedure of the automated variant simulation is shown in figure 1 on the right. At first, the possible variants of the process are defined. This is done by means of a program script. If the script is executed within the simulation system, the following steps will be executed automatically. For each variant, the 3-D model of the plant is changed according to the definitions made. This may include geometrical, electrical, and / or control aspects of the model – even complete robot programs in the manufacturer’s native programming language can be created. After changing the model, the real time simulation of the work cell is started. During the simulation run, output parameters of the simulation are gathered and can be used in a user defined way to evaluate the current variant. In most cases, the time used to fulfill a certain task is the most important aspect as it allows minimizing the overall cycle time of a process. The simulation of a current variant can be cancelled directly if it becomes clear that further progress will not produce any better results than the best variant obtained so far. This is especially true, if variants are aborted because a robot position cannot be reached at all or without collision. After all variants have been modeled, simulated, and rated automatically, the best variant remains and can be presented as a real time 3-D simulation to the engineer. By means of this visualization, the engineer does not only get some abstract output data but he can use his know-how to verify the result and thus build up confidence in the variant.

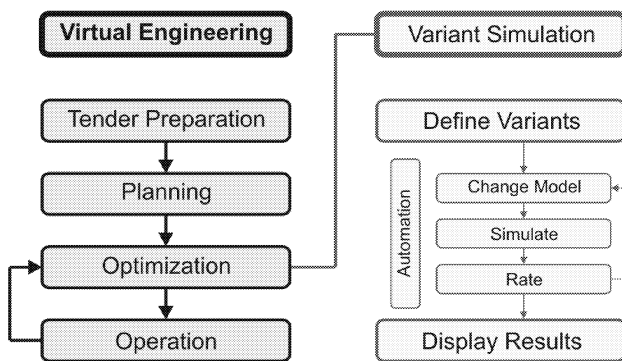


Figure 1: Integration and Workflow of the Automated Variant Simulation

## Requirements

The basic requirement to be able to use the automated variant simulation is the availability of 3-D digital models of all work cell components which may influence the sought-after optimum. Such models originate from 3-D CAD systems and must be enhanced with simulation functionality within a 3-D simulation system. For simple robot work cells it is sufficient to have 3-D models of the robot itself, the robot gripper, the robot tool (if applicable) and the work piece. Depending on the required accuracy of the results, the simulated behavior of these model components must match the behavior of the real components concerning kinematic or even dynamic aspects.

## Introducing Example

To give an example for further reference, figure 2 shows a screenshot of a 3-D model of a small flexible manufacturing work cell. The screenshot is taken from a 3-D simulation system described in (Freund et al. 2004). The robot used is a Mitsubishi RV-E4 with a Schunk gripper and tool change system. Carriers move on a Bosch MTS2 transport system and supply the robot with different work pieces. Front left of the robot, a polishing machine is mounted. The real work cell exists at the RWTH Aachen University.

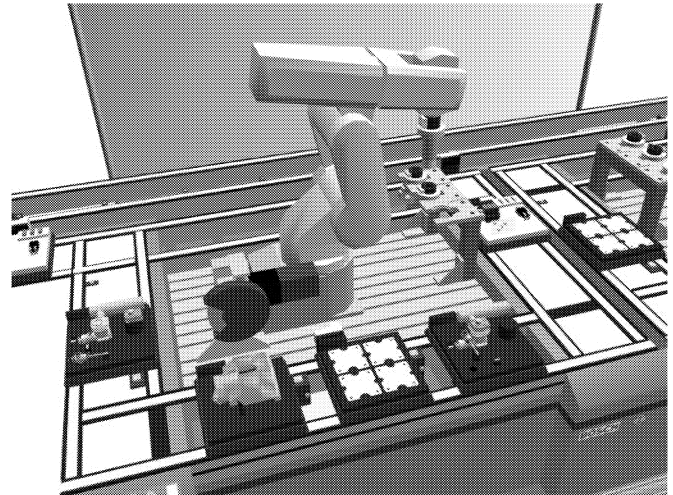


Figure 2: 3-D Model of a Flexible Manufacturing Work Cell

One task within the work cell is to polish the four butterfly-shaped work pieces. These are first gripped by the robot using a magnetic gripper and then moved smoothly around the operating point of the polishing machine. The whole outer shape of the work pieces is polished this way. To save time, this operation shall be completed without re-gripping the work piece. The operation requires more than 100 interpolation points and it would be a time consuming task to teach them to the robot manually. Moreover, robot and polishing machine must be placed relatively to each other in a way that the whole operation can be done in one move. Due to the kinematic restrictions of the robot, this placement is not trivial. Summarizing, we have two tasks to fulfill i.e. creating robot trajectories and optimizing the layout of the components. In the following we will focus on the solution of these two typical optimization tasks.

## Creating Robot Trajectories

For the automated creation of optimum robot trajectories, the program script language of the automated variant simulation offers geometrical functions to extract trajectory interpolation points on the surface of the 3-D model of a work piece. These points can be used to create simple robot programs to make the robot move along in linear or point-to-point interpolation mode. There are also mathematical

functions to manipulate the points e.g. set a required orientation of the robot's TCP (tool center point) in every point.

Figure 3 shows an example for the automated creation of a robot trajectory. The robot's tool shall move along the inner contour of a C-shaped work piece. The angle between the tool and the outer wall of the work piece shall be constant for all points. For the mathematical definition of this movement, geometrical functions of the script language are used. At first, a plane is defined which splits the outer wall of the work piece into two parts of identical height. After this, the intersection between this plane and the work piece is calculated. The result is a one-dimensional curve which already corresponds to the required robot path. Now equidistant interpolation points are defined on this curve. For each point, its orientation relative to the tangent to the curve is set to a constant angle. Finally, a MELFA BASIC IV program is created for the Mitsubishi RV-12SL robot to make it move along all the points. Figure 3 shows on the right an excerpt of the used script. Top left the work piece and the calculated points are displayed graphically to allow easy verification by the engineer. Bottom left the list of all positions is shown and right of this the automatically created robot program is visible.

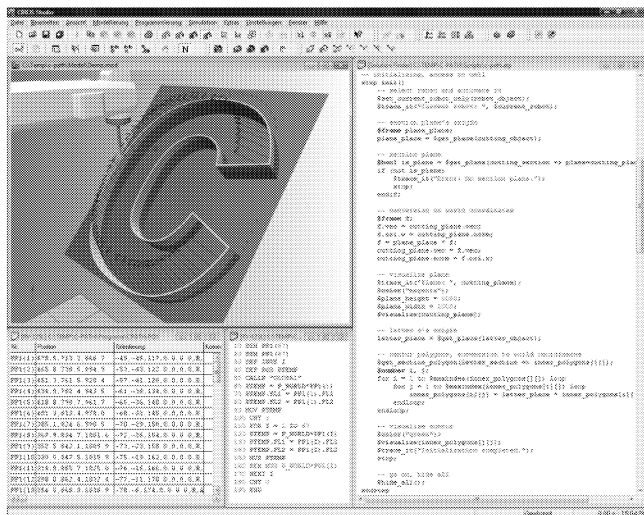


Figure 3: Automatic Generation of an Optimal Robot Path

The automated variant simulation further allows for comparison of different robot configurations (above / below, lefty / righty, flip / no-flip) or calculation of optimum approach paths.

## Layout Optimization

The automated variant simulation can also be used to find the optimum layout of a work cell by determining the best relative placement of all components. We will focus on varying the placement of a robot here. At first, fixed robot positions must be defined. In most cases, these are pick and place positions inside machines. After this, the required movement of the robot is specified using a robot program.

The possible variations concerning the robot's placement are then given as  $x_{\min}$ ,  $x_{\max}$ ,  $y_{\min}$ ,  $y_{\max}$ , and the step widths  $\Delta x$  and  $\Delta y$ . The z-component is typically not considered as a robot is usually fixed to the ground. The orientation is not varied in this example although this would be possible. The variants thus describe a grid of points onto which the robot can be placed. In every step, i.e. for every variant, the automated variant simulation repositions the robot, recalculates the fixed positions relative to the new robot placement, creates the new movement trajectories, and simulates the whole process. The aim of this optimization run is specified to find the variant with the shortest execution time in which the robot reaches all positions and all trajectories are free of collisions.

During the optimization run, the automated modifications to the model and the current state of the simulation can be visualized to the engineer. Figures 4 and 5 show two screenshots taken during the execution of an optimization script. The model shows two machines in low detail – one on the left and one on the right. The robot has to move into the left machine, pick up a part with its gripper, and move on to the right machine to put the part there. The robot's placement was programmed to be allowed anywhere between the two machines while the variation of the placement was defined by a grid width of 100 mm. In figure 4, the robot is set to one fixed placement. The spheres visualize the positions the robot has to move to: a green (light grey) sphere indicates a reachable position, a red (dark grey) sphere an unreachable position. The first position in front of the robot can be reached but already the second position inside the left machine cannot be reached. After this, the simulation of this variant is cancelled. Looking at the visualization, the engineer can directly see why certain process variants are not possible.

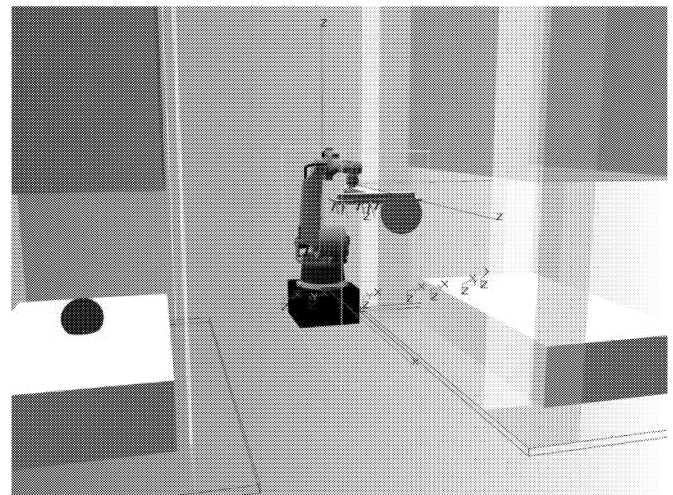


Figure 4: Invalid Robot Placement:  
The Position in the Left Machine cannot be reached

Figure 5 shows another robot placement. In this case, all necessary positions are reachable and thus visualized by green spheres. Moreover, the resulting path along the positions is free of any collisions. If this applies to more



than one variant the script will choose the variant with the shortest execution time.

The use of the automated variant simulation for layout optimization is also subject of the industrial case study presented at the end of this paper.

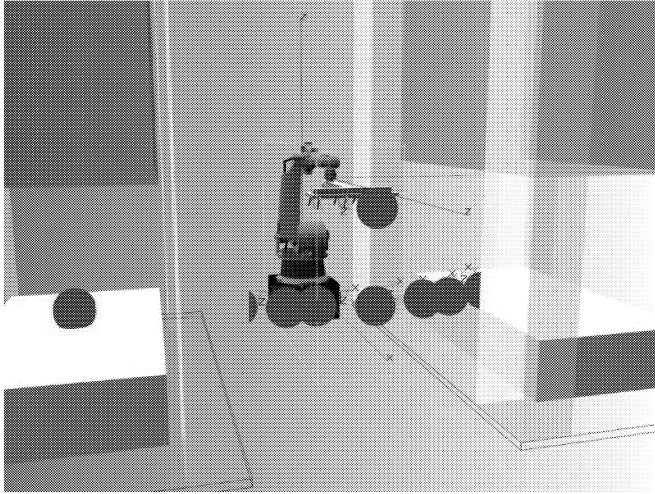


Figure 5: Valid Robot Placement:  
All Positions can be reached without Collision

### Further Applications

Besides the creation of robot trajectories and the optimization of work cell layouts, the script language also allows for handling nearly every kind of optimization task. The script language offers access to almost every mechanical and electrical detail of the model. Even simulation parameters like the current simulation time or visualization details can be read and written. During the course of the script, runtime data can be shown to or be requested from the engineer using messages and input forms.

### PRODUCT INTEGRATION

It is crucial for the practical applicability of the automated variant simulation that is integrated into a comprehensive 3-D simulation system. Such a system must provide functionality to map all relevant components of the considered manufacturing work cell. All aspects concerning mechanical, electrical, and control functions must be considered if they influence the choice of the optimum variant. The system used offers model libraries containing mechanical cylinders, axes, motors, and complete transport systems. Electrical inputs and outputs can be defined and used to connect controllers with components – just like in reality. This means that components are really controlled by I/Os and not by simulation system specific methods. For every component, and thus for the whole model of the work cell, a reality compliant I/O interface can be defined. By this means it becomes possible to use real controller programs within the simulation system. Control can then be executed using original robot programs, original PLC

programs (Siemens STEP7 for S7) or sequential flow charts as specified in IEC 61131-3. Even external controllers (via OPC) or the PLC simulator Siemens PLCSIM (via COM) can be coupled to the system.

The simulation system can be used to cover all stages of the Virtual Production from the early planning to the final commissioning (Roßmann et al. 2007). However, especially SMEs (small and medium sized enterprises) usually resign to establish all stages of the Virtual Production because they are afraid of high effort and thus high costs. In this case, the simulation system can also be used to handle isolated problems and rare special cases. The simulation system used is the commercially available CIROS Studio ([www.ciros-engineering.com](http://www.ciros-engineering.com)). The automated variant simulation can be executed using an extension module called “Solution Finder”.

### RESTRICTIONS

In this chapter we discuss restrictions of the approach i.e. the accuracy of the simulation results and their applicability to real robot work cells. The script based creation of model variants and their evaluation works deterministically and is thus free of any faults. Also the kinematic models of the robots are exact. They are created using the manufacturer’s original CAD data and data sheets. Restrictions come up concerning the simulation of robot movement commands. In reality, robot trajectories are interpolated by the real robot controller of a certain robot manufacturer. Unfortunately, the manufacturers do not reveal the algorithms used for interpolation. In the past, there has been some effort by German automotive, robot and simulation enterprises to specify a common interface for VRCs (virtual robot controllers). The projects Realistic Robot Simulation (RRS I and RRS II) were launched but did not result in freely available VRCs with an open interface. Thus, also the simulation system used in our approach cannot use original algorithms. Instead, we use a general robot controller which was first described in (Keibel 2003). Unfortunately, this results in differences between real and simulated trajectories concerning the time of movement and, especially with point-to-point control, also the exact course of the path. The magnitudes of the differences cannot be declared generally but they may depend on the scenario. Future work will give a detailed analysis of the magnitude of the error for classes of movements and some original controllers.

To minimize or even eliminate the error between simulation and reality, it is possible to couple real or virtual external robot controllers of some manufacturers to the simulation system. The engineer should keep in mind that there may be differences concerning program execution times even between the real and the virtual robot controller of one single manufacturer.

### CASE STUDY

We will now describe a case study to prove the practical industrial relevance of the automated variant simulation



which is integrated in a comprehensive simulation system. The tool is used at FRIMO Technology GmbH which is a supplier for all well-known automotive manufacturers. The division FRIMO System Automation develops concepts for the automation of manufacturing plants and puts these concepts into effect. Main areas of application are press chains and the loading of laminator machines. The automated variant simulation is applied in the construction department to optimize robot work cell layouts towards a minimum cycle time. The reachability of all robot positions is ensured by simulation. The collision detection, which is also provided with the simulation system, is not even used. Instead, a simple visual inspection of the high quality 3-D simulation visualization ensures the absence of collisions. The handling tasks executed by the robots do not require special commands of the respective native robot programming language. Because of this, programming is done using the general Industrial Robot Language (IRL) which is specified in German standard DIN 66312. The advantage of using IRL is its applicability to robots of different manufacturers. Actually, robots by manufacturers ABB and FANUC are employed at FRIMO System Automation. Another benefit is the fact that basically only one optimization script is used which has to be adapted only slightly for every new use case. Fixed parameters for all variants are the pick and place positions of the work piece. Variable parameters are x- and y-placement of the robot relative to a fixed location within the work cell. Orientation is not considered. The placement is varied within the whole possible range with a step width of 50 to 100 mm.

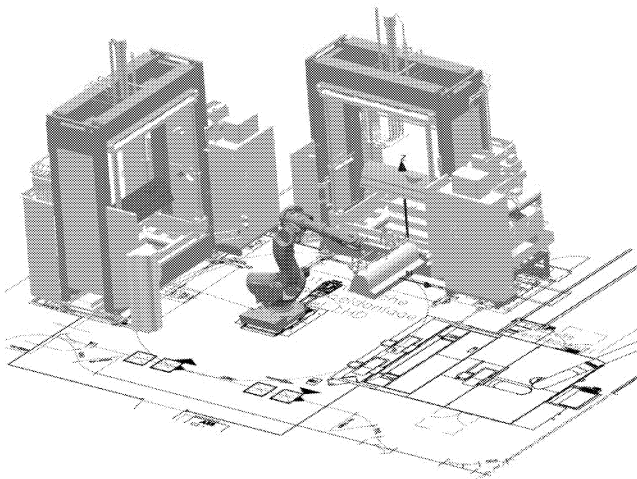


Figure 6: Example of a Layout Optimization using the Automated Variant Simulation

Depending on the number of variants and the complexity of the work cell, the execution of the real time simulation of all variants on a notebook with Intel Core 2 Quad processor may take some hours. The result of the simulation is a robot placement leading to the minimum cycle time. This placement is used for the real work cell. For example, figure 6 shows the 3-D model of a manufacturing plant with

a robot ABB IRB 6600-175/2.80 for press loading. A 2-D layout of the work cell was imported into the model and attached to the floor for easy placement of the stationary components. 3-D CAD data of the work piece and the presses existed in IGES format and were imported, too. The shape of the work piece in figure 6 is falsified for reasons of non-disclosure. An optimization script was created defining the fixed positions of the work piece inside the presses. Over all variants, the robot was placed at all free places between the presses with a grid width of 100 mm in both directions x and y. Simulation also prevented a robot movement close to a robot's singularity which would have resulted in a slower movement. The results of the simulation were directly used for the real work cell layout.

## CONCLUSION

The automated variant simulation can be used to evaluate different process alternatives in manufacturing work cells with robots. Main application areas are the creation of robot trajectories and the optimization of work cell layouts. The demands in this context are the reachability of all robot positions, the absence of collisions for all robot paths and a minimum cycle time for the whole process. After defining the range for the parameters to vary, the variants are created automatically by modifying the 3-D simulation model of the work cell. Each variant is simulated and evaluated – the optimum variant remains. A case study of a successful application in industry proves the practical relevance of the methods described.

## REFERENCES

- Freund, E. and B. Lüdemann-Ravit. 2002. "A System to Automate the Generation of Program Variants for Industrial Robot Applications". In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2002)*. Lausanne, Switzerland.
- Freund, E., R. Feist, D. Pensky, and R. Wischniewski. 2004. "3-D Graphical Simulation of Complete Manufacturing Systems in Real-Time." In *Proceedings of the 6th IASTED Conference on Control and Applications (CA 2004)*. Marina del Rey, CA, USA (March).
- Keibel, A. 2003. *Konzeption und Realisierung eines integrierten Moduls zur Simulation und Steuerung von Kinematiksystemen*. Ph.D. Thesis, University of Dortmund, Germany. In German.
- Lüdemann-Ravit, B. 2005. *Ein System zur Automatisierung der Planung und Programmierung von industriellen Roboterapplikationen*. Ph.D. Thesis, University of Dortmund, Germany. In German.
- Roßmann, J., O. Stern, and R. Wischniewski. 2007. "Eine Systematik mit einem darauf abgestimmten Softwarewerkzeug zur durchgängigen Virtuellen Inbetriebnahme von Fertigungsanlagen: Von der Planung über die Simulation zum Betrieb." *atp* 49, Issue 7, p. 52-56. In German.
- Roßmann, J., R. Wischniewski, and P. Lenk. 2009. "Optimierung industrieller Roboterarbeitszellen durch automatisierte Variantensimulation." In *Tagungsband ASIM Workshop STS/GMMS 2009 und DASS*, G. Elst (Ed.). Dresden, Germany (March). ISBN 978-3-8167-7981-0. In German.



# **PRODUCTION AND TRANSACTIONS LOGISTICS**



# UNDERSTANDING SUPPLY CHAIN PERFORMANCE BY SIMULATION

Alejandra Gomez-Padilla  
Department of Industrial Engineering, Center of Exact Sciences and Engineering  
University of Guadalajara  
Av. Revolucion No. 1500, C.P. 44430  
Guadalajara, Jalisco, Mexico  
E-mail: alejandra.gomez@cucei.udg.mx

## KEYWORDS

Industrial engineering, supply chain, decision support.

## ABSTRACT

Companies need to coordinate a wide variety of decisions that will impact their performance as well as the performance of the whole supply chain they belong. There are several simulators of supply chains available on the market; they are classified and analyzed. Then the characteristics of the simulator being developed are presented. The simulation described in this paper intends to support the decision making process by allowing to consider different situations. The performance of a supply chain strongly depends on the coordination of the decisions of the companies. These decisions are expressed by the existing flows. Among the companies of a supply chain there exist three types of flow: information, physical products and economic. These three types of flow are considered in the simulator.

## INTRODUCTION

As it was previously stated, three types of flow exist among companies of a supply chain: information, physical products and economic. Information flow may be symmetrical or asymmetrical. The first type of information flow is when all the companies in the supply chain have the same information. The second type is when at least one of the companies has more information than the others and will not share it (Gallego and Ozer 2001; Gomez-Padilla et al. 2005; Chen 2001). (Lee et al. 1997) identify information share as an important mechanism for coordination, and (Li et al. 2001) also analyze the quality of the shared information. The flow of physical products depends on how each company is organized, and is related to their production, inventory and transportation policies. (Cachon and Fisher 2000) analyze different inventory policies under different situations. (Gupta and Weerawat 2006) analyze a capacitated supply chain and (Chen et al. 2001) focus on physical flow and distribution systems. The economic flow is a result of the

prices and terms of the contract. When companies get engaged to furnish and order quantities and to pay financial amounts implies decision making. These decisions will normally be expressed and established in the contract. The decisions are made in order to achieve certain goals fixed by each company. Important decisions established in contracts are prices and volumes. The prices are fixed depending on the type of contract, and the contract determines the financial flow or transfer between the companies. Volumes normally depend on demand. (Simatupang and Shridharan 2001) relates information share to a clear contract, as well as costs reduction. (Cachon 2005) analyses five contracts for a new vendor situation. (Gomez-Padilla 2005) analyses the optimal contractual decision under a base stock inventory policy.

The beer game is a classical supply chain management game largely used in universities to teach the concepts of supply chains (Mosekilde et al. 1991). The beer game considers a supply chain formed by four companies: manufacturer, distributor, wholesaler and retailer. Each player represents one of the companies. The decision to make in this game is how many products to order to satisfy his demand. This game shows the bullwhip effect of lack of coordination in a supply chain. (Chen et al. 2000) quantifies this effect. The simulator proposed in this paper also considers four companies and each company –or each player- has to decide the amounts to be ordered. The difference from the beer game, and the originality of this simulator, is that explicitly considers the three flows that allow a supply chain to coordinate. This document explains the architecture and operation of a simulation application that considers these flows. In the next section a brief survey of existing simulators is presented, next the specifications of the simulation application are explained and finally we conclude.

## EXISTING SIMULATORS

Several simulators for supply chain exist in the market. First we present some simulators available to use them online, and then some simulators that need to be installed in a computer. Next, we present some simulators for specific

sectors. Finally, two simulators that don't have a computer-based interface are presented.

There are versions of the beer game available online for free and operate under a virtual environment. MIT offers three: Beer Distribution Game (traditional beer game), the Procurement Game (focusing mainly in buying) and the Near Beer Game. This last one considers a symmetric information situation for the beer game. A Wood Supply Game exists that represents the beer game problem transposed to a wood supply situation. HULIA centers in information and physical flow simulation for e-commerce. IESS is an internet-based simulator that considers information and physical products exchange. There are other simulators for which it is necessary to pay a user license; they usually focus on a specific industry and they may be differentiated by the elements they consider. Among these simulators we may mention: CVFISSC (for electronic sector; oriented to information and physical flow), TAC SCM (for electronic sector, centered on information flow and contracts), The Distribution Game (considers physical product flow: organizes deliveries to satisfy demand), ILMG (marketing and delivery of a product), SCMSIM (physical flow: lot size and transportation), EPSIM (economic flow). For automobile industry we found: Virtual Factory Laboratory (centered in information exchange and contract negotiation) and Net Chain (marketing and production, contract negotiation). For agro alimentary supply chains, we found LOGSME (intends to test companies integration). There are also simulators that don't have a computer interface. Two of them are: Trust and Trading (asymmetric information situation to analyze trust and opportunistic behavior) and the Strawberry Chain (highlight cultural differences).

The simulator that is being designed will have a computer interface. It is different from previous ones in that it may be played from one player and up to four. The simulator will not be for a specific sector, but will be just for one product. It is intended to simultaneously consider information, physical products and economic exchanges. It will help to explain the impact of decisions of companies in a supply chain. The player will have to decide if he shares his information, and which information to share. He will have to decide how much to ship and how much to order. The architecture of the simulator will be module-based and will operate under a virtual environment. The input and output data of every simulation are to be stocked in a database.

## SIMULATION APPLICATION

The application is being developed using free available software (Visual Studio 2008). The supply chain simulated has four stages or four companies, as in the beer game. They can play up to 4 players, who may play among them, but there will also be the option to play alone. The sequence of

events is as follows: when the player starts, he has to decide first if he will play alone or against other players. Next, he has to decide which company from the chain he wants to represent. If he decides to play alone, the simulator will play for the other companies. If it is just one player, he may start. If there are several players, once they are all set, they may start to play. Each player chooses his playing conditions regarding information, product and economic flow. Concerning the information flow, each player has to decide if he wants to share information or not. The physical flow of products is defined by the inventory policy of the company and the production policy of the first company of the chain. The first version of the simulator considers a base stock policy for each company, which is based on their costs and the production policy of the first company depends on the lot size, which is also function of his own costs. The economic flow depends of the type of contract. Each player may negotiate a specific contract among six: wholesale price, buy back, quantity flexibility, revenue sharing, capacity reservation and option contract. These contracts are defined as in (Gomez-Padilla 2005). The different prices are fixed according to the type of contract and are handled with an optimization tool. The players will also have the possibility to either fix the demand pattern from the final market to be simulated, or to use one of the five proposed by the simulator: constant high demand, constant low demand, alternating high – low demand, uniformly distributed demand and normally distributed demand. If there are playing several players, it will be the retailer who decides on the demand pattern and the number of periods to simulate. Each member will have his own information available, as shown in figure 1 for the retailer. Each player will be able to see the other player information depending on the initial information setting.

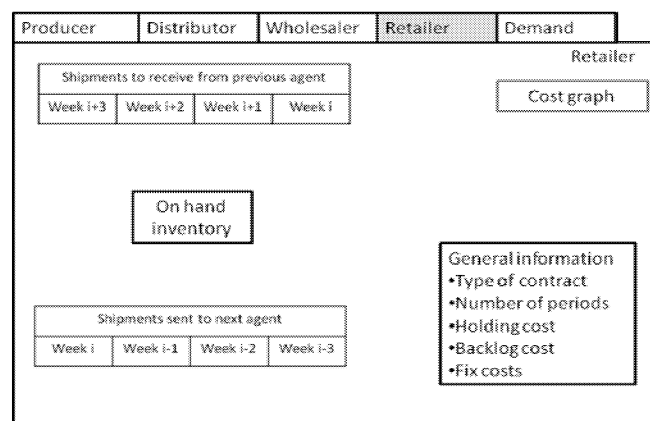


Figure 1 Interface to handle orders

All the members of the supply chain hold inventory and have a lead time that can go to up to 4 weeks. The decisions to be taken depend on the element of the supply chain represented. The retailer has lost sales and the wholesaler,

distributor and producer have back logged sales, this is, the demand that the retailer is not satisfying will not be satisfied the next period, but the demand that the wholesaler, distributor and producer don't satisfy, will be satisfied the next period. The retailer has opportunity cost and the wholesaler, distributor and producer have backorder costs. Everyone has a holding cost, handling costs and fix costs. The information on the costs and benefits are available graphically, and again, according to the information specifications (symmetric or asymmetric) this graphs will be available or not for the rest of the companies in the supply chain. This last cost and benefit graphs are very important, since they are going to show the evolution of the expenses and gains. This expenses and gains are a direct consequence of the decisions taken from the beginning of the game.

## CONCLUSION

It was explained that companies in a supply chain have to coordinate by their decisions executed by three types of flow: information, products and economic. These three types of flow were introduced. Some existing simulators were presented and from the analysis of those simulators. The innovation of this simulator is that considers the three flows for coordination. It will be computer based and the three types of flow that coordinate a supply chain are considered. The information flow may be symmetrical or asymmetrical. The information that may be shared is related to the physical flow and to the economic flow. The information related to the physical flow is: demand, shipments, passed commands and on hand inventory. The information related to the economic flow is: type of contract, parameters of the contract, costs, sales and profit. The information that will always have to be given are the orders and shipments. The physical flow will depend on: production policy for the producer and on inventory holding policy for the distributor, wholesaler and retailer. It will also depend on the information available concerning demand, on hand inventory and shipments. The economic flow will depend on the contract and its parameters, the different costs and on the ordered and shipped products.

The application being developed is for a multiple horizon period, for one product facing a stochastic demand. The player(s) will decide: the type of information flow (symmetric or asymmetric), the type of stochastic demand and the type of contract. This project will continue with the formulation of the simulation plan to validate the simulator and the analysis of preliminary results.

## REFERENCES

Cachon, G. P., 2005, Supply Chain Coordination with Contracts, in: De Kok, A.G., Grave, S. C. (eds.): (2005) Handbooks in Operations Research and Management Science, 11: Supply Chain Management: Design, Coordination and Operation, Elsevier, Amsterdam: 229-340.

- Cachon G. and Fisher M., 2000, Supply Chain Inventory Management and the Value of Shared Information, *Management Science*, 46(8), pages 1032-104.
- Chen F. et al. Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information. *Management Science*, 46(3):436-443, 2000.
- Chen, F., Federgruen A. and Zheng Y., 2001, Coordination mechanisms for decentralized distribution systems with one supplier and multiple retailers. *Management Science*, 47(5). 693-708.
- Chen., F. Information and Supply Chain Coordination. Decision, Risk & Operations. Working Paper Series. Columbia Business School, 2002.
- Lee, H., So, K. C. and Tang, C., 2000, The value of information sharing in a two-level supply chain, *Management Science*, 46. 626-43.
- Li, J., Shawn, M. J, Sikora, M. T., Tan, G.W. and Yang, R., 2001, The effects of information sharing strategies on supply chain performance", College of Commerce and Business Administration, University of Illinois at Urbana-Champaign, URL: [http://citebm.cba.uiuc.edu/B2Bresearch/ieec\\_em.pdf](http://citebm.cba.uiuc.edu/B2Bresearch/ieec_em.pdf) (30.9.2002), 34 p
- Gallego, G. and O. Ozer. 2001, Integrating replenishment decisions with advance demand information, *Management Science*, 47 (10), 1344-1360.
- Gomez\_Padilla, A., 2005, Modelisation des relations verticales: une approche economique et logistique (Modelling vertical relations: an economic and logistic approach); Ph.D. Thesis defended at the Institut National Polytechnique de Grenoble, France.
- Gomez-Padilla, A., Duvallet, J. and Llerena, D., 2005, Contract Typology as a Research Method in Supply Chain Management, dans: H. Kotzab, S. Seuring, M. Müller, G. Reiner, (eds.): Research Methodologies in Supply Chain Management, Physica, Heidelberg, p. 525-538.
- Gupta, D. and Weerawat, W., 2006, Supplier-manufacturer coordination in capacitated two-stage supply chains, *European Journal of Operational Research*, 175 (1), pp. 67-89
- Mosekilde, E., Larsen, E., and Sterman, J. D., 1991, Coping with complexity: deterministic chaos in human decision making behavior, Beyond Belief: randomness, prediction and explanation in science. Editors J. L. Casti, and A. Karlqvist CRC Press 1991.
- Simatupang, T.M. and Sridharan, R., 2001, A Characterisation of Information Sharing in Supply Chains, *Proceedings of the Twenty Naught One ORNZ Conference*, University of Canterbury, Christchurch, New Zealand, 30 November - 1 December 2001.

## BIOGRAPHIE

**ALEJANDRA GOMEZ-PADILLA** studied industrial engineering at Instituto Tecnológico y de Estudios Superiores de Occidente (ITESO) and obtained a Master's degree from the Ecole Polytechnique de Montreal and a PhD from the Institut National Polytechnique de Grenoble (France). She was product and process engineer for a company in the electronic field and since 2006 she is associate professor – researcher at University of Guadalajara.

# SIMULATION OF A MILK RUN MATERIAL TRANSPORTATION SYSTEM IN THE SEMICONDUCTORS INDUSTRY

Ricardo Raposo

University of Porto  
School of Engineering  
MITPortugal program EDAM Area  
Rua Dr. Roberto Frias  
4200-465 Porto  
Portugal  
ricardo.raposo@itarion.com

Guilherme Pereira

Luis Dias  
University of Minho  
School of Engineering  
Production and Systems Dpt  
Campus de Gualtar  
4710 057 Braga  
Portugal  
gui/lsd@dps.uminho.pt

## KEYWORDS

Engineering Systems, Internal Logistics, Industrial Engineering, Manufacturing, Discrete Event Simulation

## ABSTRACT

This paper deals with a project that consisted in the implementation of a Milk Run Lot Transportation System in Qimonda Porto Test Area, done by a multidisciplinary team formed by Qimonda Porto's workers, and the development of the corresponding simulation model. The first part of the study concerns an industrial engineering assessment of the test area, which identified sources of waste and improvement possibilities, and the implementation process of a Milk Run system in this area. Secondly, the results of the system implementation are discussed, and the construction of a simulation model in Arena® is presented. The purpose of the simulation exercise is to test different system configurations that may allow the improvement of the real-world system. Finally, some information about the simulation results and further steps to be taken regarding the improvement of the system is presented. The target of the project, framed in a Lean approach, was to reduce waste, namely transportation waste, thus optimizing the utilization of the test area human resources.

## INTRODUCTION

This paper reports the implementation of a Milk Run Lot Transportation System in Qimonda Porto (onwards referred to as QPT) Test Area, done by a multidisciplinary team formed by QPT's workers, and the development of the corresponding simulation model (Raposo 2009). Qimonda is a global memory supplier with a diversified DRAM product portfolio. At the time of the project, Qimonda had approximately 13,500 employees worldwide, accessed five 300mm manufacturing sites on three continents and operated six major R&D facilities. In Portugal, Qimonda has its major European backend production site, founded in 1996, counting with a workforce of approximately 2,000 employees at the time of the project. QPT assembles, tests and packs semiconductor products, namely DRAM memories for computers, servers and other digital applications (MP3, mobiles, digital cameras, game consoles and others), in a

plant that has a clean room area of 15,500 m<sup>2</sup>. Qimonda is a company working in an industry which is commonly designated as capital intensive (opposite to labor intensive industries), the one of semiconductors. A business process or an industry is considered capital intensive when there is a high ratio of the necessary capital to the amount of labor that is required. The process flow in a semiconductor backend factory, like QPT, can roughly be described by the following steps, having wafers as the major input, and marked chips as its output.

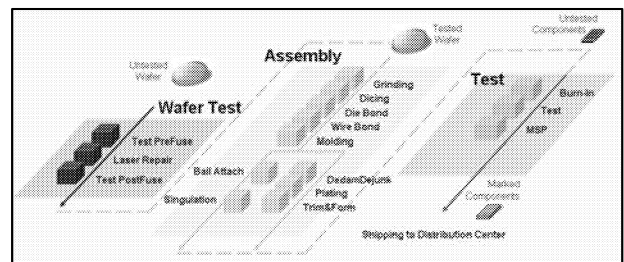


Figure 1: Qimonda Porto Backend Process Flow



Figure 2: Qimonda Porto Major Input and Output

The process area named as Test in Figure 1 actually encompasses 3 different processes:

- Burn-in: consists in an accelerated aging of the components, through thermal and electrical stress, to eliminate the components that would fail in the first years of life;
- Test: extensive electrical tests to ensure the components electrical and functional specifications;
- MSP (Mark, Scan & Pack): scanning (to ensure physical dimensions compliance), marking and packing (for expedition) the components;

This project concerns the test area of QPT. At QPT, mainly in the Test Area, the highest capital investment is associated with the equipment and the optimization of its



utilization is a very important factor, essential to keep the cost per piece below the desired limits. Facing the fact that the test area is the one that represents the largest investment at QPT, QPT's line is balanced in order that the test area may be the bottleneck of the factory.

## PROBLEM DEFINITION

Within the Lean Manufacturing approach (Womack and Jones 1996), a “waste focused” assessment of the test area was performed in 2006. The purpose of this assessment, focused on the OEE (Overall Equipment Efficiency) metric developed by SEMI (Semiconductor Equipment and Materials International), was to identify waste sources and trigger the implementation of actions to minimize these. The assessment followed a Multi Observation Study (MOS) methodology. One of the most significant inefficiencies detected in the study concerned the time spent in transportation activities by the operators, which was at the time of the study done in an ad-hoc way by all the operators. The identified action to address this waste was the implementation of an organized lot distribution system. This paper concerns the implementation of this project, whose primary purpose was to reduce the time spent in transportation activities by the operators. However, by dedicating operators to transportation activities, it was also expected that operators working with the equipment would be more focused on their specific tasks, as they would have to leave the line less frequently. From a value stream mapping perspective, the value stream consisting of all the actions (both value added and non-value added) required to transform the raw materials into final products (Rother and Shook 1999), an important metric is the time that the materials spend in non-value added activities, of which transportation is an example. This project is not expected to impact significantly the time where the “material is effectively transported”, but the time that “people spend transporting materials”. The project objectives are then stated in the following table.

|                            |   |
|----------------------------|---|
| <b>Objective Statement</b> | Reduce the time spent in transportation activities by 30%               |
| <b>Primary Metric</b>      | Time spent in transportation activities by the operators                |
| <b>Secondary Metric</b>    | Equipment Downtime states related with the operator absence, cycle time |

Table 1: Project Objectives

For this purpose, the following section will focus on the description of the implemented system and the results of its implementation. Afterwards, a section concerning the development of the Arena<sup>®</sup> Simulation Model is presented. This work will then discuss some results and state some conclusions, including some ideas for future work on this applied industrial field.

## SYSTEM IMPLEMENTATION AND RESULTS

As the lean approach birth is commonly associated with Toyota Production System (Ohno 1988), one of the approaches to minimize the waste in transportation is

framed on this approach, and consists in the creation of a new role, the *mizusumashi*, or water spider (Nomura and Takakuwa 2005). The key role of the worker with this function is to support the production activities, by performing mostly non-value added activities, so that other operators can be entirely focused on value added activities.

Optimization of the water spider transportation activities may be achieved through the definition of a programmed route for material transportation, commonly known as a milk run (Hugos 2003), in an analogy to the milk delivery systems, which would follow a well-defined route, delivering full bottles and picking up empty bottles. This is an alternative to systems based on an on-demand activity in which the transporters individually transport material for each workstation, always returning to the “purchased-parts market” (Smalley 2004; Harris and Harris 2007).

A milk run system framed in a lean manufacturing environment usually relies on *Kanban* (Smalley 2004). Among other functions, these information systems are a means of controlling the quantities of materials being transported, and the frequency with which these are transported. Kanban cards are usually physical cards, typically used to signal when a downstream process requires more material to process, working as a replenishment order to the upstream process.

Opposite to traditional “push” scheduling systems, kanban methods link and synchronize the production processes, starting as far as desired in the value chain. Although a fundamental characteristic of the concept is its simplicity, the concept has evolved over time, and although the word itself means card, today other formats exist for the same function, as “electronic” Kanban systems (Smalley 2004).

## Description of the test area organization and specificities

Test area characteristics, relevant in the definition of the system, are (Figures 3 and 4):

- For lot tracking purposes, and to ensure the correct sequence of process steps, a MES (Manufacturing Execution System) is used;
- The area does not work in FIFO (First In First Out) mode, mostly due to the high product mix, allowing to minimize the number of toolkit conversions, and to maximize the equipment productive time;
- More adequate lots to process at each moment are individually selected by the senior line operators, which imposes the need of a short period between the lot request and the lot delivery;
- Lots may undergo a different number of process steps, according to the product, as described in Figure 3;
- Between processing steps, lots can be either stored in the logistical center, at the market, or transported to another test cell directly;
- Due to the dimension of the corridors between test cells, it is not possible to circulate with large trolleys between the test cells, just in the transportation corridors;

- Different test cells and production lines may have very different throughputs;
- Lots flow from the logistical center of the area to the test cells, and vice-versa, between test cells and from the test cells to the Test Gate;
- The Test Area “purchased parts market”, is based on a vertical rotating-shelf storage solution;

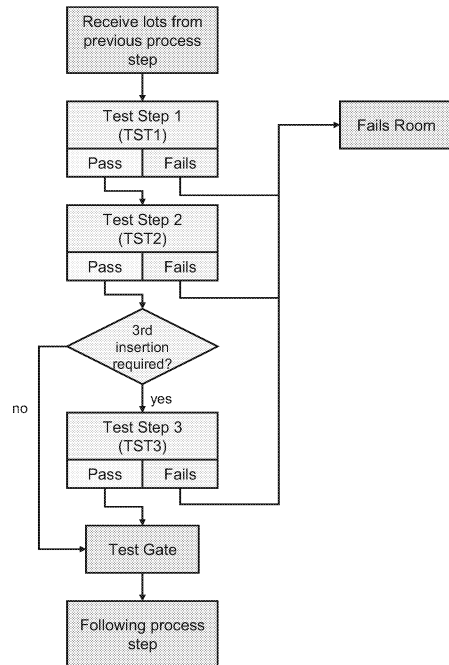


Figure 3: Test Area Simplified Process Flow

- Lots are formed by a group of packs, as presented in Figure 4, with each pack being formed by a group of trays, each tray having a capacity of 90 to 140 chips;

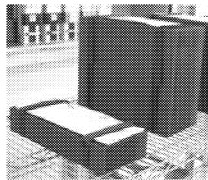


Figure 4: Packs of Trays with Straps

### Implemented system and results

To dimension the milk run system, information was collected about the type and volume of materials that had to be transported (Figure 5

Figure 5), the limitations to the transportation of materials, the standard work elements duration and the possible configurations for the milk run route, among others.

Considering the characteristics described in the previous section, and the information collected along the project, the system was dimensioned in the following way:

- A decoupled route format was selected, in which the transportation and market management tasks are performed by different resources (Harris et al. 2003);
- Dedicated human resources: one market attendant and 2 delivery route operators (onwards referred as DRO);

- Route frequency: each DRO should start a new route every 12 minutes, and DROs should be offset by 6 minutes, for a DRO to pass in each Point of Drop (onwards referred as PoD) rack every 6 minutes;
- Transport trolley capacity: the selected transportation mode were carts with a capacity for 10 packs, which the operators will transport on foot;
- Point of Drop rack capacity: a strategy based on Point of Drop (instead of Point of Delivery) racks was defined due to space limitations, with each PoD rack serving a group of test cells (the dimension of the PoD rack was not considered a critical factor and the selected PoD racks have a capacity for 15 packs);
- Lot selection method: operators request the lots via an application developed by the project team, being a physical kanban used to signal the lot destination;

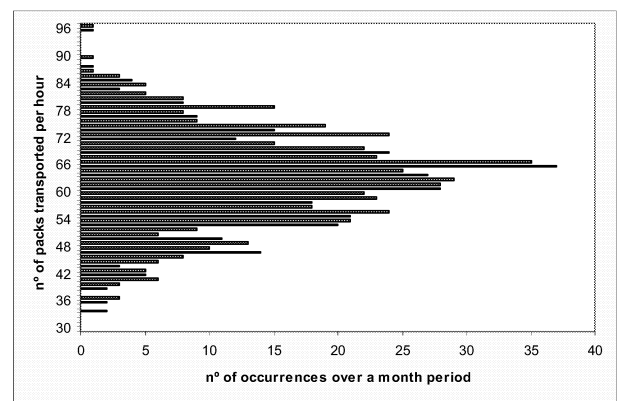


Figure 5: Distribution of Number of Packs to Transport per Hour

After the system was implemented and running a new MOS was performed, allowing to determine the results of the system implementation. Against the expectations, the overall time spent in transportation activities did not diminish, and remained approximately the same. This may indicate that the system is not optimized or is under-loaded. These hypotheses will be addressed in the simulation section.

Although not optimized, the system has brought advantages to the operators working in the equipments, focusing equipment operators on specific tasks on their work cells. The machine states that express the absence of an operator to assist the equipment improved and are presented in the following table.

| Equipment State   | Before | After |
|---|--------|-------|
| Machine Stocker is empty and WIP is available to load (test cell is idle) | 1.8%   | 0.3%  |
| Lot is finished, waiting for unloading, new lot available                 | 0.9%   | 0.4%  |
| Down, waiting for operator to assist the test cell                        | 1.7%   | 0.2%  |

Table 2: Comparison of Equipment States Affected by Absence of Operator

## ARENA® SIMULATION MODEL

At this point it is relevant to present the objectives of this simulation exercise (Shannon 1975). When dimensioning the milk run system, the team had to take several decisions, using their experience and the best available information. However, the implemented system did not bring all of the expected gains. The purpose of this simulation is to evaluate different scenarios understanding how the system responds, and find answers to remaining questions. Among these are:

- In terms of human resources, the system was dimensioned as requiring 2 DROs and one Market Attendant. Is this dimensioning correct? May different combinations of the key factors that have an impact on this decision (e.g. route frequency, route length or travelling speed) enable a system which requires 1 DRO instead of 2?
- May an electrical car solution, which affords the advantage to transport more packs (and consequently lots) simultaneously and have shorter transportation times, allow having 1 DRO instead of 2?

Simulation is a very useful tool in the analysis of manufacturing systems. Simulation can be used both during the design phase (Smith 2003), to assess the way in which the system will behave and evaluate alternatives, after the system being implemented, to measure its performance, or to evaluate the impact of system modifications and optimizations. Simulation techniques have been used to address manufacturing system topics related to lean production policies, as the utilization of kanbans (Treadwell and Herrmann 2005) or the formulation of a decision support system based on the automatic creation of Arena® simulation models representing different control strategies of materials flow in a production line (Ferreira et al. 2005). Current concerns in the development of simulation models include flexibility and user friendliness for users without specific simulation knowledge. The need of an adaptable simulation framework to compare different production control policies has already been discussed (Gahagan and Herrmann 2001). Arena® was the selected simulation tool to this exercise, due to its popularity (Dias et al. 2007), flexibility provided by its hierarchical structure and user-friendliness. "Simulation with Arena" (Kelton et al. 2007) provides not only an introduction to the concepts of the software and of the simulation process, but also a hands-on approach to model development, and, consequently, of great use in the development of this work.

The factors that characterize a specific configuration of QPT's milk run system are the following:

- The number of DROs;
- The location of the PoD racks and the associated test cells to each PoD rack, which has an impact on the route design, the travelled distances and the number of packs/lots to deliver and retrieve at each PoD rack;
- The traveling speed, or equivalently, the time that is spent in each path of the route;
- The time spent in lot loading and unloading activities;
- The transportation capacity of each DRO;
- The holding capacity of each PoD rack.

The feasibility of a certain configuration will be determined by the ability of the DROs to handle the associated workload. If the simulation run, for a specific configuration, shows frequent overflow situations at the PoD racks or at the transportation carts, with the lots accumulating at the PoD racks due to lack of transportation capacity, we can state that such configuration is not feasible. In summary, the simulation exercise will consist in testing different combinations of the previously mentioned parameters to assess the existence of overflow situations.

Meanwhile, a sensitivity analysis (Barton and Lee 2002) of some parameters will also be performed. When developing a simulation model, an important aspect is to understand how variations of the model data inputs affect the final result. Besides the need of directing data collecting efforts to the inputs the system is sensitive to, in order to have a good level of confidence on the results and an effective utilization of the data-collecting resources, a sensitivity analysis is also important for other reasons. Even if we are confident about the data, if we realize that the system is very sensitive to an input, final results should be presented with some caution, as small unforeseen differences or variations of the real system may lead to large differences between the model results and the real system results. Still, even if there are not very accurate input data, the model development and simulation may provide important insight about the interaction between the element models, and about the way the system will perform.

The diagram presented in Figure 6 exemplifies lot flow in the area. Among PoD racks the lots are transported by the DROs, and between the PoD racks and the Test Cells, the line operators transport the lots. A full-blown model of this system would consider all of these transportation steps. To develop this model, the number of packs/lots being transported would have to be estimated individually for each test cell, and the PoD racks would be modeled as transit points for the packs/lots on their way to the test cells.

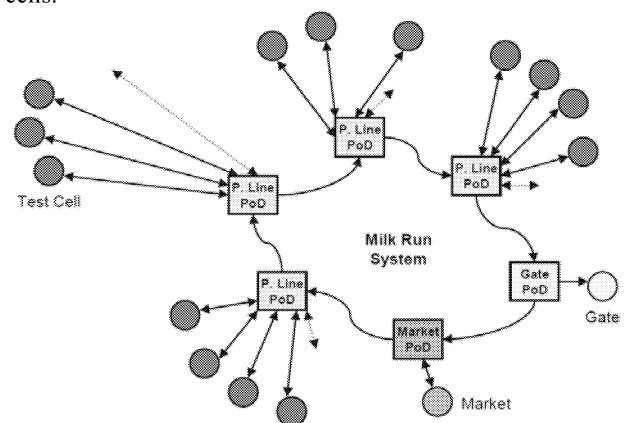


Figure 6: Diagram of Lot Flow in the Test Area

However, the purpose of this simulation exercise is to evaluate the lot transportation system, which interfaces with the production lines at the PoD racks in which the lots are dropped and picked. From the lot transportation

system perspective, after the lot is dropped in the rack, what occurs is a delay process and an attribute modification, as the lot changes from being a lot to process, to a processed lot. To fulfill the goals of this study, such a detailed model is not required, as the study focuses on the route followed by the DROs. It will be sufficient to model each PoD rack as a block that receives, processes and outputs a number of packs that corresponds to the sum of the packs assigned to each individual test cell served by the relevant PoD rack. This will greatly simplify the model, without affecting the level of detail that is required for this exercise.

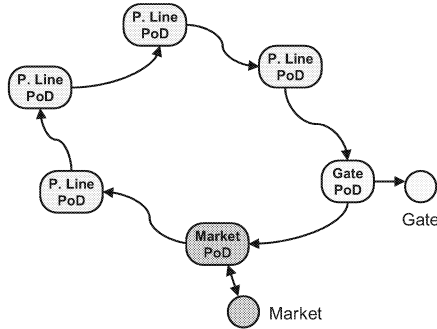


Figure 7: Simplified Diagram of Lot Flow in the Area Used in the Model

The first step in the construction of the model is to map the process in a flowchart (Figure 8). As the purpose of the simulation is to analyze the way the lots flow on the production floor under different circumstances, the flowchart will describe the way they are handled on the production floor, from the lot “perspective”. In Arena<sup>®</sup> terminology (Kelton 2007), lots are then the *entities* in this simulation model and move along the different process steps of the model.

The following aspects of the real world system are reflected in the model:

- the model addresses the transportation of packs, and not of lots, because the transportation capacity is defined in terms of the former;
- the route of each of the DROs starts with a fixed frequency, with the trigger points for each of the DROs separated by half of the period;
- if a DRO is not available when the route should start, the route will start as soon as the DRO is available;
- each DRO has a transportation capacity of 10 packs;
- after being processed or staging at the market, packs are assigned to the next destination (another test cell, the market or the test gate) following a probabilistic distribution that reflects the real-system one;
- lots that are processed in successive steps in test cells associated to the same PoD rack do not need to be transported by the DRO between the two steps.

Building the model in the more intuitive way would consist in generating packs as entities, and allocating these to transporters. However, this approach revealed some practical difficulties, namely regarding the allocation of several entities to the same transporter in the same run. This led to the development of a new approach. The underlying concept is that “DRO entities” will be created,

will never be disposed and will permanently circulate the model, representing the DROs. These entities will have an associated array attribute, which will represent the number of packs that the operator is carrying to each PoD rack and operation in each moment.

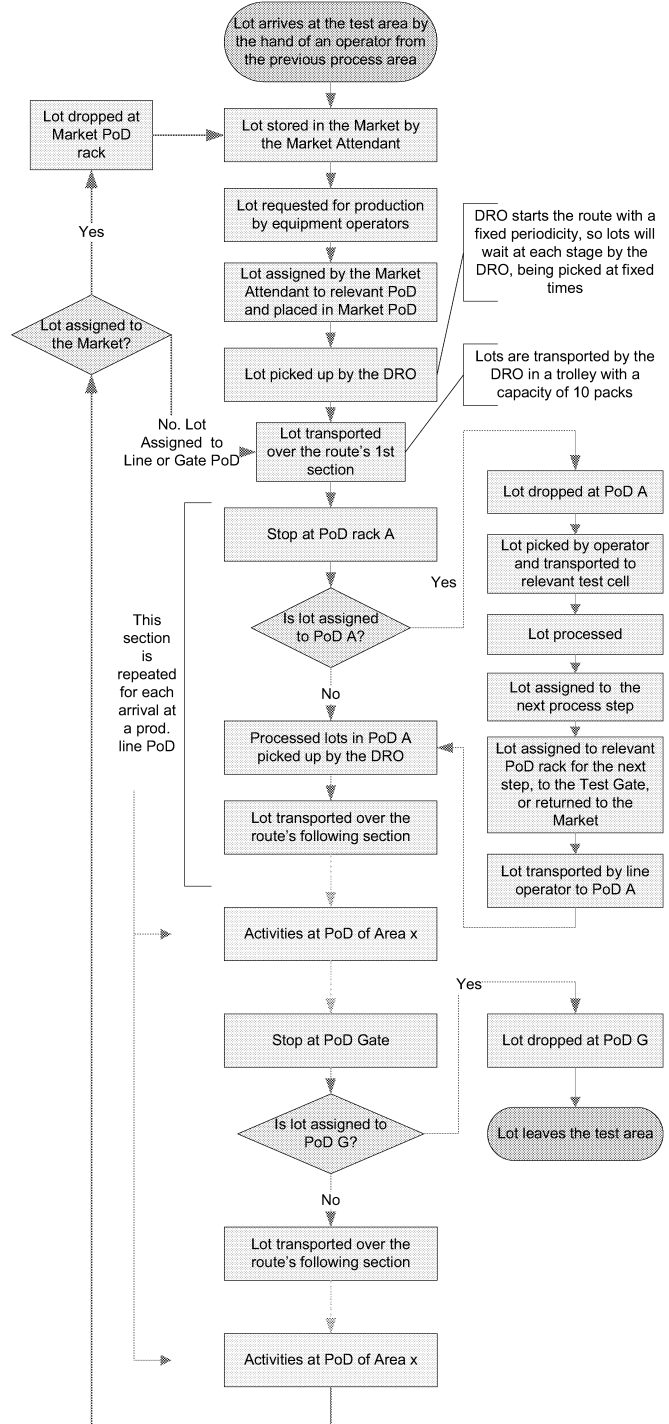


Figure 8: System Flowchart

The relatively small dimension of the Arena simulation model results from a systematic reduction, achieved by building generic “code” that can be shared by several instances of the model. However, the model shows great

complexity, using several multidimensional global variables, and using specific attributes to enable the correct identification of entities, even if placed in different stations of the model, simply running the same code (Ferreira et al. 2005). The overall model picture is presented in Figure 9.

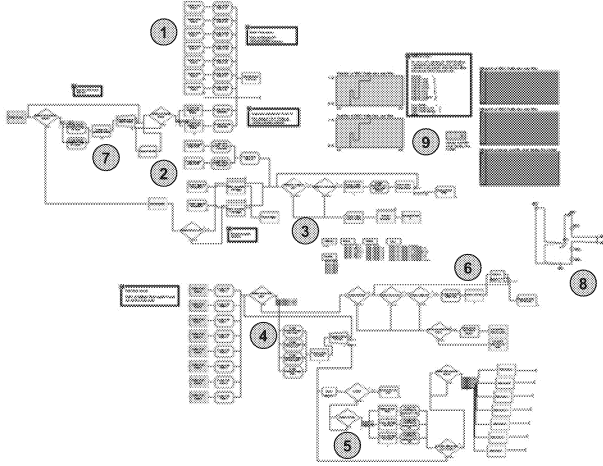


Figure 9 – Arena® Model

The purpose of each of the sections is the following:

1. Create the packs in the first process step (TST1) to be distributed to the lines;
2. Create “DRO entities” and triggers for the route start;
3. Activities at the Market PoD rack;
4. Arrival of DRO at a production line PoD rack (pack delivery);
5. Pack processing and assignment to the following step and PoD rack;
6. Pack pick-up from the PoD rack and DRO departure;
8. Model Animation;
9. Model Data visualization

The animation was kept as simple as possible, mostly for model verification and validation purposes, and was performed using Arena® guided transporters over networks – Figure 10.

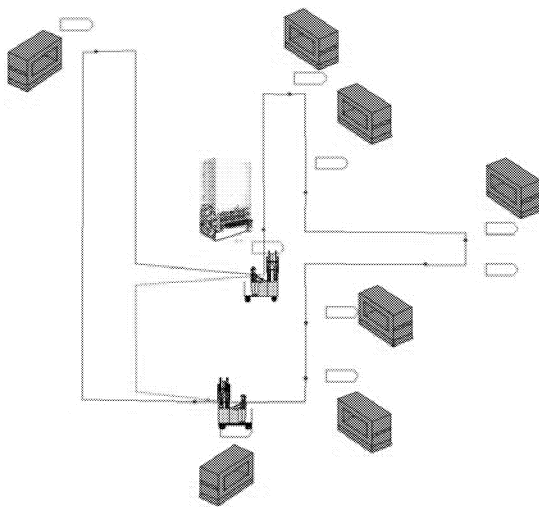


Figure 10: Animation Area Sample Screenshot

## SIMULATION RESULTS

With the purpose of answering the questions mentioned in the previous section, the following scenarios were tested.

| Reference  | DRO number | Traveling speed (m/s) | Route frequency by DRO (min) |
|------------|------------|-----------------------|------------------------------|
| Scenario A | 2          | 0.71                  | 12                           |
| Scenario B | 1          | 0.71                  | 12                           |
| Scenario C | 1          | 0.71                  | 10                           |
| Scenario D | 1          | 0.71                  | 8                            |
| Scenario E | 1          | 1.94                  | 8                            |
| Scenario F | 1          | 1.94                  | 6                            |

Table 3: Experimental Scenarios

In these scenarios the route design and length was not modified. Scenario A corresponds to the existing system, with the travelling speed resulting from a standard work study. Scenarios E and F correspond to the utilization of an electrical tow tractor. For each of the scenarios, its feasibility was analyzed. A scenario is considered feasible if the packs do not have to wait for a long time to be picked up by the DROs from the PoD racks (reference value is less than 6 minutes) and if the number of packs in the PoD racks does not exceed the current PoD rack capacity of 15 packs. The model was built so that these values are returned by Arena® after the simulation run finishes. Equally interesting to analyse is the transporter workload, which is expected to be low for the current setup, facing the results from the MOS. Several authors addressed the analysis and validity of simulation results, and also the design of experiments applied to simulation exercises (Kelton 1995; Kleijnen 1995). Arena® itself contains features aimed at reducing uncertainty regarding simulation results, such as the Replications and Warm-up period options of the Run setup, but also the half-width parameter presented in the results. Adequate values were considered for these. The results, summarized in Tables 4 and 5, allow us to draw the following conclusions:

- The current scenario (scenario A) allows low queuing times of the lots in the PoD racks, but implies an inefficient utilization of the resources;
- Keeping the actual setup in terms of route frequency and travelling speed, but reducing the number of DROs to 1 (scenario B), is not an option, as the DRO is over-loaded, departing from the Market with the cart full most of the times, and with long waiting times (Figure 11);
- A solution with 1 DRO is feasible when the route frequency is reduced (Scenarios D, E and F); if the traveling speed is increased (Scenario E) the DROs are still under-loaded and if the travelling speed is 0.71 m/s (Scenario D) the DRO is slightly over-loaded;
- Facing the fact that even only one transportation resource is still greatly under-loaded when an electrical tow tractor solution is used (speed of 1.94 m/s), the following options should be considered: using the same resource for transportation activities in the neighbor area (Burn-in); implement a coupled

route, in which the DRO performs also the market attendant functions.

| Scenario | Pack waiting time (min) |     | Packs in PoD rack |     | DRO load |
|----------|-------------------------|-----|-------------------|-----|----------|
|          | avg                     | max | avg               | max |          |
| A        | 3                       | 17  | 1                 | 8   | 57%      |
| B        | 22                      | 197 | 4                 | 22  | 65%      |
| C        | 9                       | 103 | 2                 | 15  | 74%      |
| D        | 5                       | 50  | 1                 | 13  | 87%      |
| E        | 5                       | 39  | 1                 | 9   | 38%      |
| F        | 3                       | 17  | 1                 | 7   | 44%      |

Table 4: Experiment Results

| Scenario | DROs            | Trolley capacity | Waiting times |
|----------|-----------------|------------------|---------------|
| A        | under-loaded    | adequate         | very low      |
| B        | under-loaded    | over-loaded      | very high     |
| C        | loaded          | over-loaded      | high          |
| D        | slight overload | slight overload  | low           |
| E        | under-loaded    | adequate         | low           |
| F        | under-loaded    | adequate         | very low      |

Table 5: Interpretation of Experiment Results

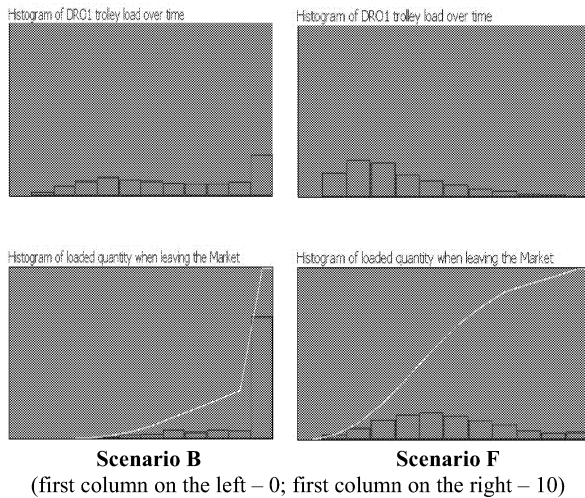


Figure 11: Arena® Model Histograms

It was also the purpose of this project to verify the sensitivity of the model to some of its inputs, namely:

- Pack load and unload times at the Point of Drop racks;
- Pack processing times in the 3 test process steps;

To perform this analysis, the following situations were simulated:

- Scenario G: identical to Scenario F, but with pack load and unload delays of 10 seconds instead of 5;
- Scenario H: identical to Scenario D, but with pack load and unload delays of 10 seconds instead of 5;
- Scenario I: identical to Scenario D, but with pack processing times following a triangular distribution (min:0.5; most likely: 3.0; max: 5.5), very different from the current one;

The system was found to be sensitive to variations in the pack loading and unloading time periods. For a situation in which the transporter was underloaded (scenario F), the increase in this parameter from 5s to 10s led to an increase of the transporter occupation from 44% to 66%, while the queue waiting times were similar (in the range of 3.1 to 3.5). For a situation in which the transporter was already significantly loaded (scenario D), the transporter occupation rose slightly from 87% to 91%, but the queue average waiting times increased sharply, from a range of 4.2 to 5.5 minutes, to a range of 25 to 30 minutes, with maximum values ranging from 120 to 210 minutes. Regarding variations in the pack processing times, the system was found to be insensitive. A strong variation of this parameter, considering scenario D as a reference, led to similar transporter occupation (87%) and queue waiting times.

## CONCLUSIONS, FINDINGS AND FURTHER STEPS

This paper addresses a project that consisted of the implementation of a Milk Run Lot Transportation System in Qimonda Portugal Test Area. The target of the project, framed in a Lean approach, was to reduce waste, namely the transportation waste, optimizing the utilization of the test area human resources. The project was triggered by a Multi Observation Study performed in the Test Area which identified that approximately 4.8% of the operators' time was spent in transportation activities, performed in a "taxi" mode. After characterizing the test area and its transportation needs, a milk run system based on a fixed route, with 2 Delivery Route Operators and a Market Attendant, was implemented. Each DRO initiates a new route every 12 minutes, with a DRO starting a new route 6 minutes after the previous DRO started his route. To support this system, and avoid long periods of time between the lot request by the equipment operator and the lot delivery by the DRO, an IT application was developed to support the lot request process. Another MOS, performed after the milk run system was implemented, revealed that the average time spent in transportation by each manufacturing team remained approximately the same, contrarily to the expectations. Nevertheless, the system allowed equipment operators to focus on their activities, as the same MOS reveals, because the indicators related with the availability of operators to assist the equipment improved. Still, the MOS revealed that the system is not optimized and this triggered the development of a simulation study to understand which may be the best setup.

The simulation study, performed in Arena®, confirmed that the current setup allows having fast deliveries of the lots to the lines, with the lots staying for a short time in the PoD racks, but also that the transportation resources are under-utilized. The simulation of other scenarios revealed that it is possible to have a system with just one DRO, under certain conditions. One option is to use an electrical tow tractor, which allows reduction of the travelling time, and increase the number of routes per hour. Simulation revealed that in this situation the transportation resource would be under-loaded as well, what opens the possibility of the same resource carrying out transportation activities

in the neighbour area (Burn-in) or of implementing a coupled route. Another option would be to increase the number of routes per hour of the current setup, reducing the interval between route starts from 12 minutes to 8 minutes. It must be taken into account that the latter option would represent a high workload to the DRO, just feasible if there is some rotation of the operators occupying the DRO function along the shift.

The results of this simulation experiment will now be shared with the decision-makers, in order to determine the steps to achieve a more optimized setup, and focus even further the operators on value added activities.

## REFERENCES

- Barton, P.I., and Lee, C.K. (2002). "Modeling, simulation, sensitivity analysis, and optimization of hybrid systems". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 12(4) (October 2002), 256-289 (ISSN:1049-3301).
- Dias, L.S., Pereira, G.B. and Rodrigues, A .G. (2007). "A Shortlist of the Most Popular Discrete Simulation Tools". *Simulation News Europe*, April 2007, 17(1), 33-36. (ISSN 0929-2268)
- Ferreira, L., Pereira, G. and Machado, R. (2005). "Geração Automática de Modelos de Simulação de uma Linha de Montagem de Auto-Rádios". *Investigação Operacional*, Junho, 25(1), 37-62. (ISSN 0874-5161).
- Gahagan, S. M., and Herrmann, J. W., (2001). "Improving simulation model adaptability with a production control framework". *Proceedings of the 2001 Winter Simulation Conference*. Arlington: ed. B.A. Peters, J.S. Smith, D.J. Medeiros, and M.W. Rohrer.
- Harris, C., and Harris, R. (2007). *Developing a Lean Workforce*. Portland: Productivity Press. (ISBN 9781563273483)
- Harris, R., Harris, C., Wilson, E., (2003). *Making Materials Flow*. Brookline: Lean Enterprise Institute. (ISBN 0-9741824-9-4)
- Hugos, M., (2003). *Essentials of Supply Chain Management*. Chichester: John Wiley & Sons. (ISBN 9780471434290)
- Kelton, W.D., (1995). "A Tutorial on Design and Analysis of Simulation Experiments". *Winter Simulation Conference 1995*: 24-31
- Kelton, W., Sadowski, R., and Sturrock, D. (2007). *Simulation with Arena*. New York: McGraw-Hill. (ISBN 9780073259895)
- Kleijnen, J.P.C., (1995). "Statistical validation of simulation models". *European Journal of Operational Research*, Elsevier, 87(1), 21-34.
- Nomura, J., and Takakuwa, S., (2006). "Optimization of a number of containers for assembly lines: the fixed course pick-up system". *International Journal of Simulation Modelling* 5 (2006) 4, 155-166 (ISSN 1726-4529).
- Ohno, T. (1988). *Toyota Production System: Beyond Large-Scale Production*. Productivity Press; 1st edition. (ISBN 0684810352)
- Raposo, R., (2009). "Implementation of a Milk Run Material Transportation System in the Test Area of Qimonda Porto Backend". Unpublished Thesis, framed in MITPortugal's EDAM TME Advanced Studies Course, supervised by professors Guilherme Pereira and Luís Dias from Universidade do Minho, Portugal.
- Rother, M., and Shook, J., (1999). *Learning to See: Value Stream Mapping to Add Value and Eliminate MUDA*. Brookline, Mass.: Lean Enterprise Institute., (ISBN 0-9667843-0-8)
- Shannon, R., (1975). *Systems Simulation: The Art and Science*. Englewood Cliffs: Prentice-Hall. (ISBN 9780138818395)
- Smalley, A., (2004). *Creating Level Pull: a Lean Production-System Improvement Guide for Production-Control, Operations, and Engineering Professionals*. Lean Enterprises Inst Inc. (ISBN 0-9743225-0-4)
- Smith, J.S., (2003). "Survey on the use of simulation for manufacturing system design and operation". *Journal of Manufacturing Systems*. 22(2): 157-171.
- Swain, J. J. (2001). "Power tools for visualization and decision making: 2001 Simulation Software Survey". *OR/MS Today* Baltimore, Maryland: INFORMS. 28(1): 52-63.
- Treadwell, M. and Herrmann, J., (2005). "A Kanban Module for Simulating Pull Production in Arena". *Proceedings of the Winter Simulation Conference*. Orlando, Fla. Dec. 4-7, 2005.
- Womack, J., & Jones, D. (1996). *Lean Thinking*. New York: Simon & Schuster. (ISBN 0684810352)

## BIOGRAPHY

**RICARDO RAPOSO** was born in 1975 in Coimbra, Portugal. He graduated in Electrical Engineering in the University of Coimbra, Portugal. Since then he has worked in telecommunications and semiconductor areas, holding positions related to network design and process engineering. He is currently finalizing the TME Advanced Studies Course framed in MITPortugal's program *Engineering Design and Advanced Manufacturing* area.

**GUILHERME PEREIRA** was born in 1961 in Porto, Portugal. He graduated in Industrial Engineering and Management in the University of Minho, Portugal. He holds an MSc degree in Operational Research and a PhD degree in Manufacturing and Mechanical Engineering from the University of Birmingham, UK. His main research interests are Operational Research and Simulation.

**LUÍS DIAS** was born in Vila Nova de Foz Coa, Portugal. He graduated in Computer Science and Systems Engineering in the University of Minho, Portugal. He holds a PhD degree in Production and Systems Engineering from the University of Minho, Portugal. His main research interests are Operational Research and Simulation.

# IMPLEMENTATION STEPS FOR DEVELOPMENT OF A SIMULATION APPLICATION SOFTWARE FOR PRODUCTION SCHEDULING AND DOOR-TO-DOOR LOGISTICS MANAGEMENT OF A MULTI-ASSEMBLY-LINES AUTOMOTIVE VEHICLE MANUFACTURING PLANTS

Hossein Hosseinzadeh  
Research Institute for Technological  
Development Studies (RITDS), Iran  
Research Organization. for Science &  
Technology (IROST),

[h\\_hosseinzadeh@irost.org](mailto:h_hosseinzadeh@irost.org)

Mohsen Bahrami  
Department of Mechanical  
Engineering, Amirkabir University of  
Technology (AUT),

[mbahrami@aut.ac.ir](mailto:mbahrami@aut.ac.ir)

Alireza Hosseinzadeh  
Department of Industrial Engineering,  
Iran University of Science & Technology  
(IUST),

[a\\_hosseinzadeh1988@yahoo.com](mailto:a_hosseinzadeh1988@yahoo.com)

## ABSTRACT

A hybrid of simulation modeling approach and value stream mapping (VSM) methodology were used for implementing lean in IKCO. Based on utilization of information technology (IT), an application software was developed as a tool for achieving the desired lean systems. This paper reflects the implementation steps for development of IKCO LeanPlanner Simulation Software. The developed models for IKCO are BOM-based and have been developed in the environment of Enterprise Dynamics of the InControl. The sample results in IKCO Group show the effectiveness of the proposed framework for implementing lean philosophy; as well as usefulness of computer aided lean Manufacturing tools.

**Keywords:** Simulation & Modeling, VSM, Logistics Systems Software Applications, Lean Manufacturing

## 1.0 INTRODUCTION

The core challenges for industries in the 21st century involve identifying and delivering value to every stakeholder. Meeting that challenges requires lean capability at the enterprise level. In this relation, lean thinking paradigm deals with the systematic approach of identifying value and eliminating waste (non-value added activities) at all levels, through continuous improvement by flowing the product or process at the pull of the customer or client in pursuit of perfection [1]. In lean manufacturing, identifying wastes and eliminating them for creating value for all stakeholders, is one of the basic pillars of the lean thinking. For studying simpler systems, the function of identifying the wastes can be considered as a process of problem-solving approach, and each of the 7-tool methodologies which is more suitable (e.g., fishbone, Pareto and Scatter Diagrams, and the rest, could help to find the best solution for a specific problem. However, for studying the behavior of large and complex systems, more advanced approaches are needed. The Value Stream Mapping (VSM) methodology is one of advanced approaches for visualizing the existing and desired situations of these kinds of systems. With this approach: first, the current flow of information and materials are mapped, and then, on this basis, as well as considering the company's constraints, a pencil and paper in hand draft of the desired state of the system is created by the responsible team. By analyzing the two mentioned maps, the related gaps are identified [2]. Theoretically, by implementing the defined improvement projects, the wastes should hopefully be removed. However, in complex environments, like the

automotive industry and its supply chains, because of their dynamic natures, advising the solutions for achieving the desired/ideal state is a matter of jeopardizing ourselves by diving in the stormy and deep dark water. Simulation-and-modeling (S&M) is a superior tool for this solution. In this relation, the process of carving away the unwanted reality from the bones of a problem is called modeling the problem. The idealized version of the problem that results is called a model [3].

Simulation is the process of developing a model out of a real system and experimenting with this model for understanding the behavior of the system, and/or evaluating the different strategies (scenarios) for the performance of the desired system [4]. VSM and simulation are two complementary tools for implementing lean manufacturing. The VSM is considered as a snapshot, and the simulation is seen as a movie [5]. It worth to mention that, among the 7 deadly wastes, at least 5 of them are related to the logistics system drawbacks.

In line with implementing total lean, at the Iran Automotive Manufacturer Company [6] the project entitled: "The Simulation of Door-to-Door Logistics System of IKCo" was defined implemented successfully.

## 2.0 SIMULATION & AUTOMITVE INDSTRY

Most of the global automotive manufacturers use the simulation as a strong decision making tool. The application of simulation in the automotive industry has wide spectrum. The compilation of a number of publications in this field shows that the main usage of its application includes:

- Logistics and supply chain management (SCM),
- Plant/office lay-out and resource allocation,
- Factory establishment for main parts,
- Assembly lines and their extensions,
- Final audit and processing of manufactured cars,
- New Products Development (NPD) and or New Sub-Systems Development (NSD),
- Design and Development of New Processes, and/or the related improvement projects,

Among the vast applications of simulation and modeling (S&M) in different activities of the automotive industry, its utilization in manufacturing and logistics systems, has a distinguishing position. In this relation, in order to achieve a pull system of production of the lean thinking philosophy it is essential that a number of changes to be made to the existing push systems of mass production



paradigm. The S&M approach can be considered as the most appropriate tool to achieve the targets of the lean manufacturing systems in dynamic environments.

### 3.0 LEAN MANUFACTURING PRINCIPLES

The main objective of the lean system is increasing benefit through decreasing or eliminating wastes in order to create value for different stakeholders. The “lean” word, meaning “free from fat”, was first coined by a graduate student, and then a researcher in the International Motor Vehicle Program (IMVP) at the Massachusetts Institute of Technology (MIT). This system of production was first invented in Toyota Engine Company in Japan, and is called Toyota Production System (TPS) as well. James P. Womack and Daniel T. Jones, postulate that five principles should be followed in order to transit from a mass production paradigm to a lean manufacturing system: 1) Specifying the value, 2) Identifying the Value Stream, 3) Making it in Motion, 4) Pull, and 5) Perfection [7].

To proceed for achieving the desired goals of lean production, a number of mechanisms have been introduced. To name a few, the following list includes the basic tools and strategies in this matter [8], [2]:

- Just in time (JIT) with small batches,
- Minimum inventory of work in processes (WIPs),
- Co-location of the assembly lines and the part manufacturers,
- Pull system of the parts with Kanban Cards,
- Leveling of production (heijunka),
- Machine and/or assembly line balancing,
- Work standardization,
- Utilization of error proofing tools,
- Autonomation (automation with human touch),
- High level of outsourcing
- Continuous improvement.

Invented by Toyota Motor Company, the utilization of the lean manufacturing principles and tools caused that the mentioned company as well as the other Japanese companies to achieve the highest productivity in their new product development (NPD) and production activities. By producing the highest quality products and articles, the Japanese proceeded the western companies, at local and global markets.

### 4.0 VALUE STREAM

The “value stream “ word is relatively a new word in the manufacturing glossary. Value stream includes all of activities (added or non-added value) which are used for manufacturing of a product. It begins with customers/consumers, and ends with the customers/consumers. It encompasses all of the life cycle system: process of raw material, generation of new ideas, design and development of new products and or processes, make/production of products, and finally putting the products in the hands of the customers. In this paper when we talk about the value stream, by looking backward, we limit it to the internal logistics system (door-to-door) of the IKCo’s Complex at Tehran; i.e., from the IKCo’s Sales Office (Parking Lots) to its Receiving Area (Goods Reception Entrance Gates).

### 4.1 Value Stream Maps of Internal Logistics

With using the value stream mapping methodology, and by looking backward, the physical flow of materials are drawn from left to right, and the flow of information is mapped from right to left. This combination, which is a process of pencil and papers in hand, deploys a number of icons, and depicts a big picture of the flow of information and materials in order to produce the demanded products from the sales office. In this project, the value stream maps (VSMs) of the mentioned scope (the internal door-to-door logistics system) was mapped, in three levels:

Level I: the VSM of the Total Logistics System

Level II: the VSMs of logistics areas of 3 assembly lines,

Level III: the VSM of the 3 assembly lines in detail.

Figure.1 shows the big picture of the Total Logistics System (Level I), at the beginning of the project. Figure.2 represents the desired state of the IKCo’s internal logistics system at this stage.

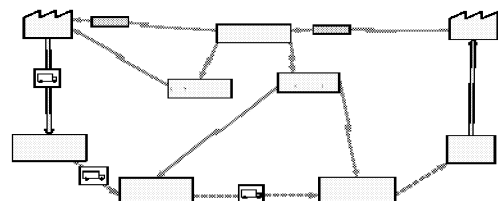


Figure.1- The First Level of the Current VSM of Logistics System (Push System)

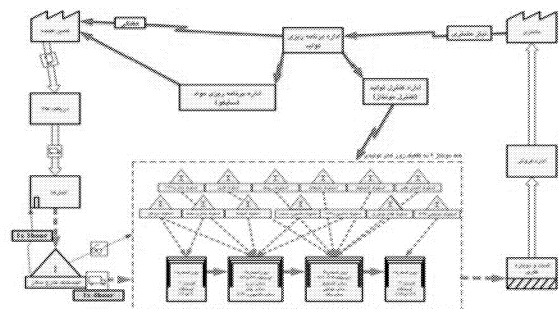


Figure.2- Current VSM of Logistics System: Overall four zones and their Inventory Stocks of the Assembly line #4 (Push System)

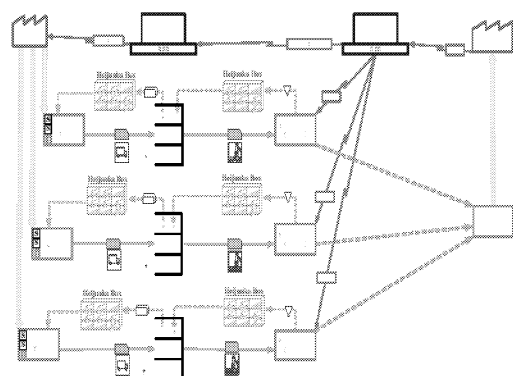


Figure.3- The VSM of the Desired State of the Logistics System (Pull System)

The desired door-to-door logistics system at this stage is very similar to the flow of materials and information in Western super-markets. To achieve these goals, the seamless information, as well as the integrity between the flow of material, and the flow of information is very crucial. In this respect, an IT based plan, Enterprise Resource Planning (ERP) Program, is under way at IKCO, to make the desired lean goals, including the desired state of the logistics system happen.

## 5.0 MAIN ACTIVITIES OF THE IKCO'S PROJECT

On this basis, in order to modeling the existing state of the logistics system and the desired system designs (scenarios), in line with the 9-stage process of proceeding a simulation project [4], [9] [10], [11], [12], [13], [14] the following activities were planned and took place accordingly:

1. the scope of the project defined,
2. the existing state of different activities of the internal logistics system of the entire complex identified,
3. the Data Flow Diagrams (DFDs) of the related logistics activities mapped,
4. the VSMs of the current logistics activities of the areas and the assembly lines mapped (in three level),
5. the future/desired state of the logistics activities of the areas and the assembly lines mapped,
6. a multidisciplinary core team of following different fields, shaped: 1) Simulation & Modeling, 2) Statistics and Data Collection, and 3) Value Stream Mapping,
7. an intensive simulation training course held for the internal stakeholders representatives,
8. exploring for the appropriate simulation software, the Taylor Enterprise Dynamics (TED), selected,
9. The distribution function (DFs) or the histograms of different transportation agents identified.
10. the door-to-door from-to chart flows of different transportation vehicles mapped,
11. on the basis of different VSMs, DFs, and the related data bases of bills of materials (BOMs), the simulation model of the assembly lines, in the environment of the TED, developed,
12. according to the desired state and/or the required system designs, the simulation models run,
13. the primary results were analyzed and after performing the necessary tune-ups, the valid modules developed,
14. According to master schedule of the enterprise, several one-week production plans of different lines were simulated, and the related criteria were tested, and the necessary modifications and optimum solutions advised.

Among the mentioned activities, the process of identifying the distribution or monographs of different transportation vehicles was a heavy task. In That process, a group of 120 short-employed individuals were assigned for a period of one week for collecting the data for 12 hours (two shifts of working time). Making a network of 15 computers, the collected data was entered and the desired DFs as well as the flow of different transportation vehicles produced. The

SAS Software was applied for that purpose. Figure.3 shows a sample the these distribution functions (DFs)

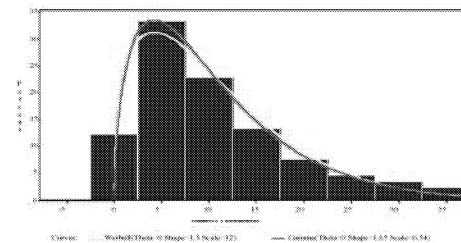


Figure.4 The Distribution Function of Waiting Times of Transporters at Warehouse #54 -Gate 3

The DFs were double checked with the AutoFit Function of the TED environment and showed that fitnesses were rational.

## 6.0 SOME REPLICATION RESULTS

Figure.5 depicts a snap shot of the 2D View of the Assembly Lines # 4. It worth to mention in addition that any of the triple assembly lines virtually could be run independently, however, the system solution has the capacity to run the total system simultaneously, i.e., all of the assembly lines working concurrently and making the logistics system shaped and cars manufactured.

The consecutive lines of rectangles in the picture mimic the work stations; where the car bodies from Painted Bodies Stack (PBS) enter the line and get their necessary parts in order to be made as complete automobiles. The models are connected online with the Bills of Processes (BOPs) databases as well as the Production Planning Center. In this way, the replications results are more close to the IKCo's real production requirements.

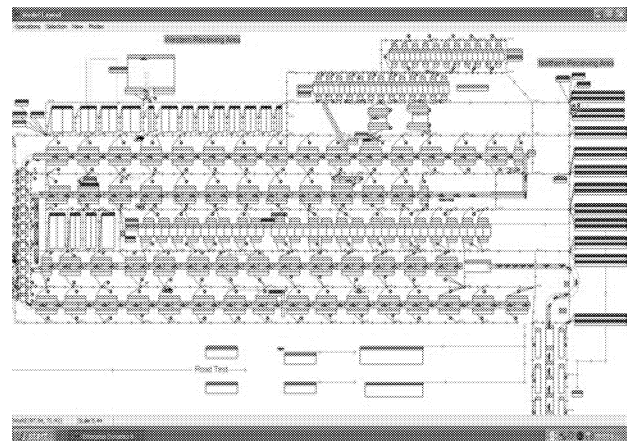


Figure.5 2D View of the Assembly Line #4

In a 15 replications of a desired production plan, for a three working shifts period (22 hours and 20 minutes of the week 3 of the September 2008), the average results are depicted in upper part of Table.1. The actual statistics after the implementation of the production plan during the mentioned period reported from the Production Planning & Control Department has been shown in the lower section of Table.1.

| Station Number  | Part Number      | StockOut(StartTime)            | StockOut(EndTime)              | Stoppage (Mintes) | Vehicle Type       |
|---|------------------|--------------------------------|--------------------------------|-------------------|--------------------|
| 4C003   | X7V7041001<br>AB | 16777.3<br>2<br>(04:39:3<br>7) | 16790.9<br>4<br>(04:39:5<br>1) | 0.22              | SorenEmo_3_5<br>54 |
| 4C008   | 9655021580       | 17252.9<br>9<br>(04:47:3<br>3) | 17744.3<br>3<br>(04:55:4<br>4) | 8.18              | SorenEmo_3_5<br>54 |
| 4SE02<br>4  | X7V3025001<br>AB | 25592.8<br>6<br>(07:06:3<br>3) | 26134.9<br>1<br>(07:15:3<br>5) | 9.03              | SorenEmo_3_5<br>54 |
| 4B054   | 5131503          | 57360.2<br>9<br>(15:56:0<br>0) | 57626.9<br>8<br>(16:00:2<br>7) | 4.44              | SorenEmo_3_5<br>54 |
| 4C008   | 9655021580       | 72729.4<br>6<br>(20:12:0<br>9) | 74873.7<br>5<br>(20:47:5<br>4) | 35.7<br>3         | EL_CNG             |
| 4C008   | 9655021580       | 78441.2<br>6<br>(21:47:2<br>1) | 78609.1<br>3<br>(21:50:0<br>9) | 2.79              | Samand_10_29<br>8  |
| Total Stops of Assemblyline during 22 Hr and<br>20 Mins |                  |                                |                                | <b>60.39</b>      | -                  |

a) Statistics resulted from replications of 15 runs

|   |                  |                                |                                |           |                    |
|---|------------------|--------------------------------|--------------------------------|-----------|--------------------|
| 4C008   | 9655021580       | 17252.9<br>9<br>(04:47:3<br>3) | 17735.3<br>9<br>(04:55:3<br>6) | 8.05      | SorenEmo_3_5<br>54 |
| 4SE024  | X7V3025001<br>AB | 25584.2<br>2<br>(07:06:2<br>4) | 26094.2<br>2<br>(07:14:5<br>4) | 8.5       | SorenEmo_3_5<br>54 |
| 4C008   | 9655021580       | 72680.5<br>1<br>(20:11:2<br>1) | 74386.3<br>1<br>(20:39:4<br>6) | 28.4<br>5 | EL_CNG             |
| Total Stops of Assemblyline during 22 Hr and<br>20 Mins |                  |                                |                                | <b>45</b> | -                  |

b) The actual statistics after the implementation of the production plan during the mentioned period

Table.1- The Results of 15 replications of a three shift working of the Assembly line #4.

As the information shows the resulted statistics are very valuable for decision makers of production planning and logistics people. First of all, the LeanPlanner pointed out that a number of parts have the potential for stopping the assembly lines (parts numbers in stations: 4C003, 4SE024, 4B054, 4C008). As the table shows the starting time and ending time of the each stoppage due to stock-out of the mentioned stations have been forecasted. For instance, simulations runs showed that the stock-out of the station number 4SE024 would start at 07:06:33 after the start of the working shift of the plant, and ends after 9 minutes; and it was expected that the Assembly line would stop for average of one hour (60.39 m).

Secondly, the implementation statistics depicted in the lower part of the table show that mentioned parts could really jeopardize the safe working of the assembly line, and three of them (parts numbers: 9655021580, X7V3025001AB, 9655021580, in stations 4C008, and 4SE024) really caused that the line to be stopped for 45 minutes. It should be mentioned that the human involvement in making decisions rather than the procedures for working normal of the equipments caused that a number of hampering factors to be removed and so the 15 minutes difference is due to those strategies.

The information evidences shows the validity of the LeanPlanner Models. Also, statistical analysis proved the equality of the results (t-paired Test), which the report of these kinds of analyses are out of the scope of this conference paper.

## 7.0 CONCLUSION

Identifying wastes and eliminating them to create value for stakeholders, as well as responding to change are the main tasks of lean thinking. For identifying the wastes and visualizing the desired situations, the VSM methodology is an appropriate tool. Although, this approach is very effective within static and normal surroundings, however, in the complex and dynamic environments, like in the context of automotive industry, the desired situations (scenarios) should somehow be tested before any real action. The simulation and modeling (S&M) approach helps in this. Scholars and professional consultants who have used a hybrid approach of dynamic simulation approach and the VSM methodology in different organizations report its effectiveness to achieve lean goals.

In line with implementing lean at enterprise level in IKCO Group [10], using this combined methodology [14] for improving the door-to-door logistics system and production scheduling of the IKCo's Tehran Complex, we came to conclusion that it is very useful tool for lean learning and teamwork in a dynamic environment like the automotive vehicle assembly plants,

To the best of our information from works done in this field, the mentioned methodology has been implemented for vehicle assembly lines of one or at most two, with a few of product types. In three assembly lines of IKCo which could be called a multi-product, and multi assembly lines surroundings; there are a product mix of more than 130 sorts of cars and there are about 30 warehouses for housing the received parts from more than 500 part and material suppliers. In this context, simultaneous working of the assembly lines which consume parts from different venues (warehouses, logistics areas, and synchronized sending of parts directly form suppliers) is very vulnerable to be stopped by different failing factors. High support of top management and associates for successful implementation of the project was a sign of the importance of the issue and the results achieved up to this point proved that the developed system solution (LeanPlanner) has helped to make sound decisions by responsible managers and the line staff. The cost-benefit analysis of the project shows positive cash flows and implies that utilization of these kinds of systems is logical and wise.

## ACKNOWLEDGEMENTS

The authors wish to thank IKCo and SAPCo's executives for their financial and managerial support. Without their associates' ambition and supports this work could not exist. Our words cannot explain the values created by our project team members. Their persistence, patience and ongoing effort in a complex and huge environment like IKCo, shows their great ambitions for national support and development.

## REFERENCES

- [1] Earl Murman et. al., Lean Enterprise Value (LEV), Insights from MIT's Lean Aerospace Initiative, The Lean Enterprise Foundation Inc., 2002, p. 3.
- [2] Don Tapping, Tom Luyster and Tom Shuker, Value Stream Management: Eight Steps to Planning, Mapping, and sustaining Lean Improvements, Productivity Press, 444 Park Avenue South, Suite 604, New York, NY 10016, USA E-mail: [info@productivityinc.com](mailto:info@productivityinc.com)
- [3] Jack R. Meredith, Samuel J. Jr. Mantel, Project Management: A Managerial Approach, John Wiley & Sons Inc /December 2002
- [4] Averill M. Law, W. Daniel Kelton, Simulation Modeling and Analysis, Third Edition, Mc Graw Hill, International Edition 2000, pp. 83-86.
- [5] Anthony J. Donatelli, Gregory A. Harris, Combining Value Stream Mapping and Discrete Event Simulation, Proceedings of the Huntsville Simulation Conference, 2001, by The Society for Modeling and Simulation International, San Diego, CA.
- [6] Hossein Hosseinzadeh, Mohsen Bahrami, The Aerospace Initiative (LAI)' Structure, the Right Choice for Cooperation of Government, University and Industry, and the LAI's Models and Products, the Right Tools for Lean Implementation, Proceedings of the 9<sup>th</sup> Congress on Cooperation of Government, Universities and Industries for National Development, Tehran/IRAN December 13-15, 2005, p. 102.
- [7] James P. Womack and Daniel T. Jones, Lean Thinking: Banishing Waste and Creating Wealth in Your Corporation, (translated into Farsi Language by Mrs. Azadeh Radnejad, Aammozeh Publications, 2003)
- [8] Michael A. Cusumano, the Limits of Lean, Sloan Management Review, summer 1994.
- [9] Introduction to Simulation and Modeling, <http://www.uh.edu/~lcr3600/simulation/contents.html>
- [10] Beau Keyte, Drew Locher, THE COMPLETE LEAN ENTERPRISE: Value Stream for Administrative and Office Processes, Productivity Press, 444 Park Avenue South, Suite 604, New York, NY 10016, USA E-mail: [info@productivityinc.com](mailto:info@productivityinc.com)
- [11] Kevin J. Duggan, Creating Mixed Model Value Streams, Productivity Press, 444 Park Avenue South, Suite 604, New York, NY 10016, USA E-mail: [info@productivityinc.com](mailto:info@productivityinc.com)
- [12] Banks, J., Carson, J. S., Nelson, B. L., and Nikol, D. M. (BCN&N), Discrete-Event System Simulation, 4th edition, Prentice-Hall, 2005, pp. 17-20.
- [13] Mike Rother and Rick Harris, Creating Continuous Flow, Version 1.0, June 2001, The Lean Enterprise Institute, Brookline, Massachusetts, USA, <http://www.lean.org>
- [14] Michelle Eileen Scullin, INTEGRATING VALUE STREAM MAPPING AND SIMULATION, A thesis submitted to the Faculty of Brigham Young University in partial fulfillment of the requirements for the degree of Master of Science, School of Technology, Brigham Young University, August 2005

# A study on transaction scheduling in a real-time distributed system

Y. Jayanta Singh

Faculty of Information Technology,  
7<sup>th</sup> October University, Misurata, Libya  
y\_jayanta@yahoo.com

Hasan Al-Saedy

British Institute of Technology and E-Commerce  
London, UK  
hlnaser@yahoo.com

S.C. Mehrotra

Dept of Information Technology and Computer Sc,  
Dr. B.A.Marathwada University, Aurangabad, India  
mehrotrasc@rediffmail.com

## Abstract:

The transaction in a real time database system has deadlines to process the workloads. Many transaction complexities are there in handling concurrency control and database recovery in distributed database systems. And complexity increase in real time applications by placing deadlines on the response time of the database system and transactions processing. Such a system needs to process Transactions before these deadlines expired. The performances depend on many factors such as traffic workloads, database system architecture, underlying processors, disks speeds, and variety of operating conditions. A series of simulation study have been performed to analyze the performance under different transaction scheduling condition such as different workloads, different environments, varying file selection time, and varying arrival rate. The scheduling of data accesses are done in order to meet their deadlines and to minimize the number of transactions that missed deadlines. A new approach "firm randomization" method is introduced to help to distribute workloads to all sites uniformly.

**Keywords-** *Transaction scheduling, Real time system, deadlines, firm randomization, throughput.*

## I. INTRODUCTION

A real time distributed computing has heterogeneously networked computers to solve a single problem. If a transaction runs across two sites, it may commit at one site and may failure at another site, leading to an inconsistent transaction. Two-phase commit protocol is most widely used to solve these problems [Silberschatz02]. To ensure transaction atomicity, commit protocols are implemented in distributed database system. A uniform commitment is guarantee by a commit protocol in a distributed transaction execution to ensure that all the participating sites agree on a final outcome. Result may be either a commit or an abort condition.

In a real time database system the transaction processing system that is designed to handle workloads where transactions have complete deadlines. Many real time database applications in areas of communication system and military systems are distributed in nature. To ensure transaction atomicity, commit protocol are implemented in distributed database system. This paper shows report of study the performance implications of supporting transaction atomicity in a real time distributed database system with the help of simulation. Experimental performances of transaction scheduling under variety of workloads and different system configuration are evaluated through this simulation.

The section II describes the concept of a real time database system. In section III, detail simulation model and simulation parameters are given. The detail experiment results and

analysis are given in section IV. The overall conclusions are discussed in section V.

## II. REAL TIME DATABASE CONCEPT

Database researchers have proposed varieties of commit protocols like Two phase commit [Gray'78, Mohan et al.'86], Nested two phase commit [Gray'78], Presumed commit [Lampson and Lomet93] and Presume abort [Mohan et al.'86], Broadcast Two phase commit [Oszu and Valduriez'91], Three phase commit [Oszu and Valduriez'91, Kohler'81] etc. These require exchanges of multiple messages, in multiple phases, between the participating sites where the distributed transaction executed. Several log records are generated to make permanent changed to the data disk, demanding some more transaction execution time [Lampson and Lomet93, Nystrom and Nolin04, Ramamritham et al'04]. Proper scheduling of transactions and management of its execution time are important factors in designing such systems.

Transactions processing in any database systems can have real time constraints. The scheduling transactions with deadlines on a single processor memory resident database system have been developed and evaluated the scheduling through simulation [Robert et al'92]. A real time database system is a Transaction processing system that designed to handle workloads where transactions have complete deadlines. A centralized timed Two-phase Commit protocol has been designed where the fate of a transaction is guaranteed to be known to all the participants of the transaction by a deadline [Davidson et al'89]. In case of faults, it is not possible to provide such guarantee. Real actions such as Firing a weapon or dispensing cash may not be compensatable at all [Levy et al'91]. Proper scheduling of transactions and management of its execution time are the important factors in designing such systems. In such a database, the performance of the commit protocol is usually measured in terms of number of Transactions that complete before their deadlines. The transaction that miss their deadlines before the completion of processing are just killed or aborted and discarded from the system without being executed to completion [Jayant et al'92].

## III. SIMULATION DETAILS

### A. Simulation model

The study follows real time processing model [Jayant et al'92] and transaction processing addressing timeliness [Han'03]. Such model consists of a database that is distributed in a non-replicated manner, over all the available sites (say 8 sites in this study) connected by a network [Jayant et al'96,90,91]. In the actual model, each of sites has six components: (i) a source: generate transactions, (ii) a transaction manager: models the execution behavior of the transaction, (iii) a concurrency control manager: implements the concurrency control algorithm, (iv) a resource manager: models the

physical resources, (v) a recovery manager: implements the details commit protocol and (vi) a sink: collects statistics on the completed transactions. A network manager models the behavior of the communications network. The study is concentrated in managing the transaction scheduling under different workloads. In this study, the part of concurrency control is not fully implemented. The modified model is shown in Fig 1. The study has concentrated to minimized numbers of the percentage of miss transactions in order to optimize the atomicity problem in a distributed database system. The definition of the components of the model is given below.

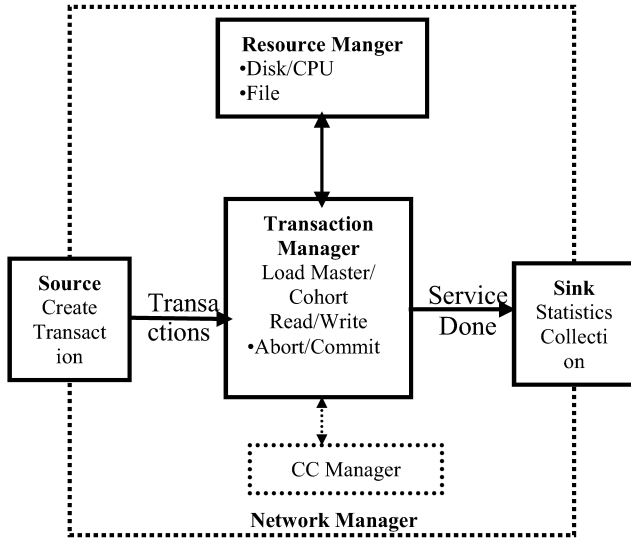


Figure 1. Real time distributed database model.

(i) The source:

This component is responsible for generating the workloads for a site. The workloads are characterized in terms of files that they access and number of pages that they access and also update of a file.

(ii) The transaction manager:

The transaction manager is responsible for accepting transaction from the source and modeling their execution. This deals with the execution behavior of the transaction. Each transaction in the workload has a general structure consist of a master process and a number of cohorts. The master resides at the sites where the transaction was submitted. Each cohort makes a sequence of read and writes requests to files that are stored at its sites. A transaction has one cohort at each site where it needs to access data.

To choose the execution sites for a transaction's cohorts, the decision rule is: if a file is present at the originating site, use the copy there; otherwise, choose uniformly from among the sites that have remote copies of the files. The transaction manager also models the details of the commit and abort protocols.

(iii) The concurrency control manager:

It deals with the implementation of the concurrency control algorithms. In this study, this module is not fully implemented. The effect of this is dependent on algorithm that chooses during designing the system.

(iv) The resource manager:

The resource manager models the physical resources like CPU, Disk, and files etc for writing to or accessing data or messages from them.

(v) The sink:

The sink deals for collection of statistics on the completed transactions.

(vi) The Network Manager:

The network manager encapsulates the model of the communications network. It is assuming a local area network system, where the actual time on the wire for messages is negligible.

B. Execution model and Simulation Parameters

The execution model is discussed below. A common model of a distributed transaction is that there is one process, called as Master, which is executed at the site where the transaction is submitted, and a set of processes, called Cohorts, which executes on behalf of the transaction at these various sites that are accessed by the transaction. In other words, each transaction has a master process that runs at its site of origination. The master process in turn sets up a collection of cohort's processes to perform the actual processing involved in running the transaction. When cohort finishes executing its portion of a query, it sends an execution complete message to the master. When the master received such a message from each cohort, it starts its execution process.

When a transaction is initiated, the set of files and data items that, it will access are chosen by the source. The master is then loaded at its originating site and initiates the first phase of the protocol by sending PREPARE (to commit) messages in parallel to all the cohorts. Each cohort that is ready to commit, first force-writes a prepared log record to its local stable storage and then sends a YES vote to the master. At this stage, the cohort has entered a prepared state wherein it cannot unilaterally commit or abort the transaction but has to wait for final decision from the master. On other hand, each cohort that decides to abort force-writes an abort log record and sends a NO vote to the master. Since a NO vote acts like a veto, cohort is permitted unilaterally abort the transaction without waiting for a response from the master.

After the master receives the votes from all the cohorts, it initiates the second phase of the protocol. If all the votes are YES, it moves to a committing state by force-writing a commit log record and sending COMMIT messages to all the cohorts. Each cohort after receiving a COMMIT message moves to the committing state, force-writes a commit log record, and sends an acknowledgement (ACK) message to the master. If the master receives even one NO vote, it moves to the aborting state by force writing an abort log record and sends ABORT messages to those cohorts that are in the prepared state. These cohorts, after receiving the ABORT message, move to aborting state, force-write an abort log record and send an ACK message to the master. Finally, the master, after receiving acknowledgement from all the prepared cohorts, writes an end log record and then forgets and made free the transaction. The statistics are collected in the Sink [Jayanta et al'90,92,96]

The database is modeled as a collection of DBsize pages that are uniformly distributed across all the NumSites sites. At each site, transactions arrive under Poisson stream with rate ArrivalRate, and each transaction has an associated firm deadline. The deadline is assigned using the formula

$$DT=AT+SF*RT \quad (1)$$

Here DT, AT, SF and RT are the deadline, arrival rate, Slack factor and resource time respectively, of transaction T. The Resource time is the total service time at the resources that

the transaction requires for its execution. The Slack factor is a constant that provides control over the tightness or slackness of the transaction deadlines.

In this model, each of the transaction in the supplied workload has the structure of the single master and multiple cohorts as described above. The number of sites at which each transaction executes is specifying by the Fileselection time (DistDegree) parameter. At each of the execution sites, the number of pages accessed by the transaction's cohort varies uniformly between 0.5 and 1.5 times CohortSize. These pages are chosen randomly from among the database pages located at that site. A page that is read is updated with probability of WriteProb. Summary of the simulation parameter is given in table I.

### C. Parameter Settings

The values of the parameter set in the simulation are given in table II. The CPU time to process a page is 10 milliseconds while disk access times are 20 milliseconds.

TABLE I. SIMULATION PARAMETERS

| Parameters             | Description  |
|------------------------|--|
| NumSites or Selectfile | Number of sites in the Database                                  |
| DbSize                 | Number of pages in the database.                                 |
| ArrivalRate            | Transaction arrival rate/site                                    |
| Slackfactor            | Slack factor in Deadline formula                                 |
| FileSelection Time     | Degree of Freedom (DistDegree)                                   |
| WriteProb              | Page update probability  |
| PageCPU                | CPU page processing time   |
| PageDisk               | Disk page access time  |
| TerminalThink          | Time between completion of 1 transaction & submission of another |
| Numwrite               | Number of Write Transactions                                     |
| NumberReadT            | Number of Read Transactions                                      |

TABLE II. VALUES OF SIMULATION MODEL PARAMETERS

| Parameters  | Set Values     | Parameters         | Set Values   |
|-------------|----------------|--------------------|--------------|
| NumSites    | 8              | FileSelection Time | 3            |
| DbSize      | vary(max.2400) | PageCPU            | 10ms         |
| ArrivalRate | 6 to 8 job/sec | PageDisk           | 20ms         |
| Slackfactor | 4              | TerminalThink      | 0 to 0.5 sec |
| WriteProb   | 0.5            |                    |              |

In this simulation mainly concentrated in processing timing of commit protocol. It is sure that if it is reduced the number of miss transactions, it can also able to reduce the atomicity problem. If the transaction's action deadline expires either before completion of its local processing, or before the master has written the global decision log record, the transaction is killed and discarded.

## IV. EXPERIMENTS AND RESULTS

The study for performance evaluation starts by first developing a base model. Further experiments were constructed around the base model experiments by varying a few parameters at a time. The experiment has been performed using different simulation language such as, in report [Jayanta et al'96] using C++Sim, and in report [Jayanta et al'92] using DeNet. For this study, GPSS World [Minutesmansoftware'01] is used as a simulator. Literatures are also collected from several recent studies [Xiong and Ramamritham'04,Datta and Son'02, Gustavsson and Andler'05,Ziong et al'05,Jan 06,Kang et al'07, Idoudi et al'08].

The performance metric of the experiments is MissPercent that is the percentage of input transaction that the system is

unable to complete before their deadline. The MissPercent values in range of 0% to 20% are taken to represent system performance under "Normal" loads, while ranges of 21% to 100% represent system performance under "heavy" loads. The study analyzes the performance of the system under different workload with varying the arrival rate of the transaction and their deadlines. Also the study analyzed the performance using this new concept of firm randomization technique (given below). Only the statistically significant results are discussed. The comparison of the performance of the system under Centralized commit and Distributed commit processing also shown. The experimental results are discussed below.

### 4.1. Normal load in Centralized and Distributed systems

This section discusses the statistical results of this simulation under different environments. The distributed systems have higher percentage of miss Transactions than centralized system. The higher miss percentage of transaction creates problem in managing atomicity of the transaction in such a system. This leads to design of a new distributed commit-processing protocol to have a real-time committing performance. The comparison of Centralize and distributed performances is shown in Fig 2.

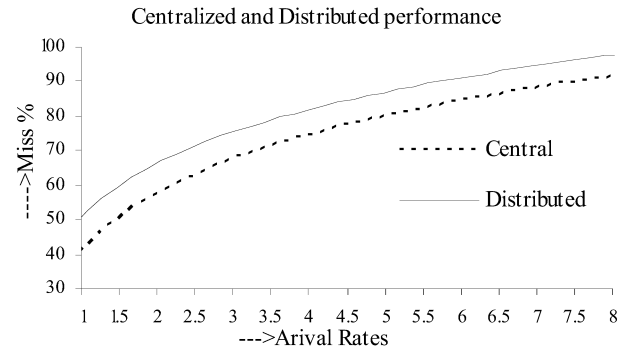


Figure 2 Comparison of Centralize and distributed performances

### 4.2. Impact of file selection time to Individual sites

The simulation was performing by taking 8 distributed sites and by varying their file selection timing. The performance analyses of each of the sites under a normal workload are shown in Fig 3. The miss percentage of transaction is reduced in increasing the file selection timing.

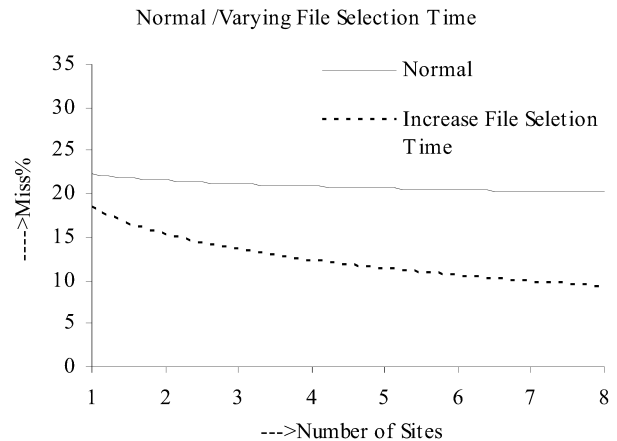


Figure 3. Effect of file selection time to Individual sites

#### 4.3. Impact of arrival rates to Individual sites

In this set of experiments, the impact of increasing the arrival rates was observed on the performances of the each of sites under normal load and heavy load. Fig4 presents the results obtained. Here s1, s2 etc are representing the 8 sites. Under both conditions, the miss percentage is reduced at the lower values of arrival rate of the transactions for each of the sites. In both the normal and heavy load the arrival rate play an important role to give a minimized miss percentage. The success ratios of the transaction are also increase by lowering the arrival rate.

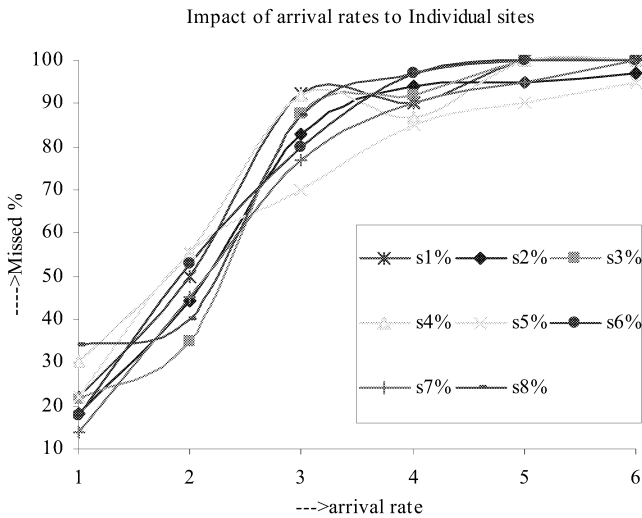


Figure 4 Impact of arrival rates to Individual sites

#### 4.4. Impact of arrival rate to Throughput

In this set of experiments, the impact of increasing the arrival rates was observed on the throughput of the system. Fig5 presents the results obtained. The throughput initially increases with increase in arrival rate. But it drops rapidly at very high loads. So the more studies are required to observe the correlation between missed % and the throughput to have an optimized performance.

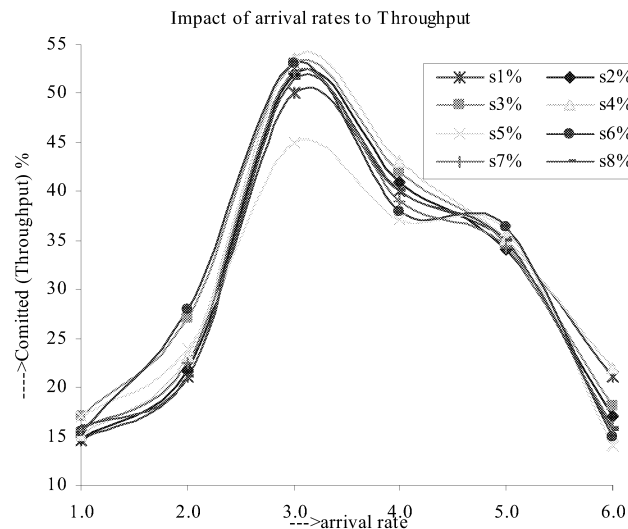


Figure 5. Impact of arrival rates to Throughput

#### 4.5. Firm Randomization

General randomization rules may give a repetition value in a short iteration. A work is assigned to a site number, which is generated by the randomization rule among 8 sites. For example, the first work may be assigned to site number 5 and the second work may be assigned site number 2. The third work may be assigned to site number 5 again, if the randomization rules generate 5. This value 5 should not be generated again at least before completion last iteration. A sites or cohort need some time to complete a previous work and to process another new workload. Here introduced a new concept of randomization for distributing workload to sites so as to avoid assigning more worked load to a busy site. There are “n” numbers of sites in the system. It is made the firm randomization in such a way that, any randomized value should not be repeated twice, at least in between “1 to n” iteration; repetition is allowed after “n+1” iteration. This helps to distribute a heavy work uniformly to all cohorts. The new techniques also give better statistics as compared to the normal randomization rule

#### 4.6. Comparison of Normal and new firm Randomization

In this set of experiments, the performance comparison under normal available randomization and new firm randomization process are observed. While distributing the works to different sites, the new firm randomization process helps in avoiding continuous assignments to some specific busy sites. It helps to distribute the works uniformly to all sites. The fig 6, shows that the performance of the system gets better under firm randomization technique by about 10 to 6%. So proper choice of the distribution or randomization are require designing a best-real time distributed processing system.

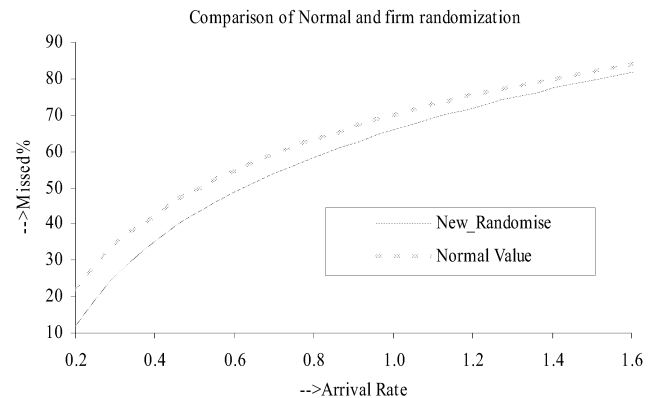


Figure 6. Comparison of Normal randomization and new firm randomization process

## V. CONCLUSION

The real time distributed processing system has been simulating under different traffic load such as normal load and heavy load. In both conditions the arrival rate of transaction plays a major role in reducing number of miss percentage and improved performance. The analysis report of the individual sites also shows that a smaller value of the arrival rate is important to have a less percentage of miss transactions which will give better performance. It also demands the larger value of File selection timing to minimize the miss percentage to give better results.



The throughput initially increases with increase in arrival rate. But it drops rapidly at very high work loads. So the more studies are required to observe the correlation between missed % and the throughput to have an optimized performance.

Time consumption on busy sites by repetitively assignment of new workload and reassigning that to other sites can be solved by proper distribution or randomization process. A new approach "firm randomization" method helps to distribute workloads to all sites uniformly so that percentage miss transactions are minimized. Once a workload is over assigned to a busy site, reassignment of works to some other free sites will take extra time. The new 'firm randomization' will prevent from taking extra time by directing to free sites in initial assignment.

#### ACKNOWLEDGEMENT

It is gratefully acknowledge Minuteman Software, P.O. Box 131, Holly Springs, North Carolina, 27540-0131, USA for providing free Study materials in their site <http://www.minutemansoftware.com>.

#### REFERENCE

- Datta A and Son S. H., 2002, " A Study of Concurrency Control in Real-Time Active Database Systems," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14(3),465-484
- Davidson S., Lee I and Wolfe V.,1989, A protocol for Times Atomic Commitment, *Proc. of 9th Intl. Conf. On Distributed Computing System*
- Gray, J,1978 "Notes on Database Operating Systems", Operating Systems: An Advanced Course, *Lecture notes in Computer Science*, 60
- Gustavsson S and Andler S, 2005, Decentralized and continuous consistency management in distributed real-time databases with multiple writers of replicated data, *Workshop on parallel and distributed real-time systems*, Denver, CO
- Han Q, 2003, Addressing timeliness /accuracy/ cost tradeoffs in information collection for dynamic environments, *IEEE Real-Time System Symposium, Cancun, Mexico*
- Idoudi, N, Duvallet, C, Sadeg, B, Bouaziz, R, Gargouri, F,2008, Structural Model of Real-Time Databases: An Illustration, *11th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing (ISORC 2008)*.
- Jan Lindstrom, 2006, "Relaxed Correctness for Firm Real-Time Databases," *rtcsa*, pp.82-86, *12th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA'06)*
- Jayant H. 1991, "Transaction Scheduling in Firm Real-Time Database Systems", *Ph.D. Thesis, Computer Science Dept. Univ. of Wisconsin, Madison*
- Jayant H, Carey M and Livney M, 1990, "Dynamic Real-Time Optimistic Concurrency Control", *Proc. of 11th IEEE Real-Time Systems Symp.*
- Jayant H., Ramesh G, Kriti R, S, Seshadri, 1996, "Commit processing in Distributed Real-Time Database Systems", Tech. Report-TR-96-01, Pro. Pro. Of 17th *IEEE Real-Time Systems Symposium, USA*
- Jayant, H, Carey M, Livney M, '92, Data Access Scheduling in Firm Real time Database Systems, *Real Time systems Journal*, 4(3)
- Kang W, Son, S., Stankovic J, and Amirijoo M, 2007, I/O Aware Deadline Miss Ratio Management in Real-Time Embedded Databases, *IEEE RTSS*
- Kohler W, 1981, A survey of Techniques for Synchronization and Recovery in Decentralized Computer System, *ACM Computing Surveys*, 13(2)
- Lampson B and Lomet D, 1993, A new Presumes Commit Optimization for Two phase Commit, *Pro.of 19th VLDB Conference*

- Levy E., Korth H and Silberschatz A.1991, An optimistic commit protocol for distributed transaction management, *Pro.of ACM SIGMOD Conf.*
- Minutesmansoftware, *GPSS world*, North Carolina, U. S. A. 2001(4E). [GPSS-Book]
- Mohan, C, Lindsay B and Obermark R, 1986, Transaction Management in the R\* Distributed Database Management Systems, *ACM TODS*, 11(4)
- Nystrom D, Nolin M,2004,Pessimistic Concurrency Control and Versioning to Support Database Pointers in Real-Time Databases, *Proc. 16th Euromicro Conf. on Real-Time Systems*, Catania
- Oszu M, Valduriez P, 1991, *Principles of Distributed Database Systems*, Prentice-Hall,1991
- Ramamritham, Son S. H, and DiPippo L, 2004, Real-Time Databases and Data Services, *Real-Time Systems J.*, vol. 28, 179-216
- Robert A and Garcia-Molina H, 1992, Scheduling Real-Time Transactions, *ACM Trans. on Database Systems*, 17(3)
- Silberschatz, Korth, Sudarshan-2002, *Database system concept*, 4th (I.E), McGraw-Hill Pub. 698-709,903
- Xiong M, Han S., and Lam K, 2005, A Deferrable Scheduling for Real-Time Transactions Maintaining Data Freshness, *IEEE Real-Time Systems Symposium, FL*
- Xiong M. and Ramamritham K., 2004, Deriving Deadlines and Periods for Real-Time Update Transactions, *IEEE Trans. on Computers*, vol. 53,(5)

#### BIOGRAPHY

DR. Y. JAYANTA SINGH is working as a Lecturer in Faculty of Information technology, 7<sup>th</sup> October University, Misurata, (Libya). He obtained his M.Sc and Ph.D in Computer Science from Dr. B.A.Marathwada University, (India) in 2000 and 2004 respectively. He had worked with Keane (Canada & India), Skyline University College (Middle East), TechMahindra (India). His research papers are published in International and National Journals and in conference proceedings. His areas of interest are Simulation modeling; Real time Database, Software Engineering, Networking, parallel and high speed computing and Mobile computing etc.

DR HASAN AL-SAEDY is working as The Director of Higher Education in the British Institute of Technology and E-Commerce; he obtained his MPhil and PhD degrees from Westminster University and Southampton University (UK). Both degrees were in computer Science in 1977 and 1981 respectively. He had worked in many universities in Middle East, Africa and UK. His research papers are published in International conferences proceedings and Journal. His areas of interest are Machine Learning and Internet Security applications

DR. S. C MEHROTRA, F.N.A,Sc., FIETE is working as a professor in Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (India). He received his master degree in Physics from Allahabad University (India) in 1970 and Ph.D. in Physics from Austin (USA) in 1975. He is recipient of Welch Foundation Fellowship (1975), Alexander Von Humboldt Foundation Fellowship (1983-85), FOM (Netherlands). He has published more than 150 papers in areas of Time Domain Reflectometry, Speech Processing, and Network Simulation



# **URBAN AND INTERMODAL TRAFFIC SYSTEMS**



# ANALYZING IN-FLIGHT AIR HOSTESS TASKS THROUGH DISCRETE-EVENT SIMULATION

Mehmet Savsar  
Esra E. Aleisa  
Department of Industrial and Management Systems Engineering  
Kuwait University  
P.O. Box 5969 Safat 13060, Kuwait  
[mehmet@kuc01.kuniv.edu.kw](mailto:mehmet@kuc01.kuniv.edu.kw); [aleisae@gmail.com](mailto:aleisae@gmail.com)

## KEYWORDS

Discrete-event simulation, in-flight, aircraft, air hostess, statistical validation, ANOVA

## ABSTRACT

Nowadays, in-flight comfort has become an essential characteristic to survive in domestic airline businesses. Air hostesses' quality of service and response time to customer demand plays a vital role in enhancing quality of the flight experience and passenger satisfaction. However, additional requirements on airhostess expose them to fatigue which could eventually result in job dissatisfaction that also will reflect onto the customer. This research studies the detailed tasks of airhostesses that take place from take-off to landing. And answers the following question: Can discrete-event simulation assist in assessing the performance measures of an airline hostess and can it improve them? Can the different activities scenarios developed through simulation provide a better organization of the multi-tasks that the airhostess provide to customers throughout the flight? This study is conducted on an airbus aircraft.

## INTRODUCTION

Airline companies are striving to improve their in-flight services to maintain sustainable competitive advantage in today's fiercely competing market. The role of the air hostesses plays a vital role in achieving this objective. In this study, we analyze the tasks and work procedures scheduled on typical domestic flights that are required from the air hostesses in an attempt to improve service levels without elevating the work burden on them. We use discrete-event simulation to model the air-hostess in-flight activities due to following advantages:

1. Simulation has the ability to identify system parameters such as utilization, flow-time and buffer sizes (Mosier 1989; Shafer and Meredith 1990; Aleisa 2005; Aleisa and Lin 2005; Savsar 1992)
2. Its ability to reveal potential problems and bottlenecks in proposed layout structures prior to implementation (Aleisa 2008)
3. Evaluate various strategies for the operation of the system (Pegden, Shannon et al. 1995; Aleisa 2008)

4. Compress or expand time, which gives the analysts the convenience to studying the layout in the long-run or under specific short-term scenarios.
5. Incorporate any stochastic behavior and uncertainty by incorporating the probability distribution that best describes the activity (Shafer and Charnes 1997)
6. Simulation model were used to generate random flow volumes to be subsequently supplied to traditional facility layout algorithms (Savsar 1991)

The aim of the simulation study is to achieve the best combination of the following objectives:

1. Reduce response time to passenger requests
2. Reduces bottlenecks in airplane corridors that often take place between airhostess serving passengers and the passengers themselves.
3. Reduce crew travel distance and time
4. Utilize the resources (hostesses) efficiently without overworking them.

Discrete-event simulation was applied requisitely in the areas of aircraft industry and airline business. Increased demand on air travel forced pioneers to investigate air-taxi services. The operations of such services were evaluated using discrete event simulation (Boyd, Bass et al. 2006; Consiglio, Williams et al. 2008). Such aircrafts were characterized with narrow bodies. Rijsenbrij and Ottjes (2007) have used discrete-event simulation to try different concepts of baggage transport to and from such narrow body aircrafts. Moreover, newly established security procedures have affected the business as usual tasks have grown in airport terminals. De Barros (2007) have used discrete-event simulation to evaluate the impact of newly established security measurements on the planning and operation of airport passenger terminals.

Manivannan and Zeimer (1996) have applied discrete-event simulation techniques to evaluate and improve the plane offloading operations in a central air cargo Hub. In addition, discrete-event simulation was applied to improve fleet maintenance schedules with respect to performance measures such as aircraft cycle time and mechanics labor utilization (Bazargan and McGrath 2003) and to improve maintenance operations of a fleet of fighter aircrafts in crises situations (Mattila, Virtanen et al. 2003). It was also applied

to test different scenarios to examine different aircraft policies in the battle field (Mishra, Batta et al. 2004).

The simulation model in this paper is applied to two classes: first/business class and economy class. After validation of the as-is system, improvements including automation, and other procedures will be suggested, modeled and compared with the as-is model. Finally the best improvements will be chosen through the application of ANOVA.

## AIRCRAFT AND AIR HOSTESS DUTIES

This research is conducted on the Airbus 300 (A300) as it is the most commonly used aircraft type in the Middle East. A300 accommodates 232 passengers, its length is 54.10 m and height is 16.54 m. A300 consists of four zones:

- Zone A (F/C) row from 1-3 18 seats
- Zone B (J/C) row from 6-8 18 seats
- Zone C (E/Y) row from 9-20 96 seats
- Zone D(E/Y) row from 21-33 100 seats

It also includes six lavatories including one for handicapped.

Before passengers enter the airplane many safety procedures and scheduling take place. First, a 1-hour briefing session takes place in the in-flight services department. In this session each crew member is assigned to specific tasks. Furthermore, the crew is tested orally on safety questions. A crew member failing to answer any of these questions is offloaded from the program, and sent back to retake a safety exam.

The crew then enters the airplane before the passengers arrive to construct routine checks on the inside of the airplane according to a checklist that they have. If the airplane is ready, the crew will give clearance to the chief, which will then report clearance to the aircraft captain in order to call for the passenger's boarding. The checklists are divided according to different classes, and the phase in which the services should be offered. Longer flights have different checklists with different services. The duties of an air hostess will differ according to class. The duties of an air hostess for the economy class are provided in Table 2.

## DATA COLLECTION

The data collection was conducted during actual flights that took place on airbus aircraft that traveled to various destinations. The data collection included the time required for scheduled in-flight air hostess tasks, as shown in Table 3 1 in minutes. In addition, the data collection included the inter-arrival times for passengers calling air hostesses for question, beverage, extra pillows and blankets, screen, chair problems and other types of assistance, and the time for an air hostess to serve a passenger. Moreover, inter-arrival times to lavatories and the times these lavatories were occupied was also collected. Also, the study considered times that the aisles were blocked by the food cart and impeded passenger service.

## MODEL ASSUMPTIONS AND DATA FITTING

The simulation study is assumed to take place during normal flight conditions. No emergency situations took place including the need for oxygen mask or CPR. Due to the randomness nature of data collected for passenger requesting air hostess assistance and using lavatories, the data were fitted into distributions using the Arena Input Analyzer statistical add-in. The data fitting is provided in Table 3 and the fitting graph for inter-arrival of service calls to economy class passengers is provided in Figure 1.

## THE SIMULATION MODELS

Two simulation models were created to mimic the activities conducted by air hostesses within the aircraft. The first model was for the First and Business classes (F/J) and the second was for the passenger class. The entities of the simulation system are passenger orders and tasks scheduled to the air hostesses. On the other hand the system resources are the air hostesses themselves and the aircraft lavatories.

For Airbus 300, there are three air hostesses of (F/J) classes and six for the economy class. As discussed earlier, the simulation model was carried assuming normal flight conditions with no emergencies. Handicapped were given priority of service within each class. The simulation model was created using Arena Rockwell Software v 11. A snapshot of the simulation model for the (F/J) class is shown in Figure 2. The economy class simulation has similar logic but is not shown in this paper to avoid repetition. The output of the simulation model is provided in Table 4.

Table 1: Average Air hostess in-flight tasks time in minutes

| Class                    | Wet towels | headsets | Immigration cards | Food cart | Beverage trolley | Tea  | Other Beverages |
|--------------------------|------------|----------|-------------------|-----------|------------------|------|-----------------|
| First and business class | 7.3        | 3.67     | 5.3               | 24.3      | 28.67            | 10.3 | 9.53            |
| Economy class            | 7.7        | 7        | 6.65              | 32.3      | 35               | 15.6 | 14.4            |

Table 4 shows that during a 4-hour flight on airbus 300 an airhostess was busy conducting 35 requests while 71 for economy class. Notice that the hostesses were able to satisfy all requests except for one in first class and 2 in economy. This was probably due to placing orders during times when services are not provided, such as take-off and landing. Also, the waiting time for service is 9.21 and 8.6 for F/J and economy classes respectively because many customers will place requests during times where hostesses are conducting longer scheduled tasks such as serving lunch or dinner to all aircraft passengers.

The utilization excludes times when air hostesses are forced to remain seated according to aviation regulations during take-off and landing.

Table 2: Detailed Air Hostess Duties in the Aircraft

|                |   |                        |  |
|----------------|---|------------------------|--|
| On Ground      | Complete checklist  | In-flight seat belt OF | Place baby cots, offer blankets, pillows & check toilets immediately after take-off. |
|                | Standard galley preparation   |                        | Offer children give-aways.   |
|                | Place juice, beverages, water on ice  |                        | Proceed with service as per procedure.   |
|                | Check cleanliness of cabin, seat pockets, hangers, headrest covers, open bins, outlets & window shades. |                        | Inspect cleanliness of toilets after each use and when needed.                       |
|                | Check pillows, blankets, baby cots, spares etc.   |                        | Prepare& distribute documentation & restow.  |
|                | Check cleanliness/ serviceability of toilets, place cosmetics & spray.                                  | Decent                 | Inform Purser of any U/S equipment.  |
|                | Check J/C headsets, J/C &E/Y & complete forms.  |                        | Clear up galleys, uplift via Purser if needed.                                       |
|                | Check toolkit, menu cards & complete forms.   |                        | Clear up cabin, restow pillows, blankets, etc...                                     |
|                | Prepare reading material & headsets.  | Seat Belt Sign On      | Distribute jackets.  |
|                | Prepare drinks trays with glasses.  |                        | Collect& restow toilet cosmetics.  |
|                | Conduct random check of seats serviceability.   |                        | Collect& restow menu cards cosmetics.  |
|                | Liaise with deck crew regarding their meals & offer beverages.  |                        | Seal dry stores, beverage trolleys, tool kit container &secure galleys.              |
| Boarding       | Welcoming, checking boarding cards, addressing pax by name, assisting in seating, coats& hand baggage.  |                        | Complete zone checks according to emergency zone.                                    |
| After Boarding | Distribute newspapers on lined trays.   | After Landing          | Give zone clearance to purser.   |
|                | Prepare & offer welcome drinks, collect glasses.  |                        | Take emergency stations for landing after passing clearance.                         |
|                | Prepare hot & cold face towels.   |                        | Distribute coats/belongings.   |
|                | Distribute face towels & collect.   |                        | Change mode selector upon announcement & cross check.                                |
|                | Close door when instructed.   |                        | Give clearance to purser.  |
|                | Change mode selector upon announcement, cross check.  |                        | Obtain clearance to open door.   |
|                | Offer dates/Arabic delights.  |                        | Assist passengers during disembarkation& farewell.                                   |
|                | Offer Arabic coffee & collect cups.   |                        | Collect headsets.  |
|                | Distribute headsets.  |                        | Check cabin for left behind pax belonging.   |
|                | Safety demonstration( special demo for blind, deaf)   |                        | Clear up cabin from CRDs , newspapers,   |
|                | Clear up cabin  |                        | Seal give aways containers & complete forms.   |
|                | Clear up& secure galleys.   |                        | Collect headsets, complete form& seal container.                                     |
|                | Complete zone clearance to zone responsible.  |                        | Seal newspaper bag.  |
|                | Give zone clearance to Purser   |                        | Check serviceability of toilets& inform Purser of any defect.                        |
|                | Take emergency stations for take-off  |                        | Collect personal belongings.   |

Table 3: Fitting random data into statistical distributions

| Task                                  | Distribution (min)       | Squared Error |
|---------------------------------------|--------------------------|---------------|
| Hostess call inter-arrival (F/J)      | TRIA(0.5, 5, 22.5)       | 0.012512      |
| Hostess call inter-arrival (Econ)     | -0.5+EXPO(4.51)          | 0.006046      |
| Hostess response time to calls (F/J)  | -0.5+5*BETA(7.91, 11.6)  | 0.001523      |
| Hostess response time to calls (Econ) | -0.5+7 *BETA(1.15, 1.83) | 0.003209      |
| Time in Lavatory (F/J)                | NORM(3.47, 0.882)        | 0.001538      |
| Time in Lavatory (Econ)               | NORM(4.56, 2.01)         | 0.011896      |

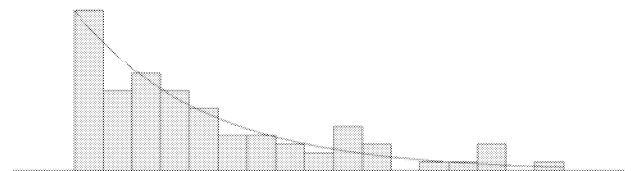


Figure 1: Data fitting of inter-arrival service call times for air hostess from passengers of the economy class

### SIMULATION MODEL VALIDATION

Statistical analyses were conducted to ensure that the model is a valid representation of reality. A t-test between the parameters of the real system and the simulation model are provided to in this section. The simulation was replicated 10

times and the results are shown in Table 5, where “R” denotes data collected from the real system and “A” denotes data collected from the simulated system in Arena. To accomplish validation, we need to compare two populations (the real and simulated), by drawing random samples from each population. Depending whether or not the sample sizes and variances are equal, different formulas need to be used.

Table 4: Output data from simulating the as-is system of in-flight airhostess activities (average values)

| Parameter                            | F/J class | Economy Class |
|--------------------------------------|-----------|---------------|
| Number of tasks served               | 35        | 71            |
| Waiting time (min)                   | 9.21      | 8.60          |
| Number orders tasks in hostess queue | 1         | 2             |
| Hostess utilization                  | 0.5110    | 0.6685        |

Let  $\bar{X}_i$ ,  $s_i$  and  $n_i$  indicate mean, standard deviation and sample size of sample  $i$  respectively. Sample values form the real system (1) for number answered calls by air hostess for the F/J class:

$$\bar{X}_1 = 25.83 \text{ calls; } s_1 = 3.25 \text{ calls; } n_1 = 6$$

Sample values form the simulated system (2) for number of answered calls by air hostess for F/J class:

$$\bar{X}_2 = 27 \text{ calls; } s_2 = 1.9 \text{ calls; } n_2 = 6$$

To conduct a statistically sound validation, the equality of two population variances needs to be verified, prior to checking the equality of means. At a 95 % confidence level we *fail to reject* that the two variances are unequal, therefore, the variance for the real and simulated system are pooled. The pooled standard deviation  $s_p=2.6615$ .

The  $t$ -distribution statistic will result in:

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{1/n_1 + 1/n_2}} = -0.76$$

The  $t$ -test 95% confidence intervals for  $m_1 - m_2$  is calculated as follows:

$$m_1 - m_2 \hat{=} \bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Substituting values in the above equation yields:

$$- 4.58 \leq m_1 - m_2 \leq 2.26$$

Also, the  $p$ -value =0.465 is larger than the significance level (1-0.95) which means that we *fail to reject* the hypothesis that means of the two populations are equal. In other words, the simulated model is a valid representation of reality.

The statistical validation procedure indicated above was repeated for all parameters provided in Table 5 for all classes F, J and Economy. The results are shown in Table 6. Again, the statistical analysis at a 95% confidence level indicates that the simulated system is a valid representation of reality.

## SIMULATION MODELS OF THE IMPROVEMENT SCENARIOS

By analyzing the production runs of the in-flight simulation model, we realized that several improvements could take place to improve the passenger service without elevated work burden on the air hostess. These improvements are based on possible automation some in-flight services that are listed below. These improvements were formulated as-is simulation model scenarios to be compared with the as-is system. Six scenarios were simulated:

**Scenario 1:** Providing the overhead passenger belongingness bins with a hydraulic closing mechanism that locks the bins automatically during take-off and landing.

**Scenario 2:** Passengers choose the meal from the menu using the screen on the seat in front of them.

**Scenario 3:** Allowing customer to place request for beverages from the touch screen. This allows air hostess to satisfy multiple orders at the same time.

**Scenario 4:** Allowing customers to preview whether or not the aisles are blocked the food cart. This attempts to facilitate both hostess and passenger movement during flight.

**Scenario 5:** Since the request for water is the most frequently task that an air hostess will answer, this scenario examines the case when bottled water is placed at every passenger's seat before passengers enter the airplane.

**Scenario 6:** Use an automatic inflating and deflating pillow that is attached to the passenger seat. This reduces the call for pillows which is also requested in a high rate.

**Scenario 7:** A combined scenario of all the previous 6.

## OUTPUT ANALYSIS OF THE IMPROVEMENT SCENARIOS

A simulation model was created for each of the seven scenarios discussed earlier. Each of which was replicated 10 times. The results for the F/J class are provided in Table 7, while those for the economy class are provided in Table 8. As shown in Table 8, combining the six scenarios have tremendously affected system performance measures. For instance, the average response time for calls for calls has reduced from 9.21 to 0.1 minutes for the F/J class and from 8.6 minutes to 0.1 minutes for the economy class too. In addition, the air hostess utilization was reduced by 12% in the F/J class and around 14% for the economy class. The number of class for hostess has been reduced due to the addition of inflatable pillows and bottled water.





Table 5: Comparisons of data from the real and simulated system over 10 replications for calls received by air hostess in the F/J class

| Run #       | Blanket  |      | Game        |      | Headsets    |      | Beverage    |      | Pillow      |       | TV          |      | Water     |       | Calls waiting time in queue |      | No. of calls waiting in queue |   |
|-------------|--|------|-------------|------|-------------|------|-------------|------|-------------|-------|-------------|------|-----------|-------|-----------------------------|------|-------------------------------|---|
|             | R  | A    | R           | A    | R           | A    | R           | A    | R           | A     | R           | A    | R         | A     | R                           | A    | R                             | A |
| 1           | 3  | 5    | 1           | 0    | 0           | 2    | 2           | 3    | 0           | 2     | 2           | 4    | 14        | 9     | 5                           | 8    | 0                             | 1 |
| 2           | 5  | 2    | 4           | 3    | 1           | 0    | 7           | 5    | 2           | 1     | 3           | 5    | 12        | 15    | 8                           | 8    | 0                             | 1 |
| 3           | 3  | 6    | 2           | 1    | 1           | 3    | 6           | 3    | 1           | 0     | 5           | 2    | 8         | 10    | 11                          | 8    | 1                             | 1 |
| 4           | 2  | 4    | 1           | 4    | 2           | 1    | 3           | 4    | 1           | 2     | 5           | 2    | 15        | 13    | 10                          | 9    | 2                             | 1 |
| 5           | 6  | 1    | 3           | 2    | 0           | 1    | 2           | 6    | 3           | 2     | 1           | 0    | 11        | 15    | 9                           | 7    | 1                             | 1 |
| 6           | 3  | 2    | 0           | 4    | 2           | 1    | 3           | 3    | 1           | 1     | 0           | 2    | 5         | 15    | 6                           | 10   | 1                             | 1 |
| 7           | 1  | 6    | 0           | 0    | 3           | 1    | 4           | 5    | 1           | 2     | 0           | 2    | 6         | 11    | 12                          | 11   | 0                             | 1 |
| 8           | 1  | 3    | 1           | 5    | 1           | 2    | 4           | 5    | 1           | 2     | 2           | 1    | 10        | 5     | 10                          | 11   | 1                             | 1 |
| 9           | 4  | 6    | 2           | 2    | 1           | 2    | 3           | 2    | 2           | 1     | 4           | 5    | 13        | 7     | 8                           | 7    | 2                             | 1 |
| 10          | 4  | 5    | 11          | 2    | 3           | 0    | 4           | 6    | 2           | 1     | 4           | 0    | 10        | 15    | 8                           | 12   | 2                             | 1 |
| Mean        | 3.2  | 4.0  | 1.5         | 2.30 | 1.4         | 1.3  | 3.80        | 4.20 | 1.400       | 1.400 | 2.60        | 2.30 | 10.40     | 11.50 | 8.70                        | 9.10 | N/A                           |   |
| StDev.      | 1.6  | 1.8  | 1.27        | 1.70 | 1.1         | 0.95 | 1.62        | 1.40 | 0.843       | 0.699 | 1.90        | 1.83 | 3.69      | 3.31  | 2.16                        | 1.79 |                               |   |
| SE-Mean     | 0.51   | 0.60 | 0.4         | 0.54 | 0.3         | 0.30 | 0.51        | 0.44 | 0.27        | 0.22  | 0.58        | 0.60 | 1.0       | 1.2   | 0.57                        | 0.68 |                               |   |
| CI on Diff. | -0.85, 2.45  |      | -0.62, 2.22 |      | -1.05, 0.85 |      | -1.02, 1.82 |      | -0.73, 0.73 |       | -2.05, 1.45 |      | -2.2, 4.4 |       | -1.47, 2.27                 |      | N/A                           |   |
| Key :       | R: Data from the real system      A: Data from the simulated system of the Arena model |      |             |      |             |      |             |      |             |       |             |      |           |       |                             |      |                               |   |

Table 6: Statistical validation between the real and simulated systems using the sample t-test ( $\alpha=0.05$ ) over 10 replications

| Parameter | $X_1$<br>(real) | $X_2$<br>(simulated) | $m_1, m_2$    | $P$ -value |
|-----------|-----------------|----------------------|---------------|------------|
| Blankets  | 4.0             | 3.2                  | (-0.85, 2.45) | 0.32       |
| Games     | 2.3             | 1.5                  | (-0.62, 2.22) | 0.25       |
| Headsets  | 1.3             | 1.4                  | (-1.05, 0.85) | 0.83       |
| Beverage  | 4.2             | 3.8                  | (-1.02, 1.82) | 0.59       |
| Pillow    | 1.4             | 1.4                  | (-0.73, 0.73) | 1          |
| TV        | 2.3             | 2.6                  | (-2.05, 1.45) | .72        |
| Water     | 11.5            | 10.4                 | (-2.2, 4.4)   | .49        |
| Wait in Q | 9.1             | 8.7                  | (-1.47, 2.27) | 0.66       |

## ANALYSIS OF VARIANCE

We applied the analysis of variance (ANOVA) to statistically identify the most influential factor in improving the performance measures of the system. Each of the previously discussed scenarios is modeled as an ANOVA factor. ANOVA analysis was conducted for both F/J and economy classes and with respect to all performance measures. ANOVA was conducted using Minitab package. The Minitab results for the F/J class with respect to order waiting time and air hostess utilization is shown in Figure 3.

Table 7: The simulation model output of the as-is model and the seven improvement scenarios for the F/J class

| F/J Class                               | As-is-Model | 1 <sup>st</sup> Scenario | 2 <sup>nd</sup> Scenario | 3 <sup>rd</sup> Scenario | 4 <sup>th</sup> Scenario | 5 <sup>th</sup> Scenario | 6 <sup>th</sup> Scenario | 7 <sup>th</sup> Scenario |
|---|-------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Number of calls requested               | 27          | 27                       | 27                       | 27                       | 27                       | 25                       | 24                       | 25                       |
| Number of calls satisfied               | 27          | 27                       | 27                       | 27                       | 27                       | 25                       | 23                       | 25                       |
| Waiting time in calls queue (min)       | 9.21        | 6.7                      | 8.95                     | 8.69                     | 6.3                      | 9.8                      | 9                        | 0.0909                   |
| Number of Orders waiting in calls queue | 1.008       | 0.73                     | 0.97                     | 0.94                     | .067                     | 0.99                     | 0.84                     | 0.5519                   |
| Hostess Utilization                     | 0.5110      | 0.4801                   | 0.4890                   | 0.4910                   | 0.4703                   | 0.5040                   | 0.5010                   | 0.4507                   |

Table 8: The simulation model output of the as-is model and the seven improvement scenarios for the economy class

| Economy Class                           | As-is-Model | 1 <sup>st</sup> Scenario | 2 <sup>nd</sup> Scenario | 3 <sup>rd</sup> Scenario | 4 <sup>th</sup> Scenario | 5 <sup>th</sup> Scenario | 6 <sup>th</sup> Scenario | Combined Scenario |
|---|-------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|-------------------|
| Number of calls requested               | 63          | 63                       | 63                       | 62                       | 63                       | 41                       | 64                       | 43                |
| Number of calls satisfied               | 63          | 63                       | 63                       | 62                       | 63                       | 40                       | 64                       | 43                |
| Waiting time in calls queue (min)       | 8.6         | 6.462                    | 8.59                     | 8.02                     | 5.83                     | 8.09                     | 8.52                     | 0.0982            |
| Number of Orders waiting in calls queue | 2.15        | 1.62                     | 2.148                    | 2.08                     | 1.46                     | 1.31                     | 2.20                     | 1.0058            |
| Hostess Utilization                     | 0.66        | 0.64                     | 0.62                     | 0.63                     | 0.62                     | 0.63                     | 0.66                     | 0.5689            |

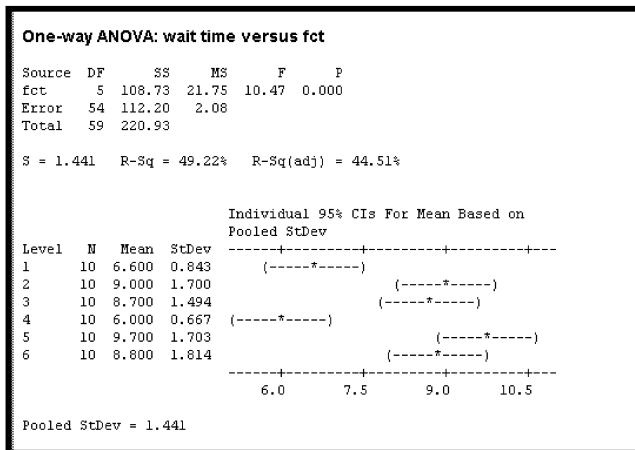


Figure 3: Minitab ANOVA output for comparing scenarios with respect to calls waiting time in queue for the F/J class

As shown in Figure 3 the  $p$ -value is very small, which indicates that there is significant difference between the effects of each factor (scenarios in this case) with respect to waiting time. In the same figure, the multiple comparison diagrams show that the first and fourth factors have the most influence since they have the least values for waiting time. However, since very strong interaction occurs between the two factors, they are both candidates for being the best factors or scenarios.

The Tukey's test shown in Figure 4 gives further analysis related to the interaction. Tukey's test is an efficient statistical method to conduct pairwise comparisons among the means of all factors in the ANOVA design. Tukey's test proves the strong interaction between scenario one and four. However, it shows that the fourth scenario has a slightly stronger effect.

Similar analysis is conducted with respect to the air hostess utilization. The results are provided in Figure 5. Again, the  $p$ -value equals to zero, which indicates significant effect between the factors. The analysis implies that the first scenario is the most effluence factor in this case.

The preceding analyses were repeated for the air hostesses of the economy class but were omitted here to avoid repetition.

## CONCLUSIONS

In this research we have conducted a discrete event simulation study for the activities of air hostesses during aircraft flights. The aim was to improve passenger service without increasing workload burden on air hostesses. Several improvement scenarios were suggested and simulated. The simulation models were statistically validated and compared to real data collected during actual flights. The outputs of the improved scenarios were statistically verified and modeled using ANOVA. The study has resulted in improved service and less fatigue on air hostesses in first, business and economy classes. The results could be extremely useful for airlines that want to improve their in-flight service operations.

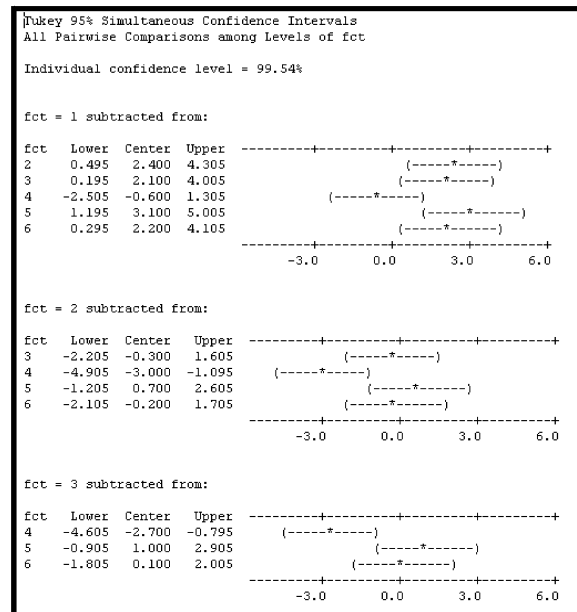


Figure 4: Tukey's pairwise comparison test of the effect of simulation scenarios with respect to waiting time in the F/J class

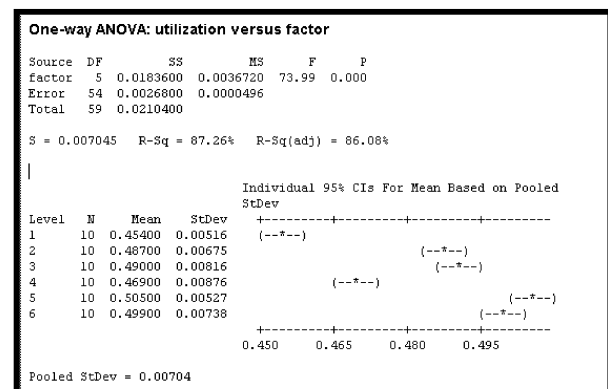


Figure 5: Minitab ANOVA output for comparing scenarios with respect to hostess utilization for the F/J class

## REFERENCES

- Aleisa, E. 2005. "Multilevel Integration of Simulation and Facilities Planning for Large-Scale Systems". *Department of Industrial Engineering* The State University of New York at Buffalo, Buffalo, NY,
- Aleisa, E. 2008. "Developing Efficient Testing and Unloading Procedures for a Local Sewage Holding Pit". *International Journal of Computer, Information, and Systems Science, and Engineering* 2, No 3, 138-145.
- Aleisa, E. 2008. "An Overview Of Multilevel And Hierarchical Methods For Discrete-Event Simulation Of Complex Systems". In *Industrial Simulation Conference (ISC08)*, J. Colloc, C. Petit and C. Dussart. Lyon, France, 249-255
- Aleisa, E. and L. Lin 2005. "For effective facilities planning: layout optimization then simulation, or vice versa?". *Proceedings of the*

- 37th conference on Winter simulation Winter Simulation Conference, Orlando, Florida, 1381-1385
- Bazargan, M. and R. N. McGrath 2003. "Discrete event simulation to improve aircraft availability and maintainability". *Annual Reliability and Maintainability Symposium, 2003 Proceedings* No 63-67.
- Boyd, J. A., E. J. Bass, S. D. Patek and D. W. Lee 2006. "Investigation of future air taxi services using discrete-event simulation". *2006 IEEE Systems and Information Engineering Design Symposium* No 25-30.
- Consiglio, M. C., D. M. Williams, J. L. Murdoch and C. A. Adams 2008. "Concept validation experiment for small aircraft transportation system higher-volume operations". *Journal of Aircraft* 45, No 2, 359-365.
- de Barros, A. G. and D. D. Tomber 2007. "Quantitative analysis of passenger and baggage security screening at airports". *Journal of Advanced Transportation* 41, No 2, 171-193.
- Savsar, M. (1990), "Flexible Facility Layout by Simulation", *Computers and Industrial Engineering*, Vol. 20, No. 1, 155-165.
- Savsar, M. (1992), "Maintenance Crew Size Determination by Computer Simulation", *Journal of King Saud University Engineering Sciences*, Special issue on Maintenance, Vol. 4, 45-66.
- Manivannan, S. and M. Zeimer 1996. "Modeling and analysis of aircraft offloading operations". *1996 Winter Simulation Conference Proceedings* No 1359-1366.
- Mattila, V., K. Virtanen and T. Raivio 2003. "A simulation model for aircraft maintenance in an uncertain operational environment". *Esm 2003: 17th European Simulation Multiconference* No 456-461.
- Mishra, S., R. Batta and R. J. Szczerba 2004. "A rule based approach for aircraft dispatching to emerging targets". *Military Operations Research* 9, No 3, 17-30.
- Mosier, C. T. 1989. "Experiment investigating the application of clustering procedures and similarity coefficients to the GT machine cell formation problem". *International Journal of Production Research* 27, No 10, 1811-1835.
- Pegden, C. D., R. E. Shannon and R. P. Sadowski 1995. *Introduction to simulation using SIMAN*. McGraw-Hill, New York.
- Rijssenbrij, J. C. and J. A. Ottjes 2007. "New developments in airport baggage handling systems". *Transportation Planning and Technology* 30, No 4, 417-430.
- Shafer, S. M. and J. M. Charnes 1997. "Offsetting lower routing flexibility in cellular manufacturing due to machine dedication". *International Journal of Production Research* 35, No 2, 551-567.
- Shafer, S. M. and J. R. Meredith 1990. "A Comparison of Selected Manufacturing Cell-Formation Techniques". *International Journal of Production Research* 28, No 4, 661-673.

## AUTHOR BIOGRAPHY

**MEHMET SAVSAR** is professor and chairman of the Industrial & Management Systems Engineering Department at Kuwait University. Prof. Savsar received his B.Sc. degree from Black Sea Technical University in Turkey, his M.Sc. and PhD. Degrees from the Pennsylvania State University, PA, USA in the areas of Industrial Engineering and Operations Research. He has been with Kuwait University since 1997. Prior to joining Kuwait University, he has worked as a faculty member in Turkey and in Saudi Arabia and as a researcher at Pennsylvania State University, USA. Prof. Savsar has taught a variety of courses in the areas of Production Planning and Inventory Control, Quality Control, Maintenance and Reliability, Operations Research, Stochastic Processes, Computer Simulation, Facilities Planning and Layout Design,

Cost Analysis, and Manufacturing Systems. His research interests include: Modeling of Production Systems, Plant Layout, Reliability and Maintenance, Flexible Manufacturing Systems, Quality Control, and Just-In-Time (JIT) Production Control. He has over 100 publications in refereed international journals and international conference proceedings. He serves in editorial boards of several international journals and is a referee to several journals, including EJOR, IJAMT, Simulation, Int. J. of Syst. Science, Int. J. of Prod. Econ., Int. J. of Prod. Research. He is a senior member of IIE, INFORMS.

**ESRA ALEISA** is an assistant professor of Industrial and Management Systems Engineering (IMSE) at Kuwait University. She has received her Masters and PhD in Industrial Engineering and production systems from the department of Industrial Engineering at the State University of New York at Buffalo in 2001 and 2005 respectively. In 1998, she has earned her B.S. degree in industrial engineering from Kuwait University.

Dr. Esra Aleisa research interests include, Planning and design of large scale facilities, simulation and improvement of manufacturing and service systems, multilevel planning and design of complex engineering design, group technology (GT), and design structured matrices (DSM). She is a member of Omega Rho, the international operations research honor society, IEEE, INFORMS, IIE, ASEE.

# A SOFTWARE ENVIRONMENT FOR MICROSCOPIC PEDESTRIAN SIMULATION

Edgar F. Esteves, Rosaldo J. F. Rossetti, Eugénio C. Oliveira  
Department of Informatics Engineering (DEI)  
Artificial Intelligence and Computer Science Lab (LIACC)  
Faculty of Engineering, University of Porto (FEUP)  
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal  
E-mail: {edgar.esteves, rossetti, eco}@fe.up.pt

## KEYWORDS

Pedestrian modelling and simulation, multimodal transport exchange interfaces, multi-agent systems, OpenSteer.

## ABSTRACT

This paper reports on first steps and preliminary results of the implementation of a platform for pedestrian simulation in multimodal exchange interfaces. Whereas most approaches in pedestrian simulation are aimed at studying interactions among people using a common environment with different purposes and under different mobility restrictions, our research is focused on how those interactions will ultimately affect transportation operation at multimodal stations. This paper begins with a general overview of pedestrian simulation and focus on the microscopic representation of individuals and their interoperability as the basis for the implementation of our model.

## INTRODUCTION

Nowadays the prediction of the influence of urban planning to people comfort and mobility inside public buildings is becoming a very important process.

Several studies were carried out in the last years as an attempt at predicting these influences. In this perspective modern pedestrian simulators become one of the most important tools to help one in achieving it. These tools have the objective of simulating and predicting pedestrian behaviours, as well as people or crowd flow in normal or in emergency conditions. Their main focus is on people movement inside public buildings such as train stations, shopping malls, airports and many others. The rising number of people attending certain public buildings and building restructuring makes the investigation of pedestrian movement very important in order to insure measures that will improve building capacity and better comfort to people using public building services by improving their mobility. For example, in a train station it can be very useful to know what kind of changes can be made in order to improve efficiency in interchange operations or even how many ticket boxes or ATM machines are necessary to reduce client queues. Another situation where pedestrian simulation can be very useful is to the design of emergency evacuation plans.

With the intention of modelling and simulating pedestrian crowd behaviours, it is often used two distinct levels of

pedestrian analysis: macroscopic level, which examines the average characteristics of pedestrian flows; and microscopic level, which analyses the movement and behaviour on an individual basis. This paper aims to clarify the reader about the types of microscopic simulation often used and its place in the overall picture of pedestrian studies. It also aims to provide a basic idea of how to conceptualise such applications.

Throughout this paper we have analysed the related literature and some of the most recent platforms and tools for this kind of simulation. Then we present a proposal for the architecture of a microscopic simulation platform, the preliminary results, as well as final remarks and conclusions.

## PEDESTRIAN MODELLING AND SIMULATION

### Basic Approaches

Pedestrian modelling studies started to gain relevance since a few decades ago. The main knowledge of pedestrian traffic systems comes from empirical and observation studies that can be consulted in greater detail in (Helbing et al. 2002). According to (Teknomo 2002) the pedestrian studies can be divided into pedestrian data collection and pedestrian analysis as illustrated in Figure 1. The data collection focuses its attention on the observation and compilation of pedestrian movement whereas pedestrian analysis is concerned with interpretation of the collected data, both to understand an observed situation and to plan enhanced management policies (e.g. urban planning).

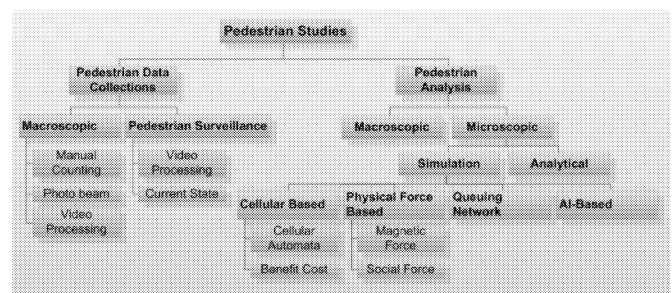


Figure 1: Pedestrian Studies (Based on Teknomo 2002)

As illustrated in the Figure 1, these studies have an analogy to vehicular traffic studies (Ferreira 2008) since they can also

be divided into two categories, namely the macroscopic and microscopic abstractions (Kl pfel et al. 2000) (Ke fel et al. 2002) (Shao 2006). The former compares the pedestrian flows with gases or liquids and studies the flow on a global basis whereas the latter studies rather focus on the individual speed and interactions of pedestrian crowds.

In this work, we focus our attention on the Microscopic Simulation Analysis Model (MSAM) that can be categorised into four distinct types, namely Cellular Based Models, Physical Force Based Models, Queuing Network and AI-Based Models (Helbing et al. 2002). An extended discussion on these approaches can be found elsewhere (Esteves et al. 2009). We have opted to follow AI-based models, namely those based on multi-agent systems (MAS), whose basic units of activity are agents (Schelhorn et al. 1999) able to autonomously steer moveable objects by applying behaviourally determined forces to their point of mass.

Reynolds (1999) and Buckland (2005) suggest several basic steering behaviours that if combined can produce more advanced behaviours, as summarised in Table 1. These behaviours were compiled and implemented into an open source library named OpenSteer (Reynolds 2004).

Table 1: Suggested Steering Behaviours

| Steering Behaviours |                      |                               |                  |
|---------------------|----------------------|-------------------------------|------------------|
| Seek                | Flee                 | Pursuit                       | Evasion          |
| Offset Pursuit      | Arrival              | Obstacle Avoidance            | Wander           |
| Path Following      | Flow Field Following | Unaligned Collision Avoidance | Separation       |
| Flocking            | Cohesion             | Alignment                     | Leader Following |

### Microscopic Simulation Platforms and Tools

We have carried out the study of different simulation platforms and tools so as to identify common characteristics and requirements. Analysing Table 2, it is possible to distinguish between two types of applications, namely the black-box ones (in this specific case they are all commercial applications with the exception of Micro-PedSim) and the open-source applications.

Table 2: Microscopic Simulation Platforms & Tools Analysis

| Software     | Open Source Software | Model         | Runtime Simulation | Emergency Conditions | 3D Visualization | Scenario Edition |
|--------------|----------------------|---------------|--------------------|----------------------|------------------|------------------|
| STEPS        | no                   | Agent Based   | yes                | yes                  | yes              | yes              |
| VISSIM       | no                   | Social Forces | yes                | yes                  | yes              | yes              |
| SimWalk      | no                   | ***           | yes                | yes                  | yes              | yes              |
| Legion       | no                   | ***           | yes                | yes                  | yes              | yes              |
| Micro-PedSim | no                   | Agent Based   | yes                | yes                  | yes              | yes              |
| AnyLogic     | no                   | Agent Based   | yes                | yes                  | yes              | yes              |
| SimPed       | no                   | Agent Based   | yes                | ***                  | ***              | yes              |
| UAF          | no                   | Agent Based   | yes                | yes                  | yes              | yes              |
| PedSim       | yes                  | ***           | yes                | yes                  | ***              | no               |
| OpenSteer    | yes                  | Social Forces | yes                | no                   | yes              | no               |

The former set was only analysed in a higher level of detail so as to gather some of the most relevant functional requirements for this type of environments. On the other

hand, the latter set was analysed in more details (e.g. architecture, modularity and performance), which was easily accomplished by analysing their source code.

So, focusing on the two last applications, the evaluation of low level parameters helped us to decide whether to use the tool as is or to adapt it by aggregating the best characteristics of both. Although the OpenSteer platform is not a pedestrian simulator, it offers a kinematic library that favours the creation of heterogeneous behaviours. Therefore, an approach that uses the most significant features of OpenSteer seemed to be a very reasonable solution for us.

## SYSTEM DEVELOPMENT

### The System Architecture

Figure 2 illustrates the conceptualised system architecture, composed of six distinct modules: the simulation editor (SEM), the model (MM), the simulation engine controller (SEC), autonomous agents (AA), 3D visualization tool (3DV), and the simulation analysis functions (SAF).

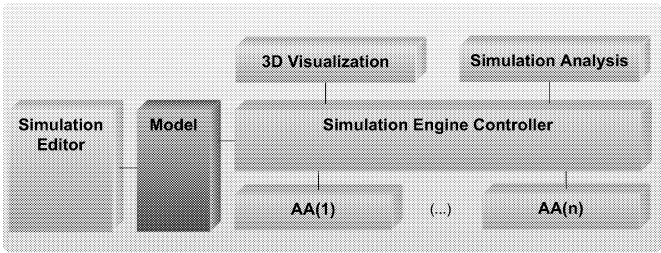


Figure 2: System Architecture Overview

The SEM can be considered as informational and manipulation software (Bret 2006) and features the interaction environment that interfaces the application and end-users. It allows the model to be easily handled by the user then, from model construction to parameters configuration. It also offers the appropriate visualisation of operations performed. Some challenges to the implementation of this module are to favour an easy integration of functionalities and to implement a WYSIWYG-based (What You See Is What You Get) editor. One imperative component of which must be focused, however, is its scenario builder component (SBC) that integrates capabilities such as to build scenarios (from very basic to more complex ones) in a structured way, from scratch. Through WYSIWYG, structuring scenarios follows a very intuitive approach, with all elements resembling their counterparts in real world (e.g. double lines for walls, single lines for sidewalks, icons for Points of Interest (PoI), and so on). Zones definition is another important feature of this component, as they basically define areas through which pedestrians can actually walk. Apart from pedestrians, all objects placed in zones will belong to the specific zone to which the object has been attached.

The MM is an essential module that serves the purpose of integrating in an associative way both the simulation model and the real-world model. Basically, parameters are

manipulated in such a way a reference model is always kept for further consultation. This module must support an easy creation and definition of new scenarios, and also allow one to navigate among different scenarios, in order to identify appropriate comparison points. Its structure contains the simulation scenarios, as well as the rules for some other components to work properly, according to the type of simulation one intends to carry out.

The SECM, on the other hand, is responsible for guaranteeing synchronization, coordination and rule compliance throughout the simulation process, among all modules of the proposed architecture. It also integrates two important components, namely the 3DV and the SAF. As for the SECM operation, four essential management elements must be present, namely the model data manager (MDM), the plug-in manager (PIM), the agent registry manager (ARM) and the simulation manager (SM).

The MDM is responsible for managing, in a structured way, all information related to the model being simulated. To improve efficiency, we apply to the structure of the model the special queries structure proposed in (Reynolds 2000). It also interacts with the ARM whenever a new agent must be registered in the simulation scenario. The PIM, on the other hand, allows plug-ins to be incorporated into the simulation environment. Its basic function is to manage the plug-ins and guarantee their proper use according to the type of simulation being carried out. The ARM is one of the most important features of the SECM, as it is responsible to manage all agents being simulated in a certain scenario. Whenever agents are transferred from zone to zone, the ARM feature guarantees agent parameters will be updated accordingly. Finally, the SM feature is responsible for the simulation whole process, including synchronisation and coordination among the various other components that compose the system architecture.

As for the agent concepts used in this work, we based our approach on the work presented in (Reynolds 1999) and (Buckland 2005), in which agents are regarded as entities embodied within a moveable structure which they control according to their reasoning capabilities. Therefore, agents are autonomous to decide where to go and which path to follow in order to reach their destinations. To implement the reasoning kernel of an agent, we adopted the BDI (beliefs, desires and intentions) architecture, initially proposed by Rao and further discussed in (Kinny et al. 1996). This is one of the most popular architectures for cognitive agents, which has been applied to the domain of traffic and transports and other applications elsewhere (Rossetti et al. 2002) (Ronald and Sterling 2005).

### The Environment Representation Model

In this section we discuss on the approach adopted for the description of a general scenario meta-model, which is instantiated accordingly whenever a new scenario is built within the simulation environment. The meta-model was developed with special account of requirements for pedestrian simulation in multimodal transport interfaces. Thus, the elements involved in any public transport exchanging area

were considered in the conceptualisation of this model, which is depicted in Figure 3, through a UML class diagram.

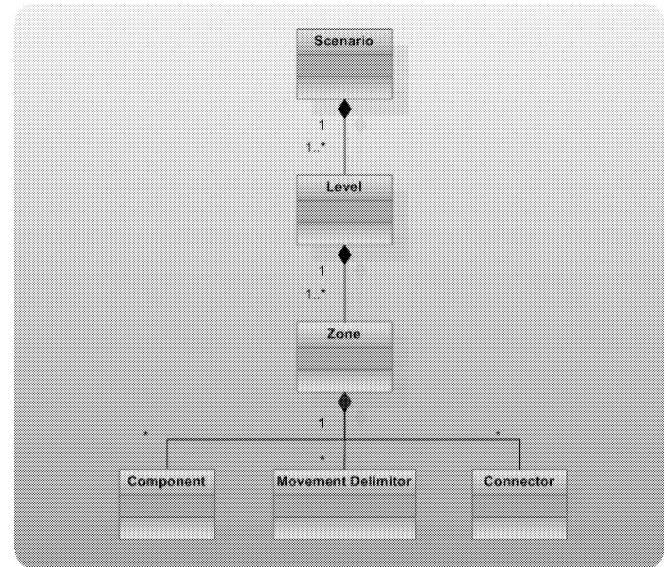


Figure 3: Buildings Case Scenario Architecture

The **movement delimiters** are physical and rigid structures basically used to build the contour of the areas through which agents can walk. Examples of such elements are the walls and sidewalks, for instance. Walls are obstacles that avoid agents to transpose the area in which they are currently situated (for instance, it will avoid an agent to go from one room to another). On the other hand, sidewalks can be used to lay out desired paths we want agents to follow (as in open areas, such as streets, gardens, etc.)

The **components/obstacles** are virtually all the other objects one can find situated in diverse environments and can be understood as physical barriers, imposed by the intrinsic structure of certain components that oblige pedestrians to go around.

Any area that can be delimited by movement delimiters are regarded as **zones** and can contain in their structure any other components. The **connectors** are to connect them, allowing pedestrians to move across several zones. The **levels**, in turn, aggregate zones in virtually independent environments, but yet can such an abstraction be connected to other levels. This is an important element of the model, especially when larger areas are being represented.

The **Points of Interest (PoI)** generalise all components that are necessary as an aid, or sometimes even imperative, for an agent to achieve its final goal. Distinguishing between being essential and only being an aid will certainly result from the agents' decision-making process. On the other hand, the **Flow Controllers (FC)** are the generalisation of all components that propitiate people coming in and going out of the simulated environment. Examples of such components would include, for instance, main entrance doors, escalators, stairs to/from different levels, as well as buses, trains and taxis). These components implement dispatching and draining

functions that must be calibrated according to the purpose of the simulation study being carried out.

Despite simple, the combination of these elements can result in very complex scenarios.

## PRELIMINARY RESULTS

A prototype was developed in C++ using the Qt framework that follows the main concepts of the OpenSteer architecture. The environment then is well suited to run pedestrian microscopic simulations in a continuous space environment. A screenshot of the editor application is depicted in Figure 4.

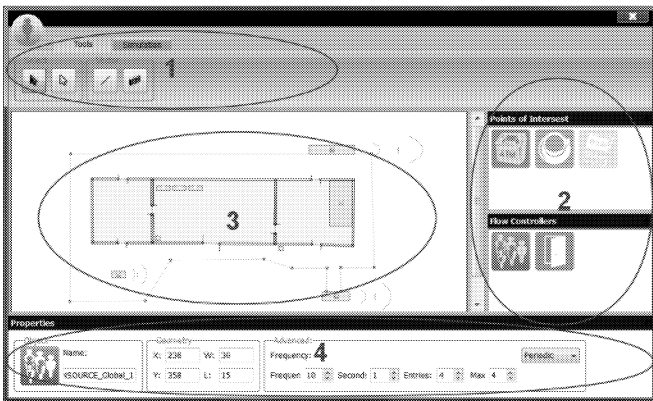


Figure 4: Simulation Editor Application

There are four zones exemplifying the most important components of this application, namely the ToolBarMenu (zone 1), the ComponentDockWidget (zone 2), EditionWidget (zone 3), and the PropertiesWidget (zone 4).

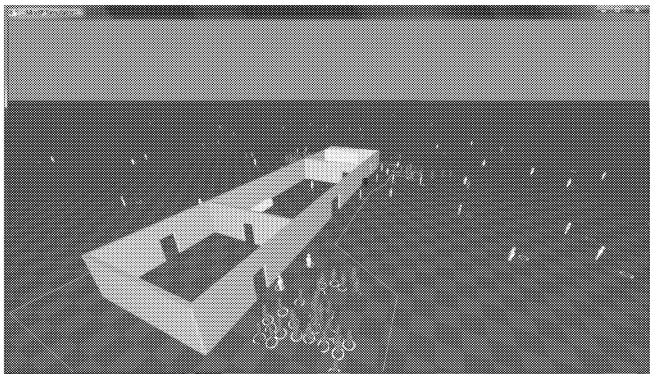


Figure 5: 3D Simulation Visualization

The simulation application provides a 3D view of the environment, using the OpenGL technology integrated into the Qt framework. In Figure 5, we present a screenshot of a simulation being executed. The scenario was built through the editor module mentioned above and is the same as the model depicted in Figure 4.

The pedestrian behaviour model used is based on a stochastic plan allocated on a stack that is dynamically adjusted according to the agents' basic needs. A probability

of choosing each of the attraction components of the scenario is assigned to each of the agents in the synthetic population. During the simulation and after the achievement of each task, a new one is assigned to the agent, according to the pre-defined task plan.

The prototype is currently in a stable version and has a very intuitive interface. Nevertheless it still does not allow setting up the objectives/rules of pedestrian behaviour in runtime. Since this feature is not yet implemented in this prototype and for us to obtain some preliminary experimental results, the following simulation parameters were defined in the generation of the agent population used in the simulation runs of this experiment:

- 20% of the created pedestrians will search and go directly toward an ATM machine.
- 50% of the created pedestrians will search and go directly toward a ticket machine.
- 30% of the created pedestrians will search and go directly toward an exit.

The pedestrians are also designed to go directly to a ticket office whenever they leave an ATM machine. After leaving the ticket office, they will go directly to an exit point. The ticket office and ATM machines were configured to allow a queue of five people. When this parameter reaches the defined limit, the pedestrians that wish to use the component will try to find one with a queue shorter than the imposed limit. If the pedestrian wishes to go to an ATM machine and all the ATM machines have a queue with five people, then it will wander the scenario until it finds one with a smaller queue too.

To better analyse the results of the simulation, a statistic component was implemented that draws two plots that represent the occupancy rates of the components and identifies the environment areas that have high occupancy rates. Simulation runs were carried out and the obtained results are depicted in Figure 6, relative to one hour of simulation. The scenario represents a room with an entrance, with three ticket offices, an ATM machine and two exit points. Some obstacles are also included.

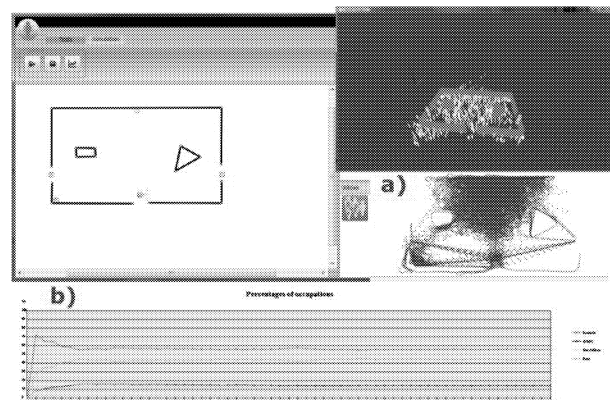


Figure 6: Example Scenario with Obstacles



Through the analysis of Figure 6, it is possible to obtain some basic conclusions. Analysing the plot that represents the occupancy rate of the environment spaces (represented by  $a$  in the image), it is possible to observe that the zones with a higher occupancy rate are the areas where the ticket offices and the ATM machines are located. It is also possible to observe by analysing the component occupancy rates (represented by  $b$  in the image) that the occupancy rate of the components (yellow for the ticket office and blue for ATM machines) stabilises after a while and approaches the inputted parameter values, freeing the exit points considerably.

This kind of analysis may help practitioners to understand if the scenario components are distributed well and if their numbers are enough to meet demand and the desired occupancy rates. Figure 6 also helps us to understand the effect that the number of components and their placement can have upon the occupancy rate inside the scenario.

From the occupancy zone plot and the occupancy rate of the components, it is possible to conclude that a single ticket office is not enough to deal with the amount of people trying to use it. This fact, associated with the inclusion of obstacles near the scenario exit points, creates a high occupancy rate in the entire scenario meaning that the exit rate after an hour of simulation is lower than the people entry rate (making the room completely crowded). Of course these are experimental results only but, despite the simplicity of the scenarios, they are good enough to demonstrate the potential of the proposed approach and framework prototype.

## CONCLUSIONS

This kind of approach for the development of a microscopic pedestrian simulation platform has proved reliable, easy to implement and with some interesting results. Only part of the OpenSteer library was actually used in this work. To further improve and extend the presented framework, we intend to make full use and profit from all advantages of the entire kinematic library. As for the environment representation, it is important to bear in mind the relevance of the objects abstraction if the biggest concern is to ensure the reuse of the implementation of the steering behaviours and of the pedestrian decision-making processes, in scenarios both with similar and with different characteristics. Also, another useful scenario component that might play a key role to turn pedestrian behaviours into more heterogeneous characters is the "Connector". In fact, if agents can regard connectors semantically, they will be able to make better decisions on their destinations and routes to follow. Finally, and really interesting is the particularity of the system to include a semantic editor. This can bring about the potential development of completely different simulations. That is, instead of defining semantics as referred before, which are almost exclusive to train stations, the user can choose to set up geographical information related to roads. From that point on, a vehicular traffic simulation looks more plausible. In addition to that, with similar scenarios and some visual changes, the convergence of such different kinds of

simulations (namely traffic and pedestrians) becomes even stronger and more feasible.

## ACKNOWLEDGEMENTS

Authors greatly acknowledge the kind support and contributions given by Prof. Álvaro Costa, from DEI/FEUP, as well as by Oana Santos and Vera Ferreira, from TRENMO Eng. Ltd. This project has been partially supported by RCM/Agência de Inovação.

## REFERENCES

- Bret, V. (2006). Magic Ink: Information Software and the Graphical Interface. Available on <http://worrydream.com/MagicInk/> (last accessed May 13, 2008).
- Buckland, M. (2005). Programming Game AI by Example. Plano, TX: Wordware Publishing, Inc.
- Esteves, E.F., Rossetti, R.J.F., Ferreira, P.A.F., Oliveira, E.C. (2009) Conceptualization and implementation of a microscopic pedestrian simulation platform. In ACM symposium on Applied Computing. p.2105-2106.
- Ferreira, P. (2008). Specification and Implementation of an Artificial Transport System. Master's Dissertation, FEUP, Portugal.
- Helbing, D., Farkas, I. J., Molnár, P., & Vicsek, T. (2002). Simulation of Pedestrian Crowds in Normal and Evacuation Situations. In Pedestrian and Evacuation Dynamics. p.21-58.
- Keßel, A., Klüpfel, H., Wahle, J., & Schreckenberg, M. (2002). Microscopic simulation of pedestrian crowd motion. In Pedestrian and Evacuation Dynamics. p.193-202.
- Kinny, D., Georgeff, M., & Rao, A. (1996). A Methodology and Modelling Technique for Systems of BDI Agents. In 7th European Worksgop on Modelling Autonomous Agents in a Multi-Agent World, LNAI 1038/1996. p.56-71.
- Klüpfel, H., Meyer-König, T., Wahle, J., & Schreckenberg, M. (2000). Microscopic Simulation of Evacuation Processes on Passenger Ships. In 4th International Conference on Cellular Automata for Research and Industry: Theoretical and Practical Issues on Cellular Automata. p.63-71.
- Reynolds, C. W. (1999). Steering Behaviors For Autonomous Characters. In Game Developers Conference.
- Reynolds, C. W. (2000). Interaction with Groups of Autonomous Characters. In Game Developers Conference. p.449-460.
- Reynolds, C. W. (2004) OpenSteer: Documentation. Available at <http://opensteer.sourceforge.net/> (last accessed May 13, 2009)
- Ronald, N., & Sterling, L. (2005). Modelling pedestrian behaviour using the BDI architecture. In IEEE/WIC/ACM International Conference on Intelligent Agent Technology. p.161-164.
- Rossetti, R., Bordini, R., Bazzan, A., Bampi, S., Liu, R., & Van Vliet, D. (2002). Using BDI agents to improve driver modelling in a commuter scenario. Transportation Research C(10):373-398.
- Schelhorn, T., O'Sullivan, D., Hakley, M., & Thurstain-Goodwin, M. (1999). STREETS: An Agent-Based Pedestrian Model. Centre for Advanced Spatial Analysis, UCL. London.
- Shao, W. (2006). Animating Autonomous Pedestrians. Ph.D. Thesis, New York University, New York.
- Tekmono, K. (2002). Microscopic Pedestrian Flow Characteristics: Development of an Image Processing Data Collection and Simulation Model. Ph.D. Thesis, Tohoku University, Japan.

# Using Heuristic Search for Initiating the Genetic Population in Simulation-Based Optimization of Vehicle Routing Problems

Anna Syberfeldt and Lars Persson  
Center for Intelligent Automation  
University of Skövde  
Box 408, 541 48 Skövde, Sweden  
E-mail: anna.syberfeldt@his.se

## KEYWORDS

Simulation-Based Optimization, Genetic Algorithm, Simulated Annealing, Tabu Search, Vehicle Routing Problem.

## ABSTRACT

Genetic algorithms are nowadays commonly used in simulation-based optimization of vehicle routing problems. These algorithms work with a population of solutions that are iteratively improved in an evolutionary process. Usually, the initial population is created randomly. In general, this is not very efficient since it takes unnecessarily long time before sufficiently good solutions have evolved. For a better performance of genetic algorithms, this work describes the use of heuristic search for creating the initial population. A new heuristic search procedure is described in the paper and evaluated using a real-world problem of garbage collection. The results from the evaluation show that the new procedure is able to improve a genetic algorithm.

## INTRODUCTION

The most well-known routing problem is probably the traveling sales person problem, which is a classical problem of finding the shortest route between a number of nodes. The problem might seem simple, but is classified as "NP-hard" in its simplest form (Santos, 2008). This means that the time it takes to solve the problem grows exponentially with the problem size (that is, the number of nodes). A common variant of the traveling sales person problem is the vehicle routing problem. In this problem, a number of trucks aim at visiting their clients by running the shortest possible distance. In a real-world vehicle 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 optimization, all the different parameters must be considered simultaneously. This is no easy task for a human to accomplish, and therefore computers are often used for this purpose.

With a computer it is possible to find the optimal route, but it may 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. As previously mentioned, *all* of these routes must be evaluated in order to find the best one (that is, the optimum). Due to the general complexity of vehicle routing problems, 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.

Genetic algorithms are a class of such inexact techniques that can be used to approach vehicle routing problems (Jozefowicz, Semet, and Talbi, 2008). Genetic algorithms are essentially based on Darwin's theories about "survival of the fittest" (Darwin, 1859). 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 an ordinary genetic algorithm, the initial population of solutions is created randomly. A random strategy often results in a good diversity among solutions in the population, but it might take unnecessarily long time before a sufficiently good solution has evolved. Chinneck (2006) discusses that the performance of the algorithm potentially can increase with a better starting population. Manipulating the initial population, however, poses a difficulty in maintaining the diversity of solutions. If some solutions outperform the other ones, the entire population will quickly converge and the search will become trapped in a local optimum.

This study describes the use of heuristic search to create the initial population for vehicle routing problems. Using such approach, the genetic algorithm is given a headstart and its performance may thereby be improved.

## BACKGROUND

The first step of a genetic algorithm is to create solutions for the initial population. The approach initially proposed to do this is based on a random strategy (Holland, 1992). Still today this is the standard way to create the initial population (Chinneck, 2006). A random approach is simple, but not very efficient in terms of algorithm performance (Chinneck, 2006). Random solutions are often poor and improving their quality takes long time since the refinement must be done over several generations (Chinneck, 2006). It has been shown that the final results of the algorithm depends (among other factors) on the quality of the solutions in the initial population (Chinneck, 2006).

Some approaches have been suggested to initialize the first population, but in general this problem has been paid relatively little attention (Johnson, 2006 pp. 270). One

approach is to use experts' domain knowledge to create one or more solutions and inject these into the population (Johnson, 2006). The main disadvantage of this approach is that it requires the existence of experts with sufficient domain knowledge. Furthermore, it also requires that the problem is relatively simple so that it is possible to develop solutions manually.

Another approach is to make use of solutions from previous optimizations (Louis and Li, 2000). The disadvantage of this approach is that it requires that the previous optimizations have been run with the same (or at very similar) variable values (for example, customer requests or vehicle set-up), which requires a static problem domain.

A third approach is to use heuristic search to create solutions for the initial population. A heuristic is a function that evaluates alternative solutions based on available information about the problem. Although heuristics do not guarantee to find an optimal solution, they can often provide decent solutions. As discussed later on in the paper, heuristics are used by search algorithms such as simulated annealing and tabu search to guide the search.

The idea of using heuristics to initialize the first population, instead of creating it randomly, was discussed by Grefenstette already in 1987 (Grefenstette, 1987). Grefenstette speculated that if valuable genetic material could be added to the genetic algorithm from the beginning, the algorithm could probably achieve better results.

## AIM OF STUDY

The heuristic approaches suggested in previous studies for vehicle routing problems generally have two limitations. First, they are designed and evaluated solely on theoretical problems. Second, they take into account only a single parameter, namely the distance between cities. In real-world problems, there are usually many more parameters to take into account (for example, load, time, environmental impact, etc.). The complexity of the problem is also considerably higher in real-world scenarios as a result of restrictions that exist (such maximum load, statutory breaks, etc.). Johnson (2006) argues that, especially for real-world problems, it is likely that there are more effective ways to initiate the population than those previously proposed (p. 270).

Using heuristic search to initialize the first population is not a trivial task. Grefenstette (1987) highlights an important problem, namely that the algorithm can be quickly trapped in a local optimum if there are a few solutions in the population who are superior to the others. These solutions will then quickly take over the population and the diversity of genes in the population will get lost. When all solutions are very similar the search will halt and not move forward.

Another important aspect in the use of heuristic search for initializing the population is, according to Johnson (2006), the solution representation. The solutions in the initial population are only useful if they use the same genetic representation as the genetic algorithm (Johnson, 2006, pp. 274). Therefore, the solution representation used by the

heuristic search and the representation used by the genetic algorithm must be synchronized.

## VEHICLE ROUTING PROBLEM

The vehicle routing problem is an extension of the traveling sales person problem. In the traveling sales person problem, the aim is to find the shortest path between a number of nodes. The name of the problem comes from the simile with a salesman traveling from town to town selling products. The salesman starts in his hometown and aims at visiting all possible cities and then go back again to where he started. In order for the profit of the sale not to be eaten up by the traveling cost, the trip should preferably be the shortest possible.

In the vehicle routing problem, a fleet of vehicles aims at visiting all customers with the lowest cost possible (Figure 1). The cost might for example be the time it takes to drive, or the distance between nodes. There may be several requirements for the design of routes (Kang, Lee, Lee and Lee 2008). For example, the vehicles may have a maximum load capacity or maximum working hours that must be considered. Furthermore, customers may have limitations in terms of periods when they are available. The demand may also vary over time.

Vehicle routing problem is classified as "NP-hard", which means that the time it takes to solve the problem grows exponentially with the number of customers (Kang et al. 2008). To deal with the time complexity, techniques that are fast but do not guarantee to find the optimal solution are often used. Such techniques include genetic algorithms, tabu search and simulated annealing. The latter two are described in the next section.

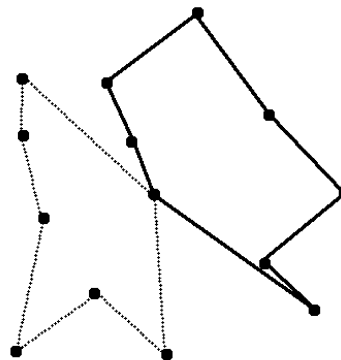


Figure 1 - Vehicle routing problem with two vehicles (black and gray) visiting a number of customers.

## HEURISTIC SEARCH TECHNIQUES

According to Hillam (2003) heuristics are "common sense", that is, things that are known but not necessarily formally researched. When it comes to finding the shortest route, common sense is to go straight towards the goal. It is not certain, however, that this will result in the shortest route, since a river may make you have to take a detour over a bridge.

Two techniques that use heuristics to solve problems is simulated annealing and tabu search (Michalewicz and Fogel, 2000). In simulated annealing, a random solution is first created and evaluated using a simulation. A neighbor of this solution is then created, that is, a quite similar solution. For example, the neighbors may have the same genetic material (same nodes), but differ in terms of the order of some of the genes (nodes). If the neighbor is better than the current solution, the neighbor replaces the current solution. If the neighbor is worse, it is not immediately discarded but there is a small chance that it is saved anyway. This is determined by a temperature  $T$  in combination with the fitness of the neighboring solution. A high temperature means that the chance of saving the solution is high, and vice versa.

After a new solution has been saved, or the old one has been kept, the temperature is slightly reduced. A new neighbor is then created and compared with the solution saved in the last iteration. The procedure continues in this way until the final goal is reached. The goal may, for example, be that the temperature is sufficiently low, or that there has been no improvement for a long time. Since the temperature drops in every iteration, it means that the search stagnates more and more during time.

Michalewicz and Fogel (2000) describe simulated annealing for a maximization problem using the following pseudo code:

#### Procedure simulated annealing

##### Begin

$t \leftarrow 0$

initialize  $T$

select a current point  $v_c$  at random

simulate  $v_c$

##### repeat

##### repeat

select a new point  $v_n$  in the neighborhood  $v_c$

if  $\text{eval}(v_c) < \text{eval}(v_n)$

then  $v_c \leftarrow v_n$

else if  $\text{random}[0,1] < e^{-(\text{simulate}(v_n) - \text{simulate}(v_c))/T}$

then  $v_c \leftarrow v_n$

until (termination-condition)

$T \leftarrow g(T,t)$

$t \leftarrow t + 1$

until (halting-criterion)

##### end

In the procedure above,  $v_c$  is the current solution and  $v_n$  is a neighbor to  $v_c$ .  $T$  is the temperature and  $t$  represents the temperature change. *simulate* is a function that calculates the fitness of a solution. *random* is a function that returns a random value between 0 and 1. The function  $g$  performs a reduction of  $T$  depending on  $t$ .

It can be noted that the heuristic as such is used in the function *simulate*. This function is problem-dependent. For example, if the problem is to find a route, it might be good if the distance is as short as possible. If the problem instead is to find the highest mountain peak in a mountain range, it

might instead be good to be as high above the water as possible.

Tabu search is another technique that also uses heuristics. Tabu search is similar to simulated annealing in that the basic procedure consists of comparing neighbors. What makes the tabu search different from simulated annealing is that some information are stored in a memory, called tabu list (Michalewicz and Fogel, 2000). The memory may, for example, keep track of which parts of a solution that have been altered. To maintain circulation, a requirement can, for example, be that the same piece may not be changed within a certain number of iterations.

## A NEW HEURISTIC SEARCH PROCEDURE FOR INITIATION OF THE GENETIC POPULATION

This section describes a new heuristic search procedure for practical vehicle routing problems which is intended to be used to create the initial population of a genetic algorithm. The overall all objective of the optimization is to collect as much value from the (heterogenous) nodes as possible, given the limitation that it is not possible to visit *all* nodes. The procedure, basically a mix of simulated annealing and tabu search, is described by the pseudo code below and explained in detail afterwards.

### Function GeneratePopulation()

while(population.size < enough)

Route = new route( $n$ )

Improve(route)

Population.add(route)

end

return population

### end

### Function Improve(route)

timeDiff = simulate(route)

##### do

if(timeDiff < 0) //Add node

node = new random node not in route and not in tabu

if(node.Value > route.averageValue)

route.add(node)

tabu.add(node)

##### else

route.simulatedAnnealingAdd(node)

else //Remove node

node = route.randomNode

if(node.Value < route.averageValue)

route.remove(node)

##### else

route.simulatedAnnealingRemove(node)

##### end

route = SelectShortestPermutation(route)

timeDiff = simulate(route)

tabu.checkSize()

tempature.update()

badWorkday = !isWithinTolerance(timeDiff)

while(highTemperature || badWorkday)

### end

In the procedure outlined above, a first solution is created that consists of a list of  $n$  nodes.  $n$  is a user-defined value

that is set to the approximate number of nodes that can reasonably be visited. When creating the list, one node at a time is added. An added node is placed in the tabu list, which means that a certain node cannot be added twice since no node within the tabu list can be used.

When the first solution has been created, this is improved using a combination of tabu search and simulated annealing. For evaluating the fitness of the solution, a simulation is used. The simulation returns the relative time consumption of the route in relation to the workday (variable *timeDiff*), which is used to decide if an additional node should be added or if a node should be removed. Depending on the problem domain, the simulation might also return other performance metrics such as time consumption between specific nodes, environmental impact, and filling degree of vehicles.

The improvement procedure consists of a do-while loop that ends when the temperature has decreased below a certain level and the time difference is small enough. In the beginning of the loop, a control is made if the time difference is above or below zero (that is, if the route takes shorter or longer time then the workday). If below zero, a new node is added, otherwise a node is discarded. A new node to be added is selected randomly with consideration to the tabu list. If the value gained by adding the selected node is higher than the average value of the nodes in the route, the node is added. If the value is lower, it might still be added if it is approved by the simulated annealing formula.

If the time difference is above zero and a node therefore needs to be discarded, one of the nodes is selected randomly. If the value of this node is lower than the average value of all nodes, it is discarded immediately. Otherwise, it is processed by the simulating annealing formula. If not discarded by this formula, the node is kept and a new node is randomly selected and checked for deletion.

In the next step of the procedure, the route is improved by performing a number of permutations of the node list. Each permutation is simulated and the one with the lowest time difference is selected. After this, the length of the tabu list is checked and if the list is too long, its oldest node is deleted. The temperature is updated by multiplying the current temperature with a constant. After this, the procedure continues from the beginning by checking if further iterations are needed. If not, the new solution is added to the population and a new random solution is created and improved. The generation of new solutions continues until the genetic population is filled.

## EVALUATION

The proposed heuristic has been evaluated on a real-world problem of garbage collection. In this problem, about 20 nodes (garbage containers) are to be visited every day by a garbage truck. With 20 nodes, there are about  $20! = 2.43290201 \times 10^{18}$  different possibilities to solve the problem. The objective is to collect as much garbage as possible during a workday.

A discrete-event simulation exists for the problem and can be used to evaluate solutions. Given a list of  $n$  nodes, the simulation returns the average filling degree of the garbage truck, and the time the route takes in relation to the workday. If the workday ends at 5 p.m., for example, and the route ends at 4 p.m., the simulation returns -1.

For the evaluation, a genetic algorithm was constructed and run for 100, 500, and 1000 generations. After each run, the best and worst fitness values in the final population were measured. The proposed heuristic procedure for creating the initial population was compared to the approach of random creation.

Results from the experiments can be seen in Table 1-3. Note that the time denotes the total time of the optimization procedure (that is, the time for creating the first population plus the time used by the genetic algorithm). Also note that a high fitness is desirable.

|                 | Random procedure | Heuristic procedure |
|-----------------|------------------|---------------------|
| Maximum fitness | 0.506778675      | 0.546950425         |
| Minimum fitness | 0.274586         | 0.3278084           |
| Time (seconds)  | 7.1175           | 13.465              |

Table 1: 100 generations (average from five runs)

|                 | Random procedure | Heuristic procedure |
|-----------------|------------------|---------------------|
| Maximum fitness | 0.54545864       | 0.574607475         |
| Minimum fitness | 0.405183         | 0.450005            |
| Time (seconds)  | 12.855           | 17.8575             |

Table 2: 500 generations (average from five runs)

|                 | Random procedure | Heuristic procedure |
|-----------------|------------------|---------------------|
| Maximum fitness | 0.554043         | 0.58216775          |
| Minimum fitness | 0.36898875       | 0.4399232           |
| Time (seconds)  | 18.2625          | 22.3825             |

Table 3: 1000 generations (average from five runs)

As can be seen from the results, there is a clear benefit of using the proposed heuristic for creating the initial population of the genetic algorithm. The largest benefit of using the proposed heuristic is when the number of generations of the genetic algorithm is small (100). This is no surprise, since the fewer the generations the more important the starting conditions of the algorithm.

As can also be seen from the results, the time consumption of the heuristic procedure is worse than that of the random procedure. The time difference decreases with the number of generations, since the larger the number of generations the larger proportion of the time is spent by the genetic algorithm. Further studies will be performed in the future in order to investigate how the time consumption of the heuristic can be reduced.

## SUMMARY

Genetic algorithms are powerful search strategies commonly used in simulation-based optimization of vehicle routing problems. A genetic algorithm works with a population of solutions that is iteratively improved in a evolutionary process. The initial population is usually created randomly, which generally results in a good diversity in the population but implies a large time-consumption before a sufficiently good solution has evolved. For increased efficiency through better starting conditions for the optimization process, this paper describes the use of heuristic search to initialize the population. A new procedure is proposed that is based on the combination of two existing heuristic procedures; simulated annealing and tabu search. The proposed procedure is evaluated on a real-world problem of garbage collection and the results from the evaluation show the benefits of using the procedure. Further studies are needed, however, to investigate how the time consumption of the procedure can be reduced.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge Avafallshantering Östra Skaraborg (a Swedish garbage collection organization) for their support in this study.

## REFERENCES

- Chinneck J. W. (2006) Practical optimization: a gentle introduction. Carleton University. Available from the Internet: <http://www.sce.carleton.ca/faculty/chinneck/po.html> [09.05.11]
- Dablanc, L. (2007) Goods transport in large European cities: difficult to organize, difficult to modernize. *Transportation Research*, 41(A), 280-285.
- Darwin, Charles (1859) *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.
- Deb, K. (2001) *Multi-objective optimization using evolutionary algorithms*. Chichester: John Wiley & sons, Ltd
- Grefenstette, J. J. (1987). *Genetic Algorithms and Simulated Annealing*, London: Pitman.
- Hillam, B. P. (2003) *Introduction to Algorithms: a java based approach*. California state polytechnic university. Available from the Internet: <http://www.csupomona.edu/~bphillam/algotext/> [09.03.25]
- Holland, J. H. (1992) *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control and artificial intelligence (2:a upplagan)*. Cambridge: MIT Press
- Johnson, D. S., (1990) Local optimization and the traveling salesman problem. *Proceedings of the 17th International colloquium on automata, languages and programming*
- Jozefowicz, N., Semet, F. och Talbi E.G. (2008) Multit-objective vehicle routing problems. *European journal of operational research*, 189, 293-309
- Kang, K. H., Lee, B. K., Lee, Y. H., och Lee, Y. H. (2008) A heuristic for the routing problem with due times. *Computers & industrial engineering*, 58, 421-431
- Louis, S. J. och Li, G. (2000) Case injected genetic algorithms for traveling salesman problems. *Information sciences*, 122, 201-225
- Michalewicz, Z. och Fogel, D. B. (2000) *How to solve it: modern heuristics (2:a upplagan)*. Berlin: Springer-Verlag
- Santos, L., Coutinho-Rodrigues, J. och Current, J. R. (2008) Implementing a multi-vehicle multi-rout spatial decision support system for efficient trash collection in Portugal. *Transportation research*, 42(A), 922-934

# TOWARDS A MICROSCOPIC TRAFFIC SIMULATION FRAMEWORK TO ASSESS VEHICLE-TO-VEHICLE NETWORKS

João F. B. Gonçalves, Rosaldo J. F. Rossetti, Edgar F. Esteves, Eugénio C. Oliveira  
Department of Informatics Engineering (DEI)  
Artificial Intelligence and Computer Science Lab (LIACC)  
Faculty of Engineering, University of Porto (FEUP)  
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal  
E-mail: {ee02123, rossetti, edgar.esteves, eco}@fe.up.pt

## KEYWORDS

Intelligent transportation systems (ITS), service-oriented architectures (SOA), vehicular *ad-hoc* networks (VANET), vehicle-to-vehicle (V2V)

## ABSTRACT

This paper presents the specification of a framework based on the concept of service-oriented architectures (SOA) to support the assessment of vehicular ad-hoc networks (VANET). A preliminary study of concepts related to SOA was carried out, as well as of those technologies that allow real-time data acquisition and dissemination within urban environments, and simulation tools to aid the simulation of the VANET. The requirements for our simulation framework were identified and a two-layered architecture was specified, which rely on the abstraction levels of services for vehicle-to-vehicle (V2V) communication. A prototypical application was implemented, which was used to demonstrate the feasibility of the approach presented through experimental results.

## INTRODUCTION

The number of people living in urban areas has increased very quickly and unquestionably. Unfortunately, most people opt for using private car instead of public transport. This trend has turned infrastructures unable to cope with the ever increasing demand, which work most of the time in saturation regime. This problem motivates the scientific community that strives to devise different solutions. Some approaches suggest the physical improvement of infrastructure by means of increasing road capacity. Others try to enhance the control system responsiveness, with relative success. More recently, though, researchers have experimented promising innovative information technologies to aid driving tasks and traveller decision-making. The latter ones underlie the concept of intelligent transportation systems (ITS), which attempts at integrating contemporary and breakthrough computer and information technologies to better manage and control modern urban transportation systems. ITS are intended to turn future urban transport (FUT) into greener, cheaper, reliable, secure, and sustainable systems, both functionally, energetically and environmentally efficient as well.

The work reported in this paper is based on our study of the communication technologies underlying vehicle-to-vehicle interactions. High-class vehicles will very soon be

equipped with short-range wireless communication interfaces, which will allow them to form an ad-hoc communication network. V2V networks of such a genre are being studied already and motivating much advances in ITS-based technology. The scientific community believes information disseminated through these networks will play an imperative role in future urban traffic network security and efficiency. This new application domain brings about major concerning issues, though. Thus, simulation tools will need to be adapted to support the assessment of V2V communication infrastructures and performance, allowing researchers and practitioners to define which architecture is most adaptable to urban traffic scenarios, for instance.

In this work, we aim at studying concepts related to SOA and ways in which such concepts can be adapted to the VANET. The work started by the specification and implementation of a simulation framework that will allow us to gain further insight into such motivating and challenging new arena. Such a framework is described in more details and some preliminary results are discussed.

This work is organised as follows. In the next section, we review some important concepts such as SOA, V2V communication networks and some simulation tools already available, in the light of concepts we are going to use in the proposed approach. In section three, the development of a simulation prototype is discussed and detailed, whereas in section four some experimental results are discussed and commented. Finally, we draw some conclusions in the last section, point up directions for further developments and opportunities for future works.

## FUTURE URBAN TRANSPORT TECHNOLOGIES

ITS-based technologies have proven a great impact and influence in future urban transport scenarios. As the automotive industry starts marketing vehicles equipped with wireless communication capabilities, some technologies are being deemed to be as potentially applicable and beneficial.

### Service-oriented Architectures

Service-oriented architectures are devised as software architectures whose main goal advocates that application functionalities must be made available in form of services. Thus, services from SOA's point of view are functions of a computational system that are made available to other systems. Such a novel approach might be well represented

by the “find-bind-execute” paradigm, which is analogous to the Deming’s cycle applied to services, involving planning, execution, monitoring and pro-active decision-making phases to improve systems’ performance. The “find-bind-execute” paradigm allows a consumer of a service to ask for registering in a service provider that suits its criteria and requirements. If the service of interest is found, the provider sends the consumer a contract and the address in which the service can be found. Such a mechanism is depicted in Figure 1.

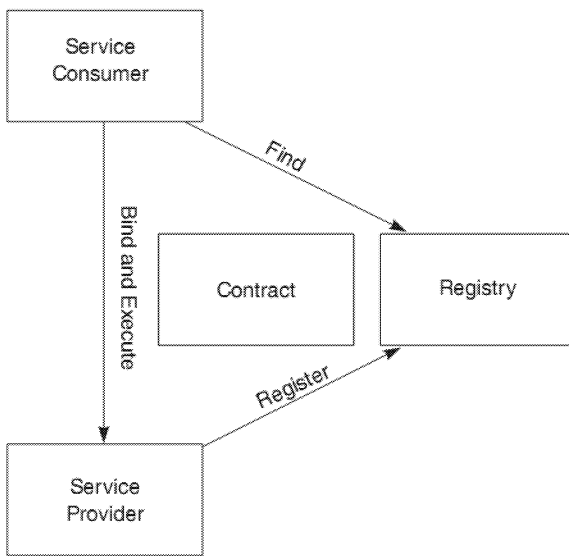


Figure 1: Find-Bind-Execute Paradigm  
(adapted from McGovern et al. 2003)

According to McGovern et al. (2003), the entities that support such a service infrastructure are:

- *Service consumer*, which might be basically an application, a service, or another type of software requiring a service. This is the entity who initiates the process, looking for a service capable of supplying its requirements. The consumer executes the service by sending a request formatted in accordance to the contract.
- *Service provider*, which is the addressable entity that accepts and executes the service. This entity might be a mainframe, a component or another type of software that executes a requested service. Service providers publish their contract in the service registry for other potentially interested consumers to access them.
- *Service registry*, which is a sort of network directory that “knows” the available services. This entity accepts and stores contracts of providers and provides consumers with these service options.
- *Service contract*, which is sort of a specification of the way in which a service consumer must interact with the provider. This entity rules the protocol for service request and respective answers from the services. A contract may require a set of pre- and post-conditions representing the state of a service to be deemed acceptable to execute certain functions. Contracts may also contain information concerning the quality of services, as well as certain conditions to which consumers must comply.

## Vehicle-to-vehicle networks

Vehicle-to-vehicle communication networks are, as its designation suggests, networks formed by several vehicles equipped with wireless communication devices that can communicate with each other. In V2V networks, each vehicle analyses, within a certain radius, the other vehicles that are in range, and can inform its position, velocity, direction and other characteristics. This kind of communication has been one of the fields of interest in telecommunications that grew very quickly lately. Thus, vehicles with such capabilities can form a special type of mobile ad-hoc network with particular applications, known as Vehicular Ad-hoc Networks (VANET). VANET are a special type of Mobile Ad-hoc Networks (MANET) that supports communications between vehicles. According to Mello and Endler (2006), the VANET inherit some characteristics from the MANET, but also improve the former with new characteristics, which differentiate it from other ad-hoc mobile networks. These characteristics include high mobility, open network with dynamic topology, limited connectivity, potential to achieve larger scale, all nodes are providers, forwarders and consumers of data and the wireless transmission can suffer from much noise and interferences.

These characteristics make the VANET sufficiently different from other networks and significantly affect their properties. For example, in (Chen et al. 2001) is demonstrated that the movement of vehicles has a significant effect on the latency of messages delivery. Most applications that can be implemented on a MANET require a certain type of data dissemination. Studies argue, however, this must be implemented through specific routing protocols, as long as VANET are concerned, basically due to their particular characteristics, such as in (Blum et al. 2004). Nonetheless, there are other studies that suggest it is possible to use the available MANET protocols. Some of those protocols are designed to support dedicated short range communications (DSRC), which is already implemented in USA, such as VITP and PAVAN. Due to its topology, the MANET already have a large set of protocols.

There are several ways to classify routing protocols for such networks, some of which are listed below:

- According to the range, they can be either *unicast* or *multicast*.
- According to the route discovery, they can be either pro-active, reactive or hybrid.
- According to the search algorithm they are based on, they can be either Distance Vector, Link State, based on geographic information, or Zone based.

The *unicast* protocols are those that transmit information from one transmitter to one receiver. In contrast, multicast protocols are those in which information is sent to a previously created group of nodes.

In what refers to route discovery, the pro-active algorithms are those that periodically update the network routes. These particular algorithms have, at every moment, the knowledge of the network topology, trying to get the optimal route for the time that is necessary to send information. The exchange



of control packets and the update of routing tables are continuous and all nodes are known in advance. To the contrary, there are reactive algorithms (also known as on-demand ones) that only discover the route to a destination when they need to send information. In these algorithms, the route discovery is done on-demand, eliminating the permanent routes. Hybrid algorithms combine the formers' good characteristics using, for instance, pro-activity in the node neighbourhood and discovering the route to distant nodes at the exact moment it is need only.

Now, as for the type of algorithms on which the routing protocols are based, Distance Vector algorithms require that all nodes in the network exchange their distance vectors periodically, meaning each node knows the cost information for each destination; each router (node) maintains a table (a vector) containing the cheapest known route to each destination. These tables are updated by exchanging messages with neighbours and comparing the received table with their own table. In the case of a better route is found, the router updates its table and stores the source of this information too. In Link State algorithms, each node monitors the state of the link with its neighbours and disseminates this information. Each node knows the status of all links, and determines the complete topology of the network as well as the locally shortest path. On the other hand, geographic information based algorithms are able to estimate the location of the nodes and the information is sent in which direction. In turn, in zone based (source routing) algorithms, one route is only created when the source node demands it, after which creation the route is maintained by a maintenance protocol and there is no need for periodic updates.

In the Figure 2, the protocols previously mentioned are related, as well as are their classification according to the characteristics mentioned above.

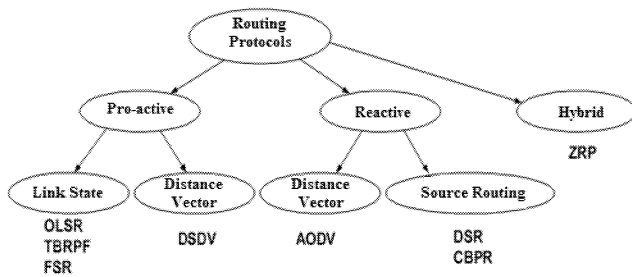


Figure 2: MANET Routing Protocols Classification

### Simulation tools for V2V networks

The development of applications and protocols associated to VANET can be studied through simulation, especially when a real traffic network in urban environments, which must involve a large number of nodes, is the subject of study. Through simulation models, the performance of V2V networks, as well as other characteristics can be assessed and improved. Indeed, conducting experiments and studies on such a large scale within the real scenario has proven extremely difficult and expensive. Thus, simulation has become an indispensable and even an imperative tool.

Basically, the simulation of V2V networks requires two different components, namely a communication networks simulator, capable of simulating the properties of a wireless network, and a vehicular traffic simulator, able to monitor and represent the kinematic aspects of mobility throughout the VANET nodes. Recent studies (Choffnes and Bustamante 2005) suggested that the vehicular mobility model is very important to obtain significant results and should be well integrated with the wireless communication networks model. Other authors further suggest that the use of an inappropriate model, such as the popular “random waypoint model” (which can work very well for some Mobile ad-hoc networks, but very likely is not an appropriate representation of mobility in wireless vehicular networks) can lead to erroneous results (Choffnes and Bustamante 2005; Saha and Johnson 2004).

Nevertheless, traffic simulators have been subjected to enormous developments and greatly improved in recent years to include communication between vehicles. Several ways to achieve such an advanced feature have been proposed and actually implemented. The greatest trend in most studies moves towards the creation of simulators that include traffic and wireless communication simulation models in one single simulation tool. Some examples of such tools include GrooveNet and Divert.

However, other studies prefer to use two independent simulators, combining stand-alone traffic and communication networks simulators, which are interconnected by an application that ensures the exchange of information between them. Among traffic simulators used for this purpose, one can mention the popular SUMO, as well as VISSIM and CARISMA. Wireless communication networks simulators include GloMoSim, QualNet and the NS2.

## ARCHITECTURAL PROTOTYPING

### The Layered SOA Structure

Currently, the services that both drivers and travellers in general can use on-board in journey time demand a great deal of hardware and software. Each new feature must be implemented in a new device to be embedded in the car. Such an approach has proven very expensive, with little flexibility, which contradicts the increasing trend of services made available on-demand, for instance. On the other hand, services made available as software to be executed in an onboard unit (OBU) based on an embedded computer with considerable processing and memory, as well as communication capabilities seem to be very promising and present great potentials and advantages. According to such not-so-much futuristic scenario, we intend to specify and implement an extensible architecture based on OBU computers and featured with communication capabilities to the level of services.

Bearing in mind such architecture was intended to promote services throughout V2V communication networks. We then opted for a layer-based structure. Thus, the proposed architecture is divided into two main levels, namely the network services level and the end-user services level. This

approach is illustrated in Figure 3, in which the two different levels are identified. The first level in such a structure is responsible for network-related tasks, such as building network topology, as well as discovering and exchanging services among vehicles (the nodes of the communication network). The second level, on the other hand, implements the necessary basis underlying the management of the so called high-level services, to be made available to end-users.

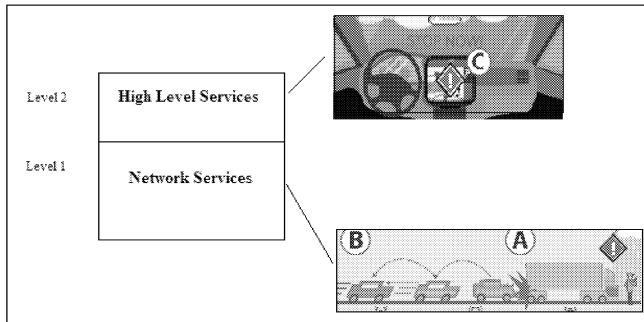


Figure 3: SOA Layered Architecture for V2V Networks

For the first level, and accounting for its functions within the proposed structure, we adopted for the solution suggested in (Halonen and Ojala 2006). The scenario illustrated in Figure 4 is a good representation of the adopted solution, as explained below.

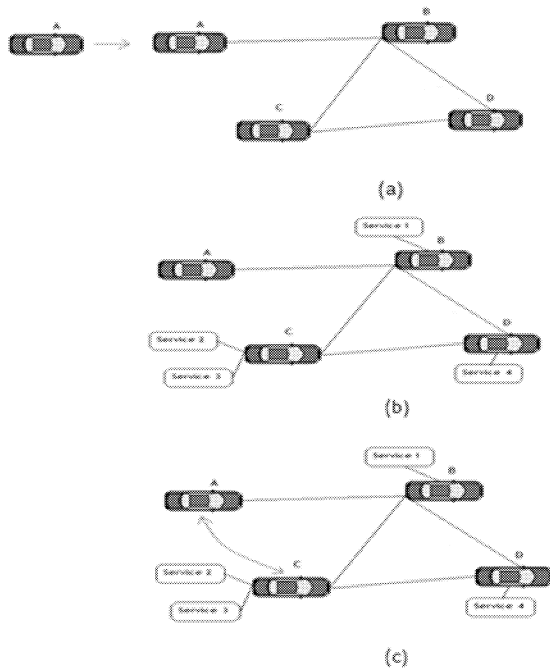


Figure 4: SOA-Based V2V Communication Interactions

Let us consider the communication network formed up by the vehicles as illustrated in Figure 4. In (a), the network is actually set up. All nodes (A to D) are vehicles equipped with mobile wireless V2V communication devices. Initially, the A node is not connected for there is no other vehicle within its range vicinity. As soon as it finds other vehicles within its range, then they get connected. All connected nodes thus form a VANET, allowing V2V communication.

After all the necessary automatic configurations are set up, the VANET is ready for routing any messages sent by any node and addressed to any other node of the network.

After the VANET is then ready, the SOA-based features must be deployed too. All nodes that provide any service must publish it alongside a comprehensible description, so that other nodes will be able to discover the service and use it as needed. The service discovering capabilities of each node must be implemented in such a way that the node may be able to find the needed service throughout the VANET whichever configuration must apply. Thus, in (b), every provider node publishes its services, whereas any consumer will be able to find them whenever necessary.

An illustrative service use example is demonstrated in (c). The A node gets connected to the service interface of service whose ID is 3, hosted in node C. All information necessary to message exchange between consumer (A node) and provider (C node) is available in the description of service 3, needing A no further information about node C. As A and C nodes are not directly connected to each other, messages must be routed through the B node.

Two basic components are necessary for the example presented above to be possible, namely a routing protocol for the VANET and the implementation of SOA functionalities. One of most important SOA features is the service discovering mechanism. Two approaches are then possible for us to implement such a mechanism. First, we can integrate it within the VANET routing protocol or, alternatively, we can implement it on top of the network layer, as a separate functionality. In the current work we opted for the first solution, and integrated it in the VANET routing mechanism. Such a decision proved advantageous as both the routing tables and protocol routing techniques are used in the service discovering process. For this purpose, we chose the optimised like state routing protocol (OLSR) for it is a pro-active protocol, which facilitates the propagation of services description throughout a network of known topology. Among pro-active protocols, OLSR has the advantage of using multipoint relaying (MPR) and offers an open source application (OLSR *Daemon*) allowing it to be extended.

Routing in MANET is based on the cooperation of participating nodes; all nodes must run the protocol in such a way they collectively achieve routing goals. Such a behavior is also desirable in VANET, as all nodes are expected to collaborate in services discovery as well. In practice, it means all nodes manage message routing even though they may not know its content. This way all service provider nodes may use the entire network and routing protocol.

A node intending to publish its services must generate and propagate SOA messages periodically, carrying the service information, according to certain transmission interval. Whenever a node receives a SOA message, it must resend it and, if it supports SOA services, process it too.

The format of a SOA message is presented in Figure 5. The format has two basic fields, namely the message length and the service description, which describes the service as a unique service (no other service will have the same description). This sort of message is carried within OLSR packages and propagated throughout the network.

Nonetheless, there is no specific format for the content of a service descriptor, which will basically depend on the type of service and the way it is implemented. XML files might be used for this purpose, though.



Figure 5: SOA Message Format

In the second layer of our proposed architecture, high-level services are defined. These services are made available to the end-users, which mean they can implement any kind of application the user might opt to use. In other words, they are the applications to which the traveller can access during the journey. Thus, every hardware and software components a vehicle must feature in order to provide such services are specified in this layer. As previously mentioned, we consider the vehicle will be equipped with an OBU.

## Development Environment

The prototypical implementation herein presented is conceived on the basis of extending the MAS-T<sup>2</sup>er Lab framework (*Laboratory for MAS-based Traffic and Transportation Engineering Research*). MAS-T<sup>2</sup>er Lab is an integrated multi-agent system that applies a methodological approach based on the agent metaphor to devise, test and implement intelligent transportation solutions. A more detailed discussion on the main principles underlying such a framework, as well as a full description of its architecture and implementation can be found elsewhere (Rossetti et al. 2007; Ferreira et al. 2008).

For the present work, it is enough to mention that the MAS-T<sup>2</sup>er Lab framework is conceived in terms of three main subsystems, which are multi-agent systems themselves, namely the *real world* (RW), the *virtual domain* (VD), and the *control strategies and management policies inductor* (CSMPI). The RW represents the real urban transportation system, in which physical entities such as vehicles, control systems, intelligent transport solutions and travellers in general cohabit and interact. These components are replicated within the VD and modelled by means of agents. Software agents in the VD are expected to emulate the individual behaviour of their counterparts in the RW. Finally, the CSMPI is formed up by expert agents, both humans and virtual ones, which observe the synthetic population within the VD and are able to directly interfere upon their behaviour so as to test different control strategies and management policies.

The VD subsystem relies on a multi-paradigm traffic simulator and is the core of the MAS-T<sup>2</sup>er Lab framework. The high-level architecture devised for its microscopic traffic model, depicted in Figure 6, is structured in layers where the most important component is undoubtedly the simulator engine controller (SEC).

The SEC component is responsible for managing all applications connected to the VD's simulation environment and their interaction with the urban network data structure. It also controls the simulation process, ensuring the necessary synchronisation of parts that can be run as distributed processes. All VD components and their interactions are detailed presented elsewhere (Ferreira 2008).

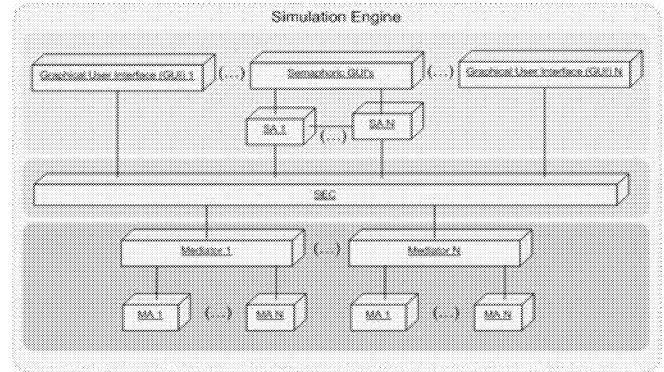


Figure 6: High-Level Architecture of VD Subsystem

In this work, the SEC component was extended with additional features especially designed to support modelling and simulating V2V communication networks, as well as their application in urban scenarios. The extended version of the SEC component is presented as an UML component diagram, as depicted in Figure 7, in which the V2V interface is emphasised in the red ellipse.

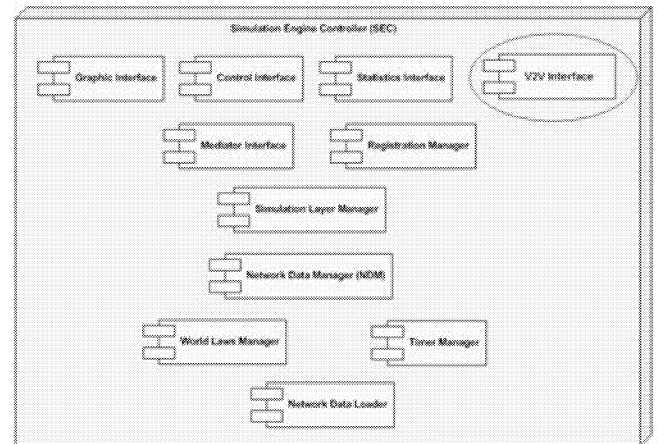


Figure 7: UML Diagram Illustrating the Physical Architecture of the SEC Component

The V2V interface is the SEC communication GUI that allows, whenever it is demanded, information on the vehicle-to-vehicle communication simulation process to be sent. At the current version of our prototype, such information includes global positioning of vehicles, their communication radius and the other vehicles within their range, as well as all messages that are sent and received by different vehicles of the synthetic population of the VD subsystem. The V2V interface also manages interaction with the end-users, periodically sending structured information on the simulation process and network status to the GUI modules attached to the VD simulation engine controller. Through this interface

module, end-users can also interact with the model in simulation time, changing parameter values and other simulation variables.

The prototype has been implemented in the C++ object-oriented programming language, with the Qt framework that allows easy GUI development. All vehicle references are stored in a dynamic structure, in which each car can be identified unequivocally by their IP address, for instance. In this structure we also keep the global position of each vehicle, as well as other characteristics such as the communication range and messages log. As mentioned before, end-users can interact with the model in simulation time through dialog boxes, through which a list of all cars currently using the network is rendered. From this list it is possible to access each vehicle's current status. A circle is drawn for the selected vehicle by double clicking its ID, which represents its communication range. Through a similar mechanism, the user can send messages on behalf of any vehicle travelling throughout the network. They can be sent either through unicast or multicast protocols, with vehicles recursively relaying the messages to their destinations. Addressees are identified through their global positioning and vehicles receiving sent messages are also emphasised in the graphical interface (a circle representing their communication range is drawn as well). Users can distinguish between different messages sent in different ways, namely through the messages log in the vehicles list or through differently coloured circles surrounding cars in the graphical visualisation of the network.

Figure 8 depicts a screenshot of the simulator GUI, with the vehicles list and the network visualisation windows. Different coloured circles represent different messages propagation throughout the network. In addition, the simulation environment also features graph-based representation of some performance measures, which can be dynamically updated in simulation time.

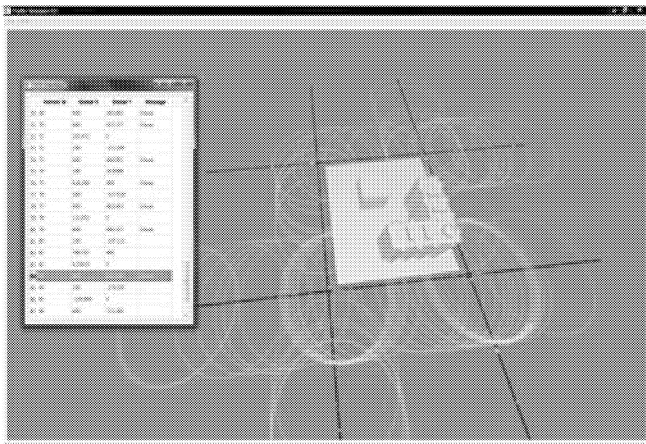


Figure 8: GUI for the V2V Communication Extension

An external tools interface is also available, allowing third-party tools to interact with the Simulator in run-time. Such an interface was implemented through socket-based streams via TCP. Any external application can then interact with our simulation model in run time and interact with the vehicles travelling throughout the network according to a protocol

designed for this specific purpose. For instance, different graphical interfaces can be attached to the simulator, rendering alternative visualisations or graphs to ease performance assessment. Also, other tools can be implemented to improve the services available via the V2V communication network.

## EXPERIMENTAL FRAMEWORK

### Scenarios Set-up

The experimental framework carried out was based on a simple network, as depicted in Figure 8. Despite its simplicity, results showed promising potentials of the proposed approach. The network was coded in the XML format supported in the simulation engine of MAS-T<sup>2</sup>er Lab, as defined in (Ferreira 2008).

Three different traffic flow configurations were set up, which allowed us to test the system behaviour under different flow regimes. The effects of these flow regimes on the V2V communication performance were analysed then.

*First scenario:* a low traffic flow regime was implemented in this scenario, in which source nodes were set to yield 100 to 250 vehicle/hour traffic flows, representing a free-flow regime.

*Second scenario:* this scenario is intended to represent an average traffic flow regime. In this scenario, source nodes were set to yield 250 to 850 vehicle/hour traffic flows, allowing us to analysis V2V performance under average conditions.

*Third scenario:* in the last scenario, source nodes were set to yield 850 to 1300 vehicles/hour traffic flows, representing traffic under saturation conditions. Such a scenario is quite common in must urban areas, during peak hours, for instance, or during the occurrence of some incidents strangulating road capacity, such as accidents.

Each simulation run was set to last for 10 minutes' time, corresponding to 30 minutes in real time. The time scale can be easily modified, though. To test the communication interaction between vehicles, a car approaching an intersection sends a message (this is done with a 2 minutes' frequency). Intersections are selected sequentially, following a clockwise order. Allowing cars to send messages at intersections is required as an attempt at maximizing the neighbouring cars within the range of the sender's wireless range. Indeed, as intersections are spots of converging traffic streams, then it is very likely the number of cars at intersections, especially in the second and third scenarios, will be considerable high.

### Results Discussion

Preliminary results are plotted in the graph of Figure 9. The graph shows how much of the network is covered by the V2V communication mechanism over time. The results are very interesting and demonstrate the ability of the implemented prototype to cope with both communication and vehicular traffic simulation. As we expected, the results of different scenario set-ups suggest different behaviours for the system under varying traffic conditions.

In the low traffic flow scenario, message dissemination through the V2V infrastructure is quite poor, basically due to the discontinued coverage of the network. Only a small part of the network, *circa* 27%, can be affected by the message propagation. In such circumstances, it is possible that even none of the vehicles will receive the message sent.

Scenarios in which traffic flow follows an average flow regime, such as in the second experimental set-up above, a larger part of the network can be easily covered. In our case, about two thirds of the network was covered in the most appropriate circumstances, meaning the increasing number of cars propitiated a better coverage and improved communication.

Only in the third scenario it was possible to achieve a full coverage of the network, which is quite expectable too. Indeed, in nearly saturated traffic conditions, network links tend to work in their full capacity, meaning the density is very high and vehicle headways tend to the average vehicle unit. In these circumstances, it is most probable that neighbouring vehicles will be within the range of other vehicles' wireless sensors, improving the connectivity of the network.

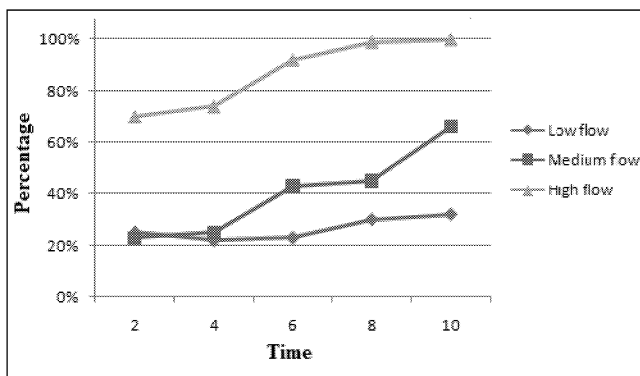


Figure 9: Comparison of the Three Different Scenarios, Relating Simulation Time and the Percentage of Network Covered by the V2V Communication

In all simulation scenarios, nonetheless, coverage increases as time evolves, which can be associated to the small network used and to the semaphore control at all junctions. Indeed, controlled intersections tend to group vehicles together at red lights, especially if traffic flow is higher than the saturation flows of each phase of the control plan. In such a situation, intersections will never be empty improving the connectivity of the V2V communication network.

The three different scenario set-ups were inspired in a typical daily flow profile of an urban network, in which the first scenario might represent a free flow regime, e.g. at high night and dawn, the second scenario might represent the average situation of off-peak hours during the day, whereas the third one might represent the morning and afternoon peak hours. Nonetheless, the first scenario might also be associated to rural areas, whereas the third scenarios might well be representative of a bottleneck caused by an incident, such as accident, for instance.

In either case, results corroborate the idea that V2V communication presents a great potential and is a promising

technology of future urban transport, whereas implementing SOA beneath such communication infrastructure can contribute a great deal for a better transport service and quality of life in urban areas.

## CONCLUSIONS

In this work, a V2V architecture based on the concept of services was specified. A layered architecture based on different levels of abstractions of services providing SOA in V2V networks was specified, as well as was a prototype implemented for testing and evaluating the approach proposed. The proposed architecture is structured into two main levels, namely the network services level and the consumer/service higher level. The former is responsible for building the network topology and for discovering and exchanging services among vehicles. The latter consists of the architecture to support service distribution, coined high-level services, to serve consumers. For the first level, we have implemented a SOA-based feature that allows service discovering within the routing protocol of VANET networks. This resulted in the specification of a dynamic and adaptable routing structure compliant with the abstract nature inherent in any SOA-based applications. The SOA features are rather decentralised in the first level, meaning there is no registry or central repository for services.

For the second level, on the other hand, we presented a modular and easily extensible architecture that allows services to be made available in vehicles. We have specified the vehicular architecture that underlies the implementation of on-board services, as well as the adequate means to request services from exogenous sources and other vehicles too.

Also, we have conceived and implemented a prototype of the proposed framework and extended the MAS-T<sup>2</sup>er Lab original structure to support V2V communications within its microscopic simulation model. Besides V2V communications, such an extension allowed us to integrate external applications within the framework's traffic simulation engine.

In general terms, the developed prototype and experimental results demonstrated the feasibility of the proposed approach. The simulation environment resulted from the improvements implemented is an important asset for testing and experimenting new generation intelligent transportation systems, which will strongly rely on V2V communication capabilities. Further developments will include the implementation of more complex scenarios and adequate tools to support the assessment of a whole urban network. The introduction of other concepts, such as multi-agent systems are equally envisaged and will certainly contribute to the improvement of the way the decision-making mechanisms of drivers are affected under the influence of such a novel informational paradigm. The information visualization is another concern being addressed at the moment. Service utilisation and dissemination in large networks sometimes represents a tricky piece of information to visualise. Concepts of visual

interactive modelling and simulation will be used to address such issues.

## REFERENCES

- Blum, J.J., Eskandarian, A., Hoffman, L.J., "Challenges of inter-vehicle ad hoc network", *IEEE Transactions on Intelligent Transportation Systems*, 2004. p.347-351.
- Chen Z. D., Kung H., Vlah D., "Ad hoc relay wireless networks over moving vehicles on highways". In Proceedings of the 2nd ACM International Symposium on Mobile Ad Hoc Networking and Computing. 2001. p.247-250.
- Choffnes D.R., Bustamante F.E., "An Integrated Mobility and Traffic Model for Vehicular Wireless Networks", 2nd ACM International Workshop on Vehicular Ad Hoc Networks (VANET), 2005. p.65-78.
- Ferreira P., "Specification and Implementation of an Artificial Transport System." Master's Dissertation. FEUP, Porto, 2008.
- Ferreira P., Esteves E., Rossetti R., Oliveira E., "A Cooperative Simulation Framework for Traffic and Transportation Engineering," in CDVE2008: The 5<sup>th</sup> International Conference on Cooperative, Mallorca, Spain, 2008. p.89-97.
- Halonen T., Ojaja T., "Cross-Layer Design for Providing Service Oriented Architecture in a Mobile Ad hoc Network." In Proceedings of the 5th international conference on Mobile and ubiquitous multimedia. Paper 11. 2006.
- McGovern J. et al., "Java Web Services Architecture", 1st Ed, Elsevier Science and Technology Books, 2003.
- Mello, H., Endler, M., "Identificação de Região de Congestionamento através de Comunicação Inter-veicular." In 24th Brazilian Symposium on Computer Networks, 2006.
- Rossetti, R.J.F., Oliveira, E.C., Bazzan, A.L.C., "Towards a specification of a framework for sustainable transportation analysis." In 13th Portuguese Conference on Artificial Intelligence, Guimarães, Portugal, 2007.
- Saha A., Johnson D., "Modelling Mobility for Vehicular Ad-hoc Networks", ACM International Workshop on Vehicular Ad Hoc Networks (VANET), 2004. p.91-92.

# Development and Evaluation of Traffic Management Strategies for Personal Rapid Transit

Pengjun Zheng  
David Jeffery  
Mike McDonald

Transportation Research Group, School of Civil Engineering and the Environment  
University of Southampton  
Highfield, Southampton, SO17 1BJ, UK  
E-mail: (p.zheng; d.j.jeffery; mm7@soton.ac.uk)

## KEYWORDS

Personal Rapid Transit, Optimisation, Traffic Management, Simulation

## ABSTRACT

This paper describes a simulation study of traffic management opportunities with an extended Personal Rapid Transit (PRT) network at Heathrow airport. The investigation was based on the Citymobil reference scenario of the Heathrow airport PRT Demonstrator between Terminal 5 and a car park, but modified to allow investigation of a wider area including links to Terminal 1 to 3 and car parks. A predictive demand management strategy was identified to be incorporated in both local vehicle dispatching and wide network operation. The impacts of the proposed management strategy were evaluated using a microscopic PRT simulation model – Hermes which was specially modified for this research by the developer. Simulation results suggested a significant reduction in average waiting time across all stations in the test network as a result of the implementation of the proposed traffic management strategy. Despite the fact that PRT is a demand responsive service with its operation optimised through a control algorithm, this investigation has shown that the service level can be further improved by incorporating traffic management measures such as demand prediction.

## INTRODUCTION

PRT (Personal Rapid Transit) is a public transport system that uses small automated vehicles running on a fully segregated guide-way and with off-line stations (e.g. Lawson, 2003; Lawson, 2005; Cottrell, 2008). The vehicles can run directly from origin to destination with no intermediate stops. PRT represents an automated transport system that has evolved from the point of view of a public transport operator rather than an automobile manufacturer.

The primary objective of PRT is to provide an alternative PT (public transport) system that is demand responsive and particularly attractive to users because it offers minimal waiting times and a travel experience that is very close to travelling by private car or taxi. In principle, PRT systems can also be used to transport goods.

Past research on PRT has mainly been focused on the control of PRT vehicles and optimisation of PRT networks (e.g. Won, et al., 2006; Kai, et al., 2005; Anderson, 2003), i.e.,

operational management. Research on the traffic management issues of PRT system has hardly been reported. This paper presents a simulation study of the impacts of potential traffic management strategies for PRT systems. It is revealed that traffic management is a viable approach to further enhance operation efficiency of PRT systems.

## TRAFFIC MANAGEMENT OPPORTUNITIES

The traffic management opportunities with PRT are related to the efficient operation of vehicles in a segregated and closed network. They are similar to metro systems, which also do not interact with conventional traffic; but are different because the stations and stops are off-line in a PRT system, so that vehicles can travel direct to the destinations specified by the passengers without making any intermediate stops.

There are many types of traffic management measures for traditional transport system such as incident management, route guidance and priority control (e.g. Wren, et al., 1996; Gartner, et al., 2002; Bretherton, et al., 2002). PRT systems have their own unique routing control algorithms and all PRT vehicles are treated with equal priority. Thus, some traditional traffic management measures are not applicable to the PRT systems. However, one of the most important traffic management measures, that of demand management, is relevant because PRT systems are driven by the demands at stations across the networks. Although demand is determined by external factors, it can be predicted based on historical patterns and demand generation mechanisms. Demand management based on prediction allows the PRT control system to actively respond to changing demands thus improving operational efficiency.

Based on the analysis of traffic management opportunities, a predictive demand management strategy was proposed and traffic management strategies incorporating both local and network predictive demand are developed. The objective was to reduce waiting times at stations by guiding the PRT operations based on predictive demand rather than responding in a reactive way. As PRT is a traffic responsive system, time delay is unavoidable in response to a local demand arising at a station or overall network demand change. By running the system based on predictive demand, it is possible to reduce system response delay. It is expected that if the actual traffic demand is close to the predicted traffic demand, then optimised vehicle dispatching and relocation can be realised with minimum response time.



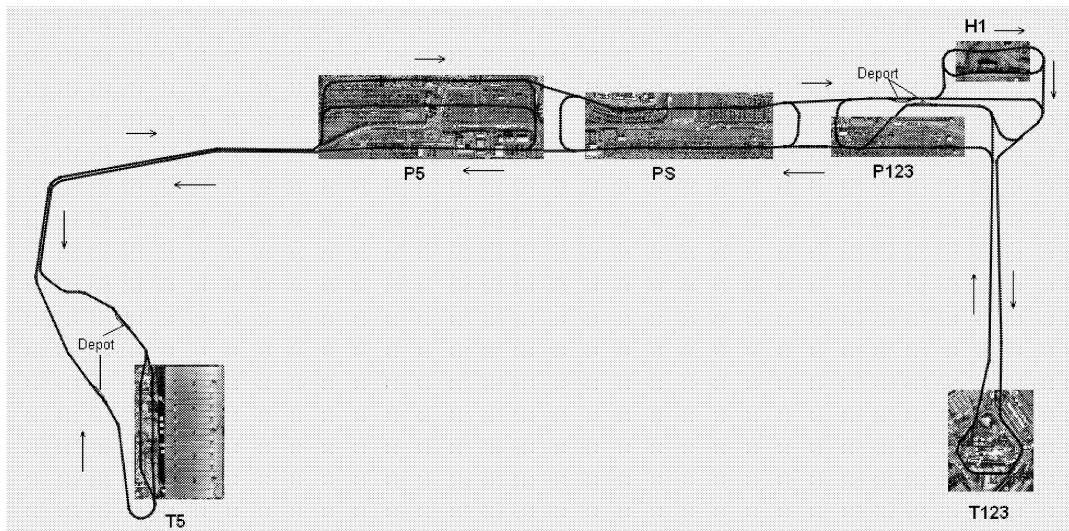


Fig. 1: PRT network investigated in Simulation

However, if actual demand is not consistent with the predicted demand, then either resources may be wasted (demand lower than predicted) or the quality of service may not be improved (demand greater than predicted) because part of the demand will have to be met in a responsive mode.

Two demand management strategies were considered in this investigation. The first is a 'local demand prediction tactic' which predicts local demands at a station within a time window of travel time for an empty vehicle to be dispatched from a depot (which provides empty vehicles to the station in question) to the station. The second is a 'long term demand prediction strategy' which estimates traffic demand changes for near future for all stations (5~10 minutes) based on long term trends in demand.

## SIMULATION STUDY

A simulation study was carried out to investigate predictive demand management strategies for PRT operation. Traffic management strategies incorporating the predictive demand with a demand responsive traffic service were tested. The investigation was carried out at two different levels. At the vehicle dispatching level, a local demand prediction tactic was used. At service level, a predictive demand program was examined.

A dedicated PRT simulation model, PRT Hermes (Xithalis, 2008) was used in the investigation, as it allowed the implementation of a demand program to simulate different demand inputs. Additionally, Hermes has the functionalities necessary to model PRT networks with different system specifications.

The investigation required some minor modifications to the software in order to implement the demand management strategies. Following correspondences with the developer of Hermes, necessary modifications to the simulation model were made to allow demand to be specified in two separate demand programs:

- 1) Demand: the PRT operation (vehicle dispatching and relocation etc.) within the simulator is calculated based on this demand
- 2) Actual demand: the actual generation of travel

demand is calculated based on this demand.

The developer has also modified Hermes to allow an OD-matrix representation of demand which allows travel demand between stations to be conveniently represented in the simulation. The modified version of Hermes is capable of reproducing PRT operations under both demand responsive and predictive demand responsive modes, and the modified Hermes was used for the investigation of PRT operation under the proposed traffic management strategies.

A 'local demand prediction tactic' was incorporated in the Hermes simulation, which is used for empty vehicle prediction in vehicle dispatching. The general logic, as described by the author of the Hermes simulation model, can be summarised as follows:

- Each station has a source capacitor (depot) which provides empty vehicles.
- Assuming it takes  $x$  minutes for an empty vehicle to travel from the source capacitor to a station.
- The prediction algorithm constantly predicts the needs for empty vehicles at the station  $x$  minutes in advance taking the following into account:
  - The number of vehicles (full and empty) currently present in the station berths
  - The number of vehicles (full and empty) currently underway towards the station that will arrive within the next  $x$  minutes.
  - The number of passengers that are expected to arrive within the next  $x$  minutes, based on current average rate of arrivals
- The number of spare vehicles that should be available at the station after all passengers get serviced

The demand prediction for vehicle dispatching is very much for the short term only, i.e., within the travel time between a depot and a station. The operation of PRT with and without this 'local demand prediction tactic' was investigated by comparing passenger waiting times at stations

The 'long term demand prediction strategy' was focused on the daily changes in demand, e.g. demand rising before the morning peak and falling after the evening peak. By running a PRT system on predictive demand, resources such as



relocation of vehicles (e.g. at different depots) can be pre-arranged to meet an anticipated increase or decline in the demand, which may improve the responsive service offered by PRT. It is expected that if the actual passenger demand is close to the predicted demand, then the optimized vehicle relocation should be best. If the actual demand is not consistent with the predicted demand, then either resources will be wasted (demand lower than predicted) or the quality of service will be poor (demand greater than predicted). The operation of PRT based on both predictive and non-predictive demands was examined in the simulation study.

The PRT network investigated in the research is shown in Figure 1. An extended network was used which consists of 23 stations serving four terminals, the main car parks and a hotel to the north of the airport. The network consists of 16505 m of guide-way. The distribution of the stations at different locations is summarised in Table 1. Terminal 5 and Terminals 1-2-3 together have four and three stations respectively, whilst the car parks have 14 stations. There are 4 depots, two close to Terminal 5 and another two close to Terminal 1-2-3 (as shown in Figure 1).

Table. 1: Stations in the Simulated Network

| Label on map | Stations               | Description                        |
|--------------|------------------------|------------------------------------|
| T5           | 1, 2, 3, 4             | Terminal 5                         |
| T123         | 5, 6, 7                | Terminal 1, 2 and 3                |
| P5           | 8, 9, 10, 11, 12       | Car park close to Terminal 5       |
| PS           | 13, 14, 15, 16, 17, 18 | Shared car park by all             |
| P123         | 19, 20, 21             | Car park close to Terminal 1, 2, 3 |
| H1           | 22, 23                 | Hotel                              |

A travel demand profile with typical morning and afternoon peaks was constructed to mimic travel demand patterns between terminals, car parks and the hotel. The base demands between them are shown in Table 2. The average total demand over the network is 5355 vph.

Table. 2: Base Demand between Destinations (vph)

|      | T5   | T123 | P5   | P123 | PS  | H1  | Sum  |
|------|------|------|------|------|-----|-----|------|
| T5   | 0    | 360  | 1000 | 24   | 12  | 40  | 1436 |
| T123 | 120  | 0    | 15   | 900  | 135 | 60  | 1230 |
| P5   | 900  | 30   | 0    | 60   | 30  | 10  | 1030 |
| P123 | 48   | 990  | 60   | 0    | 36  | 60  | 1194 |
| PS   | 24   | 135  | 30   | 36   | 0   | 30  | 255  |
| H1   | 24   | 66   | 10   | 60   | 30  | 20  | 210  |
| Sum  | 1116 | 1581 | 1115 | 1080 | 243 | 220 | 5355 |

The total demand over the network across a day is shown in Figure 2, which is expressed in number of vehicles rather than number of passengers. Depending on the capacity of the vehicle (2 passengers in Hermes), a group of passengers with the same destination may share a ride. In this investigation, only vehicle demand was considered. It is assumed that a vehicle will be dispatched no matter whether one or more passengers request it.

The maximum total demand is about twice the base demand. The demand profile is constructed to mimic a typical two-peak everyday traffic profile. It should be noted that the

profile is indicative and is not intended to reflect actual timings of passenger demand peaks. Overall, the profile covers low demand, high demand and transitional demand scenarios as are to be expected in real life.

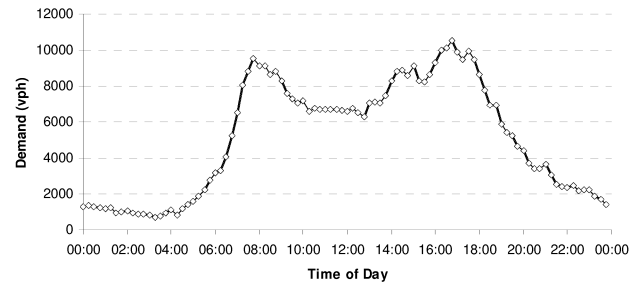


Fig. 2: Total Demand Profile in Simulation

For the investigation of the predictive demand management strategy, simulations were run for 6 different demand situations as follows:

- Low Demand 00:00~04:55
- Rising Demand 05:00~06:55
- Morning Peak 07:00~08:55
- Medium Demand 09:00~15:55
- Afternoon Peak 16:00~17:55
- Falling Demand 18:00~21:55

This was considered to be necessary as the predictive demand management strategy might perform differently at different demand situations. For example, it can be expected that a predictive demand strategy is not effective when travel demand is stable because then, predictive demand is always similar to the current demand. Predictive demand should be more effective if clear trends such as a rising or falling demand are evident

The basic settings used for the simulations are:

- Vehicle speed 10m/s (36 km/h), which is believed to be close to the operation speed of the Heathrow PRT system
- The minimum headway between PRT vehicles is 0.5 second, which is believed to be practical technically, and is able to accommodate the demand profile proposed in this investigation.

The effects of both the local demand prediction tactic for vehicle dispatching and the long term demand prediction strategy were then investigated based on the above simulation settings for the PRT network, traffic demand and traffic management strategy.

## RESULTS

For the local demand prediction tactic investigation, a with-and-without prediction approach was taken over the whole period of the demand profile. The results are shown in Figure 3. The average waiting time across all stations was reduced from 29 seconds to 26, e.g. by 11.5% with the local demand prediction tactic implemented. The difference is statistically significant ( $t=2.80$ ,  $p<0.05$ ), so the introduction of demand prediction in vehicle dispatching can lead to a small but significant reduction in passenger waiting times at stations,

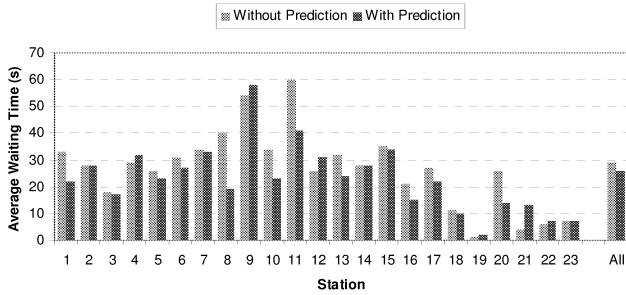


Fig 3: Average Waiting Times at Stations

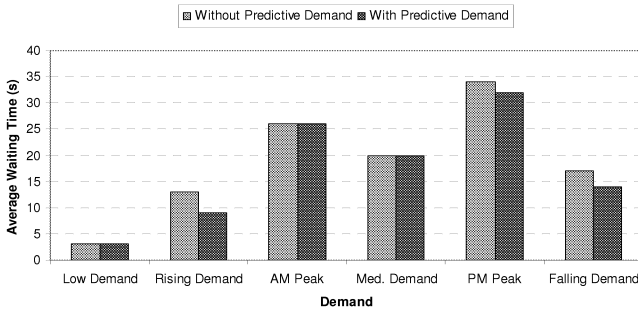


Fig 4: Average Waiting Times for Different Demand Periods

For the long term demand prediction strategy investigation, a with-and-without predictive demand approach was adopted based on the six demand situations described above. The results are shown in Figure 4. The maximum number of vehicles used for AM and PM peaks are 1020 and 1117 vehicles respectively. It can be observed from the figure that the average waiting times remain unchanged for low demand, medium demand and AM peak situations. This may be explained by that fact that travel demand in these periods is relatively stable, and predictive demand is basically the same as the current demand. In these conditions, no advantages in terms of waiting time reduction can be expected from predictive demand. On the other hand, reductions of waiting times can be found from applying the predictive demand strategy in rising, falling and PM peak demand situations. This is clearly an indication that the long term demand prediction strategy is helpful to the operation of PRT when traffic demand is varying and showing a clear trend. It should be noted that the PM peak demand (16:00~17:55) shows an overall trend of falling demand in this investigation. This may explain why the predictive demand strategy is effective for the PM peak scenario. However, the t-test reveals that the reduction in average waiting time for this scenario is not statistically significant ( $t=0.61$ ,  $p>0.05$ ).

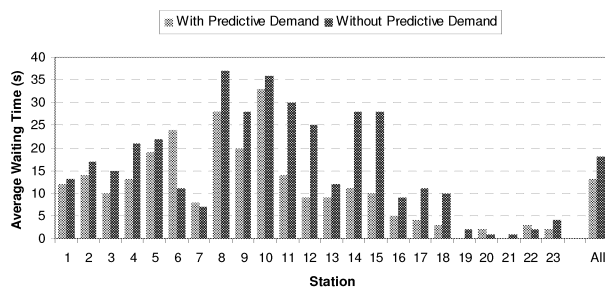


Fig 5: Comparison of Waiting Times for Rising Demand

Station-by-station comparisons of waiting times for simulation runs with and without the long term demand prediction strategy are shown in Figure 5 and Figure 6 for rising and falling demand situations respectively. In both situations, average waiting time is significantly less. For the rising demand situation, the average waiting time at all stations is about 5 seconds or 26% less with the predictive demand strategy than without. The difference is statistically significant ( $t=3.46$ ,  $p<0.05$ ). For the falling demand situation, the average waiting time at all stations is about 4 seconds or 20% less with the predictive demand strategy. The difference is statistically significant ( $t=2.80$ ,  $p<0.05$ ).

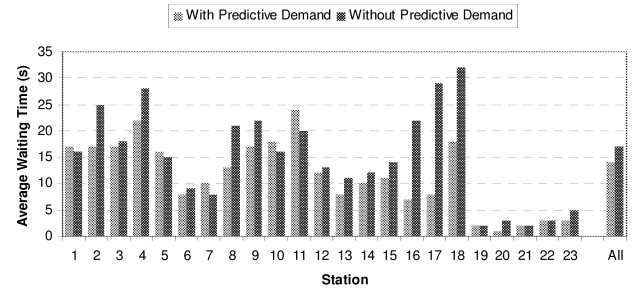


Fig. 6: Comparison of Waiting Times for Falling Demand

It is clear from the simulation results that introduction of a local demand prediction tactic in vehicle dispatching and the introduction of a long term demand prediction strategy in resource distribution can significantly reduce passenger waiting time at stations, and thus improve PRT operation and level of service overall.

## CONCLUSION AND DISCUSSION

PRT is a demand-responsive transportation service where vehicles are dispatched in response to demand in real-time. Whilst the operational efficiency of the PRT system can be improved by proper network design (e.g. topology of tracks, arrangement of stations and depots) and optimal vehicle dispatching and control algorithms, traffic management strategies for PRT system are explored and investigated in this work package.

A traffic management strategy for PRT operation – based on predictive demand management - was proposed and tested using the Hermes simulation model. The investigation was carried out at both local level for vehicle dispatching (using a local demand prediction tactic) and network operational level (using a long term demand prediction strategy) to quantify the effects of the 2 methods on the operation of PRT system based on an extended Heathrow network.

With the local demand prediction tactic for vehicle dispatching, demand for empty vehicles at a station were predicted using a local demand prediction algorithm. This tactic would allow vehicles to be dispatched in advance to cover the travel time between a depot and a station and thus reduce passenger waiting time at a station.

With the long term demand prediction strategy for network operation, passenger demand for the PRT network was predicted using a global demand prediction algorithm. This

strategy would allow the network operation (e.g. relocation of vehicles, number of vehicles in operation) to be ready for rising/falling demands in advance and thus reduce passenger waiting time across the network.

Simulation results revealed a significant reduction of 11.5% in average waiting time across all stations in the test network as a result of using the local demand predication tactic. The effects of using the long term demand prediction strategy are significant for scenarios with both rising and falling travel demands, with reductions in average waiting time of 26% and 20% respectively. When overall network demand is stable, waiting time is not significantly different for scenarios with and without the long term demand prediction strategy. This is expected as demand and predictive demand for the network are basically identical when network demand is stable.

It can be concluded, based on the simulation results that both the local demand prediction tactic for vehicle dispatching and the long term demand prediction strategy for network operation can enhance the service level of a PRT system by significantly reducing the average waiting time at stations. Despite the fact that PRT is a demand responsive service, the service level can be improved by traffic management measures such as those investigated in this research.

It should be noted that the investigation was based on the assumption that there are no limits on the number of vehicles which can be put into service. In a real system, it may not be possible to provide this maximum number of vehicles required because of constraints in initial investment. However, in the long run, it is usually the network capacity rather than number of vehicles which limit the capacity of a transport service, especially in situations where demands are high (e.g. as in this investigation). The assumption of unlimited vehicle provision may therefore be realistic for a well developed PRT system. As the average waiting time for PRT system is low (with sufficient vehicle supply), the absolute reduction in average waiting time is relatively small although the percentage is high. However, it should be noted that such reduction in average waiting time could be manifested as big waiting time reductions in some extreme cases (e.g. long waiting passengers). The simulation investigation may help understand and design a system so as to limit the worst case experiences.

It is not possible to implement the proposed traffic management strategies independent of the PRT operation as predictive demands need to be incorporated into the PRT core operation algorithm in order to influence operations such as vehicle dispatching, vehicle relocation and vehicle provision etc. Meanwhile, significant deviation between predictive and actual demands can result in reductions of operational efficiency of the PRT system. In this investigation, the network demand is predicted using a historical profile which is believed to be realistic. There may be occasions when abnormal network demands occur. In such cases a supplementary traffic demand sensing mechanism and prediction algorithm may be required to ensure travel demand can still be predicted with reasonable accuracy.

Whilst traffic management strategies may have wider impacts on the PRT operation, certain impacts, such as journey time, safety, accessibility, vehicle occupancy and trips made, are more associated with system and service design rather than traffic management. Journey time will be determined by design speed (with the synchronous control system used in HERMES, congestion is not possible on the track. excessive demands will be kept on stations and revealed as increased waiting time). Safety is a combination of vehicle design such as protection to passengers, and system design such as operating speed and hardware/software reliability. Accessibility is determined by network design and station/vehicle design. Vehicle occupancy and trips made are mainly determined by vehicle capacity and demands at stations. These are unlikely to be affected by the proposed traffic management strategy.

## ACKNOWLEDGEMENT

We would like to thank Christos Xithalis, the developer of the Hermes simulation model, for providing and making relevant modifications to the model to accommodate the simulation reported in this paper.

## REFERENCES

1. Anderson, J. E. 2003. "Control of Personal Rapid Transit Systems." *Teletronikk*, Vol. 99, No. 1, 108-116
2. Bretherton, D., Bowen, G., Wood, K. 2002. 'Effective urban traffic management and control - SCOOT VERSION 4.4'. *Proceedings of European Transport Conference Proceedings*. Cambridge
3. Christos Xithalis, 2008, PRT Hermes ([http://students.ceid.upatras.gr/~xithalis/index\\_en.html](http://students.ceid.upatras.gr/~xithalis/index_en.html))
4. Cottrell, W. D. 2008. "New-generation Personal Rapid Transit Technologies – Overview and Comparison" *Transportation Research Record*, Vol. 2042, 101-108
5. Gartner, N. H.; Pooran, F. J.; Andrews, C. M. 2002. 'Optimise policies for adaptive control strategy in real-time traffic adaptive control systems: implementation and field testing.' *Transportation Research Record*. 2002. Vol. 1811, 148-156.
6. Johnson, R. E. 2005. "Doubling Personal Rapid Transit Capacity with Ridesharing" *Transportation Research Record*, Vol. 1930, 107-112
7. Kai, C.; Matsuda, R.; Yano, Y.; Hamamatsu, Y. 2005. "An Analysis of Double Queue on PRT System and a Control Strategy for Merging" *Transactions of the Institute of Electrical Engineers of Japan*, Vol. 125-D, No. 6, 645-651
8. Lowson, M. 2003. "New Approach to Effective and Sustainable Urban Transport." *Transportation Research Record*, Vol. 1838, 42-49.
9. Lowson, M. 2005. "Personal Rapid Transit for Airport Applications" *Transportation Research Record*, Vol. 1930, 99-106
10. Won, J.; Choe, H.; Karray, F. 2006. "Optimal Design of Personal Rapid Transit" *Proceedings of IEEE Intelligent Transportation Systems Conference*, pp. 6
11. Wren, A. and Laughlin, K. 1996. 'Hampshire County Council, UK, ROMANS: road management system for Europe.' *Proceedings of European Transport Conference*

# MICROSIMULATION MODELS IN INTERMODAL CONTAINER TERMINALS: ORDINARY AND PERTURBED CONDITIONS

Vincenzo Assumma

Antonino Vitetta

Department of Computer Science Mathematics Electronics and Transportation

University Mediterranea of Reggio Calabria

Via Graziella Feo di Vito – 89060 Reggio Calabria

Italy

E-mail: vitetta@unirc.it - vincenzoassumma@unirc.it

## KEYWORDS

Model, Microsimulation, Design.

## ABSTRACT

This paper proposes a discrete event simulation model to simulate and design loading/unloading operations in an intermodal container terminal by means of microscopic stochastic models; the method consists in specifying the models represented by nodes, links and times for the single operations, so as to be able to control critical events, namely the possibility of breakdown or failure that affects ordinary operations in an intermodal container terminal. Each model is specified by definition of the variable and is described in terms of input and output. From the simulations of the models, we derive the output, essential to assess system utility, expressed by loading/unloading operation times.

## STATE OF THE ART

The significant development in recent years of container transport, in terms of technology and organization, has led to profound changes in international transportation.

The basic aim in intermodal container terminal management is to reduce train waiting times from arrival to departure by minimizing loading/unloading times. This problem has already inspired many researchers to develop innovative loading/unloading systems in container terminals.

In the last few years discrete-event simulation models in the object-oriented approach have been proposed as decision support systems for intermodal container terminals. The object-oriented approach has been particularly useful to describe logistic integration of the main system terminal modules.

Methodology to represent the behavior of real systems can be traced back to two different approaches: simulation or "what if" (what happens if); design or "what to" (what to do for). Simulation differs from design because in design the various scenario hypotheses are produced through models and are analyzed and assessed with an optimization procedure in respect of certain constraints. Further analysis can be found in Assumma et al. (2006).

Simulation models can be classified by variables such as (Kelton et al. 2004):

- static/dynamic: in dynamic models time is an independent variable;

- continuous/discrete: in a continuous model the system state can continually change in time, while in a discrete model changes occur only in certain instants; discrete models can be classified into discrete time or discrete event. In discrete time models, time is a succession of intervals of fixed length; in discrete event models the state's changes are recorded as events happen;
- deterministic/stochastic: in stochastic models, random variables and probability distributions are also inputs.

Simulation models (Lieberman and Rath 1997) can also be classified by level of detail:

- macroscopic models simulate network performances by means of aggregate variables (speed, density, flow) and use a continuous representation of traffic;
- mesoscopic models simulate network performances (aggregate variables with explicit capacity are used), but differ in terms of traffic representation; the peculiarity of mesoscopic models consists in a discrete flow representation for groups of vehicles (users);
- microscopic models: individual trajectories of all the vehicles or users are simulated by using disaggregate variables with implicit capacity, and a discrete traffic representation.

High-fidelity microscopic models faithfully reconstruct reality but are expensive to develop and run due to the complexity of the logical relationships and number of parameters.

In the literature models are also classified by intermodal terminal operation:

- arrivals models (Legato and Mazza 2001; Gaudio et al. 2001; Kia et al. 2002; Dai et al. 2004; Imai et al. 2001; Imai et al. 2003; Nishimura et al. 2001);
- management models (Holguin-Veras and Jara-Diaz 1999; Kim and Kim 1998; Kim and Kim 2002; Fortino et al. 2001; Kozan and Preston 1999; Kozan 2000; Ng 2005; Yun and Choi 1999; Bielli et al. 2005; Degano 2003);
- loading and unloading models (Bontempi et al. 1997; Zaffalon et al. 1998; Mastrolilli et al. 1998; Kim et al. 2002; Kim and Park 2004; Corry and Kozan 2005);
- integrated models (Bruzzone et al. 1999; Rizzoli et al. 2002; Gambardella et al. 1996; Gambardella et al. 2001; Gambardella and Rizzoli 2002; Shabayek and Yeung 2002; Liu et al. 2002; Hartmann 2004).

These models can be identified by their characteristics (tables 1-2); they are mostly dynamic, discrete and stochastic.

Table 1: Characteristics of terminal simulation models

|                              | Static<br>Dynamic | Continuous<br>Discrete | Deterministic<br>Stochastic | Microscopic<br>MAcroscopic |
|------------------------------|-------------------|------------------------|-----------------------------|----------------------------|
| Bielli et al. 2005           | D                 | D                      | S                           | MI                         |
| Bruzzone et al. 1999         | D                 | D                      | S                           |                            |
| Carteni et al. 2005          | D                 | D                      |                             | MI                         |
| Degano 2003                  | D                 | D                      | D                           |                            |
| Fortino et al. 2001          |                   |                        |                             |                            |
| Legato and Mazza 2001        | D                 | D                      | S                           |                            |
| Gaudioso et al. 1999         |                   |                        |                             |                            |
| Gambardella et al. 1996-2001 |                   |                        |                             |                            |
| Bontempi et al. 1997         | D                 | D                      | S                           |                            |
| Mastrolilli et al. 1998      |                   |                        |                             |                            |
| Hartmann, 2004               | D                 |                        | S                           |                            |
| Kia et al. 2002              |                   | D                      | S                           |                            |
| Liu et al. 2002              |                   |                        | D                           | MI                         |
| Rizzoli et al. 2002          |                   |                        |                             |                            |
| Gambardella and Rizzoli 2002 | D                 | D                      | S                           |                            |
| Shabayek and Yeung 2002      |                   |                        | S                           | MA                         |
| Yun and Choi 1999            | D                 |                        | D                           |                            |

Table 2: Characteristics of terminal design models

|                         | Exact<br>Approximate | Deterministic<br>Stochastic |
|-------------------------|----------------------|-----------------------------|
| Corry and Kozan 2005    | A                    | S                           |
| Imai et al. 2001        |                      |                             |
| Nishimura et al. 2001   | A                    | S                           |
| Imai et al. 2003        | E                    | D                           |
| Kim and Kim 1999-2002   | A                    | S                           |
| Kozan and Preston 1999  |                      |                             |
| Kozan 2000              | A                    | S                           |
| Preston and Kozan 2001  |                      |                             |
| Murty et al. 2005       | A                    | S                           |
| Zaffalon et al. 1998    |                      |                             |
| Gambardella et al. 1998 |                      |                             |
| Mastrolilli et al. 1998 | E                    | S                           |
| Rizzoli et al. 1999     |                      |                             |

Recently, a state-of-the-art container terminal was proposed by Vis and Koster (2003) and Steenken et al. (2004), with a description and classification of the main logistic processes and operations in container terminals and some methods for their optimization. Microscopic, statistic and discrete events models are most widely used for the experimental case proposed.

After a short introduction to the problem and an updated review of recent models in the literature, our paper contains:

- simulation and design methodology and models used;
- application to land-side and sea-side operations in a RO-RO High Speed Services intermodal container terminal;
- application to land-side operations in the intermodal container terminal at the Termini Imerese freight village;
- conclusions and potential developments.

## MODELS

A port container terminal constitutes a fundamental connection in the chain of a transport container. The characteristics of the port container terminal are determined by the development of innovative loading/unloading systems, such as integration between road transport and RO-RO High Speed Service transport ships by using rail systems for loading/unloading containers. Minimization of operation times in the maritime node is a fundamental factor for justifying the use of RO-RO High Speed Shipping Services

for freight transport, as against the vessels currently used. The characteristics of loading/unloading areas also depend on ship design, Loading Units (LUs) and freight type.

The simulation model of a port container terminal consists of a set of submodels, namely:

- road network entry/exit model;
- entry buffer model;
- storage area model;
- berth model;
- ship arrival/departure model.

This paper proposes:

- a land-side berth operations model to simulate loading/unloading operations in ordinary and perturbed conditions;
- a sea-side berth operations model to simulate loading/unloading operations in perturbed conditions, making reference to the ordinary conditions simulated in Serranò and Vitetta (2006).

The methodologies to represent the behavior of real systems can be traced back to two different approaches: simulation or "what if"; design or "what to". Construction of such models allows system operations to be simulated and verified both under ordinary and emergency conditions. This is possible because the approach followed is microscopic and it is therefore possible to represent every activity and every functional relationship with a high level of detail and intervene on the single element to verify particular conditions.

## Simulation model

Microscopic simulation methodology is based on a "what if" approach, schematized in Serranò and Vitetta (2006). In this paper dynamic, discrete events and stochastic models are specified.

While the advent of management information services and data processing has greatly improved the ability of terminal managers to control the whole process, raw data still have to be analysed to provide some insight into the performance of terminal operations.

In many cases simulation models have proved to be a reliable and convenient tool to support decision makers in their daily operations. They provide a test-bed to assess the validity of management policies and can be used to point out problems such as conflicts in resource allocation and terminal space management. Such a simulation scheme is represented in figure 1.

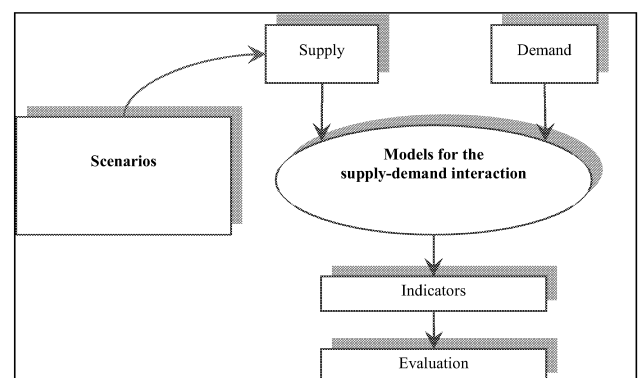


Figure 1: Simulation scheme ("what if")

The objective of the model is to determine a generalized cost function (total distance of the container for loading/unloading operations):

$$f(X, Y) = \sum_{j=1}^M \sum_{i=1}^N \left[ \left| (X_{ij} - X_{t_{ij}}) \right|_{loading} + \left| (Y_{ij} - Y_{t_{ij}}) \right|_{loading} + \left| (X_{ij} - X_{c_{ij}}) \right|_{loading} + \left| (Y_{ij} - Y_{c_{ij}}) \right|_{loading} \right] + \sum_{j=1}^M \sum_{i=1}^N \left[ \left| (X_{ij} - X_{t_{ij}}) \right|_{unloading} + \left| (Y_{ij} - Y_{t_{ij}}) \right|_{unloading} + \left| (X_{ij} - X_{c_{ij}}) \right|_{unloading} + \left| (Y_{ij} - Y_{c_{ij}}) \right|_{unloading} \right]$$

The variables are:

- number (j = 1, ..., M) of arriving/departing trains;
- number (i = 1, ..., N) of loading units (LUs);
- positions (X<sub>ij</sub>, Y<sub>ij</sub>) of the container in the storage area for loading/unloading operations.

The constraints are:

- positions (X<sub>t<sub>ij</sub></sub>, Y<sub>t<sub>ij</sub></sub>) of LUs in arriving/departing trains (fixed);
- positions (X<sub>c<sub>ij</sub></sub>, Y<sub>c<sub>ij</sub></sub>) of LUs in arriving/departing trucks (fixed);
- type, number and velocity of vehicles used to move containers in the storage area;
- configuration of the storage area.

## Design model

In design models the various scenario hypotheses are produced through models and are analyzed and assessed with an optimization procedure in respect of certain constraints.

The objective of the model is minimization of a generalized cost function (total distance of the container for loading/unloading operations):

$$F.O. = \operatorname{argmin} f(X, Y)$$

where the positions (X<sub>ij</sub>, Y<sub>ij</sub>) of the container in the storage area for loading/unloading operations are not fixed.

## Perturbed conditions

Perturbed conditions concern the possibility of breakdown or failure that affects ordinary operations in the intermodal container terminal. In the case of delays due to decelerations of operations there is a variation in loading/unloading times; in the case of delays due to breakdown on, for instance, a railway, the loading/unloading conditions must be varied, hence the storage area organization.

The simulation model proposed in this paper contains an explanation, by means of a microscopic stochastic disaggregate model, with time represented by a Gamma distribution function, of perturbed conditions to control critical events in loading/unloading operations in a RO-RO High Speed Services intermodal container terminal.

Let total time T calculated under perturbed conditions for loading/unloading operations be

$$T = T_{det} + T_{stoc}$$

where

- T<sub>det</sub> is the deterministic time calculated under ordinary conditions;
- T<sub>stoc</sub> is the stochastic time calculated under perturbed conditions; this time is represented through a stochastic

variable distributed according to a density probability function, f(T, α, β), Gamma

$$f(T, \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} T^{\alpha-1} e^{-\frac{T}{\beta}}$$

$$\alpha \beta = E(T)$$

$$\alpha \beta^2 = \operatorname{var}(T)$$

with

- α a real positive form parameter
- β a real positive scale parameter
- $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$
- average = αβ
- variance = αβ<sup>2</sup>

A density probability function Gamma is used since:

- by definition it is not negative;
- for α = 1 Gamma distribution is the exponential distribution.

T<sub>stoc</sub> can assume different values as a function of coefficient of variation C<sub>v</sub>. This coefficient, defined by the relation between standard deviation and average, describes the variability of the model, i.e. the possibility of breakdown or failure that affects ordinary operations in the port container terminal.

In particular, three cases are examined:

- C<sub>v</sub> = 0.1
- C<sub>v</sub> = 1.0
- C<sub>v</sub> = 2.0

In the case of delays due to decelerations of operations there is a variation in loading/unloading times, hence a low variation coefficient; in the case of delays due to breakdown, for instance on a railway, the loading/unloading conditions must be varied, hence the storage area organization, resulting in a high coefficient of variation.

## APPLICATION

This paper proposes two applications for the discrete event simulation model: a RO-RO High Speed Services intermodal container terminal and the Termini Imerese freight village.

### Ro-Ro high speed services intermodal container terminal

Occupation of the storage area by trains and trucks must be established. The necessary input data are the following:

- position of trucks (a queue of trucks is hypothesized as waiting for loading/unloading, all at the same point);
- position of the trains on which the containers are loaded/unloaded by trucks;
- number of unloaded containers and destination trains;
- position of every container on the destination train.

The following are hypothesized:

- absence of congestion among the vehicles in the storage area;
- only one vehicle to move the containers in the storage area (reach stacker);
- the runs actually performed by the reach stacker are approximated to straight lines consisting of segments parallel and orthogonal to the berth.

In this paper we adopt a Cartesian orthogonal system to represent a hypothesis of the storage area occupied by trucks

and trains waiting for loading/unloading: the center O coincides with the center of the trucks and the ordinates are parallel to the trains; it is hypothesized that vehicles traveling on the rails are arranged in only one "train" with a capacity of  $12 + 10 + 8 = 30$  containers. Every free slot on the train will therefore have a distance ( $X_a$ ,  $Y_a$ ) from the truck.

For the simulation the following parameters were fixed:

- number and position of trucks and trains;
- number and distribution of containers;
- type and number of the vehicles required to move the containers.

Such variables constitute input data and the basic hypotheses above are non-binding. They can be modified during the various simulations in relation to the various organizational hypotheses of the terminal.

In this paper we consider a parallel sequence of loading/unloading operations, from arrival to departure of containers by ship and road.

To reproduce the real trial to which the container is subject once it arrives in the terminal we use a procedure implemented by commercial Rockwell Arena software. Arena is a microscopic discrete events simulator based on SIMAN language. It combines the utility of high-level simulators with the flexibility of general-purpose programming languages (Visual Basic, C). The step-by-step procedure implemented in Arena performs a series of instructions providing, at the end of simulation, the allocation of containers loaded/unloaded from trucks and trains.

The time for loading/unloading containers under ordinary conditions in land-side operations is:

$$T_{det}^{loading} = 33 \text{ min}$$

$$T_{det}^{unloading} = 33 \text{ min}$$

The time for loading/unloading containers under perturbed conditions in land-side operations depends on  $C_v$  values considered. Table 3 gives times under ordinary and perturbed conditions for loading/unloading containers in land-side operations.

Table 3: Time (min) in ordinary and perturbed conditions

|             | Loading   |            |     | Unloading |            |    |
|-------------|-----------|------------|-----|-----------|------------|----|
|             | $T_{det}$ | $T_{stoc}$ | T   | $T_{det}$ | $T_{stoc}$ | T  |
| $C_v = 0,1$ | 33        | 3          | 36  | 33        | 3          | 36 |
| $C_v = 1$   | 33        | 34         | 67  | 33        | 32         | 65 |
| $C_v = 2$   | 33        | 69         | 102 | 33        | 63         | 96 |

The time for loading/unloading containers under ordinary conditions in sea-side operations is:

$$T_{det}^{loading} + T_{det}^{unloading} = 26 \text{ min}$$

According to the  $C_v$  values considered, there are different values of  $T_{stoc}$ . Table 4 gives times for loading/unloading containers under ordinary and perturbed conditions in sea-side operations; another hypothesis introduced is the breakdown of a railway in the berth, such that there is only one rail platform useful for unloading trains from the ship.

Table 4: Time (min) for ordinary and perturbed conditions

|             | Ordinary  |            |    | Railway breakdown |            |     |
|-------------|-----------|------------|----|-------------------|------------|-----|
|             | $T_{det}$ | $T_{stoc}$ | T  | $T_{det}$         | $T_{stoc}$ | T   |
| $C_v = 0,1$ | 26        | 3          | 29 | 26                | 17         | 43  |
| $C_v = 1$   | 26        | 32         | 58 | 26                | 53         | 79  |
| $C_v = 2$   | 26        | 64         | 90 | 26                | 93         | 119 |

### Termini Imerese freight village

In the Termini Imerese freight village, occupation of the storage area by trains and trucks must be established. The necessary input data are the following:

- position of trucks (a queue of trucks is hypothesized as waiting for loading/unloading, all at the same point);
- position of the trains on which the containers are loaded/unloaded by trucks;
- number of unloaded containers and destination trains;
- position of each container on destination trains.

The following theoretical formulation is hypothesized:

- absence of congestion among the vehicles in the storage area;
- only one vehicle to move the containers in the storage area (reach stacker);
- the runs actually performed by the reach stacker are approximate to straight lines consisting of segments parallel and orthogonal to the berth.

For the simulation the following parameters were fixed:

- number and position of trucks;
- number and position of trains;
- number and distribution of containers in trucks (loading 40 LUs + unloading 40 LUs);
- number and distribution of containers in trains (loading 40 LUs + unloading 40 LUs);
- type and number of vehicles required to move the containers (only one reach stacker).

Such variables constitute input data and the basic hypotheses above are non-binding. They can be modified during the various simulations in relation to the various organizational hypotheses of the terminal.

To reproduce the real trial to which the container is subject once it arrives in the terminal we use a procedure implemented by commercial "Arena" software created by Rockwell Software.

In the simulation model the time for loading/unloading containers under ordinary conditions is:

$$T_{det}^{loading} = 326 \text{ min}$$

$$T_{det}^{unloading} = 323 \text{ min}$$

$$T_{det}^{loading} + T_{det}^{unloading} = 649 \text{ min}$$

The time for loading/unloading trains (loading 40 LUs + unloading 40 LUs) and for loading/unloading trucks (loading 40 LUs + unloading 40 LUs), using only one reach stacker, is 649 min. Hence the movement per hour of the reach stacker is 15 LUs/h.

In the design model the time for loading/unloading containers under ordinary conditions is:

$$T_{\text{det}}^{\text{loading}} = 303 \text{ min}$$

$$T_{\text{det}}^{\text{unloading}} = 300 \text{ min}$$

$$T_{\text{det}}^{\text{loading}} + T_{\text{det}}^{\text{unloading}} = 603 \text{ min}$$

The time for loading/unloading trains (loading 40 LUs + unloading 40 LUs) and for loading/unloading trucks (loading 40 LUs + unloading 40 LUs), using only one reach stacker, is 603 min. Hence the movement per hour of the reach stacker is 16 LUs/h.

After optimization of container positions in the storage area, the time for loading/unloading containers is lower than 49 minutes, hence around 10%.

## CONCLUSIONS

This paper proposed a discrete event simulation and design model for loading/unloading operations in an intermodal container terminal by means of stochastic models. Some interesting results emerged from combining deterministic and stochastic times under different scenarios. Total time for completing loading/unloading operations of containers inside the terminal varies according to any perturbed condition and the variation coefficient considered.

These results concern the simulation hypotheses, and are thus still subject to revision and integration, say, by modifying the terminal structure and handling equipment or type and quantity of containers.

## REFERENCES

- Assumma V., Serrano B. and Vitetta A. 2006. "Microsimulation models in a RO-RO High Speed Services intermodal container terminal." Internal Report no. 1-2006-LAST-DIMET-UNIRC
- Bielli M., Boulmakoul A. and Rida M. 2005. "Object oriented model for container terminal distributed simulation." *European Journal of Operational Research*
- Bontempi G., Gambardella L.M. and Rizzoli A.E. 1997. "Simulation and Optimization for Management of Intermodal Terminals." *European Simulation Multiconference 1997*, Istanbul, June 1-4
- Bruzzone A. G., Giribone P. and Revetria R. 1999. "Operative requirements and advances for the new generation simulators in multimodal container terminals." *Proceedings of the 1999 Winter Simulation Conference P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds*
- Carteni A., Cantarella G. E. and De Luca S. 2005. "A simulation model for a container terminal". *Proceedings of the 2005 European Transport Conference, Strasbourg*
- Degano C. 2003. "Modelling and Control of freight handling in Intermodal Terminals". Ph.D. Thesis, Research Doctorate in Automatics and Computer Sciences for Transportation Systems
- Fortino M. and P. Legato 2001. "Un simulatore dei processi di accumulo e rinvio dei container presso il terminale marittimo di Gioia Tauro". *Laboratorio di Logistica, DEIS, UNICAL 4/01*
- Gambardella L.M., Bontempi G., Taillard E., Romanengo D., Raso G. and Piermari P. 1996. "Simulation and forecasting in intermodal container terminal", *IDSIA*
- Gambardella L.M., Rizzoli A.E. and Zaffalon M. 1998. "Simulation and planning of an intermodal container terminal", *Simulation* 71 (2), 107-116
- Gambardella L.M. and Rizzoli A.E. 2002. "Agent-based Planning and Simulation of Combined Rail/Road Transport", *IDSIA*
- Gaudio M., Giallombardo G. and Legato P. 2001. "Ottimizzazione e simulazione nella gestione di un terminal container portuale: il caso Gioia Tauro". in G. E. Cantarella and F. Russo (Eds.), *Metodi e tecnologie dell'ingegneria dei trasporti*, Collana Trasporti, Franco Angeli
- Hartmann S. 2004. "Generating Scenarios for Simulation and Optimization of Container Terminal Logistics". *OR Spectrum* 26, pp.171-192
- Kelton W. D., Sadowski R. P. and Sturrock D. T. 2004. "Simulation with Arena", Third Edition, McGraw Hill
- Kia M., Shayan E. and Ghotb F. 2002. "Investigation of port capacity under a new approach by computer simulation". *Computers and Industrial Engineering* 42, 533-540
- Kozan E. and Preston P. 1999. "Genetic algorithms to schedule container transfers at multimodal terminals". *International transactions in Operational Research* 6 (3), 311-329
- Kozan E. 2000. "Optimising container transfers at multimodal terminals". *Mathematical and Computer Modelling* 31 (10-12), 235-243
- Legato P. and Mazza R. M. 2001. "Berth planning and resources optimization at a container terminal via discrete event simulation". *European Journal of Operational Research* 133, pp. 537-547
- Lieberman E. and Rathi A. K. 1997. "Traffic Simulation". In *Traffic Flow Theory*, Oak Ridge National Laboratory
- Liu C., Jula H. and Ioannou P. A. 2002. "Design, Simulation, and Evaluation of Automated Container Terminals". *Fellow, IEEE*, Vol. 3, no. 1
- Mastrolilli M., Fornara N., Gambardella L.M., Rizzoli A.E. and Zaffalon M. 1998. "Simulation for policy evaluation, planning and decision support in an intermodal container terminal". In *Merkuryev Y., Bruzzone A. and Novitsky L. (Eds.) Proceedings of the International Workshop Modeling and Simulation within a Maritime Environment*, Riga, Latvia, pp. 33-38
- Rizzoli A.E., Gambardella L.M. and Fornara N. 2002. "A simulation tool for combined rail/road transport in intermodal terminals". *IDSIA*
- Rockwell Software 2004. "Arena User's Guide". ID ARENA-UM001A-EN-P, United States of America
- Serrano B. and Vitetta A. 2006. "Microsimulation models in a RO-RO High Speed Services intermodal container terminal: sea-side operations", Internal Report no. 2-2006-LAST-DIMET-UNIRC
- Shabayek A.A. and Yeung W.W. 2002. "A simulation model for the Kwai Chung container terminals in Hong Kong". *European Journal of Operational Research* 140, pp. 1-11
- Steenken D., Voss S. and Stahlbock R. 2004. "Container terminal operation and operations research: a classification and literature review". *OR Spectrum* 26, 3-49
- Vis I.F.A. and De Koster R. 2003. "Transshipment of containers at a container terminal: an overview". *European Journal of Operational Research* 147(1), 1-16
- Yun W.Y. and Choi Y.S. 1999. "A simulation model for container-terminal operation analysis using an object-oriented approach". *Int. J. Production Economics* 59, pp. 221-230



# **HOSPITAL INFRA- STRUCTURE LOGISTICS**



# Quality of Service in Transplantation via The Electronic Medical Record

David Belo, Miguel Miranda, António Abelha, José Machado and José Neves  
Universidade do Minho, Departamento de Informática, Braga, Portugal  
{davidbelo,miranda,abelha,jmac,jneves}@di.uminho.pt

## ABSTRACT

Medical Informatics may sets itself as a new area of research open to share its practices, but with a predisposition to consider different but complementary computational paradigms and new methodologies for problem solving. Indeed, healthcare providers, namely the institutions on the public sector, with its physicians, nurses, administrative staff and patients present the right universe to consider and study. It is in this context that the use of Electronic Health Records changed the workflow of facilities which use heterogeneous, scheduled and unstructured methods of recording, in order to solve patient problems and to improve medical and clinical research and education. In this paper the Electronic Health Record of the Centro Hospitalar do Porto (one of the major portuguese healthcare facilities in the public sector) is presented. The focus is made on the integration and recording of different processes, being studied the case of the hepatic transplantation surgery department.

## Introduction

In the middle of the 20th century, the Healthcare Informatics Systems (HIS) were based in the communication between services. The clinical role was not directly involved. Lawrence Weed [1] gave the name of Problem Oriented Medical Record (POMR) to the first information indexation problem-oriented. The HIS with this feature permitted to access patient records much easier and faster than before. Three years later, the United States started to record its ambulatory services. As a result, the Electronic Medical Record (EMR) was born in healthcare facilities. Electronic devices permitted the use of structured forms to obtain standards in reports, making information easier to be dealt with [1][2][3][4]. With the increase of complexity and level of knowledge that can be extracted from medical data, paper medical records (PMR) obsolete and unproductive for education and research purposes. In the last decades, the enhancement of the accessibility and suitability of data have been provided by different implementations of EMR. This major breakthrough in medical records derives directly from the features it may provide, such as epidemiological studies, statistical analyses, financial inquiries and integrate different sorts decision support

[3][5][6].

As supporters of electronic reporting, the American College Surgeons (ACS) developed a database for analyses and comparison of surgical outcomes, the National Surgical Quality Improvement Program (NSQIP). This system was based on a Department of Veterans Affairs (VA) application that started in 1980s. The ACS NSQIP collects data on 135 variables, including preoperative risk factors, intraoperative variables, and 30-day postoperative mortality and morbidity outcomes for patients undergoing major surgical procedures in both the inpatient and outpatient setting [7].

Transplantation services could benefit with this method. The lack of organs and the difficulties of maintaining patients alive and with a reasonable quality-of-life while a viable organ arises, can be very expensive, bearing a great cost in economical and survival terms. Time is crucial and all the paperwork and candidate processing efforts should be minimized and properly automatized to accelerate the process of transplantation surgery. As an example, the time for the liver to be unfit for transplantation is approximately between 12 to 18 hours while the kidney between one or two days. Although the procedures and policies related to transplantation are designed to reduce patient and graft waiting time, EMR architectures enable a just in time and more reliable behavior, bearing even more efficient results if well implemented and integrated. The level of quality of service in this sort of services has a direct connection with patient survival [8].

As an example of a successful an effective implementation of an EMR architecture, the Centro Hospitalar do Porto E.P.E. (CHP) in Oporto has its EMR implementation (Processo Clínico Electrónico (PCE)) expanding throughout all its departments. One of the areas and objectives in development in the PCE project is to ensure that the liver transplantation services improve their overall quality evaluation, embedding in this solution the ACS-NSQIP variables and method of obtaining data [9][6].

## What is the EMR?

The EMR is an information system that uses electronic devices to generate structured and different views of the medical data. This information is acquired by healthcare staff and its value is automatically stored and dis-

tributed to where it is needed. Every document created within a specialized service (e.g. private clinic, hospital) respect this rules, making different and individualized departments closer [4][8].

### **Coding and Ordering**

This feature is very useful to link different data to one specific problem, as coded data is much easier to access and it is recommended for decision support using Artificial Intelligence (AI) [4][8].

The electronic ordering embedded in EMR can be used not only to obtain medical equipment or pharmacological prescriptions, but also for acquiring laboratory and imaging studies outside the service where it is used. Furthermore, it may enable the centralization of exam display, allowing different services to share results concerning the same patient, diminishing costs on unnecessary exams, and above all, improving the quality of service being provided [4][8].

### **Viewing and Access**

There are different access permissions when dealing with medical data. Although it can only be viewed by the authorized personnel from any terminal inside the healthcare facility or even on its own laptop or PDA, the access must be flexible in order to enable the professionals to access it when needed. In other words, the access to the medical information of the patient is as important in terms of privacy as in terms of significance for medical situations. On the other hand, interfaces must be intuitive and easy to use [4][8].

### **Messaging and Integration**

Messaging enables one to create, send and retrieve messages online. It may be very useful for handling data, images or even file exchange. Encryption and the right protocols of trading are also paramount [4][8]. Messaging systems are extremely important not only for the internal workflow in an healthcare institution, but as well as an essential component for the development of group work, namely in the area of diagnostic that is supported by decision support systems [6].

The EMR must, and indeed is a very flexible contender. It is a system that evolves and adapts by itself to the current and future demands. New features, interfaces and updates must be made without disrupting what have been built previously and the data must be read at any-time and anywhere. It must embrace the possibility of attaching files, such as images, shared databases, even web services or decision support systems on the fly[4][8].

Integration of the information from the different departments and services within healthcare institutions in order to make it available for the EMR system is also an

important prerequisite for an efficient EMR. If the generated information is not organized in order to prevent inconsistencies and gaps, the information entropy generated will draw back the efficiency and quality properties enabled by the EMR

### **Decision Support Systems**

If the EMR combines the input of data with a decision support system, for help in diagnosis, automatic prescriptions or treatment options, the data should be structured and organized. This feature is increasing in popularity so it is viable to consider. This technology is based on the use of AI based technologies and methodologies for problem solving, as well as those as Logistic Regression (LR) or even Artificial Neural Networks (ANN) for data analyses and pattern recognition [4][8]. Therefore, any free text that makes almost impossible for the application of AI techniques to perform any kind of data mining or knowledge extraction, may open the way to structured information, which can be afterwards used for medical and management studies.

### **Problem-Oriented**

One property of the EMR is that it can be easily associated with the POMR method. This connection is based on the development of an array of problems integrating every report, where therapy and diagnosis are associated with a specific medical record. POMR uses the SOAP philosophy [10], i.e. Subjective (patient's observations), Objective (Doctor's observations and tests), Analyses (Doctor's understanding of the problem), and Plan (Doctor's actions). This type of thinking revolutionized the way that physicians look at healthcare processes and made a major breakthrough in quality of healthcare delivery [2][10][11].

### **Implementation and Objective**

There is a description on how to implement and evolve the EMR, as it is described by Elberg [12], i.e. (1) Establishing the system basis, (2) Learning the organization and policies through EMR, (3) Linking the EMR and the active clinic systems and, (4) Maturing the EMR. The final objective for the EMR is to be extended to the whole medical facility [4][12].

### **Quality in the Healthcare Facility**

In order to ensure quality in surgery, it is imperative to reduce medical errors. 'Medical Error' is a term that applies to problems that occur in the healthcare environment, and it does not apply only to therapists or pharmaceutical errors. Its scope is much wider and it consists in human, logistical or machine miscalculation or mistake. An accurate and trustful reporting system

is the key to minimize these problems. Knowing the surgical outcomes and clinical practices, it is possible to envisage a way to eliminate systematic errors [13][14].

### Surgical Indicators

For an accurate analysis of the quality-of-service within an healthcare facility, it must be measured and evaluated. The famous sentence 'you can't improve what you can't measure' represents the use of quality indicators. There are three main groups of indicators to consider in this specific case, namely [15]:

**Structural indicators** , which denote the equipment and other resources in a healthcare environment;

**Process indicators** , which establishes a link between a specific process and its outcome; and

**Direct outcome indicators** , which stand for scores and rates of mortality, morbidity, outcome/expected outcome (O/E), length of stay in hospital or even costs.

### Risk Adjustment Models

After assuring data accuracy and reliability the next step is to analyze it. The objective of risk adjustment is to predict surgical outcomes from early patient and surgery data. The results can be used to score and be compared with other surgical outcomes. In this way it may be assured which are the best practices, healthcare institutions, departments or most common mistakes [16].

There are several risk adjustment models like Charleston Morbidity Index, more commonly called CCI, DxGC or even NSQIP risk adjustment model. To obtain the score it is necessary a statistical preparation, such as regression algorithms [17][18].

The NSQIP risk adjustment score is based on direct outcomes of O/E ratio in eight sub-specialties of surgery. Transplantation is not yet included. Anyway, the NSQIP is based on forms that apply a structured data that is computable with the analysis system. Information is input by a specified nurse. The results can be viewed in semi-annual reports or on-line reviews [7][19].

### Implementation

The CHP's EMR is based on the POMR methodology and it is being implemented in the transplantation service. The process considers different records, including [2]:

1. The creation of an individual data-base (according to the problem/patient);
2. A problem list with all the information about complications and patient healthcare progress;

3. A plan for investigation and therapy, according to the SOAP philosophy;
4. Extra records that are gathered with the previous SOAP.

The implemented EMR is developed using the Object Oriented Architecture and Agent Oriented Architecture and it is based on .Net framework web application. The forms are constructed in a specific XML format that is compatible with the existing system. Each field is identified with a specific tag and presents several attributes like title, subtitle, or type of field. Several fields can be created, such as textbox, combobox, checkbox and check-textbox (if 'true' a textbox appears). A friendly and intuitive view was created for easier report. If it is too complicated or not attractive, than the selected personnel will not use it and, as a result, data consistency may be lost.

In this project two stages were considered:

1. The former one is concerned with the creation of four forms according to each phase of the transplantation process; and
2. The last one was based on the migration of the responsible Doctor database in MS Access to the constructed forms build in the first stage.

### Form Creation

Some meetings with the transplantation staff and the Physician in charge of the working group took place, either to select the relevant factors that will drive the whole process, or to define how to present them to an user-friendly interface. Indeed, from a working document (created by the Physician) the selected factors (here understood as variables) were matched into four different views:

1. The former is referent to the patient admission on the hepatic surgery services;
2. The second refers to the laboratory data and the inconvenient (adverse conditions) that may arise during the transplantation surgery;
3. The third form involves the number, type and gravity of the complications that may occur during the 3 (three) postoperative months, and;
4. The last records the patient history of complications.

Once the creation of the database concerning the patient admission to the hospital has been accomplished, it is possible to index an admission to transplantation. At this stage of the process, it is possible to input values according to the organ receptor (Fig. 1). It includes

Paciente internado, - na CAMA:  
 Episódio-Modulo 9999999 -XXX N° Proc. \_\_\_\_\_  
 Nome: TESTE  
 Data de Nascimento 01-01-1900 Idade 107 Sexo Masculino

Nota de Admissão ao receptor de um transplante

Transplante hepático a: \_\_\_\_\_  
 Tipo: - escolha - ▼ Prior: \_\_\_\_\_ Abdo: - escolha - ▼ Compatibilidade: - escolha - ▼

Indicação  
 PAP ☐ Perda de estômago ☐ Transfusão Hepática-Crônica ☐ Falência Aguda ☐  
 Colangite ☐ HCC and Tbc ☐ HCC + de Tbc ☐ HCC - outros casos ☐  
 Outra ☐

Etiologia  
 VHB ☐ Búfalo-Clon ☐ Colangite Evolues ☐  
 Colangite isquêmica ☐ Alcool ☐ Exposição Química ☐ VHC ☐ TANI ☐ At Auto-Imune ☐ Outra ☐

Outro ☐ CMV ☐  
 Doença associada: - escolha - ▼ ☐  
 Atividade: \_\_\_\_\_ TR: \_\_\_\_\_ Lúria: \_\_\_\_\_

Dados laboratoriais  
 BT: \_\_\_\_\_ Creatinina: \_\_\_\_\_ BUN: \_\_\_\_\_  
 Cálculo no dia anterior? Não ☐ Sim ☐

Medic na internação: \_\_\_\_\_  
 Médico Responsável: \_\_\_\_\_

Figure 1: Admission to a hepatic transplant

number of re-transplants, if any, weight, type of blood, compatibility with organ, the causes and the state of the organ for transplantation. Several relevant preoperational laboratory values are input as well. Some values can only be input after checked, if necessary, while others must be input in all cases. Textboxes are used for data that must be input for accuracy of the information and to make the study relevant, while comboboxes are the best way to guaranty consistence in text values. Normalized text values are easier to compare and analyze. If the value does not exist than it is possible to select 'other' from the list and click the checkbox near the combobox. A textbox will appear for the new value insertion.

The patient admission permits the designated nurse to access the intraoperative form (Fig. 2). During the transplantation surgery new data must be input. New laboratory data will be used to compare with the admission one. There is a designated field for days of waiting for transplant, because it can be crucial to the operation success. Some specific data like new problems and causes that were undetected can be input. Technical data involving the operation process must also be considered in order to enforce the best practices and evaluated the surgeon team. Other data, such as intraoperative complications and abdominal findings must also be checked. Postsurgical complications are very common. For a patient survival it is imperative that problems are identified in the first months and constant monitoring is applied. The report in Fig. 1 includes the drugs used for immunosuppression, responsible for opportunis-

Paciente internado, - na CAMA:  
 Episódio-Modulo 9999999 -XXX N° Proc. \_\_\_\_\_  
 Nome: TESTE  
 Data de Nascimento 01-01-1900 Idade 107 Sexo Masculino

Dados Intra Operatórios da Cirurgia de Transplantação

Transplante hepático a: \_\_\_\_\_

Dados laboratoriais  
 BT: \_\_\_\_\_ Creatinina: \_\_\_\_\_ BUN: \_\_\_\_\_  
 Cálculo no dia anterior? Não ☐ Sim ☐ Med: \_\_\_\_\_  
 Tíning: - escolha - ▼ Número de dias de apelo ☐

Achados Abdominais  
 Cirrose: ..... Não ☐ Sim ☐  
 Ascite: ..... Não ☐ Sim ☐  
 Hipertensão Portal: Não ☐ Sim ☐  
 Neoplasia: ..... Não ☐ Sim ☐

Dados Técnicos  
 Pyrexia: Sim ☐ Não ☐ Tipo de incisão: - escolha - ▼ Técnica: - escolha - ▼ ☐  
 Reconstrução Arterial: - escolha - ▼ Via Sâlv: - escolha - ▼

Cirurgias Responsáveis  
 Reconstrução arterial: - escolha - ▼ Hepatectomia: - escolha - ▼  
 Reconstrução biliar: - escolha - ▼ Implante: - escolha - ▼

Sistema de perfusão: Não ☐ Sim ☐  
 Tempo da Cirurgia: \_\_\_\_\_ Número da unidade de: \_\_\_\_\_ Tempo em ice fria: \_\_\_\_\_ Tempo em ice quente: \_\_\_\_\_  
 Tempo de operação: \_\_\_\_\_ Tempo sem anestesia: \_\_\_\_\_

Complicações Intra-operatórias  
 Paragem cardíaca: ..... Não ☐ Sim ☐  
 Hemorragia: ..... Não ☐ Sim ☐  
 Anestesia: ..... Não ☐ Sim ☐  
 Choque: ..... Não ☐ Sim ☐  
 Coagulopatia: ..... Não ☐ Sim ☐  
 Trombose Vena Porta: Não ☐ Sim ☐  
 Outras ☐

Figure 2: Intra-operative data input

tic diseases and other problems, but also for protecting the organ from rejection. The number of days of internment and concentration peaks in the first three days are also input. Complications and infections that attack the patient, possibility of rejection and the number of rejection episodes are also controlled. Common values are present in the designated combobox values. Finally, a summary of discharge (i.e. sending a patient out of the hospital because he/she is fitted to go home) with medical notes, is provided, which it is quite useful for the auditory system used by surgeons to report anomalies or for simple observations.

The last form (Fig. 4) is used for visits after discharge. A patient with a transplanted organ must be observed during the rest of he/she life time. The detection of biliary complications, infections and evolution of the graft are important to detect possible rejections. It is possible to report if the organ is rejected or new late complications occur, if this happens, than a new admission for re-transplant should be created.

## Migration of Data

The surgeon in charge of this project provided his own personal database. This database was created under a MS Access structure. To incorporate this information

Figure 3: Transplant postsurgical data

Figure 4: External consultation after dispatch

into the system, the conversion to the specific XML conversion was made. In order to perform this task the MS Access had to be converted into a text file using the SQL language.

PL/SQL was used to integrate information into the EMRs specific database. A specific structured transfer language was defined to store the data and metadata for information transfer. The first line contained the column names and in the next lines the values for each database line in the respective order. For example, for a hypothetical case with one database with one line only, the text file that will appear takes the form:

1. 'name','number','retransplant'
2. 'Joan','1111111', '1'

On the other hand, text values like 'yes' or 'no' are represented by numbers (e.g. 'yes'=1 and 'no'=2). This type of data, given in terms of numbers, was supposed to be converted to its respective text value, in order to be compatible with the previous Xml files. Several similar problems were found, namely some cell values had more than one number and separated by symbols

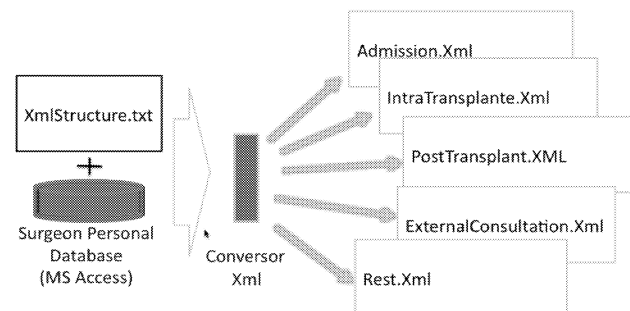


Figure 5: Diagram of the surgeon database for data migration

(e.g. '2+3+4' or '2+cort+4'). Therefore, a strategy to convert this type of situation had to be created.

ConversorXml is a tool for this specific automatic migration. This application had to be automatic because the database had a large volume of cases. Its language base is Java and was created to convert the data from the text file previously described to the structured Xml format. This program had two input files (in .txt) and five output files (in .xml), as represented in Fig. 5.

The first input file was the one with the results from the SQL query and the second one contained the structured information. To be more precise, this file was created to link each column to the respective tag, or tags because it could have more associated fields, and type of input/output data. For example, for fields 'yes/no' that could have the value '1' or '2', the attribute 'createbool' was associated, that means that the output file will have the value 'true' for the field 'yes' and 'false' for the 'no' field. There are seven possible attributes: 'hip', 'bool', 'createbool', 'createbool2', 'createbool2+' and 'hipcreatebool+'. If the field does not have any, it is because the values are automatic (e.g. laboratory values that only have a associated value for a specific tag/title). The result translates into four files with the surgeon data. The structure of these Xml is quite simple: the specific tag, title and associated value. This structure is necessary for an accurate extraction of data.

## Conclusion

This project was shaped according to the demands on the level of the information quality and service effectiveness required by different healthcare institutions. An organized and intuitive way of reporting is a good start to use or archive this data, i.e. the introduction of a structure report going from the admission, passing through the transplant itself and dispatch of receptors, until the monitoring in the external consultation, will be beneficial to the transplantation staff. In this way the reviewing through visualization, research and risk adjustment analysis will be possible and easier.

The professionals were enthusiastic about the new sys-

tem and it had a good reception within the department, being the collected data now linked to the one gotten at different services in the healthcare facility. Statistical analysis of mortality, morbidity, retransplantation outcomes are now possible with the tools provided in the EMR.

This work open new doors in service quality enhancement. From now on, it may be possible to integrate a compatibility program that with the existing data can track the organs available, in real-time, inside or outside the institution. Assessment on the patient status is also enhanced, once images from the imagiology service and laboratory data present in the patient EMR record can now be accessed in real-time (e.g. using mobile technology like PDAs).

## REFERENCES

- [1] Weed, L.L.: Medical records that guide and teach. *New England Journal of Medicine* (1968) 278:593–599
- [2] Abelha, A.: O Processo Clínico Electrónico – Sistemas Multiagente de Apoio ao Trabalho Cooperativo em Unidades Hospitalares. PhD thesis, University of Minho, Portugal (2004)
- [3] Ginneken, V., Stam, H., Moorman, P.W.: A multi-strategy approach for medical records of specialists. *International Journal of Bio-Medical Computing* (1996) 42:21–26.
- [4] Goradia, V.K.: Electronic medical records for the arthroscopic surgeon. *The Journal of Arthroscopic and Related Surgery* (2006) 22:219–224
- [5] Stausberg, J., Koch, D., Ingenerf, J., Betzler, M.: Comparing paper-based with electronic patient records: Lessons learned during a study on diagnosis and procedure codes. *J Am Med Inform Assoc* **10**(5) (September 2003) 470–477
- [6] Miranda, M., Abelha, A., Santos, M., Machado, J., Neves, J.: A group decision support system for staging of cancer. *Electronic Healthcare* (2009) 114–121
- [7] Khuri, S.F.: The nsqip: A new frontier in surgery. *Surgery* **138**(5) (2005) 837 – 843
- [8] van Bommel, J.H., Duisterhout, J.S.: Education and training of medical informatics in the medical curriculum. *International Journal of Medical Informatics* **50**(1-3) (1998) 49 – 58
- [9] Rigor, H., Machado, J., Abelha, A., Neves, J., Alberto, C.: A web-based system to reduce the nosocomial infection impact in healthcare units. Webist 2008: Proceedings of the Fourth International Conference on Web Information Systems and Technologies, Vol 1 (2008) 264–268 Cordeiro, J Filipe, J Hammoudi, S 4th International Conference on Web Information Systems and Technologies MAY 04-07, 2008 Funchal, PORTUGAL.
- [10] Weed, L.: Medical records, medical education and patient care. the problem-oriented record as a basic tool. Case Western Reserve University Press (1969)
- [11] Tange, H.: How to approach the structuring of the medical record? towards a model for flexible access to free text medical data. *International Journal of Bio-Medical Computing* **42**(1-2) (1996) 27 – 34 Papers presented at the 8th European Health Records Conference.
- [12] Elberg, P.B.: Electronic patient records and innovation in health care services. *International journal of medical informatic* **64**(2) (2001) 201–205
- [13] Kalra, J.: Medical errors: overcoming the challenges. *Clinical Biochemistry* **37**(12) (2004) 1063 – 1071
- [14] Duff, F.L., Daniel, S., Kamendjé, B., Beux, P.L., Duvauferrier, R.: Monitoring incident report in the healthcare process to improve quality in hospitals. *International Journal of Medical Informatics* **74**(2-4) (2005) 111 – 117 MIE 2003.
- [15] Birkmeyer, J.D., Dimick, J.B., Birkmeyer, N.J.O.: Measuring the quality of surgical care: structure, process, or outcomes? . *Journal of the American College of Surgeons* **198**(4) (2004) 626 – 632
- [16] Fink, A.S.: Evidence-based outcome data after hernia surgery: A possible role for the national surgical quality improvement program. *The American Journal of Surgery* **188**(Supplement 1) (2004) 30 – 34
- [17] Rosen, A., Wu, J., Chang, B.H., Berlowitz, D., Rakowski, C., Ash, A., Moskowitz, M.: Risk Adjustment for Measuring Health Outcomes: An Application in VA Long term Care. *American Journal of Medical Quality* **16**(4) (2001) 118–127
- [18] Atherly, A., Fink, A.S., Campbell, D.C., Mentzer, R.M., Jr, Henderson, W., Khuri, S., Culler, S.D.: Evaluating alternative risk-adjustment strategies for surgery. *The American Journal of Surgery* **188**(5) (2004) 566 – 570
- [19] Rowell, K.S., Turrentine, F.E., Hutter, M.M., Khuri, S.F., Henderson, W.G.: Use of national surgical quality improvement program data as a catalyst for quality improvement. *Journal of the American College of Surgeons* **204**(6) (2007) 1293 – 1300



# SYSTEM DYNAMICS APPROACH FOR MODELLING COMPLEX HEALTHCARE SYSTEMS

Ruby Wai Chung Hughes and Terrence Perera  
Simulation and Modelling Group,  
Faculty of Arts, Computing, Engineering and Sciences,  
Sheffield Hallam University,  
Sheffield, S1 1WB, United Kingdom  
E-mail: r.w.lau@shu.ac.uk

## KEYWORDS

Simulation, System Dynamics, Healthcare System, Decision Support Systems, Dynamic Modelling

## ABSTRACT

Simulation is a well-known Operational Research technique, and it has been widely used in the healthcare sector for clinical decision-making, facility location planning, and resource allocation. Discrete-event simulation (DES) and System Dynamics (SD) are the most popular simulation techniques used for these purposes, mainly because of their abilities to model uncertainty and variability.

This paper describes a case study of using SD approach to investigate the dynamic complexity of the National Health Service (NHS) 18-weeks pathway. Since the 18 weeks pathway involves a large population and with complex definition of patient status, SD was chosen for this purpose. This study applied both the qualitative aspect (by using causal loop diagram) and the quantitative aspect (by using stock and flow model). Healthcare managers found this SD approach very easy to apply, and very helpful for enhancing the understanding to the dynamic nature of a complex system.

## INTRODUCTION

Simulation is a well-known Operational Research technique, and it has been widely used in the healthcare sector for clinical decision-making, facility location planning, and resource allocation. Discrete-event simulation (DES) and System Dynamics (SD) are the most popular simulation techniques used for these purposes, mainly because of their abilities to model uncertainty and variability. Software providers introduced easy to use simulation packages with graphical interfaces to make these kinds of simulation techniques become more popular. Healthcare systems can be affected by different influences (human behaviour, technological changes and organisational policies), which are difficult to represent on a piece of paper. Simulation has the ability to capture these influences on the healthcare systems, to enhance the understanding of the comprehensive structure, identify bottlenecks and to support decision-making.

DES and SD are two different approaches in modelling healthcare systems. DES, models each patient as an individual entity flows around a simulation model. Each patient can be modelled to have different characteristics (age, gender, level of illness etc.) , the simulation model is based on these characteristics modelling patients in appropriate queues (waiting lists) for appointments, examination and treatments. Therefore, this kind of simulation model often requires high-quality data with complex logical rules, in order to model the real behaviour of a healthcare system. Because of the ability of DES to model complex systems with details, it is commonly used at operational level for capacity planning, to identify bottlenecks and redesign patient pathways.

Comparatively, SD is often used at a more strategic level. In SD patient is no longer identified as an individual entity, patients become a “Stock” within the system which flows around the model. Compared with DES, SD cannot model patient details. Therefore, this kind of simulation often involves less time in data collection. On the other hand, SD concerns with the dynamic complexity and the feedback effects within the whole system, often SD tend to have broader boundaries than DES models. Since SD has the ability to include a variety of realistic causal factors, policy issues, and feedback loops within a model. Therefore, SD becomes advantageous to use for understanding the dynamic structure of healthcare systems and for supporting strategic solutions for a complex system.

This paper describes a case study of using the SD approach to investigate the dynamic complexity of the 18 weeks pathway. Since the 18 weeks pathway involves a large population and with complex definition of patient status. This study applied both the qualitative aspect (by using causal loop diagram) and the quantitative aspect (by using stock and flow model) of SD. This paper focuses on two main issues. Firstly, introduces SD approach to model healthcare systems. Secondly, uses the 18 weeks target as a case study to show how a SD approach can be used to understand the dynamic structure of a complex system.

## SYSTEM DYNAMICS IN THE HEALTHCARE

Brailsford (2007) states there are three main reasons that SD is ideal for modelling healthcare systems. Firstly, healthcare

systems are characterised by uncertainty and variability, it is important to understand the behaviour of the system as a whole. SD is concerned with feedback and unanticipated effects, which is useful for decision-makers to identify the dynamic complexity of the system. Secondly, SD models are not dependent on precise data (patient arrival time, process time, resources) therefore models can be used at a more strategic level, for larger populations. The time for data collection is often highly reduced compared to other healthcare simulation models which usually require high-quality data. Thirdly, without complex logical rules and individual entities in SD models, the "run time" is generally very fast (less than a minute for 5 weeks simulation), so the real time result can become a very good communication aid for decision-makers.

There are increasing examples of SD models in literature. Brailsford, S.C. and Hilton, N. (2000) used the SD approach to understand how NHS contract policy can influence the long waiting lists for cardiac surgery. They utilised both qualitative and quantitative aspects in their study to understand the changes on the waiting lists with different contract levels. This study used a simple example to apply SD model to represent the behaviour of a cardiac system, and to understand the dynamic nature of the system as a whole in reacting to the changes in contract policy.

Emergency Care-On Demand (ECOD), a regional level SD model commissioned by Nottingham Health Authority, was the first to review a regional emergency and urgent care system as a whole (Lattimer et al 2004). The SD model from the study enabled policy-makers to review the current performance, identify bottlenecks, and investigate the scenarios for change in terms of demand and patient flows. Lane et al (2000) also developed a SD model studied an A&E department, Land describes SD model as a "visual learning environment" which can help users understand why structure produced behaviours (the Base Case), and how behaviour varied under different conditions (the Policy Analysis).

Su, Y. and Jin, Z. (2008) also used simulation modelling for redesigning health service centres in Beijing. Their project involved a discrete-event simulation (DES) model for the newly designed facilities. However, Su and Jin found that DES model alone could not help in understanding the effects caused by internal system feedback of the new system. Therefore, they built a SD model which includes other influences (patient satisfaction, friends and family, lag time, word of mouth) to the new system. The SD model showed these influences could cause important effect to the patient population (demand) and the waiting time in the new system.

## **CASE STUDY: THE NHS 18-WEEKS TARGET**

In June 2004, the government stated that "By 2008, no one will have to wait longer than 18 weeks from General Practitioner (GP) referral to hospital treatment" (DH, 2004). The 18 weeks pathway starts from the point of initial GP referral to first outpatient, to follow-up outpatient and

diagnostic test, until the start of any treatment (if necessary). Hospitals have to review these patient pathways and identify bottlenecks, to improve the overall efficiency and reduce the unnecessary waiting time.

One of the NHS strategies in targeting these waiting times is increasing the capacity (staff and clinics) to discharge the increasing demand. However, some research comments that the problem of the long waiting times in NHS is mostly caused by a lack of attention and understanding to the demand variation and the inappropriate responses to the queues, rather than a capacity shortage (Walley et al. (2006)). Gunal & Pidd (2007) critics some forms of waiting in service sector are unavoidable unless the systems can resource at the levels well above those needed for average demand. Gunal & Pidd also comment that the seasonal nature of the demand in healthcare means providing resources to meet average demand will lead to further queues for service. Therefore, it is importance to understand the demand variation and how the whole system reacts to the changes.

## **METHOD AND DATA**

System dynamics (SD) is chosen to investigate the feedback effects and the dynamic structure within this 18 weeks pathway. This is a pioneer study to use SD to understand the dynamic complexity of the 18 weeks target. Data for this study is collected via interviews with simulation analysts and a group of healthcare staff and practitioners (from a local hospital in Sheffield, Department of Health and Primary Care Trust). Additional data is collected via internet documents from the NHS official website ([www.18weeks.nhs.uk](http://www.18weeks.nhs.uk)).

### **Interview analysis**

The Interview focused on the following issues: NHS challenges, recent NHS targets, their responsibilities, experience and opinion in simulation techniques, 18 weeks target challenges and existing solutions and future improvements. Based on the interviewees, the NHS challenges are mainly regarding four issues: financial, patient safety, patient waiting time, and resources. Apart from the on-going targets to improve quality of patient care and efficiency, the core NHS target is currently the 18 weeks target.

### **NHS Strategy**

A series of actions have already started since the target launched in June 2004. Figure 1 is a fish-bone diagram showing the strategies which these interviewees were (or starting to) focusing on.

Based on the list of existing NHS strategies (Figure 1), four main issues are identified (Budget, Capacity, Demand and Waiting time) as the main components, which play an important role in affecting the long waiting times problem. Therefore, it is necessary for managers and policy-makers to understand the interrelationship between these components within the entire system.

Solutions targeting the 18-weeks target

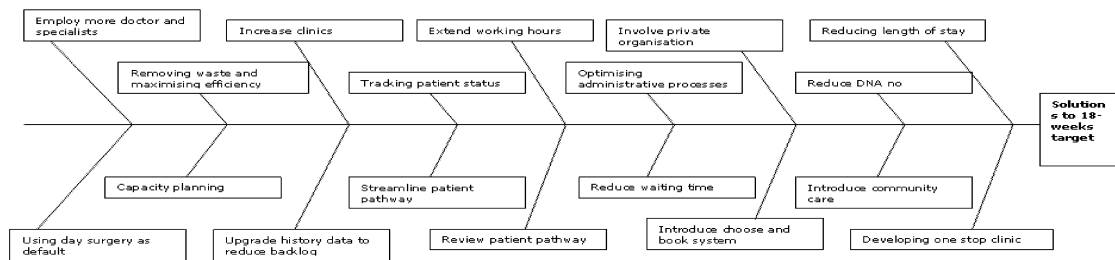


Figure 1: Fish-bone diagram -Actions in targeting the 18 weeks wait target

## QUALITATIVE ASPECT OF SYSTEM DYNAMICS

### Causal loop diagram - Level 1

Four fundamental elements - "Capacity", "Waiting time", "Demand" and "Budget" were identified as the main components from a list of NHS strategies. Healthcare managers and policy-makers are necessary to fully understand how these components influence each other and the entire system. System dynamics approach was selected for this purpose. The SD diagram replicates the behaviour of the real system which shows the interactions of all elements in a model at the same time. Figure 2 shows a causal loop diagram which represents the dynamic structure of these four main components. The diagram includes elements and arrows (which are called causal links) linking these elements together in the same manner and a sign (either + or -) on each link.

Link 1. A causal link from "Increased budget" (a) to "Increased capacity" (b) is a *positive* (that is, +) link; which a change in (a) can produce a change in (b) in the same direction.

Link 2. A causal link from "Increased capacity" (b) to "Waiting List" (c) is a *negative* (that is, -) link; which a change in (b) can produce a change in (c) an opposite direction.

Loop 1. This is a Balancing (B) feedback loop which indicated the behaviour of the system is goal-seeking. "(B) Demand pressure leads more capacity" represents the purpose of the strategy.

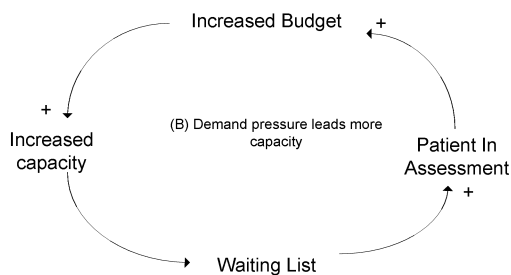


Figure 2: Causal loop diagram - the dynamic structure of a health care system

The causal loop diagram shown in Figure 2 simplifies the dynamic structure of a strategy for targeting the long waiting time problem within a complex system and provides a "sketch" to the later stages of the SD model development.

### Causal loop diagram - Level 2

Figure 3 is an advance causal loop diagram which modifies from the initial diagram (Figure 2). This diagram includes external elements which influence the main four elements. Three new feedback loops are included in this level:

Loop 1. "(R) Growth capacity" is a reinforcing feedback loop which represents that the initial element leads to further change. In this case, when the NHS launches new targets, it increases budgets into the system, which increases the amount of capacity and leads to increasing clinics and staff, then eventually resource could be increased.

Loop 2. "(B) Control waiting" is a balancing feedback loop. It represents a change in "Resource Increased" would produce a change in the "Waiting List". Likewise a change in "Waiting List" would produce a change back to "Increased capacity". In this case, if resource increases, patient waiting (in waiting list) would reduce. Natural demand growth is another element which could influence a change to the waiting list.

Loop 3. "(R) Growth demand" is a reinforcing feedback loop. It represents the initial "Patient complexity" leads to a direct change in "Patient in Assessment", and to "length of assessment". In this case, more complex patient cases would influence an increase of the demand and to the overall length of assessment, thus the waiting list would increase.

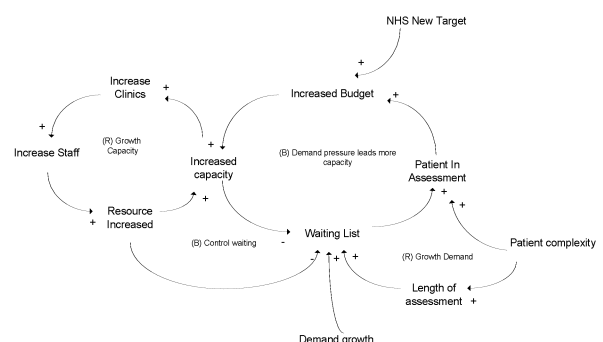


Figure 3: Causal Loop Diagram to show a complete picture of a healthcare strategy

In Figure 3, the final causal loop diagram represents the full dynamic structure of the four main healthcare components and involves its external factors (demand growth, patient complexity, and NHS new target), which provides a dynamic view of this complex system. Healthcare managers found this diagram very impressive, and very useful to understand the feedback effects between these main components and the external influences. This final diagram was also used as the basis of a quantitative SD model, which is discussed in the next section.

## QUANTITATIVE ASPECT OF SYSTEM DYNAMICS

A quantitative SD model was developed by using the software package STELLA®, which included mathematical elements to show how the 18 weeks patient pathway can be affected by changes and influences through time. Figure 4 shows the conceptual map of this stock and flow model.

This SD model represents the 18 weeks pathway, which different patient statuses were modelled as “Stock” (patient waiting for 1<sup>st</sup> outpatient appointment, patient finished 1<sup>st</sup> outpatient, patient waiting for follow-up, patient finished follow-up, and discharge), and different influences were represented as feedback loops within the pathway (nurse capacity, patient DNA, consultant capacity, transfer rate, discharge rate).

The model was run interactively with the managers from a local hospital. The key findings of the model were that if current capacity allocation and current practices of discharging patients continued, the hospital can discharge 96% of its current demand within the 18 weeks target (Figure 5). Sensitivity testing was run to simulate if there was a 10% increase to the patient demand for the first outpatient appointment. The result shows that the hospital can still discharge 95% of the increased demand. The only difference was the discharge peak time shifted from week 10 to week 12 within the 18 weeks pathway (Figure 6).

Another sensitivity testing was simulated based on the staff capacity, with two more consultants on duty to handle follow-up appointments. The simulation result shows that the hospital can discharge almost 98% of its current demand within the target.

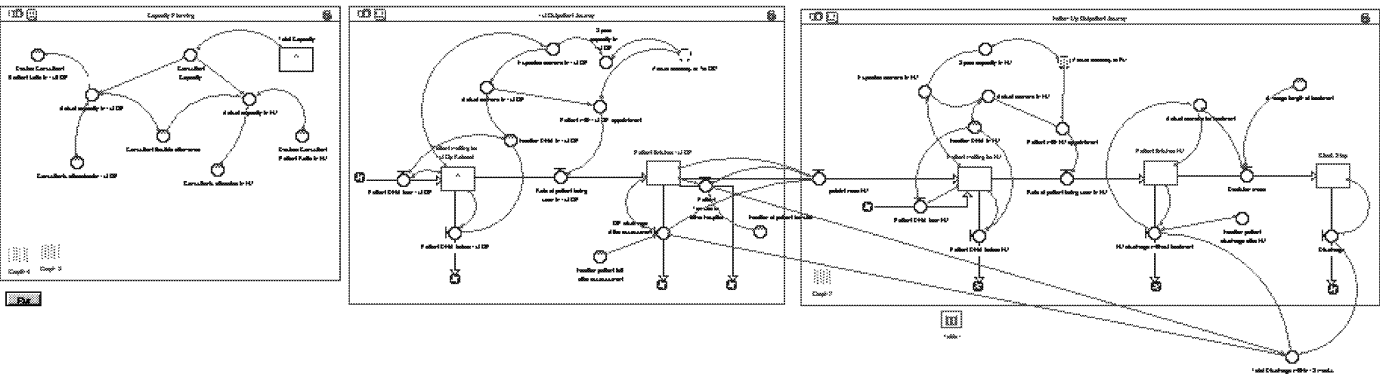


Figure 4: Stock and flow diagram for the 18 weeks patient pathway

The local hospital managers found this model very useful, which enhanced their understanding on how the pathway could react to different changes. The numerical results of the sensitivity testing were found very useful to support further strategic decisions, and to facilitate their communication within the whole team.

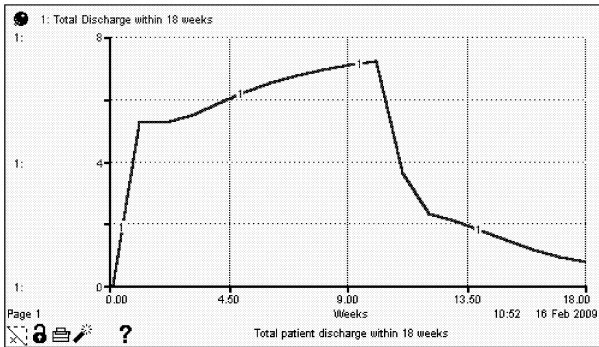


Figure 5: Simulation Result: Patients discharged within the 18 weeks

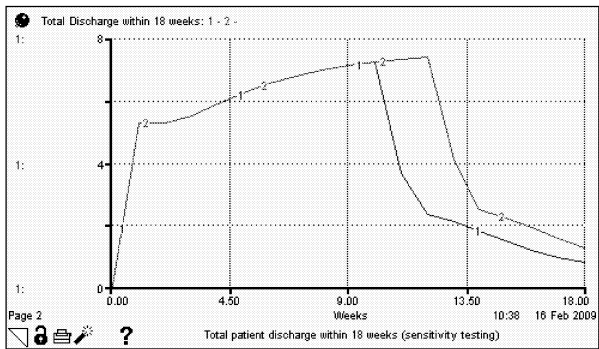


Figure 6: Sensitivity testing with 10% increased in the total demand: patients discharged within the 18 weeks

## CONCLUSION AND FUTURE WORK

This paper described a SD approach to investigate the dynamic complexity of the 18 weeks pathway and understand the dynamic nature of this complex system. This study first identified the main components within a complex healthcare system. Then, applied these components into a qualitative SD diagram (Causal loop diagram), which showed the feedback effects between these components. Healthcare managers

from a local hospital found this diagram very easy to understand, and very helpful to gain insight to the dynamic structure of the system.

The quantitative aspect of this study was found to be very useful for investigating the scenarios for change in terms of patient demand and capacity allocation. Although this SD model did not produce precise numerical simulation results (compared to DES model), it found that the stock and flow diagram is also a very useful strategic tool for facilitating discussing with healthcare managers.

Further work will use a different simulation approach (Discrete-event simulation) to simulate the patient pathways within the 18 weeks target. The study will look at two main issues. Firstly, compare the strengths and weaknesses of each approach for investigating the same healthcare system. Secondly, develop a framework to guide potential users about the selection of an appropriate simulation approach for looking at a complex healthcare system.

## REFERENCES

- Audit Commission. 2003. National Report: Waiting for elective admission. Report available online at <http://www.audit-commission.gov.uk/reports/NATIONAL-REPORT.asp?CategoryID=&ProdID=C98CB150-9FF3-11d7-B304-0060085F8572>
- Brailsford, S. C., Lattimer, V.A., Tarnaras, P., Turnbull, J.C. 2004. "Emergency and on-demand health care: Modelling a large complex system." *Journal of the operational research society*, 55: 34-42.
- Brailsford, S.C. 2007. "Tutorial: Advances and challenges in healthcare simulation modeling." *Proceedings of the 2007 Winter Simulation conference, winter 2007*, 1436-1448.
- Choose and Book Referrals. Available online at <http://www.heart.nhs.uk/18weeks/referrals/choose.html>
- Chaerul, Mochammad, Tanaka, Masaru and Shekdar, Ashok V. 2008. "A system dynamics approach for hospital waste management." *Waste management*, 28 (2), 442-449.
- Department of Health. 2004. NHS Improvement Plan: Putting people at the heart of public services. London, Her Majesty's Stationery Office
- Gunat, M. M. 2007. "Interconnected des models of emergency, outpatient, and inpatient departments of a hospital." *Proceedings of the 2007 Winter Simulation conference, winter 2007*, 1461-1466.
- King's Fund. 2007. 18-week waiting times target, London: King's Fund, 2007. Available online at [http://www.kingsfund.org.uk/publications/briefings/18week\\_waiting\\_times.html](http://www.kingsfund.org.uk/publications/briefings/18week_waiting_times.html)
- Lane, D.C., Husemann, E. 2008. "System dynamics mapping of acute patient flows." *Journal of the operational research society*, 59 (2): 213.
- Lane, D.C., Monefeldt, C. and Rosenhead, J. V. 2000. "Looking in the wrong place for the healthcare improvements: A system dynamics study of an accident and emergency department." *Journal of the operational research society*, 51: 518.
- Lewis, R., Appleby, J. 2006. "Can the English NHS meet the 18-week waiting list target?" *Journal of the royal society of medicine*, 99: 10-13.
- Lattimer, V., Brailsford, J., Turnbull, J., Tarnaras, P., Smith, H., George, S., Gerard, K., Maslin-Prothero, S. 2004. "Reviewing emergency care systems I: Insights from system dynamics modelling." *Emerg Med Journal*, 21: 685-691.
- Oxford Radcliffe Hospital. 2008. Operations Now among Quickest at Oxford Hospitals. Available online at <http://www.oxfordradcliffe.nhs.uk/news/newsrecords/0804/08041518week.aspx>
- Walley, P., Silvester, K. and Steyn, R. 2006. "Managing variation in demand: Lessons from the UK national health service." *Journal of healthcare management*, 51 (5): 309-320.

## BIOGRAPHY

**RUBY WAI CHUNG HUGHES** is currently a PhD student at Sheffield Hallam University UK. She received her BSc Degree in Information Technology and Engineering Management (First Class) in 2002 and an MPhil Degree for her work on Embedding simulation Technologies in Business Processes. Her current research focuses on system dynamics, and its use in modelling complex system such as healthcare system. Email to <r.w.lau@shu.ac.uk>

**TERRENCE PERERA** is a Professor of Manufacturing systems and the Assistant Dean (Academic Resources) at the Faculty of Arts, Computing, Engineering & Sciences, Sheffield Hallam University UK. He also leads the Systems, Modelling and Integration Research Group. His research interests include practices, integration and implementation of simulation modelling within the industrial sector. Email to <t.d.perera@shu.ac.uk>

# PAEDIATRIC PALLIATIVE CARE PLANNING: MODELS AND SIMULATION

Giada Aspergh, Paola Facchin,  
Anna Ferrante, Laura Visonà Dalla Pozza  
Epidemiology and Community Medicine Unit,  
Department of Paediatrics - University of Padova  
Via Pietro Donà 11, 35129 Padova, Italy  
e-mail: [epi@pediatria.unipd.it](mailto:epi@pediatria.unipd.it)

Giorgio Romanin-Jacur  
Department of Management and Engineering  
University of Padova  
Stradella San Nicola 3, 36100 Vicenza, Italy  
e-mail: [romjac@dei.unipd.it](mailto:romjac@dei.unipd.it)

## KEYWORDS

Health care, discrete simulation, decision support systems.

## ABSTRACT

Paediatric palliative patients are children suffering for incurable pathologies which cause them a lot of human, clinical, psychological, ethical and spiritual problems and require special assistance. Only in recent times they have been considered as specific patients needing qualified cares. In this paper we have built up a simulation model describing paediatric palliative patients movements among the interested health facilities, more precisely hospital departments, hospice, integral home care and simple home care, and the interactions of such patients with ordinary patients in seizing places in hospital departments. The model, implemented on a personal computer by means of a specific simulation language, permits to evidence the lacks of the existing assistance network and to suggest suitable adaptations. The model has been usefully applied to Veneto region in North-East Italy.

## INTRODUCTION

The increasing numbers of children who need paediatric palliative cares have made necessary a reorganization of the existing assistance network. The present organization is in fact unable to correctly satisfy the demand of this kind of cares. Children who need palliative cares are often hospitalized in non specialized acute departments where the therapies are only addressed to pain's control and reduction without paying attention to the real needs of these small patients.

An accurate bibliographic research pointed out that many papers have been written about the matter, (Frager 1996, Kenny 2003, Morrison and Meier 2004, Salas Arrambid, et al. 2004, Browning and Solomon 2005, Byock et al. 2006) but none of them aims at describing and analyzing in detail the existing organization with the scope of pointing out its insufficiencies in order to improve the offered assistance. Our work aims at filling up this gap by building up and implementing a simulation model which describes both the existing assistance network and the patients behaviour, according to the logical principles stated by the assistance policy and depending on the existing parameters. The same model can describe patients behaviour in any future modified network, according to system and policy evolution. The structures included in the model are all the ones required by a paediatric palliative patient, i.e., acute hospital, hospice (patient's house where specialized assistance is supplied) but also patient's house where only

generic assistance is supplied. A patient moves among the various assistance structures on the basis of his/her own clinical history and the resources at disposition. Some model structures are utilized by both paediatric palliative patients and "ordinary" patients, who enter in competition with palliative patients whenever they ask for admission in hospital acute departments.

The model has been applied to Veneto Region situation in a given time interval, but it can be easily employed for other Italian or foreign regions presenting similar problems.

Model's runs highlight possible assistance network insufficiencies and help us to directly propose the necessary adaptations.

The paper's organization is the following.

In Section 1 we introduce and analyze the paediatric palliative care system trying to point out its currently deficiencies. In Section 2, the adopted model and its implementation are described, in Section 3 we speak about data sources and data processing. Eventually in Section 4 the model's results, related both to current situation and to foreseen future situations, are reported; the model's results may be usefully utilized to find out a better service organization.

## PAEDIATRIC PALLIATIVE CARE

Palliative care (from Latin *palliare*, to cloak) is any form of medical care or treatment that concentrates on reducing the severity of disease symptoms, rather than striving to stop, delay, or reverse progression of the disease itself or provide a cure. The scope of this kind of cares is to prevent and relieve suffering and to improve life quality for people with incurable illness. Palliative care is not only orientated to pain reduction as its main aim lays in improving patients' life.

By now, palliative care has been mainly addressed to adults. No suitable importance is given to the fact that children too may suffer of incurable pathologies which cause them a lot of human, clinical, psychological, ethical and spiritual problems. In the literature we learn that the amount of children who need paediatric palliative care is not negligible. In 1997, the Royal College of Paediatric and Child Health estimated that, in the UK, 10 per 10,000 children aged 0-19 years require palliative care each year (ACT, 1997). More recently, the Association for Children's Palliative Care reviewed the UK data and suggested that at least 12 in every 10,000 children have life-limiting conditions requiring palliative care (ACT 2003). Based on the projections of UK estimations, we may value the presence in Italy of about 11,000 children aged 0-17 suffering from incurable pathologies. Recent studies carried

out in Veneto Region by the analysis of the Hospital Discharge Records (SDO) and of the paediatric patients' death records in the time interval 2000-2004 state that about 15,000 children need palliative care. These data shall be taken into consideration as, thanks to medical and technological progress, it is now possible to ensure to small patients both a longer and a better life; as a consequence that causes an increase of both patients' number and pathologies' variety, therefore the real demand of palliative care becomes more and more remarkable thus increasing the needs to be satisfied by the existing organizations for paediatric palliative cares.

Till now, in general these small patients have been improperly and repeatedly admitted to hospital in non specialized acute departments, where they have been subjected to series of therapies aiming at pain reduction but paying no attention to other needs, above all the need of staying at home and trying to lead a normal life. In fact repeated hospital admissions uproot the patient from his/her everyday life which is already hard.

Therefore the goal to reach consists in organizing home cares with possible short admissions in specialized structures dedicated to small patients.

In order to reach this objective it is necessary to design a new organization of paediatric palliative cares.

## **PALLIATIVE CARE ORGANIZATION PRESENT AND FUTURE**

In the last years, several assistance models have been tested to give a concrete answer to small patients, family, equipe and institutions needs (Benini et al., 2007). Till now experiences, carried out in different countries and in several Italian Regions, are based on the coexistence of four assistance systems, two residential ones (hospitals and hospices) and two house ones (home-based hospitalization and integrated home-based care): i) according to the first residential system, the patient is admitted to hospital in acute departments, where some palliative cares are given but the main aim is limited to the pain control and reduction, which is not always the correct care for a palliative patient; ii) according to the second residential system the patient is admitted in a dedicated structure, the hospice, where specific palliative cares are supplied and moreover a relief can be offered to the family which is heavily oppressed by the assistance supplied to the child; in fact hospice staff attention is addressed to both the patient and the family; iii) according to the first house assistance system (at home hospitalization) the patient is followed at home by a hospital equipe; iv), according to the second house assistance system the patient is followed by the territorial equipe (family doctor, generic and specialized nurses).

None of the above-mentioned assistance models, if singularly considered, ensures a suitable assistance quality; in fact the application of one only system shows many benefits but also serious faults.

Hospitalization is generally unable to offer to small patients an ideal place for paediatric palliative cares both for mission and vocation, as its organization and the human and technical resources at disposition are not specifically prepared to supply palliative care.

Moreover, a high number of improper hospital stays in acute departments, which amounts to 10 admissions per patient per year with a mean of 150 days of hospital stay per year, implies beds occupation for long periods in general, non specialized, semi-intensive or intensive care departments, repeated complex instrumental investigations, intensive high technology interventions; such a way a lot of resources of high value, which would be useful for other patients, are badly used; as a consequence we have a high cost without obtaining a sufficient pain reduction and without supplying a sufficient assistance level and care intensity; moreover a heavy engagement of the family in human, social and economical terms is required.

The admission in a hospice, or in another dedicated structure, where sometimes a relative may be admitted too, has the advantage of being assisted by a highly qualified personnel with an important experience, able to warrant the suitable knowhow level, necessary to manage also infrequent pathologies; in the mean time a dedicated structure ensures a catchment area large enough to justify and make economically sustainable a high number of dedicated resources with specific competence. However this assistance system has the disadvantage to uproot the small patient from him/her life contest, thus violating the hopes of the child and the family to stay at home; that makes illness management more difficult in case of long hospice stays. Then hospice is to be considered as a temporary solution useful to temporarily relieve the family of its heavy assistance task.

At home hospitalization does not uproot the patient from him/her daily life but it presents some limits: on one hand hospital personnel is not trained to a palliative approach, on the other hand the limited resources availability makes this type of organization unable to ensure an adequate territorial covering.

Integrated home-based care permits patient everyday routine but it suffers for a lack of specific resources: that obstructs care continuity and creates difficulties to the family for what concerns patient management.

Since none of the four assistance systems has revealed without limits, the majority of actual paediatric palliative care assistance systems have been based on the combination of all above-reported assistance systems, used together.

All remarks above highlight the necessity of an organization of paediatric palliative care based on a network that includes all the four systems by applying one at a time according to individual situation and disease evolution.

Home care remains the main aim to reach for paediatric age but it may be interrupted for short time intervals: i) when an admission in an acute hospital department is absolutely necessary for special cares, ii) when an admission in a hospice is useful to relieve the family from its heavy assistance task. It is evident that a complete territorial assistance network is to be set up, including assistance both by generic territorial resources and by hospital specialized personnel, with effectiveness and efficiency. Whenever such integrated home care cannot be supplied, a short period of home care with territorial assistance only may be accepted.

By considering actual applications paediatric palliative care networks are totally new.

In order to create a new network it is necessary to plan: a set of human interventions, both specific by a hospital equipe (doctors and nurses) and generic by the territorial staff (family doctor, nurses, psychologist and social supporters) and a set of specific instrumentation, like special beds, ventilators, etc.

## THE MODEL

In the present paper we build up a simulation model, which describes patients state and movements, as a decision support to plan the above assistance network. The model reports in detail the existing (or planned) assistance network and describes with accuracy the patients behaviour according to the logical principles stated by the assistance policy; depending on the stated parameters, the same model may describe patients behaviour in any future modified network, according to system and policy evolution. The structures reported by the model are all the ones which may be required by a paediatric palliative patient, in particular hospital departments, hospices, patient's house where integrated assistance is supplied, patient's house where generic basic assistance only is supplied. A patient who gets into the system moves among the various assistance structures up to his/her death; he/she gets access to the above structures according to a choice performed on the basis of patient's past clinical history and resources at disposition. Patients considered by the model are not paediatric palliative patients only, but also ordinary patients who may ask for admission in hospital acute departments and therefore compete with them. In the model a generator simulates new patients arrivals in a random independent way; generated patients are classified as either palliative or ordinary; then to every palliative patient attributes reporting his/her clinical status are defined at random, according to statistics obtained by analyzing current health service data. More precisely for every ordinary patient the model defines whether he/she requires admission to an intensive care department; for every paediatric palliative patient the model defines: i) the pathology group (onchological pathology, degenerative and progressive pathologies, malformations, disabilities), ii) the expected life duration. Newly created patients, mixed with old ones (patients already moving inside the care network), compete to enter the existing structures; ordinary patients compete with paediatric palliative patients to get admission at hospital acute departments; old and new paediatric palliative patients compete for other structures (hospice, home with integrated assistance, home with generic basic assistance). Every patient is addressed to one of the structures on the basis of his/her current status (attributes and clinical history) and of the resources at disposition (hospital beds, hospice places, home assistance staffs, family doctors, specialized clinical staffs, etc.), depending on the health care service policy, which defines priorities. After using a structure for a period of time which is obtained by analyzing health care service statistics, a patient is reconsidered for a new address; during the stay in a structure any patient may die and therefore leave the system.

## MODEL IMPLEMENTATION

The model was implemented in language MicroSaint 3.2 on a personal computer. The choice of the language was motivated by the fact that it was at disposition and its graphical representation is easily readable also by non expert users, like for instance medical doctors. Moreover it is similar to other frequently used languages like SIMAN-Arena; mutual translation is straightforward.

MicroSaint model graphical representation is based on four essential elements, tasks, represented by ellipses, describing activities, paths, represented by arrows, describing sequences among activities, decisions, represented by rhombuses, describing alternative ways among activities and finally queues, not used here.

The model graphical representation is reported in Figure 1. It includes 18 tasks; task 1 gives the simulation start to task 2 and 4, which generate in a random independent way respectively ordinary and paediatric palliative patients. For every ordinary patients task 3 decides whether he/she needs an intensive department or an ordinary one and then address all of them to task 8. To every paediatric palliative patient the pathology group is randomly assigned by task 5 and the expected life duration is assigned by task 6. Task 7 acts as a dispatcher: according to health policy and paediatric palliative patient history it decides where to address him/her: hospital (task 8), hospice (task 11), home integral care (task 13), home simple assistance care (task 15). Task 11 has the function of analyzing whether there are free places in the corresponding facility 12: if no free places are at disposition, patients are addressed to another facility; according to places at dispositions the chosen facilities are, in the order, home integral care, hospital and home simple care. Task 13 acts analogously to task 11 for home integral care. Task 8 manages both paediatric palliative patients and ordinary patients for what concerns admission to ordinary and intensive hospital departments: paediatric palliative patients preempt ordinary patients for a place in ordinary departments while ordinary patients preempt paediatric palliative patients for a place in intensive departments; preempted paediatric palliative patients come back to the system through task 18 while preempted ordinary patients (task 17) leave the system. At the end of each care cycle (hospital, hospice, integral home care and simple home care) patients either return to task 7 for next address or exit the system because of death (task 16).

The model runs on a personal computer and requires about one minute to simulate an interval of five years.

## DATA SOURCE AND DATA ELABORATED

The Epidemiology and Community Medicine Unit manages a large amount of national and regional data concerning population health and health assistance.

Data necessary for the present work are contained in the SDO (Hospital Discharge Records). The SDO is a standard form used to record a set of data relating to every person discharged from hospital. The SDO contains synthetic informations about personal and clinical data, hospital admission, interventions and discharge.

For this work, all the Hospital admissions records occurred in Veneto Region in the years 2000-2004 have been



processed. First, we have selected a list of diseases eligible for paediatric palliative care using classification ICD IX cm, according to ACT criteria (ACT 2003), and then we have used it to filter the SDO in order to obtain two different groups, one concerning palliative patients, one concerning ordinary patients. Palliative patients' SDO have been again filtered in order to select only more characterized patients. The first choice has picked out the hospital admission occurring to people 0-17 years, with one of the selected life-limiting and life-threatening illness at least, while further screening criteria has excluded from the remaining records all the children with residence out of Veneto Region, all the patients died within the first month of life (not easily eligible to paediatric palliative care), all the patients with a single hospital admission in the nursery (low care complexity) and total hospital stay shorter than 80 days (low care complexity). The two dataset obtained have been elaborated by the software SAS which has given in output all the data necessary to implement our model: patients' number, admission's number, stay time, stay departments and care time.

## MODEL RESULTS AND APPLICATION

The model has been usefully applied to Veneto region in North-East Italy, a region with about 4,700,000 inhabitants, where the paediatric palliative care network is in phase of construction, with only one hospice and an integral home care organization which is just beginning its operations with about 40 places. As expected the model simulation pointed out the insufficiency of specific care supplying with corresponding excess of hospital admissions. By now the model is employed to plan an accurate network resizing, compatible with actual abilities of local health care organization.

## CONCLUSION

A model describing paediatric palliative patients movements among specific health facilities has been built up and implemented. It has been applied to the Veneto region situation with some interesting results. Its use can be extended to similar situations in other both Italian and foreign regions by changing only data concerning local children population and care organization.

## REFERENCES

- ACT 1997. *A guide to the development of children's palliative care services*. Report by a joint working party of the Association for Children with Life-Threatening and Terminal Conditions and their families and the Royal College of Paediatrics and Child Health, London.
- ACT 2003. *A guide to the development of children's palliative care services*. Update of a report by the Association for Children with Life-Threatening and Terminal Conditions and their families and the Royal College of Paediatrics and Child Health, London.
- Benini, F.; A. Ferrante; and P. Facchin 2007. "Palliative care for newborns and children." *Medico e Bambino*, 26, 449-452.
- Browning, D.M. and M.Z. Solomon; for the Initiative for Paediatric Palliative Care (IPPC) Investigator Team 2005. "The initiative for pediatric palliative care: an interdisciplinary educational approach for healthcare professionals." *Journal of pediatric nursing*, United States, 20, 5, 326-334.
- Byock, I.; J.S. Twohig; M. Merriman; and K. Collins 2006. "Promoting excellence in end-of-life care: a report on innovative models of palliative care." *Journal of palliative medicine*, United States, 9, 1, 137-151.
- Fager, G. 1996. "Pediatric palliative care: building the model, bridging the gaps." *Journal of Palliative Care*, Canada 12, 3, 9-12.
- Kenny, L.J. 2003. "An evaluation-based model for palliative care education: making a difference to practice." *International journal of palliative nursing*. England, 9, 5, 189-194.
- Morrison, R. Sean and D.E. Meier 2004. "Palliative Care". *The New England Journal of medicine*, June 17, 350, 25, 2582-2590.
- Salas Arrambide, M.; O. Galbadon Poc; J.L. Mayoral Miravete; E. Gonzalez Perez-Yarza; and I. Amayra Caro 2004. "Pediatric palliative care: a comprehensive model of care for children with life-threatening conditions and their families." *Anales de pediatria*, Spain 61, 4, 330-335.

## AUTHORS BIOGRAPHIES

**GIADA ASPERGH** obtained a degree in Management Engineering at the University of Padova and is student of the PhD school in Developmental Medicine and Planning Sciences.

**PAOLA FACCHIN** obtained a medical degree at the University of Padova, was specialized in Paediatrics and in Public Health, obtained a PhD degree in Developmental Science; she is now Associated Professor of Paediatrics; she is also Coordinator of the Epidemiology and Community Medicine Unit.

**GIORGIO ROMANIN-JACUR** obtained a degree in Electrical Engineering at the University of Padova where he is now Full Professor of Operations Research.

**ANNA FERRANTE** obtained a medical degree at the University of Padova. After a period at the Padova Local Health Unit, where she worked on territorial health assistance, she works now at the Epidemiology and Community Medicine Unit.

**LAURA VISONÀ DALLA POZZA** obtained a degree in Statistics at the University of Padova and obtained a PhD in Epidemiology. She is now working at the Epidemiology and Community Medicine Unit.

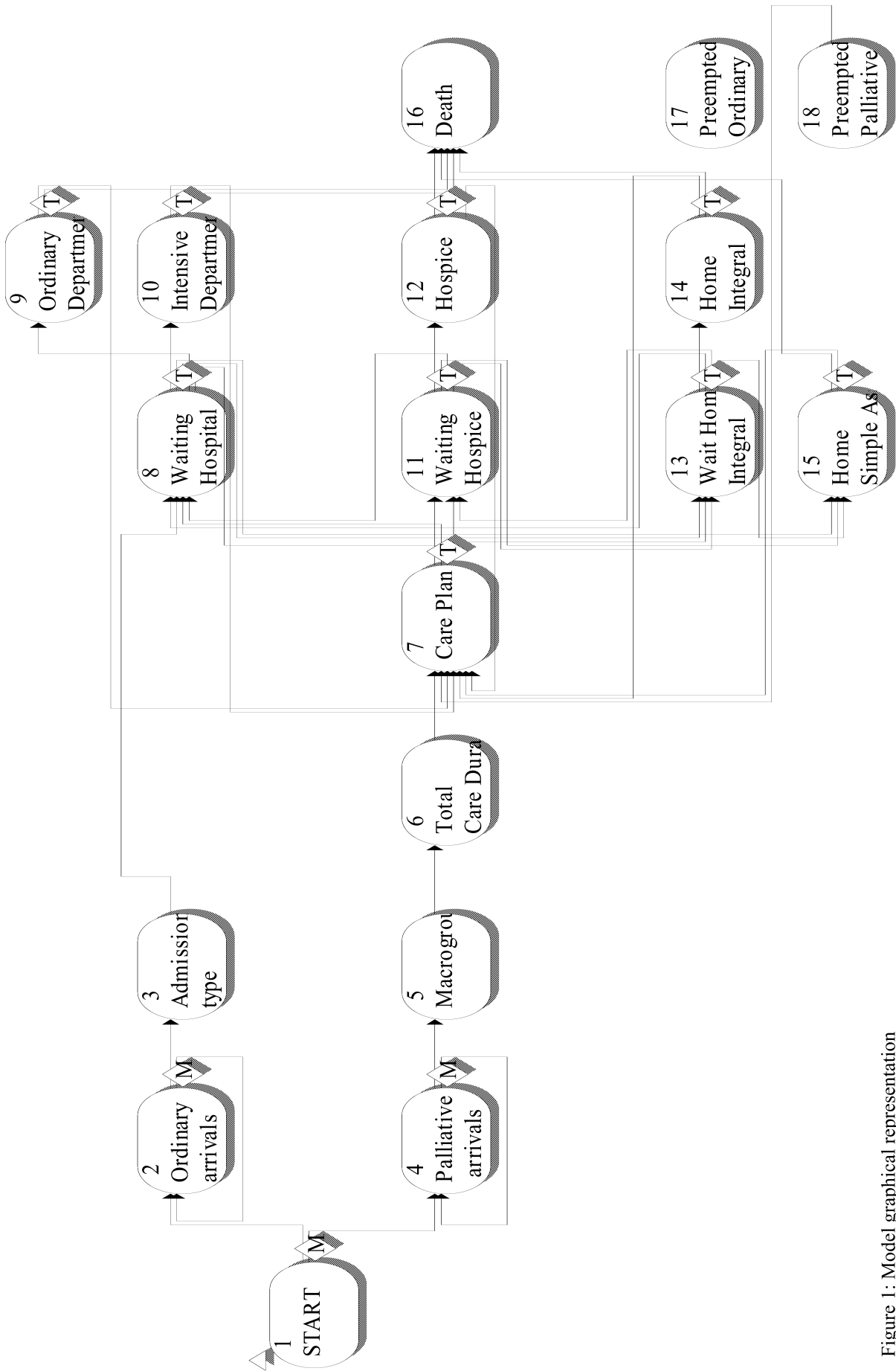


Figure 1: Model graphical representation

# BIRTH AND PERINATAL ASSISTANCE NETWORK ON THE TERRITORY: MODEL AND SIMULATION OF SERVICE DYNAMICAL BEHAVIOUR

Paola Facchin, Anna Ferrante,  
Elena Rizzato, Laura Salmaso  
Epidemiology and Community Medicine Unit,  
Department of Paediatrics - University of Padova  
Via Pietro Donà 11, 35129 Padova, Italy  
e-mail: [epi@pediatria.unipd.it](mailto:epi@pediatria.unipd.it)

Giorgio Romanin-Jacur  
Department of Management and Engineering  
University of Padova  
Stradella San Nicola 3, 36100 Vicenza, Italy  
e-mail: [romjac@dei.unipd.it](mailto:romjac@dei.unipd.it)

## KEYWORDS

Health care, decision support, discrete simulation.

## ABSTRACT

We consider a geographical region where population is distributed in health districts. In the same region there exists a three-tiered neonatal care network, which includes birth centres for supplying basic assistance for delivery, neonatal care and neonatal intensive care, with a design such that all higher-level centres also provide the lower levels of assistance. Each mother-to-be is admitted to a facility serving the level corresponding to the expected newborn conditions, based on pregnancy course; newborn transfers from a lower to a higher-level facility are feasible if the newborn conditions worsen. A simulation model describing mothers and newborns movements from districts to birth centres and among birth centres has been built up, with the aim of revealing lacks and inadequacies in supplying the required assistance in birth centres close enough to patients' homes. The model has been implemented in MicroSaint student on a personal computer and applied to Veneto region in North-East Italy but its use may be extended to other similar situations.

## INTRODUCTION AND PROBLEM STATEMENT

In recent years numerous clinical and technological advances have brought about dramatic improvements in neonatal care, enabling newborns that are severely premature or born in critical conditions to be kept alive. For this small group of newborns few specialized facilities, able to supply neonatal intensive care, are activated to serve large catchment areas. Alongside this restrict group of patients, there is a large group of newborns coming into the world in physiological or mildly pathological conditions, who can be cared for at less specialized facilities. It is convenient to distribute such facilities on the territory so to serve smaller catchment areas, so as to contain the distances between the facilities and the population residential areas they serve. To satisfy all the above requirements demands the creation of birth assistance network with facilities on different levels; moreover transfers needed from lower level facilities to higher level facilities for patients whose initial conditions deteriorate shall be provided. Three-tiered perinatal care networks have been planned in Europe and in USA (Brann et al. 1980; Le Roy et al. 2006; Van Reempts

et al. 2007). The three levels correspond to basic neonatal care on the first level, intermediate care for neonatal diseases on the second and neonatal intensive care on the third. Each perinatal care network has to be accurately planned in logistic terms, providing suitably sized and connected facilities for ensuring the maximum efficiency of each level of care.

In this paper we simulate patients movements from the territory towards the assistance network and inside the network; we assume that the mothers-to-be live in health districts dotted all over a given territory. Every mother-to-be is admitted to a birth centre serving the level corresponding to the needs of her pregnancy. Therefore each district has a certain demand for each of the three perinatal care levels previously mentioned. Within the territory considered, there is a service network that includes facilities for each level of neonatal care, and a characteristic of the network lies in that every higher level care facilities also provides the lower levels of care. Each facility is characterized by a capacity, i.e., the amount of patients who can be simultaneously admitted; more precisely the capacity of the third level of care is given by the number of ventilated incubators available and the capacity of the second level by the number of incubators; the capacity related to all the three care levels combined emerges from the numbers of beds available for the mothers. As seen above, every mother-to-be should be admitted to a facility on a level appropriate to the foreseen newborn conditions; if places are available, she is admitted to the closest facility with respect to her district, otherwise to other facilities in increasing distance order. At birth the newborn may present unexpected complications and require urgent transfer to a centre providing a higher level of care, at the same location or elsewhere, depending on the availability of levels of care and vacant places.

The aim of the simulation lies in i) checking the patients distribution among the existing facilities, ii) checking the assistance network ability of supplying a suitable service, in terms of admitting all patients sufficiently close to their home, and iii) possibly suggesting convenient adaptations in case of inadequate service.

In health care literature, and particularly for what concerns health care services planning, the problem of facilities location-allocation was discussed in (Toregas et al. 1971; Branas et al. 2000; Takinawa et al. 2006; Sahin et al. 2007; Mitropulos et al. 2006; Ratick et al. 2008;) with an optimization approach. The problem of perinatal facility

planning was studied by (Galvão et al. 2002; Boffey et al. 2003; Galvão et al. 2006). All above papers solve the planning problem by means of optimization but the solution is given as the average, in other words the mean of admitted patients is considered. In this paper the detailed movement of newborns is obtained revealing all peak conditions. The results report: patients admitted far from home, patients admitted out of the region, transfers among hospitals because of missing places. From the results useful suggestions may be obtained about suitable network resizing in order to improve service quality and reduce trouble due to distance and related costs.

## THE MODEL.

Admission requests rise in a random independent way for every level and for every district, according to statistics extrapolated from past data. Every request is dispatched to a facility of the corresponding level, more precisely the closest one with available places both for the mother and the newborn. If the newborn conditions worsen then he/she is transferred to the closest higher-level facility, without moving the mother who remains in the first admission facility. The ratio of newborns needing transfer between different levels is statistically known from the past. The lengths of stay distributions are shown to be gamma functions with parameters obtained from past data.

All distributions parameters are obtained by elaborating data from patient discharge papers, which are compiled in correspondence to every patient exit from hospital and therefore for every newborn.

The model evidences the saturation of all facilities and therefore their bed occupancy ratio. Moreover all network malfunctions may be revealed, more precisely the amount of patients that are admitted far from home and the amount of newborns that are transferred to another facility because of missing places.

sequences, rhombuses, representing decisions and striped rectangles, representing queues.

Our model implementation is reported in Figure 1, where task 1 is employed to start simulation, tasks 2-4 report all parameters utilized by the model, more precisely demand rate for each level and each district, all facility capacities and all distances (both district to birth centre and centre to centre). Tasks 5-7 are demand generators for the three levels; the arrival rate is the sum of all districts arrival rates while the district is probabilistically stated. Tasks 8-10 decide the dispatching of mothers and newborns and the possible newborns transfers. Task 11 represents mothers and newborns admitted to facilities out of the region, tasks 12-13 manage the dispatching of transferred newborns. Task 14-16 are necessary for functional aims and, finally, tasks 17-19 respectively describe mothers admissions and second and third level newborns admissions.

Note that patients' movements from all districts towards all facilities and among facilities have been described by means of a small model.

## MODEL APPLICATION AND RESULTS

The model was applied to the Veneto region in North-East Italy, a region with a population of about 4,700,000 and about 43,000 newborns per year. In the region there exists 52 health districts and 41 birth centres, all of which have both first facilities, for a total of 632 places, and second level facilities, with a total of 245 places, while only 9 of them have the third level facilities with a total of 57 places. The length of stay is a gamma distribution with a mean of respectively 3.6 days for the first level, 6.5 days for the second level, 17.6 days for the third level. All the above data have been obtained from the regional statistical office. The simulation reveals a lot of interesting results about the patients' movements and beds saturation. The saturation of beds for the mothers is 58.5%, the saturation of second

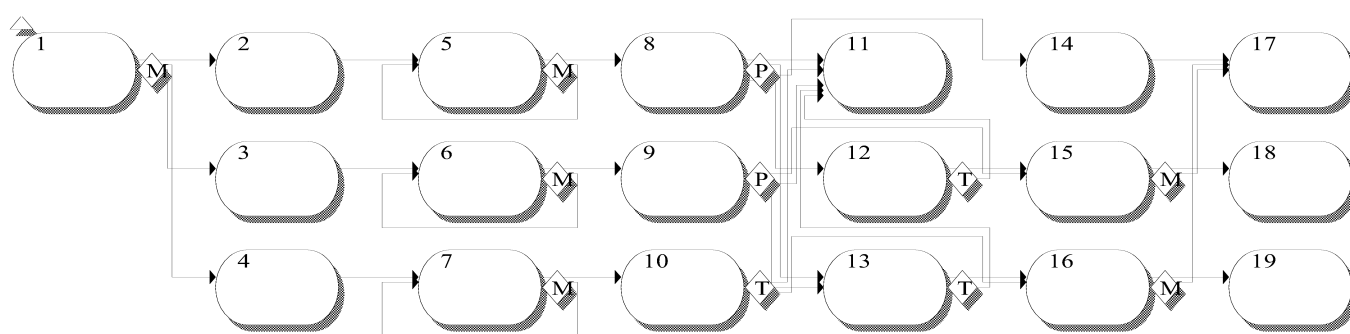


Figure 1 – The model

## MODEL IMPLEMENTATION

The model is quite simple, therefore it can be implemented using many simulation languages. Here it was implemented in MicroSaint Student, which was at disposition, but it may be easily translated in any language similar to SIMAN-Arena. MicroSaint implementation has a graphical representation that is based on four basic elements: tasks, describing activities, arrows, representing activity

level places is 61% and the saturation of third level places is over 99%. No first level newborn needs to be admitted out of the region, but the 97.6% are not admitted to the first choice birth centre and the 85% are constrained to be admitted to a facility over 50 km far from home: that evidences that the first level assistance is over dimensioned but badly distributed on the territory; the same happens for the second level, for which only the 1.6% needs to be admitted out of the region. For what concerns the third

level, we see that 64% are admitted out of the region and that evidences that the third level assistance is absolutely insufficient. If an improvement is to be obtained that shall be performed by means of an increase of places for the third level; a mild reduction of beds for the mothers and of second level places may be accepted; the first action requires an expansion of existing units, that means an increase of both ventilated incubators and specialized personnel; the second action may be simple. Obviously new distributions of birth centre on the territory would be desirable and may be easily simulated but it is much more complex in practice.

## CONCLUSION

We have built up and implemented a simulation model describing in detail the movements of mothers-to-be and newborns among health districts and assistance facilities, where assistance facilities are classified according to three different levels corresponding to the severity of newborn conditions. The aim of the model lies in checking system effectiveness and efficiency in providing adequate care to patients close enough to their home. Simulation results permit to reveal assistance lacks and to suggest suitable correcting actions. The model has been implemented on a personal computer and applied to Veneto region but it can be easily applied to other Italian or foreign regions.

## REFERENCES

- Boffey B, Yates D, Galvão RD.2003 "An algorithm to locate perinatal facilities in the municipality of Rio de Janeiro." *Journal of the Operational Research Society*;54:21-31.
- Branas CC, MacKenzie EJ, ReVelle CS. 2000 "A trauma resource allocation model for ambulances and hospitals." *Health Services Research* ;35(2):489-507.
- Brann AW, Hall RT Harper RG, Maisels J, Poland RL, Rhodes PG et al. 1980 "Level II neonatal units." *Pediatrics*; 66(5):810-1.
- Galvão RD, Espejo LGA, Boffey B.2002 "A hierarchical model for the location of perinatal facilities in the municipality of Rio de Janeiro." *European Journal of Operational Research* ;138:495-517.
- Galvão RD, Espejo LGA, Boffey B.2006 "Practical aspects associated with location planning for maternal and perinatal assistance." *Ann Oper Res*;143:31-44.
- Le Roy C, Carayol M, Zeitlin J, Breart G, Goffinet F 2006. "Level of perinatal care of the maternity unit and rate of cesarean in low risk nulliparas." *Obstetrics & Gynecology*; 107(6):1269-1277.
- Mitropoulos P, Mitropoulos I, Giannikos I, Sissouras A. 2006 "A biobjective model for the locational planning of hospitals and health centers." *Health Care and Management Science*;9:171-79.
- Ratick S, Osleeb JP, Hozumi D.2008 "Application and extension of the Moore and ReVelle hierarchical maximal covering model." *Socio-Economic Planning Sciences*;30:1-10.
- Sahin G, Sural H, Meral S. 2007 "Locational analysis for regionalization of Turkish red crescent blood services." *Computer & Operations Research*; 34: 692-704
- Tanikawa T, Ohba H, Terashita T, Uesugu M, Jiang G, Ogasawara K, Sakurai T. 2006 "Model analysis for optimal allocation of pediatric emergency center." *AMIA Annual Symposium Proceeding*;1115.
- Toregas C, Swain R, ReVelle C, Bergman L. 1971 "The location of emergency service facilities." *Operations Research*; 19(6): 1363-1373
- Van Reempts P, Gortner L, Milligan D et al. 2007 "Characteristics of neonatal units that care for very preterm infants in Europe: result from the Mosaic study." *Pediatrics*;120(4):815-25.

## BIOGRAPHIES

**PAOLA FACCHIN** obtained a medical degree at the University of Padova (Italy) in 1977. She obtained the diploma in Paediatrics in 1980, the diploma in Public Health in 1984 and the Phd in Developmental Sciences and Epidemiology in 1988 at the same University. She taught many courses of general Paediatrics, Epidemiology, Community Medicine and Child Maltreatment at many Italian Universities. She is now Associated Professor of Paediatrics and coordinator of the Epidemiology and Community Medicine Unit at the University of Padova. Her current research activities concern Health Service Planning, Community Health and Health Needs, Injury Prevention, Child Abuse.

**ANNA FERRANTE** obtained a medical degree at the University of Padova (Italy) in 1980. She obtained the diploma in Public Health in 1984. Since 1985 she has worked in Prevention Department and in Medical Districts. She is now senior doctor at the Epidemiology and Community Medicine Unit of the University of Padova. Her currently research activities concern Epidemiology in the mother-infant field and activities coordination between hospitals and territorial assistance units.

**ELENA RIZZATO** obtained a degree in Management Engineering at the University of Padova in 2006 and is now attending the PhD school in Developmental medicine and Programming Sciences (research line: health planning models/systems planning). Her current research activities concern Epidemiology in the mother-infant field and home assistance.

**GIORGIO ROMANIN-JACUR** was graduated in Electrical Engineering at the University of Padova in 1970. He was fellow and then research associate at the National Council of Researches until 1980. Since 1980 he was lecturer of Operations Research at the School of Engineering of the University of Padova, then Associated Professor and finally Full Professor. His current research activities concern modelling and optimization of health services and of transportation services.

**LAURA SALMASO** was graduated in Statistics at the University of Padova in 2000. She obtained the PhD in Developmental medicine and Programming Sciences in 2004. Since 2004 she works as a researcher at the Epidemiology and Community Medicine Unit of the University of Padova. Her current research activities concern Epidemiology in the newborn assistance field.

# CRITICAL NEWBORN ASSISTANCE IN INTENSIVE CARE UNITS: MODEL AND SIMULATION

Monica Da Frè, Paola Facchin, Elena Rizzato,  
Laura Salmaso, Laura Visonà Dalla Pozza  
Epidemiology and Community Medicine Unit,  
Department of Paediatrics - University of Padova  
Via Pietro Donà 11, 35129 Padova, Italy  
e-mail: epi@pediatria.unipd.it

Giorgio Romanin-Jacur  
Department of Management and Engineering  
University of Padova  
Stradella San Nicola 3, 36100 Vicenza, Italy  
e-mail: romjac@dei.unipd.it

## KEYWORDS

Health care, Decision support systems, Dynamic modelling, Discrete simulation

## ABSTRACT

High risk newborns mortality has been dramatically reduced by recent progress in perinatal care. Related cares are organised by local regions in assistance structures classified according to three hospital levels; third level structures supply neonatal intensive and sub-intensive care. A simulation model describing critical newborns movements among birth centres and third level hospitals has been built and implemented, with the scope of checking regional third level neonatal care capacity and possibly suggesting suitable adaptations. The model has been usefully applied to Veneto region in Italy.

## INTRODUCTION

Advances in perinatal care in the last decades are able to decrease mortality of high risk newborns, including very preterm babies (<32 weeks of gestational age) (Cooke 1991; de Kleine et al. 2007) and very low birth weight (<1500 grams) (Harper et al. 2002; Fanaroff et al. 2007). Mortality and morbidity of high risk newborns are associated with the initial level of perinatal care (Sanderson et al. 2000; Empana et al. 2003; Lee et al. 2003).

Since 1970's, in many European countries and in North America, public health programmes introduced the concept of regionalization of perinatal care and health services for high risk pregnancy were organised to ensure these babies to be born in few hospitals equipped with the expertise and technology needed for newborn optimal care (Zeitlin et al. 2004; Stark et al. 2004).

Simulation models are widely used to evaluate public health decisions including patients movements, hospital beds requests and the resource allocation. They are also an instrument of representation and analysis that allows to study the evolution of the system and describe the environment. Simulation models allow to make predictions to understand future evolution of various phenomena (Geller and Yochmowitz 1975; Kozan and Gillingham 1997; Ridge et al. 1998; Bortolato et al. 2008; Barella et al. 2008). None of the above papers describes high risk

newborn assistance in a regional network of intensive care units.

In the present paper we build up a simulation model describing high risk newborn generation and movements among existing assistance structures with the aim of detecting whole critical points of the assistance network and to suggest suitable adaptations obtainable by resizing assistance structure capacities, both in the current situations and in future scenarios.

In the first section organisation of perinatal cares is analysed; in the second section data sources and data elaboration methods are examined; in the third section the simulation model and its implementation is discussed in detail; finally some applications are reported.

## ORGANIZATION OF PERINATAL CARES

Perinatal cares are organized by local regions in assistance structures classified according to three hospital levels, corresponding to structural characteristics and to offered professional abilities; the same three levels are defined for obstetric units. First level neonatal assistance units take care of newborns without ascertained pathologies, born after more than 34 gestation weeks. Second level units assist newborns characterized by mild pathologies or weight over 1500 grams, or born after more than 32 gestation weeks. Third level units generally include an intensive care unit and a sub-intensive care unit; the intensive care units assist newborns needing artificial ventilation, surgical newborns before and after surgical operations, newborns with less than 32 gestation weeks, newborns with complex malformations and/or needing invasive or specific diagnostic procedures; the sub-intensive care units assist newborns after admission in the intensive care units, after their reaching better health conditions, but not yet admissible in a second level unit. Each third level intensive or sub-intensive care unit is characterised by the number of places at disposition, depending on the number of special incubators and on the active specialised doctors and nurses.

In a hospital we always have obstetric and neonatal assistance units of the same level, and moreover if an upper level unit is activated, all lower level units are activated too; thus we have most hospitals with first level units, some hospitals with first and second level units and few hospitals with first, second and third level units.

Birth should be generally planned in a birth centre which

corresponds to expected newborn conditions, on the basis of pregnancy course; however newborn conditions may present a sudden unforeseen worsening, and therefore require a quick transfer to an upper level neonatal unit.

Let us consider a geographical region where urban nuclei are distributed on the territory; such urban nuclei originate perinatal assistance requests, which can be satisfied by the existing assistance structures. Organization problems arise when the assistance network is not correctly sized, as an incorrect size may oblige a newborn to be admitted in a structure far away from the parents' residence nucleus, and sometimes in a structure outside the region.

## THE MODEL

In the following we present the model we have built up, which describes admissions and possible movements of third level newborns, i.e., high risk ones, requiring admission in a third level unit. Third level newborns are generated a random independent way in all birth centres according to estimated means. Then every newborn is admitted in the closest third level intensive care unit where idle places are at disposition; if no places are idle, then the second closest unit is considered, and so on, until an idle place is found. If unfortunately no idle places can be found in any of the existing units, the admission is performed out of the region. The admission implies one bed (ventilated incubator) occupation; the stay may be interrupted by death. High risk newborns admitted in intensive care units, at the end of their stay, in the majority (with given probability) compete for a place in the sub-intensive care unit: if such a place is not available, then the stay in the intensive care unit is extended but the bed occupation is incorrect; at the end of admission in sub-intensive care unit admission patients get out of the system. A small given fraction of newborns admitted in intensive care unit, at the end of their stay, are transferred to a second level hospital unit, thus getting out of the system.

Model data were supplied by the Epidemiology and Community Medicine Unit of the University of Padova: all of them were obtained from the hospital discharge papers (SDO) which are released to every patient admitted in a hospital of the Veneto region, where the Unit is sited.

Model data permitted us to define statistical distributions, which are respectively exponential for the birth interarrival time and for length of stay in the intensive care unit and gamma for the length of stay in sub-intensive care unit. The related parameters are mean time between successive births in every birth centre, mean length of stay in the intensive care units, mean and variance of length of stay in sub-intensive care unit. Length of stay in intensive care units is significantly shorter for newborns who die during the hospital stay with respect to the surviving ones. Other estimated data are respectively the probability of death during stay and the probability of admission to second level hospital unit at the end of stay in intensive care unit.

## THE IMPLEMENTATION

The model we have built up may be implemented in any discrete simulation language without difficulties. For the

present study the model was implemented by means of language MicroSaint 3.2; with small and quick adaptations may be translated into other languages similar to SIMAN-ARENA. The model ran on a personal computer requiring few seconds to simulate the activity of an actual regional neonatal third level assistance network related to a period of two years (one year for initialization and one year to get statistics), concerning about 1,800 newborns per year.

Model implementation uses a graphical model representation where essential elements are: ellipses (tasks) representing activities, directed arcs representing activity sequences, diamond representing (multiple, probabilistic or deterministic) decision and striped rectangles representing queues.

The model graphical representation, related to the application reported below, includes 126 task, of which task 1 gives the initial start to a series of 40 identical sets of tasks, like the one reported in Figure 1, each one representing a birth centre; newborns are generated in a random independent way by task 2, while task 52 manages the address to existing intensive care units in order of increasing distance, according to idle places (tactical decision); if no places are found an admission out of the region is performed by task 121 (the last one, not reported).

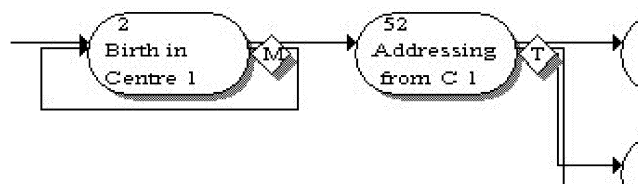


Figure 1: Generation of newborns and addressing to the closest Neonatal Intensive Care Unit with idle places

Generated newborns are addressed to a hospital equipped with intensive and sub-intensive care units: each hospital is represented by a block of five tasks reported in detailed in Figure 2 where task 94 represents admission, task 104 represents hospital stay in neonatal intensive care unit (NICU) for survivors, task 130 represents hospital stay in neonatal sub-intensive care unit, task 151 represents exit towards other hospitals and task 105 represents the stay in NICU for non survivors.

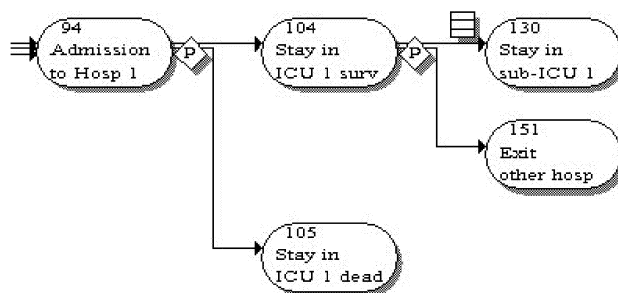


Figure 2: Hospital stay in Neonatal Intensive and sub-Intensive Care Unit

## THE RESULTS

The simulation main results are the following. We may evidence: i) the amount of admissions out of the region, ii) the amount of admissions in a site different from the one where birth took place, even when a neonatal intensive care unit is active there; iii) the amount of admissions in sites at large distance from the one where birth took place; iv) the amount of refuses to admission requests, due to busy beds, concerning newborns coming respectively from the same site and from outside; v) the amount and the duration of incorrect stays in the neonatal intensive care units before admission in the sub intensive care units; vi) mean bed occupancy in all units.

If simulation results are not satisfying, i.e., the above i) to v) amounts are judged to be too large, a network adaptation may be suggested. Adaptations to be considered concern the number of beds in the various units: may be some units need to be enlarged while some others can be conveniently reduced. In order to know which variations are to be effected, a simulation is performed after a strong increase of beds in all units: those units which increase mean bed occupancy can be suitably enlarged, while those which decrease mean bed occupancy can be suitably reduced. Obviously, the network behaviour after adaptation shall be tested after parameter adjustment on the same model.

## ACTUAL APPLICATION

The model has been usefully applied to the third level neonatal assistance network of the Veneto region in North-East Italy. The region has about 4,800,000 inhabitants and about 48,000 births per year of which 1,800 need third level assistance. In the region there are 40 birth centres of which 9 offer third level assistance, with a total of 40 places in the intensive care units and 63 in the sub-intensive care units. The model reveals that the interested assistance network is absolutely insufficient with respect to the required service. Therefore a complete reorganization is suggested to compensate such insufficiencies, which requires to increase places in all sites, even if at different ratios; the effects of suggested adaptations were tested and revealed satisfactory.

## CONCLUSIONS

A simulation model of neonatal intensive care units behaviour in a geographical region has been built up, implemented and applied to Veneto Region in North-East Italy. The model may be usefully applied to other Italian regions and to regions of foreign countries, in current situation or in future scenarios, for which the assistance requirements may be computed by statistical elaborations.

## REFERENCES

- Barella, A.; P. Facchin and G. Romanin-Jacur. "Pregnancy-birth course in Veneto Region: development of a flexible simulation model for description and forecasting" *Proceedings of the 6<sup>th</sup> International Industrial Simulation Conference ISC'2008* (Lyon, France; Jun.9-11). EUROSIS-ETI, 348-352.
- Bortolato, F.; A. Ferrante; G. Romanin-Jacur and L. Salmaso. "Critical newborn transport in Veneto Region: model and simulation." *Proceedings of the 7<sup>th</sup> International Workshop on Modelling and Applied Simulation MAS 2008* (Rende CS, Italy, Sept.17-19), DIPTM University of Genoa, 187-192.
- Cooke, R.W. "Trends in preterm survival and incidence of cerebral haemorrhage 1980-9." *Arch Dis Child*. 1991 Apr;66(4 Spec No):403-7.
- de Kleine, M.J.; A.L. den Ouden; L.A. Kollée; A. Ilsen; A.G. van Wassenae; R. Brand and S.P. Verloove-Vanhorick. "Lower mortality but higher neonatal morbidity over a decade in very preterm infants." *Paediatr Perinat Epidemiol*. 2007 Jan;21(1):15-25.
- Empana, J.P.; D. Subtil and P. Truffert. "In-hospital mortality of newborn infants born before 33 weeks of gestation depends on the initial level of neonatal care: the EPIPAGE study." *Acta Paediatr*. 2003;92(3):346-51.
- Fanaroff, A.A.; B.J. Stoll; L.L. Wright; W.A. Carlo; R.A. Ehrenkranz; A.R. Stark; C.R. Bauer; E.F. Donovan; S.B. Korones; A.R. Laptook; J.A. Lemons; W. Oh; L.A. Papile; S. Shankaran; D.K. Stevenson; J.E. Tyson; W.K. Poole and NICHD Neonatal Research Network. "Trends in neonatal morbidity and mortality for very low birthweight infants." *Am J Obstet Gynecol*. 2007 Feb;196(2):147.e1-8.
- Geller, N.L. and M.G. Yochmowitz. "Regional planning of maternity services." *Health Serv Res*. 1975 Spring; 10(1): 63-75.
- Harper, R.G.; K.U. Rehman; C. Sia; S. Buckwald; R. Spinazzola; J. Schlessel; J. Mestrandrea; M. Rodgers and R.A. Wapnir. "Neonatal outcome of infants born at 500 to 800 grams from 1990 through 1998 in a tertiary care center." *J Perinatol*. 2002 Oct-Nov;22(7):555-62.
- Kozan, E. and S. Gillingham. "A simulation model to increase the efficiency of a hospital maternity system." *Proceedings of the International Congress on Modelling and Simulation Modsim 97* (Hobart, Australia, Dec.8-11) University of Tasmania, Hobart, Vol.3, 1035-40.
- Lee, S.K.; D.D. McMillan; A. Ohlsson; J. Boulton; D.S. Lee; S. Ting and R. Liston. "The benefit of preterm birth at tertiary care centers is related to gestational age." *Am J Obstet Gynecol*. 2003 Mar;188(3):617-22.
- Ridge, J.C.; S.K. Jones; M.S. Nielsen and A.K. Shahani. "Capacity planning for intensive care units." *European Journal of Operative Research*. 105 (1998) 346-355.
- Sanderson, M.; W.M. Sappenfield; K.M. Jespersen; Q. Liu and S.L. Baker. "Association between level of delivery hospital and neonatal outcomes among South Carolina Medicaid recipients." *Am J Obstet Gynecol*. 2000 Dec;183(6):1504-11.
- Stark, A.R. and American Academy of Pediatrics Committee on Fetus and Newborn. "Levels of neonatal care." *Pediatrics*. 2004 Nov;114(5):1341-7. Erratum in: *Pediatrics*. 2005 Apr;115(4):1118.
- Zeitlin, J.; E. Papiernik; G. Bréart and EUROPET Group. "Regionalization of perinatal care in Europe." *Semin Neonatol*. 2004 Apr;9(2):99-110.

## BIOGRAPHY

**MONICA DA FRÈ** was graduated in Statistics at the University of Padova, obtained a Master in Biostatistics and Epidemiological Methodology at University of Pavia and is student of the PhD school in Developmental Medicine and Planning Sciences at University of Padova. She is now working at the Epidemiology Unit of the Regional Health Agency of Tuscany.



# **MULTIBODY SIMULATION**



# Methodology for Flexible Modelling of Escalator Multibody Systems

Juan D. Cano-Moreno  
José M<sup>a</sup> Cabanellas Becerra  
Carlos Labajo Tirado  
Jesús Félez Mindán  
Universidad Politécnica de Madrid  
Escuela Técnica Superior de Ingenieros Industriales.  
Centro de Investigación de Tecnologías Ferroviarias  
c/José Gutiérrez Abascal,nº 2.  
28006, Madrid, Spain  
E-mails: {citef.jdcano|jmcabanellas}@etsii.upm.es

## KEYWORDS

Escalator simulation, Multibody dynamics, SIMPACK,  
Dynamic Modelling, Robust Design

## ABSTRACT

This paper presents some particular escalator modelling features and methodologies developed to dramatically reduce time cost regarding two aspects: computation and implementation.

CITEF (Railway Technologies Research Centre) has been modelling escalators for three years. During this time, several static, kinematic and dynamic escalator models have been developed and improved. In parallel, automation tools mainly intended for saving time cost have been described in a piecemeal fashion. These tools are based on the repetitiveness of the bodies, and a definition of the joints, forces and loops, and on the cyclic movement of most of the bodies involved. In addition, noise signals have been programmed from MATLAB to simulate them in SIMPACK software in order to apply robust design methods for studying and optimizing certain parameters.

## INTRODUCTION

Escalator design has been carried out successfully for more than a century. An exhaustive analysis of patents and viability studies of static, kinematic and dynamic models has been the starting point for developing a methodology to simulate and analyze the overall behavior of this Multibody system from three points of view: static, kinematic and dynamic. All of these kinds of models have produced coherent results. Some representative results have been presented in past papers (Cabanellas et al. 2008a; Cano et al. 2008)

The methodology and models have been developed in parallel form as a feedback system.

## STATE OF THE ART

One and a half centuries have passed since the first escalator patent was taken out in 1859, shown in Figure 1. The history of escalators (Cabanellas et al. 2008b; Miravete and Larrode 2007) shows that no drastic change in their basic mechanism and working form has taken place.

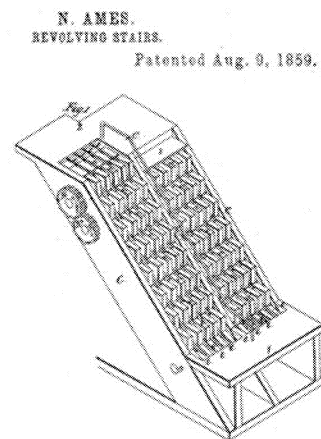


Figure 1: Nathan Ames Revolving Stairs (1859)

## Escalator Simulation Models

The number of escalator simulation models developed has been very small so far. When searching for publications in the bibliography related to this field, few papers have been found. LG Industrial Systems (LGIS) has used computer simulation to develop escalators that improve performance over conventional designs using DADS Multibody dynamics software (Sug 1999a).

Despite no other full escalator dynamic simulation models having been found, there are other partial dynamic models related to the study of escalator device handrail systems (Sug 2005) and vibration reduction using robust design (Sug 1999b).

## ESCALATOR MODELLING

Mechanically, an escalator is a Multibody system with a considerable number of bodies with their respective joints, contact forces, restrictions, etc. As previously stated herein, there is hardly any bibliography or information on escalator modelling and simulation. CITEF has had to research and select some tools to simulate the static, kinematic and dynamic behavior.

Advanced and complex dynamic models of more than 1000 degrees of freedom have been developed and simulated successfully using SIMPACK (Simulation of Multibody System Package) software. CITEF has had long experience in modelling multibody dynamics behavior using this software. In addition, as conventional escalators are mainly moved using two roller chains linked to each step with a revolute joint, a specific chain modelling SIMPACK module has been used to develop some models. Figure 2 shows a conventional escalator model moved using traction gear wheels. In this model, all chain bodies like chain links, tensioner and traction systems have been modelled by the chain module.

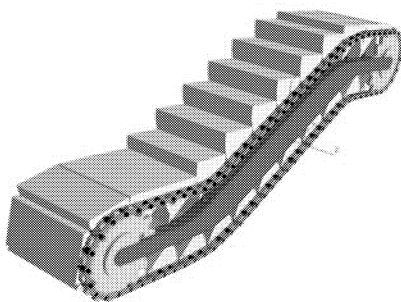


Figure 2: Conventional Escalator modelled using SIMPACK Chain Module

However, this process has meant confronting some difficulties, some of which have been overcome with the methods described in this paper.

For a start, roller-guide contact has been solved as a force between two mobile markers named parent and child markers, as Figure 3 shows. Each parent and child marker has an assigned geometry in which they can move. The child marker is always located at the minimum distance from the parent marker, moving along following its geometry. The contact force has been defined with a spring-damper parallel force element.

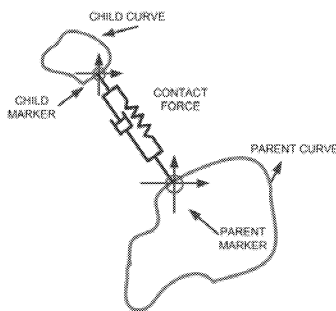


Figure 3: Roller-Guide Contact Force Definition

On the other hand, two types of traction systems have been simulated. A conventional traction system designed using SIMPACK CHAIN module (Cabanellas et al. 2008c), where chain link bodies, their contacts and traction and tensioner system can be created. In addition, linear traction has been simulated using a proportional control system that applies the longitudinal traction force on each chain link in order to maintain its velocity, as the reference velocity curve indicates, during the definite length control.

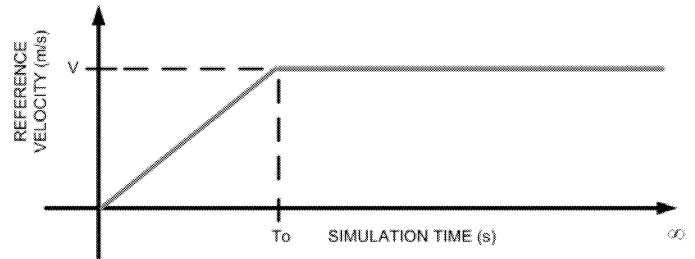


Figure 4: Reference Velocity for Control System

A tensioner system has been created in SIMPACK basic software by doubling the mobile markers and the contact forces existing between them.

In addition, some parameterized programs have been created to facilitate the implementation of some specific guide shapes like pulse-free curves or guides with an inconstant radius, taking into account that these bodies are not made of basic geometries.

Finally, if the time cost is analyzed, there are two clear fronts with a potential high time investment:

1. More than 500 bodies are defined in a model of an escalator with a height of around 4 m. and with a 30° inclined angle.
2. For the most complex models it can take more than a week to simulate a full cycle in some cases. Therefore, time integration is a decisive factor in escalator modelling.

Thus, it is clear that automation in the modelling process is a necessity if greater efficiency is to be obtained. No less important is the time integration, because the main objective of Multibody simulation is to save time and cost compared to research that uses real prototypes to test improvements.

This paper shows how to save integration time and obtain all cycle results in certain cases, under specific hypotheses.

## AUTOMATION TOOLS

As we have already described, the simulation of escalator dynamics requires an enormous cost in modelling process and simulation time. CITEF has developed some tools and methodologies in order to reduce both of these drawbacks.

The overall dynamic behavior of escalators is dictated by the traction roller chains. These Multibody systems are made up of some repetitive bodies: Chain links (inner and outer), rollers and others auxiliary bodies. In addition, almost all of these bodies have a cyclic movement. Most of the tools developed are based on these characteristics.

## Superposition Process

In a permanent regime and with a stationary state, the kinematic and dynamic outputs of each roller or chain link are the same as the previous roller or chain link when they reach the same position. In other words, the values of these outputs are scalar fields because they depend on their position,  $x$  and  $y$  for instance, for a 2D model. Then, a time variable can be expressed by these spatial variables. This assumption is clear in static and kinematic analysis, however, dynamic simulation adds some oscillations in position and velocity, which have been considered negligible when the hypothesis described can be carried out.

$$F(t) = F(t + nT_c) = G(x, y) \quad (1)$$

$n$ , number of cycles  
 $T_c$ , period of a cycle

Therefore, an output variable like velocity along the time of a chain roller, represented in Figure 5, can be associated with its guide geometry.

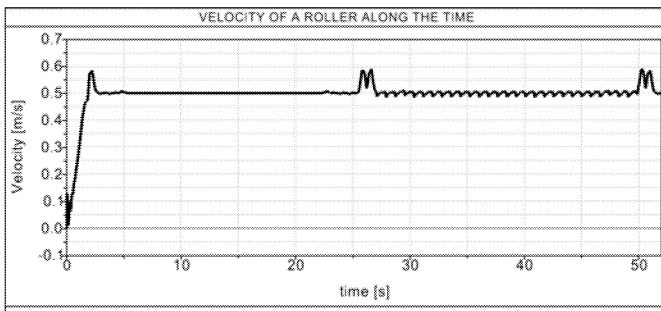


Figure 5: Velocity Output along the Time

By way of illustration, Figure 6 shows the previous velocity output as a scaled offset of the geometry that the corresponding roller has followed. Thus, each colored line length is proportional to velocity values, and all points of the geometry have an associated velocity. Other important dynamic variables such as longitudinal chain link force, roller-guide contact force, step, roller or chain link acceleration, etc., have the same property.

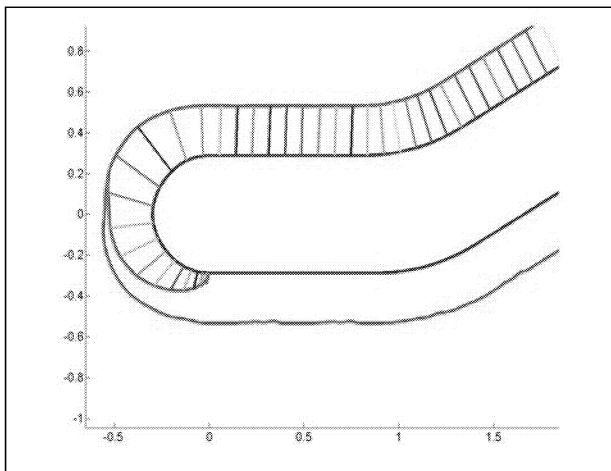


Figure 6: Velocity Polar Diagram

As a result, each output function of any variable related with the bodies which move along the guides could be obtained

from the different output functions measured during a time period corresponding to a chain link pitch.

## Code Generation

Escalator dynamic model generation involves a hard and heavy task of implementation that supposes a time cost increase. Several MATLAB programs have been created in such a way that there are a lot of bodies, markers, joints, forces, restrictions, etc, with similar functions and definition.

These programs developed produce and export to a separate text document the corresponding code that has been used in SIMPACK files to define each repetitive feature of the model. Therefore, the number of chain links in the roller chain of the model is irrelevant.

On the other hand, full SIMPACK files have been generated from MATLAB software in order to define some parameterized geometries, like guide geometries, single input functions and input functions arrays. These files are fully legible by SIMPACK software.

Input functions are temporary functions that can be used to define a lot of SIMPACK variables. Thus, some applied states have been reproduced (empty or full escalator, an escalator becoming full, etc). One of the advantages of using this implementation method is that any programmable function in MATLAB can be used in SIMPACK simulations.

## SIMULATION AND RESULTS ANALYSIS

All the tools and methodologies previously described have been successfully tested with several models for reducing time integration and to make a robust analysis. Some applications of the tools that have been described in this paper are explained.

### Incremental Construction of Time Responses

The superposition process has been mainly used in advanced and complex models where time integration is a critical factor. It is necessary to select a range of time corresponding to the time in which each roller chain reaches the next one. Although the starting time point could be any, for example, second 23, as the following figures show, the less the initial time, the greater the time cost reduction.

Figure 7 shows the absolute linear velocity output of a dynamic escalator model simulated in SIMPACK. This model is characterized by chain links of 0.405 m, a reference absolute linear velocity of 0.5 m/s and linear traction system located at the upper part of the right inclined zone. All bodies have been modelled as rigid bodies and friction force has not been taken into account. In addition, the case simulated is an "empty escalator".

Four rollers, called sequentially, have been selected in order to exemplify how the superposition tool works. From starting time, a range of 0.81 seconds has been plotted for all rollers. The lower plot shows the continuous simulation of the first roller, checking stretch by stretch the coherence between both reconstructed and continuous signals.

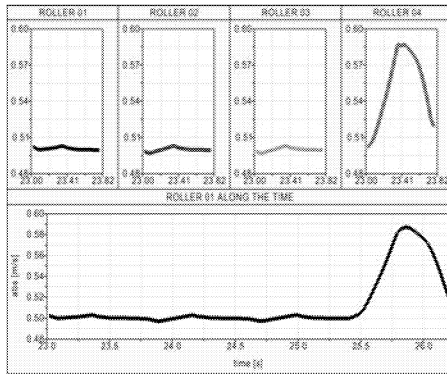


Figure 7: Dynamic Absolute Linear Velocity Superposition

Hereafter, the efficiency and consistency of this tool will be statistically quantified for some different outputs of this model by correlation coefficient.

### Robust Design

Taguchi Methods (Wu 1997) have been applied to study the most robust tensioner parameters design that minimizes velocity standard deviation. The tensioner system is defined by three parameters: damping, stiffness and pre-load force. Three levels have been selected for each parameter. The equation below shows the formula used to calculate the signal-to-noise ratio by applying the minimization criterion.

$$S/R = -10 \cdot \log_{10} \left[ \frac{1}{n} \sum_{i=1}^n Y_i^2 \right] \quad (2)$$

The noise signal is made up of random loads that represent the weight of passengers distributed by a statistical probability distribution with a mean and a pre-selected standard deviation. In addition, the probability of a passenger going up on the escalator is 50%. A code generation tool has been used to create the input functions arrays that define one noise signal for each step, for any number of cycles, chain link pitch or number of steps. All of these characteristics are parameterized in MATLAB code.

As an example, Figure 8 shows five noise signals of five consecutive steps for 20 cycles of an escalator with 60 steps, a distance between step roller of 0.405 m and a reference velocity of 0.5 m/s. The probability distribution used is a uniform distribution with a mean of 97 kg and a standard deviation of 5.625 kg.

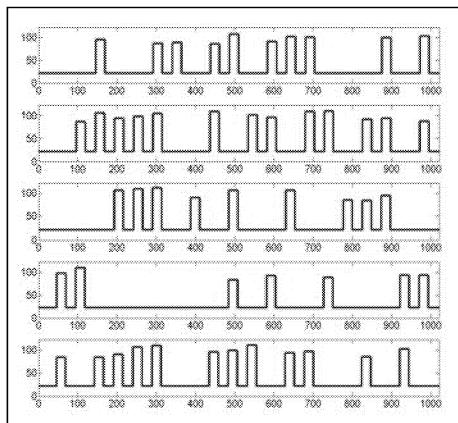


Figure 8: Noise Signal Example

This technique could be used to optimize any other parameter in a robust way, saving the costs of real experiments that involve a prototype construction.

### DISCUSSION

Different outputs related to bodies that are moving in a closed curve have been generated using this superposition process. Thus, static, kinematic and dynamic outputs along a cycle have been reconstructed. The most critical reconstruction is the dynamic one, so, this paragraph is dedicated to analyzing the coherence between real and reconstructed signals.

Dynamic reconstruction requires simulating for around 5 seconds. Theoretically, less than a second would be enough; however, there is a transitory regime at the beginning and constant velocity is reached after a few seconds, depending on how the reference velocity for the control system is defined.

Although characteristic escalator dynamic outputs have been selected, only some of them are suitable for applying this tool: step velocity and acceleration, longitudinal chain link force and roller-guide contact force have been used to analyze this tool.

Longitudinal chain link force has been selected to illustrate how this tool works. This output belongs to a model of an empty escalator with a height of 4.5m powered with a linear traction system. As Figure 9 shows, both represented curves, the obtained curve and the real curve, are close to being superposed. The Real Curve has been obtained by a cycle of simulation, in order to compare with the curve created through the superposition process.

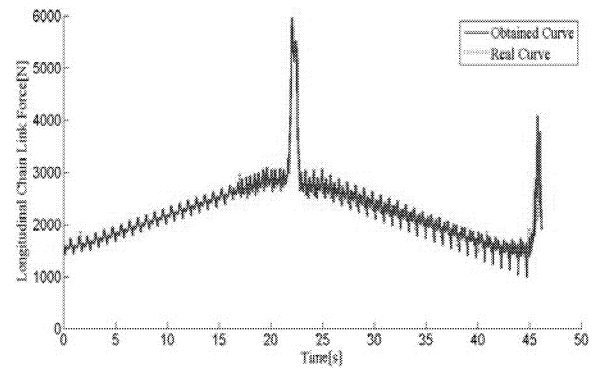


Figure 9: Longitudinal Chain Link Force Comparison

The resemblance between both signals has been analyzed statically using regression analysis. Thus, straight-line equations have been obtained for each variable, as the symbolic equation below shows.

$$V(real) = a + b \cdot V(reconstructed) \quad (3)$$

with  $a \rightarrow 0$  and  $b \rightarrow 1$

The Observed-Predicted graphic is shown in Figure 10, distinguishing a clear straight line that divides the first quadrant, which means the independent parameter of equation (3) tends to be null and the other tends to the unity.

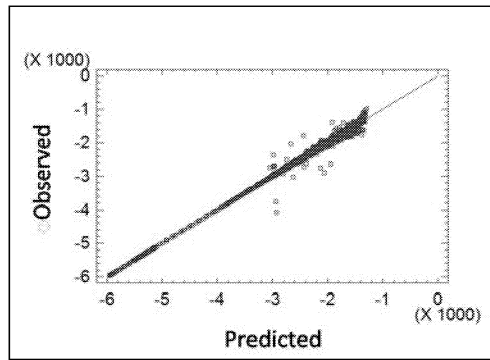


Figure 10: Observed-Predicted Relation for Longitudinal Chain Link Force Output

The rest of the reconstructed outputs present similar statistical results to the example described in this work. The working form of this tool has been evaluated by correlation coefficients, represented in Figure 11 as a bar diagram.

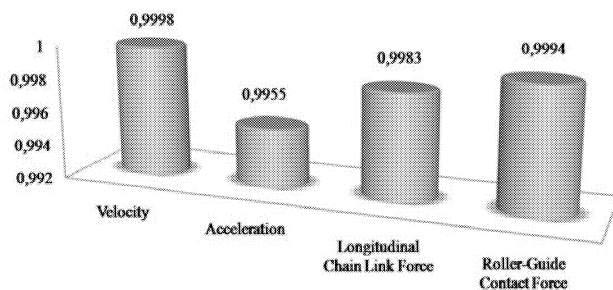


Figure 11: Correlation Coefficients

As the minimum correlation coefficient exceeds 0.99, this tool presents an efficient, accurate, consistent and powerful method for reconstructing any signal under the assumptions previously detailed. Thus, time integration cost is dramatically reduced. As Table 1 shows, the time cost reduction can reach 90% in relative terms for a full load case, and a range of 18 to 45 days per full cycle simulation.

Table 1: Time Simulation Comparison

|              | Days to Obtain one Cycle Results |   |
|--------------|----------------------------------|---|
|              | Simulation<br>(50 seconds)       | Superposition<br>Process<br>(5 seconds) |
| Without Load | 20                               | 2,5                                     |
| Full Load    | 50                               | 5,5                                     |

## CONCLUSIONS

All the aforementioned tools and methodologies described have been successfully tested with several models, obtaining a complete and flexible methodology to model and simulate overall escalator behavior, reducing time cost and resolving some modelling problems.

CITEF is developing and implementing new concepts in escalators, a product that has become stagnant or constant since its conception. Some lines of research are being

followed by CITEF with simulation software always being the option for testing and analyzing any conventional or innovative model.

Furthermore, CITEF is working to develop a full design cycle methodology in the escalator field that will be completed with a dynamic model experimental validation.

## REFERENCES

- Cabanellas Becerra, J. M., Cano Moreno, J. D., Suárez, B., Chover, J. A., & Félez, J. (2008). "Advanced Dynamical Models for Escalators Simulation". *Elevator World*, 136-141.
- Cano Moreno, J. D., Cabanellas Becerra, J. M., Suárez, B., & Chover, J. A. (2008). "Simulación y Modelización de las Escaleras Mecánicas como Vía para la Mejora del Confort de los Pasajeros y la Disminución de Costes". *IV Congreso Nacional de Innovación Ferroviaria*.
- Cabanellas Becerra, J. M., Cano Moreno, J. D., Suárez, B., Chover, J. A., & Félez, J. (2008). "Mejora de un diseño de más de 100 años. Nuevos conceptos en escaleras mecánicas". *Anales de Ingeniería Mecánica, Revista de la Asociación Española de Ingeniería Mecánica*, 1, 233-239.
- Miravete, A., & Larrode, E. (2007). "Elevadores: Principios e Innovaciones". Reverté
- Sug Kwon, Y. (1999). "Computer Simulation Helps Build Better Escalators". *Elevator World*, 120-121.
- Sug Kwon, Y., Scott, G., & Park, N. (2005). "A Multibody Dynamic Model for Escalator Handrail Systems and its Application to Dynamic Characteristics". *Multibody System Dynamics*, 13 (2), 253-266.
- Sug Kwon, Y., Won Park, T. (1999). "Computational Modelling and Vibration Reduction of an Escalator System Using Robust Design". *Computer Modelling & Simulation in Engineering*; Nov99, Vol. 4 Issue 4, 255-5p
- Cabanellas Becerra, J. M., Cano Moreno, J. D., Suárez, B., Chover, J. A., & Félez, J. (2008). "Methods for Improving Escalators". *Elevator Technology 17, Proceedings of ELEVCON 2008*, 22-33.
- Wu, Y., & Wu, A. (1997). "Diseño Robusto Utilizando Los Métodos Taguchi". Díaz de Santos.

## BIBLIOGRAPHY

**JUAN DAVID CANO-MORENO** works as a Research Engineer for CITEF. He is specialized in escalator simulation and design. He received his Master's Degree in Mechanical Engineering from the Universidad Politécnica de Madrid. At present he is preparing his Ph. D Thesis.

**JOSÉ MARÍA CABANELLAS BECERRA** is Associate Professor in the Mechanical and Manufacturing Engineering Department of the Universidad Politécnica de Madrid. His Ph.D. Thesis was on System Modelling and Simulation like the major projects that he manages in CITEF.

**CARLOS LABAJO TIRADO** works as a Research Engineer in CITEF. He is specialized in escalator instrumentation and signals processing. He received his Master's Degree in Mechanical Engineering from the Universidad Politécnica de Madrid.

**JESÚS FÉLEZ MINDÁN** is Full Professor at the Universidad Politécnica de Madrid. He is currently the director of CITEF. He has wide experience in System Modelling and Simulation and has published a large number of technical papers on the subject.

# SPATIAL KINEMATICS OF GEARS IN ABSOLUTE COORDINATES

Dmitry Vlasenko and Roland Kasper  
Institute of Mobile Systems (IMS)  
Otto-von-Guericke-University Magdeburg  
D-39016, Magdeburg,  
Germany  
E-mail: Dmitri.Vlasenko@ovgu.de

## KEYWORDS

Dynamics, Multibody, Gears, Jacobian, Absolute Coordinates, Concave-convex, Rack, Pinion

## ABSTRACT

In this article the description of some types of spatial gear constraints (spur gears, bevel gears, etc.) in absolute coordinates is considered. Instead of the numerical expensive calculation of constraint Jacobian matrix we propose to use the transformed Jacobian matrix, which can be calculated much more efficiently. The proposed methods of description of gear constraints were implemented for the simulation of dynamics of CAD model of KUKA KR 15/2 industrial manipulator.

## INTRODUCTION

Gears and gearing systems are fundamental mechanical components, widely used in the design of machines and mechanical systems for the transmission of motion and forces. On the other hand the detailed description of gear constraints is usually out of interest of multibody literature. In a few books can be found the description of planar kinematics of gears (Haug 1989, Shabana 2001) or the description of the spatial gear kinematics in joint coordinates (Schweiger and Otter, 2003).

In this article the description of some types of spatial gear constraints (spur gears, bevel gears, etc.) in absolute coordinates is considered. The use of absolute coordinates helps to integrate dynamic simulation tools with CAD systems, widely used for design of mechanical systems.

In order to better understand the issues involved, it is useful to consider the equations of motion in absolute coordinates. Let  $\mathbf{q}^i = (\mathbf{x}^{iT}, \mathbf{e}^{iT})^T$  be the vector of absolute coordinates of the  $i$ -th body consisting of position coordinates  $\mathbf{x}^i = (x_1^i, x_2^i, x_3^i)^T$  and of orientation coordinates  $\mathbf{e}^i$ . Orientation coordinates can be defined in different ways (e.g. Euler angles, Bryan angles, Rodriguez parameters, Euler parameters, etc.). The vector of generalized velocities  $\mathbf{v}^i = (\dot{\mathbf{x}}^{iT}, \dot{\mathbf{e}}^{iT})^T$  includes linear velocity  $\dot{\mathbf{x}}^i$  and angular

velocity  $\dot{\mathbf{e}}^i$ . The first derivative of  $\mathbf{q}^i$  is proportional to the vector  $\mathbf{v}^i$ :  $\dot{\mathbf{q}}^i = \mathbf{T}^i(\mathbf{q}^i)\mathbf{v}^i$ , where  $\mathbf{T}^i$  denote the relation matrix.

Let  $\mathbf{q} = (\mathbf{q}^{1T} \dots \mathbf{q}^{nT})^T$  be the vector of absolute coordinates of a multibody system. By  $\mathbf{g}(\mathbf{q})$  denote the vector of constraints, describing joints, connecting bodies in the simulated mechanical system. By  $\mathbf{G}(\mathbf{q})$  denote the Jacobian matrix of  $\mathbf{g}(\mathbf{q})$ :

$$\mathbf{G}(\mathbf{q}) = \frac{\partial \mathbf{g}(\mathbf{q})}{\partial \mathbf{q}} \quad (1)$$

Differentiating  $\mathbf{g}(\mathbf{q})$ , we get the equations of joint constraints on the velocity level:

$$\mathbf{g}_v(\mathbf{q}, \mathbf{v}) = \mathbf{G}(\mathbf{q}) \cdot \dot{\mathbf{q}} = \mathbf{G}(\mathbf{q}) \cdot \mathbf{T}(\mathbf{q}) \cdot \mathbf{v} \quad (2)$$

Let  $\widehat{\mathbf{G}}(\mathbf{q}) = \mathbf{G}(\mathbf{q}) \cdot \mathbf{T}(\mathbf{q})$  be the transformed Jacobian matrix. Then (2) can be written as

$$\mathbf{g}_v(\mathbf{q}, \mathbf{v}) = \widehat{\mathbf{G}}(\mathbf{q}) \cdot \mathbf{v} = 0 \quad (3)$$

We proposed (Vlasenko and Kasper 2009) the method of the simulation of multibodies, based on the Newton-Euler equations of motion

$$\mathbf{M}(\mathbf{q})\dot{\mathbf{v}} + \widehat{\mathbf{G}}^T(\mathbf{q})\boldsymbol{\lambda} = \mathbf{f}(\mathbf{q}, \mathbf{v}) \quad (4)$$

where  $\mathbf{f}(\mathbf{q}, \mathbf{v})$  is the vector of external forces,  $\mathbf{M}$  is the mass matrix,  $\boldsymbol{\lambda}$  is the vector of Lagrange multipliers. The main advantage of this method is that it uses the matrix  $\widehat{\mathbf{G}}(\mathbf{q})$ , which can be calculated much easily than the original Jacobian matrix  $\mathbf{G}(\mathbf{q})$ . Furthermore, if the non-minimal set of coordinates is used (e.g. Euler parameters), then the size of  $\widehat{\mathbf{G}}$  is less than the size of  $\mathbf{G}$ , that is important for the reduction of the simulation numerical costs.

In this article is shown the generation of constraints equations  $\mathbf{g}$  and of transformed Jacobian matrix  $\widehat{\mathbf{G}}$  for some types of gear joints, commonly used in mechanical systems (spur gears, bevel gears, etc.). In the description of gears we assume that constraints, generated by gear joints, limit only the relative rotation of gears. All other limitations on the relative motion of connected gears (e.g. constant distance between axes in the spur gear joint, etc.) are achieved as the result of connection of gears by other joints (usually by revolute joint) to some basement. This art of definition of gear constraints looks natural and similar to the definition of



gear constraints in CAD-like systems (Autodesk Inventor, etc.).

The proposed methods of description of gear constraints were implemented for the simulation of dynamics of CAD model of KUKA KR 15/2 industrial manipulator.

## SPATIAL KINEMATICS OF GEARS

### Spur gear joint

#### Equation of constraint on the coordinate level

Let us consider the gear  $i$  and the gear  $j$ , shown in Figure 1, which roll relative to each other about parallel axes  $\bar{\mathbf{a}}^i$  and  $\bar{\mathbf{a}}^j$ . Let  $C^i$  and  $C^j$  be the points on axes  $\bar{\mathbf{a}}^i$  and  $\bar{\mathbf{a}}^j$ , lying on the line, perpendicular to  $\bar{\mathbf{a}}^i$ . We assume that the motion of gears is constrained in such way that only the relative rotation of them is allowed, i.e.  $\bar{\mathbf{a}}^i$  remain parallel to  $\bar{\mathbf{a}}^j$ , the vector  $C^i C^j$  remains perpendicular to  $\bar{\mathbf{a}}^i$  and the distance between  $C^i$  and  $C^j$  remain constant (equal to the sum of the gears' radii).

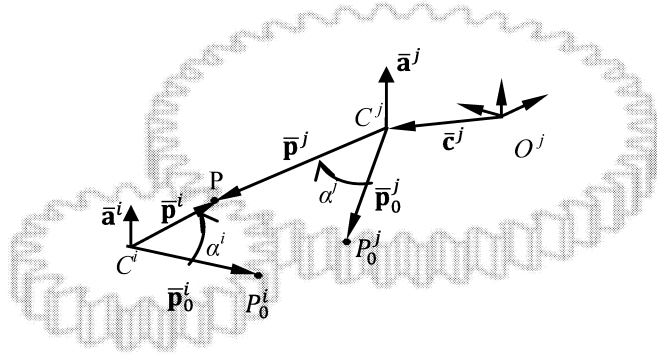


Figure 1: Spur Gear Joint

By  $\bar{\mathbf{c}}^i$  and  $\bar{\mathbf{c}}^j$  denote the local position vectors of  $C^i$  and  $C^j$ , respectively. Let  $P^i$  and  $P^j$  be the points of contact on bodies  $i$  and  $j$ , lying on the line  $C^i C^j$ . By  $P_0^j$  and  $P_0^i$  denote the points of contact on gears at initial stage.

Gears roll relative to each other without slip, therefore, the arc length  $P_0^i P^i$  and  $P_0^j P^j$  of contact on the gears must be equal. Then we get the equation of constraint

$$g = \alpha^j r^j + \alpha^i r^i \quad (5)$$

where  $\alpha^i$  is the angle between  $P_0^i$  and  $P^i$ ,  $\alpha^j$  is the angle between  $P_0^j$  and  $P^j$ ,  $r^i$  and  $r^j$  are the radii of gears.

Now we need to find the formula for the calculation of  $\alpha^j$  and  $\alpha^i$ . Let  $\bar{\mathbf{p}}_0^i, \bar{\mathbf{p}}^i$  be the vectors  $C^i P_0^i$  and  $C^i P^i$  expressed in body  $i$ . Let  $\bar{\mathbf{p}}_0^j, \bar{\mathbf{p}}^j$  be the vectors  $C^j P_0^j$  and  $C^j P^j$  expressed in body  $j$ . From the coincidence of points  $P^i$  and  $P^j$  follows that

$$\mathbf{x}^i + \mathbf{R}^i \bar{\mathbf{c}}^i + \mathbf{R}^i \bar{\mathbf{p}}^i = \mathbf{x}^j + \mathbf{R}^j \bar{\mathbf{c}}^j + \mathbf{R}^j \bar{\mathbf{p}}^j \quad (6)$$

where  $\mathbf{x}^i$  and  $\mathbf{x}^j$  are the vector of position coordinates of bodies  $i$  and  $j$ , respectively;  $\mathbf{R}^i$  and  $\mathbf{R}^j$  are the transformation matrices of the two bodies. This can be also written in another form

$$\mathbf{x}^i + \mathbf{c}^i + \mathbf{p}^i = \mathbf{x}^j + \mathbf{c}^j + \mathbf{p}^j \quad (7)$$

where  $\mathbf{c}^i = \mathbf{R}^i \bar{\mathbf{c}}^i$ ,  $\mathbf{c}^j = \mathbf{R}^j \bar{\mathbf{c}}^j$ ,  $\mathbf{p}^i = \mathbf{R}^i \bar{\mathbf{p}}^i$ ,  $\mathbf{p}^j = \mathbf{R}^j \bar{\mathbf{p}}^j$  are the vectors  $\bar{\mathbf{c}}^i, \bar{\mathbf{c}}^j, \bar{\mathbf{p}}^i, \bar{\mathbf{p}}^j$ , expressed in the global frame.

Let  $\mathbf{l}$  denote the vector of the constant length from  $C^i$  to  $C^j$ , calculated as

$$\mathbf{l} = \mathbf{x}^j + \mathbf{c}^j - \mathbf{x}^i - \mathbf{c}^i \quad (8)$$

Then (7) can be rewritten as

$$\mathbf{p}^i - \mathbf{p}^j = \mathbf{l} \quad (9)$$

Let  $\mathbf{e}^i, \mathbf{e}^j$  be the units vectors along  $\mathbf{p}^i$  and  $\mathbf{p}^j$ , correspondently

$$\mathbf{p}^l = r^l \mathbf{e}^l \quad l = i, j \quad (10)$$

From the definition of vectors  $\mathbf{p}^i$  and  $\mathbf{p}^j$  follows that  $\mathbf{e}^i$  and  $\mathbf{e}^j$  can be calculated as

$$\mathbf{e}^i = -\mathbf{e}^j = \frac{\mathbf{l}}{\|\mathbf{l}\|} \quad (11)$$

In practice the joint is usually defined by the gear ratio

$$k = r^j / r^i \quad (12)$$

Then  $r^i, r^j$  can be calculated from  $k$  using the formula

$$r^i = \|\mathbf{p}_0^i\| = \frac{1}{1+k} \cdot \|\mathbf{l}\| \quad (13)$$

$$r^j = \|\mathbf{p}_0^j\| = \frac{k}{1+k} \cdot \|\mathbf{l}\| \quad (14)$$

If the modulus of  $\alpha^j$  and of  $\alpha^i$  are less than  $\pi/2$ , then  $\alpha^j$  and  $\alpha^i$  can be calculated as

$$\alpha^l = \text{asin}[\mathbf{a}^{lT} \cdot (\mathbf{e}_0^l \times \mathbf{e}^l)] \quad l = i, j \quad (15)$$

where  $\mathbf{a}^l = \mathbf{R}^l \bar{\mathbf{a}}^l$  is the axis  $\bar{\mathbf{a}}^l$ , expressed in the global frame.

In practice we can guarantee that  $\alpha^j$  and  $\alpha^i$  are less than  $\pi/2$  if during the simulation we always move the points  $P_0^j$  and  $P_0^i$  to the tops of last contacted teeth. The number of teeth of gears can be defined manually during the gear design or automatically by the simulation pre-compiler.

#### The equation of constraint on the velocity level

Let us show how we can find the equation of constraint on the velocity level  $\mathbf{g}_v$  and to derive from  $\mathbf{g}_v$  the formula for  $\dot{\mathbf{G}}$ .

The equation of constraint on the velocity level are calculated as the time derivative of (5)

$$\mathbf{g}_v = \dot{\alpha}^j r^j + \dot{\alpha}^i r^i = 0 \quad (16)$$

Let us define the relative orthogonal system of coordinates  $C^i \mathbf{e}^{bx} \mathbf{e}^{by} \mathbf{e}^{bz}$  where the origin of the relative system is rigidly connected to  $C^i$ ,  $\mathbf{x}^b$  is the axes lying on the line  $C^i C^j$ ,  $\mathbf{e}^{by}$  lies on the axes  $\bar{\mathbf{a}}^i$  (i.e.  $\mathbf{e}^{by} = \bar{\mathbf{a}}^i$ ) and the vector  $\mathbf{e}^{bz}$  is chosen in such way that the  $C^i \mathbf{e}^{bx} \mathbf{e}^{by} \mathbf{e}^{bz}$  will be right-handed. Let  $\mathbf{u}_b^{Pi}, \mathbf{u}_b^{Pj}$  be the velocities of points  $P^i$  and  $P^j$  relative  $C^i \mathbf{e}^{bx} \mathbf{e}^{by} \mathbf{e}^{bz}$ , expressed in the global coordinate system, respectively. From the definition of  $C^i \mathbf{e}^{bx} \mathbf{e}^{by} \mathbf{e}^{bz}$  follows that

$$\boldsymbol{\omega}_b^l = \dot{\alpha}^l \mathbf{a}^l \quad l = i, j \quad (17)$$

where  $\omega_b^l$  be the relative angular velocity of  $l$ -th body, expressed in the global coordinate system, calculated as

$$\omega_b^l = \omega^l - \omega^b \quad l=i,j \quad (18)$$

where  $\omega^b$  is the angular velocity of the frame  $C^i \mathbf{e}^{bx} \mathbf{e}^{by} \mathbf{e}^{bz}$  and  $\omega^l$  is the angular velocity of the  $l$ -th body.

Cross-multiplying both parts of (17) by the vector  $\mathbf{e}^l$ , we get

$$\omega_b^l \times \mathbf{e}^l = \dot{\alpha}^l \mathbf{a}^l \times \mathbf{e}^l \quad l=i,j \quad (19)$$

Let  $\mathbf{s}$  be a vector perpendicular to the axis of rotation  $\mathbf{a}^i$  and to  $\mathbf{p}^i$ :  $\mathbf{s} = \mathbf{a}^i \times \mathbf{e}^i$ . Then from (19) follows that

$$\begin{aligned} \dot{\alpha}^i &= \mathbf{s}^T (\omega_b^i \times \mathbf{e}^i) \\ \dot{\alpha}^j &= -\mathbf{s}^T (\omega_b^j \times \mathbf{e}^j) \end{aligned} \quad (20)$$

Substituting  $\dot{\alpha}^i, \dot{\alpha}^j$  in (16), we get

$$\mathbf{g}_v = -\mathbf{s}^T (\omega_b^j \times \mathbf{p}^j) + \mathbf{s}^T (\omega_b^i \times \mathbf{p}^i) \quad (21)$$

Using the formula (18) for  $\omega_b^j, \omega_b^i$ , we obtain

$$\mathbf{g}_v = \mathbf{s}^T (\omega^i \times \mathbf{p}^i - \omega^j \times \mathbf{p}^j + \omega^b \times (\mathbf{p}^j - \mathbf{p}^i)) \quad (22)$$

Let  $\mathbf{u}^{Ci}, \mathbf{u}^{Cj}$  be the absolute velocities of points  $C^i$  and  $C^j$ , respectively, calculated as

$$\mathbf{u}^{Cl} = \dot{\mathbf{x}}^l + \omega^l \times \mathbf{c}^l \quad l=i,j \quad (23)$$

From the definition of  $C^i \mathbf{e}^{bx} \mathbf{e}^{by} \mathbf{e}^{bz}$  follows the relation

$$\omega^b \times (\mathbf{p}^i - \mathbf{p}^j) = \mathbf{u}^{Cj} - \mathbf{u}^{Ci} \quad (24)$$

Substituting this equation in (22) and using (23), we get:

$$\begin{aligned} \mathbf{g}_v &= \mathbf{s}^T ([\dot{\mathbf{x}}^i + \omega^i \times (\mathbf{c}^i + \mathbf{p}^i)] \\ &\quad - [\dot{\mathbf{x}}^j + \omega^j \times (\mathbf{c}^j + \mathbf{p}^j)]) \end{aligned} \quad (25)$$

The physical meaning of this equation is the equality of projections of velocities of  $P^i$  and  $P^j$  on the axis  $\mathbf{s}$ . Using the triple product formula

$$\mathbf{a}^T (\mathbf{b} \times \mathbf{c}) = (\mathbf{c} \times \mathbf{a})^T \mathbf{b} = (\mathbf{a} \times \mathbf{b})^T \mathbf{c} \quad (26)$$

we get from (25)

$$\begin{aligned} \mathbf{g}_v &= \mathbf{s}^T \dot{\mathbf{x}}^i - [\mathbf{s} \times (\mathbf{c}^i + \mathbf{p}^i)]^T \omega^i - \mathbf{s}^T \dot{\mathbf{x}}^j \\ &\quad + [\mathbf{s} \times (\mathbf{c}^j + \mathbf{p}^j)]^T \omega^j = 0 \end{aligned} \quad (27)$$

Now, using (3), we get the matrix  $\widehat{\mathbf{G}}(\mathbf{q}^i, \mathbf{q}^j)$

$$\widehat{\mathbf{G}}^T = \begin{pmatrix} \mathbf{s} \\ -\mathbf{s} \times (\mathbf{c}^i + \mathbf{p}^i) \\ -\mathbf{s} \\ \mathbf{s} \times (\mathbf{c}^j + \mathbf{p}^j) \end{pmatrix} \quad (28)$$

Clear, that the calculation of  $\widehat{\mathbf{G}}$  from (28) is much easier than the calculation of Jacobian  $\mathbf{G}(\mathbf{q}^i, \mathbf{q}^j)$  from (1).

### Bevel gear joint

Let us consider the gear  $i$  and the gear  $j$ , shown in Figure 2, which roll relative to each other with intersecting axes  $\mathbf{a}^i$  and  $\mathbf{a}^j$  in such way that the angles of rotation  $\alpha^i, \alpha^j$  are related as

$$\alpha^i = -\alpha^j k \quad (29)$$

where  $k$  is the gear ratio defined by the quotient between the number of gear teeth.

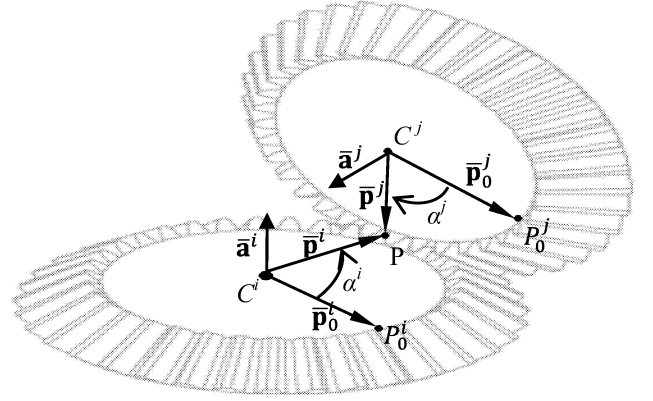


Figure 2: Bevel Gear Joint

In practice the joint is defined by the definition of bodies-fixed axes  $\mathbf{a}^i$  and  $\mathbf{a}^j$  and by the local position vectors  $\mathbf{c}^i$  and  $\mathbf{c}^j$  of some points  $C^i$  and  $C^j$  on the axes  $\mathbf{a}^i$  and  $\mathbf{a}^j$ , respectively.

By  $\mathbf{l}$  denote the vector of the constant length from  $C^i$  to  $C^j$ , calculated from (8). Let us define for each body  $l$  some vector  $\mathbf{p}^l$  as perpendicular to  $\mathbf{a}^l$ , which begins at  $C^l$  and ends at the point  $P^l$ , lying on the line of contact. Let  $r^i$  and  $r^j$  be the modulus of  $\mathbf{p}^i$  and  $\mathbf{p}^j$ , respectively (similarly to the radii of  $r^i$  and  $r^j$  of spur gears).

Let us assume that  $C^i$  and  $C^j$  are chosen in such way that points  $P^i$  and  $P^j$  are coincident, i.e.

$$\mathbf{p}^i - \mathbf{p}^j = \mathbf{l} \quad (30)$$

Then from the definition of  $\mathbf{p}^i$  and  $\mathbf{p}^j$  follows that

$$k = \frac{\|\mathbf{p}^j\|}{\|\mathbf{p}^i\|} = \frac{r^j}{r^i} \quad (31)$$

It can be shown that

$$\begin{aligned} r^i &= \frac{|\mathbf{a}^j \mathbf{l}|}{\|\mathbf{a}^i \times \mathbf{a}^j\|} \\ r^j &= k r^i \end{aligned}$$

Now we can reformulate the equation of bevel gear constraint (29) in the form similar to the equation of the spur gear constraint

$$\mathbf{g} = \alpha^j r^j + \alpha^i r^i \quad (32)$$

Let us show now how to calculate the angles  $\alpha^j$  and  $\alpha^i$ . By  $P_0^j$  and  $P_0^i$  denote the points  $P^j$  and  $P^i$  at initial stage and by  $\mathbf{p}_0^i, \mathbf{p}_0^j$  the start values of vectors  $\mathbf{p}^i, \mathbf{p}^j$ , respectively. Obviously, the angle of relative rotation  $\alpha^i$  is the angle between  $\mathbf{p}_0^i$  and  $\mathbf{p}^i$  and  $\alpha^j$  is the angle between  $\mathbf{p}_0^j$  and  $\mathbf{p}^j$ .

From the numerical point of view it's easily to calculate  $\alpha^l$  as the angle between unit vectors  $\mathbf{e}^l = \mathbf{p}^l / r^l$ ,  $\mathbf{e}_0^l = \mathbf{p}_0^l / r^l$  ( $l=i,j$ ).

From the definition of vectors  $\mathbf{p}^i$ ,  $\mathbf{p}^j$  follows that  $\mathbf{e}^i = \mathbf{s} \times \mathbf{a}^i$  and  $\mathbf{e}^j = -\mathbf{s} \times \mathbf{a}^j$ , where  $\mathbf{s}$  is a unit vector, perpendicular to  $\mathbf{a}^i, \mathbf{a}^j$ , calculated as

$$\mathbf{s} = \frac{\mathbf{a}^i \times \mathbf{a}^j}{\|\mathbf{a}^i \times \mathbf{a}^j\|} \quad (33)$$

If the modulus of  $\alpha^j$  and of  $\alpha^i$  are less than  $\pi/2$ , then  $\alpha^j$  and  $\alpha^i$  can be calculated from (15).

Using the same procedure for the generation of equation of constraint on the velocity level, as it was used above for the spur gear case, we obtain that the transformed Jacobian matrix  $\hat{\mathbf{G}}$  can be calculated from (28).

### Concave-convex gear joint

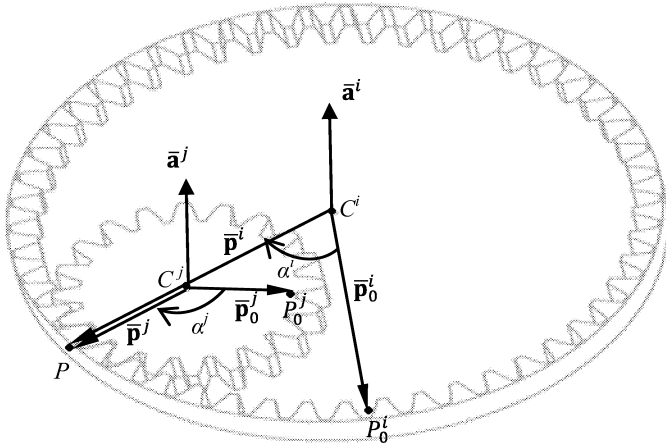


Figure 3: Concave-Convex Gear Joint

Let us consider a concave-convex gear joint, describing the rolling contact of smaller gear  $j$  makes inside the larger interior gear  $i$ , shown Figure 3. Using the definitions from the spur gear case, we obtain the following relation between angles of rotations

$$\alpha^j r^j - \alpha^i r^i = 0 \quad (34)$$

Clear, that unlike convex gear case now the vectors  $\mathbf{e}^i$  and  $\mathbf{e}^j$  are equal and are calculated as

$$\mathbf{e}^i = \mathbf{e}^j = \frac{\mathbf{l}}{\|\mathbf{l}\|} \quad (35)$$

The formula for the calculation of radii  $r^i$  and  $r^j$  from the gear ratio  $k = r^j/r^i$  also changes as

$$r^i = \|\mathbf{p}_0^i\| = \frac{1}{1-k} \cdot \|\mathbf{l}\|$$

$$r^j = \|\mathbf{p}_0^j\| = \frac{k}{1-k} \cdot \|\mathbf{l}\|$$

The angles  $\alpha^j$  and  $\alpha^i$  in (34) can be calculated from (15)

Using the same procedure, as it was used above for the spur gear case, we obtain the matrix  $\hat{\mathbf{G}}$

$$\hat{\mathbf{G}}^T = \begin{pmatrix} -\mathbf{s} \\ \mathbf{s} \times (\mathbf{c}^i + \mathbf{p}^i) \\ \mathbf{s} \\ -\mathbf{s} \times (\mathbf{c}^j + \mathbf{p}^j) \end{pmatrix} \quad (36)$$

### Rack and pinion joint

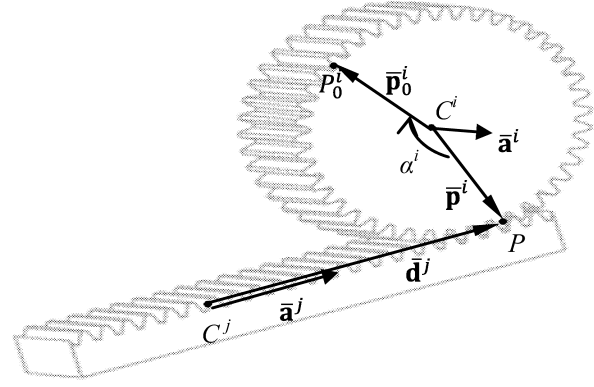


Figure 4: Rack and Pinion Joint

Let us consider a circular gear  $i$  (called rack) and a flat bar  $j$  (called pinion) constituting the constraint kinematic pair, shown in Figure 4. The rack rolls relative to the pinion about the axis  $\mathbf{a}^i$  whereby the pinion axis  $\mathbf{a}^j$  is situated on the line of contact. It's assumed that  $\mathbf{a}^i$  and  $\mathbf{a}^j$  are perpendicular in space, i.e.  $\mathbf{a}^{iT} \mathbf{a}^j = 0$ . Let  $C^j$  be the point of contact on pinion at initial stage and  $C^i$  be the points on axis  $\mathbf{a}^i$ , chosen in such way that  $C^i C^j$  is perpendicular to  $\mathbf{a}^i$ . As before, we denote by  $P^i$  and  $P^j$  the points of contact on bodies  $i$  and  $j$ , and by  $P_0^i$  the point of contact on body  $i$  at initial stage.

The equation of our constraint can be written as

$$g = d^j - \alpha^i r^i \quad (37)$$

where  $\alpha^i$  is the angle between  $P_0^i$  and  $P^i$ ,  $d^j$  is the distance between  $P_0^j$  and  $P^j$ . Let us show now how to calculate  $d^j$  and  $\alpha^i$ .

Let  $\mathbf{d}^j$  be the vector from  $P_0^j$  to  $P^j$ . It is obvious that

$$\mathbf{d}^j = d^j \mathbf{a}^j \quad (38)$$

Let  $\mathbf{p}_0^i, \mathbf{p}^i, \mathbf{c}^j$ , be the local positions of  $P_0^i, P^i, C^j$ , respectively. From the coincidence of points  $P^i$  and  $P^j$  follows that

$$\mathbf{x}^i + \mathbf{c}^i + \mathbf{p}^i = \mathbf{x}^j + \mathbf{c}^j + \mathbf{d}^j \quad (39)$$

From the definition follows that  $\mathbf{p}^i$  is perpendicular to  $\mathbf{a}^j$ . Therefore, multiplying (39) by  $\mathbf{a}^j$ , we get the formula for the calculation of  $d^j$

$$d^j = \mathbf{a}^{jT} (\mathbf{x}^i + \mathbf{c}^i - \mathbf{x}^j - \mathbf{c}^j) \quad (40)$$

Let  $\mathbf{e}^i = \mathbf{p}^i / r^i$  be the unit vector along  $\mathbf{p}^i$ . Let  $\bar{\mathbf{e}}^j$  be the vector of coordinates of  $\mathbf{e}^i$  in the pinion  $j$ . Obviously,  $\bar{\mathbf{e}}^j$  is constant. The angle  $\alpha^i$  can be calculated from (15), whereby the vector  $\mathbf{e}^i$  can be calculated as  $\mathbf{e}^i = \mathbf{R}^j \bar{\mathbf{e}}^j$ .

In the same way as in the spur gear case, we obtain  $\hat{\mathbf{G}}$

$$\hat{\mathbf{G}}^T = \begin{pmatrix} \mathbf{a}^j \\ \mathbf{a}^j \times (\mathbf{p}^i - \mathbf{c}^i) \\ -\mathbf{a}^j \\ \mathbf{a}^j \times (\mathbf{x}^i + \mathbf{c}^i - \mathbf{x}^j - \mathbf{p}^i) \end{pmatrix} \quad (41)$$

where  $\mathbf{p}^i$  can be calculated from the equation  $\mathbf{p}^i = \mathbf{r}^i \mathbf{e}^i$ .

### SIMULATION EXAMPLE: INDUSTRIAL MANIPULATOR KUKA KR 15/2

In the last years we developed a component-oriented simulation software **Virtual Systems Designer (VSD)**, integrated with CAD-like tool Autodesk Inventor (Kasper et. al 2007, Vlasenko and Kasper 2007). The proposed methods of description of gear joint constraints were implemented in VSD. As a test example we used an Autodesk Inventor model of the industrial manipulator KUKA KR 15/2, shown in Figure 5. This is a six-axis robot with articulated kinematics for all continuous-path controlled tasks. The main areas of application of KR 15/2 are handling, assembly, machining, etc. (Specifications of Robots)

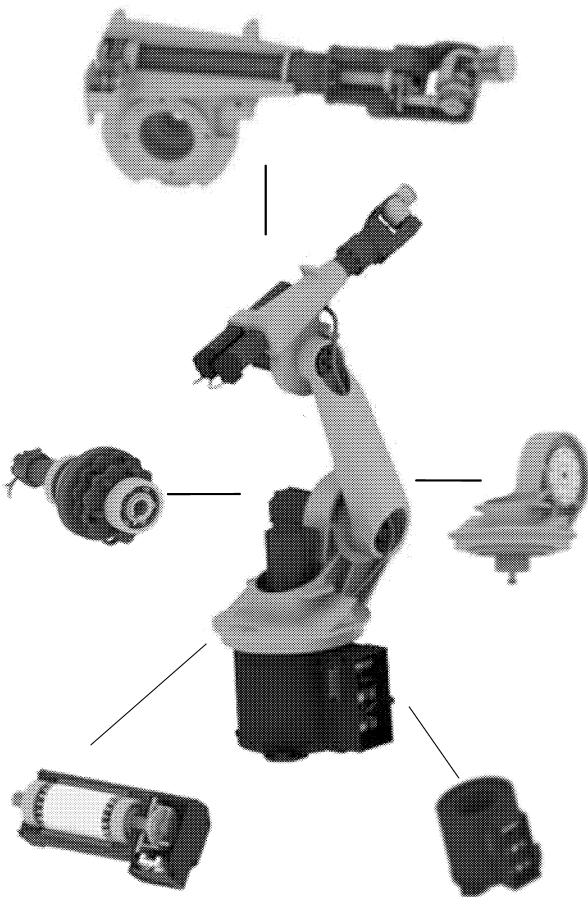


Figure 5: CAD Model of KUKA KR 15/2

The complete Autodesk Inventor model includes 1036 parts. The correspondent VSD model consists of 43 bodies connected by 95 joints (including 7 spur gear joints, 3 bevel gear joints and 10 concave-convex gear joints). Some of model constraints are redundant because of the model's design in Autodesk Inventor (e.g. the definition of stiff connection as three plane-to-plane joints leads to the generation of three redundant constraints).

The dynamics of the manipulator under the action of gravitational force and of torques in motors is simulated. The

numerical error of gear constraints on the coordinate and on the velocity levels are equal to the accuracy of used numerical methods. The analysis of monitored values of bodies' velocities and accelerations show the correctness of proposed methods for the generation of gear constraints.

### CONCLUSION AND FUTURE WORK

In this article is considered the spatial kinematics of most commonly used types of gear joints (spur gears, bevel gears, concave-convex gears and rack and pinion joints) in absolute coordinates. We show the methods of generation of equations of correspondent constraints on the coordinate and on the velocity level.

In standard methods of simulation of multibodies the calculation of constraint Jacobian matrix  $\mathbf{G}$  is needed, which in the case of gears is a numerically expensive procedure. That is why we propose to use the simulation algorithm based on the calculation of transformed Jacobian matrix  $\hat{\mathbf{G}}$ . In this article is shown that the generation of matrix  $\hat{\mathbf{G}}$  is easy in use and require a very small amount of computational effort. Moreover, the additional advantage  $\hat{\mathbf{G}}$  is its reduced size in comparison with  $\mathbf{G}$  when the non-minimal set of orientation coordinates is used.

The proposed methods of description of gear constraints were implemented for the simulation of dynamics of KUKA KR 15/2 industrial manipulator. Test results show the correctness of proposed algorithms.

In future we plan to improve the area of implementation of proposed method by the description of other motion transmission elements (e.g. belt joints, cam-followers, etc.).

### REFERENCES

- Haug, E. 1989. *Computer Aided Kinematics and Dynamics of Mechanical Systems Volume I: Basic Methods*. Allyn & Bacon, Boston.
- Kasper, R.; D. Vlasenko and G. Sintotskiy. 2007. "A Component Oriented Approach to Multidisciplinary Simulation of Mechatronic Systems". *Proceedings of the EUROSIM Congress on Modelling and Simulation (EUROSIM 2007)*, September 9-13 Ljubljana, Slovenia.
- Specification of Robots KR 6/2, KR 15/2, KR 15 L6/2, by KUKA Robot Group.
- Schweiger, C. and Otter, M. 2003. "Modeling 3D Mechanical Effects of 1D Powertrains", in *Proceedings of the 3rd International Modelica Conference*, P. Fritzson, ed., Linköping, November 2003, The Modelica Association and Linköping University, pp. 149-158
- Shabana, A.A. 2001. *Computational Dynamics*, Wiley, New York, 2001.
- Vlasenko, D. and R. Kasper. 2007. "Integration Method of CAD Systems" *Proceedings of the ASME 2007 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2007* September 4-7, 2007, Las Vegas, Nevada, USA
- Vlasenko, D. and R. Kasper. 2009. "Successive projection method for the simulation of spatial dynamics of multibodies". *Proceedings of Multibody Dynamics 2009 (ECCOMAS Thematic Conference)*, Warsaw, Poland, June 29-2 July, 2009 (accepted, to appear)

# MODELING AND SIMULATION OF GAIT FOR *MERO* MODULAR WALKING ROBOT CROSSING AN UNARRANGED TERRAIN

Ion Ion<sup>1</sup>, Alexandru Marin<sup>2</sup>, Cristian Doicin<sup>1</sup>, Constantin Chirita<sup>3</sup>  
University "Politehnica" from Bucharest, Romania  
Technology for Manufacturing Department<sup>1</sup>, Hydraulics Department<sup>2</sup>;  
Technical University "Gh. Asachi" Iasi, Romania, Fluid Power Department<sup>3</sup>  
[ioni51@yahoo.com](mailto:ioni51@yahoo.com), [marin@hydrop.pub.ro](mailto:marin@hydrop.pub.ro), [cristian.doicin@cont-edu.pub.ro](mailto:cristian.doicin@cont-edu.pub.ro),  
[constantin.chirita@disahp.net](mailto:constantin.chirita@disahp.net)

## Abstract

Modern methods of drawing up machines and tools necessarily include a simulation stage of their functioning, which means to use a mathematical model of the real, original system.

The main characteristic of the modular walking robots is that they are able to move away on not arranged, horizontal and rough terrains. The performance of walking robots is closely related to the adopted gait.

These activate the functioning simulation that encompasses several rules and specifications whose enactment generates behavior data and the instructions operating on the pattern's description variables.

The veracity and validity of a real system depend on the compliance degree.

The real system and the model are different by the fact that whereas for the former the manner to generate conduct data can be completely unknown, for the latter they mean a group of rules or specifications, whose enactment puts out the conduct data, namely instructions operating of the model's description variables.

In the paper, a robust model of the direct and inverse analysis of the open kinematic chains of the walking robot legs is constructed.

Finally, this strategy is verified by using computer graphics simulations.

**Keywords:** Modular Walking Robot, Walk, Gait, Support phase, Transfer phase, Stability margin.

## INTRODUCTION

Since long ago, many research teams worldwide have been focused on goals such as creating an autonomous walking robot equipped with functions like handling objects, locomotion, perceiving, navigation, learning, judgment, information storage and intelligent control, and that can carry out tasks like altering the multitude of the parts belonging to a dynamic universe.

Scientists of all times have been permanently mesmerized and have studied the simplest but the most important movement, namely the mechanical movement of humans and animals. Mankind is so much anthropomorphism addicted that it is almost

impossible for it to conceive or imagine automatic systems, even having artificial intelligence, and that are not anthropomorphic.

How do animals walk, how do birds fly, have always been and it will probably remain the inspiration sources and tasks of the scientific research.

The animals using their feet to walk can move freely, with substantive easiness in comparison with what the man - made robot can do.

Studying the simplest but the most important movement forms, such as the man's and the animals' movement, has been the scientists' most ancient preoccupation since the beginning of time.

The walk is defined by the manner the waking robot moves between two points, under specific circumstances. To achieve and guide a walking robot requires thorough knowledge about all walking possibilities because choosing the number of legs and their structure depends very much on the selected walk.

The selection of the walk type depends on several elements such as: shape and consistency of the ground the robot walks on, walk stability driving and controlling the movements of the elements of the walking systems, speed and mobility movement requires.

It is very difficult to select the type of walk, mainly during real walking.

Therefore, it is necessary that the ground surface to be defined before selecting the walk.

The walking robot's steps are a sequel of movements of the legs, coordinated with a succession of movements of the body for the purpose of moving the robot from one place to another.

Theoretical research of robot movement assumes next important stages:

- the establishing of the mathematical model of the movement of the kinematic and dynamic system;
- the process and built-up the results obtained through simulation with a view to determining the system's conduct;
- the structural and geometrical optimization of the movement systems compounds;
- the movement of system elements;
- the establish the functional block scheme to calculate the compounds;

- the establishing of the test methodology to perform the system's functions that is need and the identity of the functions generators walking requires;
- the design of the guidance system's software;
- the determination of the initial parameters and data characteristics of the system's structure and state;
- the conduct of the simulation program.

### MATHEMATICAL MODELING OF GAIT FOR MODULAR WALKING ROBOTS. THE WALK AS SEQUENCE OF STATES

A cycle of the movement of the leg of a modular walking robot has two phases: *the support phase* and *the transfer phase*. In the first phase, the leg's support part has a direct contact with the walking surface area.

In the transfer phase, the leg of the robot is above the walking surface and is moving so that it realizes the stability state on the whole of the walking robot.

The walk of the robot is characterized by the order *raise* and *seat* of the legs and by the trajectory form of the theoretical support point in comparison with the platform.

To establish the walking order it is needed to number the legs. The state of the leg ( $i$ ) at a given time [1], [4], [6] is described by a state's function  $q^i(t)$ , that has only two values, 0 and 1, as it follows:

$$q^i(t) = \begin{cases} 0 & \text{for the support phase;} \\ 1 & \text{for the transfer phase.} \end{cases} \quad (1)$$

On the interval  $[0, t_1]$ , the leg is in the support phase. On the interval  $[t_1, t_2]$ , the leg is not leaning upon the support surface and it is in the transfer phase. On the interval  $[t_2, t_3]$ , the leg is on the support surface again etc. At a moment of time, the state of the walking robot with  $N$  legs is defined by a  $N$ -dimensional vector  $\mathbf{q}$ , named *the vector of the legs states*. The vector's components  $q^i$ ,  $i = 1, N$ , are formed by the functions of the legs' states, ordered by their numbering:

$$\mathbf{q} = [q^1, q^2, \dots, q^N]^T, \quad (2)$$

so that the first component of the vector define the state of the leg 1, the second one is the state of the leg 2 etc.

It is assumed that in any finite interval of time there is a finit number of moments that defines the values of the functions  $q^i(t)$ .

The  $\mathbf{q}$  states, that appear at every change of the value of the function  $q^i(t)$ , are numbered chronologically as they are carried on.

As a result, the walk of a walking robot is described by a succession of states  $(\mathbf{q}_j)$ ,  $j = 1, 2, \dots$ . An example of the succession of the states for a walking robot with 4 legs is:

$$\begin{aligned} \mathbf{q}_1 &= [0,0,0,0]^T, \mathbf{q}_2 = [1,0,0,0]^T, \mathbf{q}_3 = [0,0,0,0]^T, \\ \mathbf{q}_4 &= [0,1,0,0]^T, \mathbf{q}_5 = [0,0,0,0]^T, \mathbf{q}_6 = [0,0,1,0]^T, \\ \mathbf{q}_7 &= [0,0,0,0]^T, \mathbf{q}_8 = [0,0,0,1]^T, \mathbf{q}_9 = \mathbf{q}_1 \end{aligned} \quad (3)$$

It is assumed that, at the initial moment, all the walking robot's legs are in the support phase. After this, the leg number 1 is raised and moved down, followed by the raise of the leg number 2 and its move down etc. The walk is cyclically if the succession of the states  $(\mathbf{q}_i)$  is periodical.

The total amount of the realized states in a time period is named *the walk cycle*. In the above example, the walk is *cyclic* if after the state  $\mathbf{q}_9$ , determined above, follows the state  $\mathbf{q}_2$ , state  $\mathbf{q}_3$  etc, in that order.

### DETERMINATION OF STABILITY CONDITIONS WITH THE MOVEMENT OF THE MODULAR WALKING ROBOT ON ROUGH TERRAIN

An important feature of the walking robots is that the ground's configuration never influences or affects their movement too much. Such a feature makes this locomotion type turn into an attractive solution for many applications that requires, movement on an unplanned land, namely having an uneven configuration [1], [4], [6].

As uneven land covers an endless variety, it is difficult to match all the types and cases of walking on such grounds.

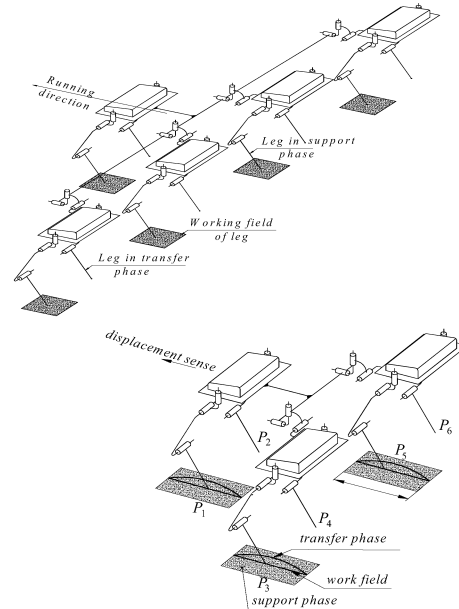


Fig.1 Models of walking modular robot

To study the issue we need a simplification of the geometrical features of the real terrain. For this, we introduce the notion *obstacle*, or *hurdle*.

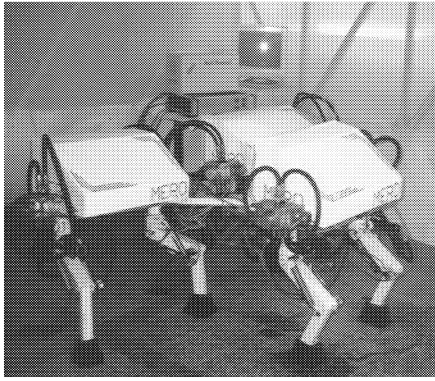


Fig.2. The walking modular robot MERO 2

The obstacle is *isolated* if it can be included in the area between a vertical plane and a perpendicular one on the walking direction, besides there are no forbidden area for the walking robot to step forward in the area of its movement.

The support area separated through the mentioned planes is called the *obstacle's area (zone)*, where it is forbidden to place the support points of the robot's legs and feet.

The width of the obstacle's area depends on the ratio of the obstacle's dimensions, on the geometrical parameters and the technical features of the walking robot.

The presence of the obstacle on the support area can lead to the change in the movement direction, in the height of the movement of the body of the walking robot or its orientation in space as well as to the appropriate redefining of the steps' order and sequence and in the regime of the legs' movement.

Figure 2 shows the model of a modular hexapod robot MERO2.

### STEPPING OVER ISOLATED OBSTACLE WHILE PRESERVING STATIC STABILITY

The presence of the obstacles on the support surface may lead to changing the direction, the height of the body's movement and its orientation in space as well as the appropriate reorganization of the sequel of watching the regime of the legs' movement.

This paragraph will follow, in compliance with [5], [7], [8], the case when to prevent the obstacles, it is enough only to change the sequel of following and maybe the height of the body's center of gravity (mass point) preserving the same the other movement features of the robot.

Like before we will assume that the legs' suspension points are symmetrically placed to the vertical plane including the body's center and the speed parallel to it.

The paths followed for the right and left legs of the robot we assume to be rectilinear and parallel to the speed vector. We are given the body's mass point to be projected on the center line between the follow paths.

We will name the obstacle 'isolated' if we can include it in the field between two vertical planes,

erect on the body's movement direction so that the follow paths beyond this field lack the points forbidden to advance.

On the support area the above mentioned planes separate the area called the *obstacle's zone* where it is forbidden to place any follow points. Its ends are erect on the follow paths.

The zone's width depends on the ratio of the obstacle's dimensions and the geometrical parameters characteristic to the robot as well as its control system.

It occurs the problem of establishing the sequence of follow, which enables the robot to step over the area without disturbing its static stability.

### MOVEMENT SIMULATION BY DENAVIT – HARTENBERG FORMALISM

Let us a modular walking robot consist of three modules [2], [5]. Each module has two 3-DOF legs, symmetrically arranged on the platform axis (fig.3) [3], [5].

The legs on the right – onto the movement direction are superscript marked with  $2i$ ,  $i=1,6$  whereas the legs on the left with  $2i-1$ . Each platform of the rear modules is connected to the platform of first module by a 3-DOF kinematic chain with two links and three rotational pairs. The axes of these pairs are concurrent and perpendicular two by two.

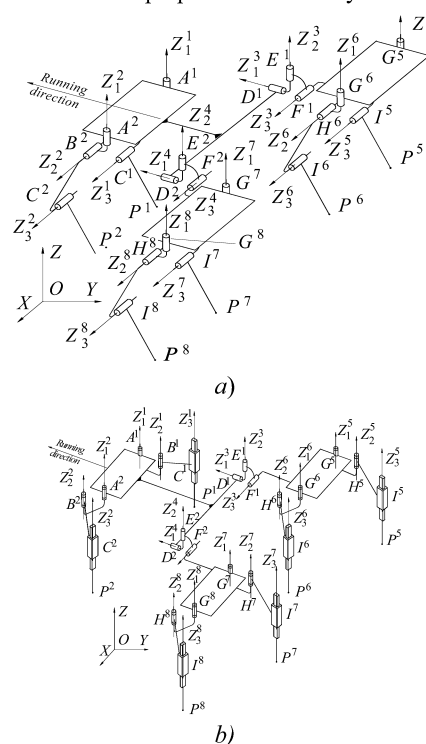


Fig.3. The Denavit –Hartenberg axis system attached of modular walking robot it is suggested support of technological equipments, for a leg mechanism RRR a) and RRP b)

In order to be carried out the movement simulation of a leg, a coordinate axes system is attached to each link, with the Denavit – Hartenberg rule [2].

This formalism may not only simplify the problem formulation, but can also yield considerable advantage in the solution of simulation problem. The pairs of each leg are numbered consecutively from  $A$  which is pair number 1 to  $C$  which is pair number 3.

The Denavit - Hartenberg systems attached to each link are subscript numbered as the pairs respectively. The platform is designed as link number (0) and the remaining links are numbered consecutively. All pairs of the leg mechanism are rotational and actuated ones.

#### DENAVIT HARTENBERG SYSTEMS ATTACHED TO THE LEG LINKS OF THE WALKING ROBOT

The characteristic axis  $Z_i$  of each pair should be defined. The positive sense of each of these axes is defined arbitrarily. If the axes  $Z_i$  and  $Z_{i-1}$  are skew with respect to each other, then there is one common perpendicular between them. The perpendicular is designed as the  $X_i$  axis. If the  $Z_i$  and  $Z_{i-1}$  axes are parallel, the  $X_i$  axis may chosen as any common perpendicular. The positive direction of the  $X_i$  axis is designed as proceeding from  $Z_{i-1}$  to  $Z_i$ . If the  $Z_{i-1}$  and  $Z_i$  intersect, the positive sense of  $X_i$  axis is arbitrarily. When the  $X_i$  axes are all defined, then are define both the  $Y_i$  axes and the origin of each right hand coordinate system.

So, a coordinate system defined is attached to each link. The parameter  $a_i$  is defined as the distance from  $O_i Z_i$  to  $O_{i+1} Z_{i+1}$  axes, measured along  $O_{i+1} X_{i+1}$ . Because of the orientation of the  $O_{i+1} X_{i+1}$  axis,  $a_i$  is always positive.

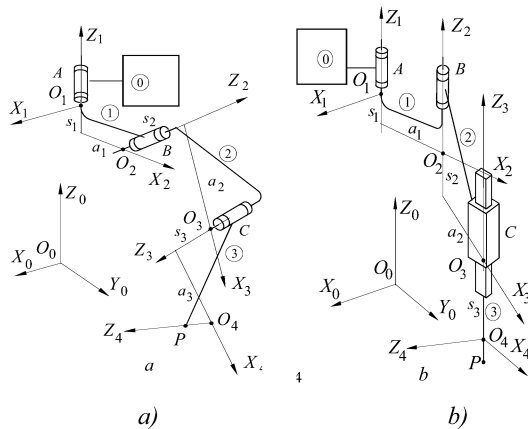


Fig. 4. The Denavit – Hartenberg coordinate axes systems for a leg mechanism RRR a), RRP b)

The parameter  $\alpha_i$  is defined as the angle between the positive  $O_i Z_i$  and the positive  $O_{i+1} Z_{i+1}$  axes, as

seen from positive  $O_{i+1} X_{i+1}$ . The parameter  $\theta_i$  is the angle between positive  $O_i X_i$  and the positive  $O_{i+1} X_{i+1}$  axes, as seen from positive  $O_i Z_i$ .

The parameter  $s_i$  is defined as the distance from  $O_i X_i$  to  $O_{i+1} X_{i+1}$  axes, measured along the  $O_i Z_i$  axis.

Under this definition, the Denavit – Hartenberg transformation matrix  $A_i^j$  has the well-known form:

$$A_i^j = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_i^j \cos \theta_i^j & \cos \theta_i^j & -\cos \alpha_i^j \sin \theta_i^j & \sin \alpha_i^j \sin \theta_i^j \\ a_i^j \sin \theta_i^j & \sin \theta_i^j & \cos \alpha_i^j \cos \theta_i^j & -\sin \alpha_i^j \cos \theta_i^j \\ s_i^j & 0 & \sin \alpha_i^j & \cos \alpha_i^j \end{bmatrix} \quad (1)$$

To mould the walking robot's moves is assumed that:

1. The kinematical length of the binary link (1) is null and it is connected to the platform (0), by pair  $A$  and to the link (2) by pair  $B$ ; the axis of pairs  $A$  and  $B$  are perpendicular.
2. the binary link (2) is connected to the link (1) by the pair  $B$  and to link (3) by the pair  $C$ ; the axis of pairs  $B$  and  $C$  are parallel.

The  $A_3^j$  matrix performed the coordinate transformation of a point belonging to link (3) from  $O_4^j X_4^j Y_4^j Z_4^j$  system to  $O_3^j X_3^j Y_3^j Z_3^j$  system attached to link (2). In a similar manner, the coordinates of lower end point  $P$  belonging to link (3) from  $O_4^j X_4^j Y_4^j Z_4^j$  system to  $O_1^j X_1^j Y_1^j Z_1^j$  system attached to the platform (0) is performed by the equation

$$\begin{bmatrix} 1 \\ X_{1P}^j \\ Y_{1P}^j \\ Z_{1P}^j \end{bmatrix} = \prod_{k=1}^3 A_k^j \begin{bmatrix} 1 \\ X_{4P}^j \\ Y_{4P}^j \\ Z_{4P}^j \end{bmatrix}, j = 1, 2. \quad (2)$$

This matrix equation described the geometrical model of the leg 1 and 2 of the walking robot. The goal of the direct kinematic analysis is to calculate the position, velocity and acceleration of the end point  $P$ , in terms of the pair variables  $\theta_i^j$ ,  $i = 1, 3$ . In inverse kinematic analysis, matrix equation (2) is solved with respect to the pair variables  $\theta_i^j$ ,  $i = 1, 3$ .

The positions of the point  $P$  and the positions of the platform with respect to the reference coordinate axes system  $OXYZ$  fastened to the ground are considered as known. Therefore, the position of the point  $P$  with respect to the platform coordinate axes system are known.

The movement of the legs of the rear modules are controlled by the following equations.

Each of these matrix equations is equivalent with three nonlinear equations and has six unknowns, namely variables of the pairs.



$$\begin{aligned}
\begin{pmatrix} 1 \\ X_{1P}^5 \\ Y_{1P}^5 \\ Z_{1P}^5 \end{pmatrix} &= \mathbf{A}_1^3 \mathbf{A}_2^3 \mathbf{A}_3^3 \mathbf{A}_1^5 \mathbf{A}_2^5 \mathbf{A}_3^5 \begin{pmatrix} 1 \\ X_{4P}^5 \\ Y_{4P}^5 \\ Z_{4P}^5 \end{pmatrix}, \\
\begin{pmatrix} 1 \\ X_{1P}^6 \\ Y_{1P}^6 \\ Z_{1P}^6 \end{pmatrix} &= \mathbf{A}_1^3 \mathbf{A}_2^3 \mathbf{A}_3^3 \mathbf{A}_1^6 \mathbf{A}_2^6 \mathbf{A}_3^6 \begin{pmatrix} 1 \\ X_{4P}^6 \\ Y_{4P}^6 \\ Z_{4P}^6 \end{pmatrix}, \\
\begin{pmatrix} 1 \\ X_{1P}^7 \\ Y_{1P}^7 \\ Z_{1P}^7 \end{pmatrix} &= \mathbf{A}_1^4 \mathbf{A}_2^4 \mathbf{A}_3^4 \mathbf{A}_1^7 \mathbf{A}_2^7 \mathbf{A}_3^7 \begin{pmatrix} 1 \\ X_{4P}^7 \\ Y_{4P}^7 \\ Z_{4P}^7 \end{pmatrix}, \\
\begin{pmatrix} 1 \\ X_{1P}^8 \\ Y_{1P}^8 \\ Z_{1P}^8 \end{pmatrix} &= \mathbf{A}_1^4 \mathbf{A}_2^4 \mathbf{A}_3^4 \mathbf{A}_1^8 \mathbf{A}_2^8 \mathbf{A}_3^8 \begin{pmatrix} 1 \\ X_{4P}^8 \\ Y_{4P}^8 \\ Z_{4P}^8 \end{pmatrix} \quad (3)
\end{aligned}$$

In the inverse kinematic analysis three out of six unknowns must be imposed from independent conditions.

In the figure 5 shown the schemes of the sequence of the Denavit – Hartenberg coordinate systems attached to the links of the left leg of the first module (a) and to the links of the leg of the left rear module (b). In the movement simulation program, the parameters of the Denavit – Hartenberg transformation matrices have the values in Tables 1 and 2.

Table 1.

| Pairs                 | $A^2$                     | $B^2$                     | $C^2$                     |
|-----------------------|---------------------------|---------------------------|---------------------------|
| Coordinate systems    | $O_1^2 X_1^2 Y_1^2 Z_1^2$ | $O_2^2 X_2^2 Y_2^2 Z_2^2$ | $O_3^2 X_3^2 Y_3^2 Z_3^2$ |
| Transformation matrix | $\mathbf{A}_1^2$          | $\mathbf{A}_2^2$          | $\mathbf{A}_3^2$          |
| $a$ [mm]              | 0                         | 350                       | 350                       |
| $\alpha$              | $-\pi/2$                  | 0                         | $-\pi/2$                  |
| $\theta$              | $\theta_1^2$              | $\theta_2^2$              | $\theta_3^2$              |
| $s$ [mm]              | 150                       | 0                         | 0                         |

Table 2

| Pairs              | $D^4$                     | $E^4$                     | $F^4$                     | $G^8$                     | $H^8$                     | $I^8$                     |
|--------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| Coordinate systems | $O_1^4 X_1^4 Y_1^4 Z_1^4$ | $O_2^4 X_2^4 Y_2^4 Z_2^4$ | $O_3^4 X_3^4 Y_3^4 Z_3^4$ | $O_1^8 X_1^8 Y_1^8 Z_1^8$ | $O_2^8 X_2^8 Y_2^8 Z_2^8$ | $O_3^8 X_3^8 Y_3^8 Z_3^8$ |
| Trans. matrix      | $\mathbf{A}_1^4$          | $\mathbf{A}_2^4$          | $\mathbf{A}_3^4$          | $\mathbf{A}_1^8$          | $\mathbf{A}_2^8$          | $\mathbf{A}_3^8$          |
| $a$ [mm]           | 0                         | 0                         | 670                       | 0                         | 350                       | 350                       |
| $\alpha$           | $\pi/2$                   | $\pi/2$                   | $\pi/2$                   | $\pi/2$                   | 0                         | 0                         |
| $\theta$           | $\theta_1^4$              | $\theta_2^4$              | $\theta_3^4$              | $\theta_1^8$              | $\theta_2^8$              | $\theta_3^8$              |
| $s$ [mm]           | 200                       | 0                         | 0                         | 380                       | 0                         | 0                         |

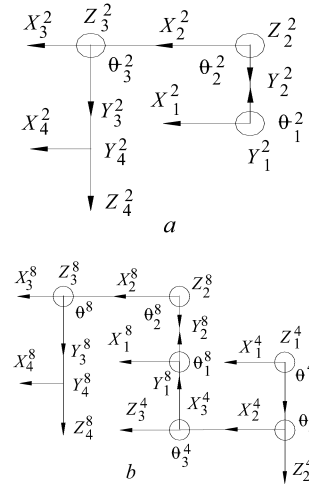


Fig 5. The Denavit – Hartenberg axes systems attached to the links of the left leg of the first module (a) and of the rear module (b)

#### First and second – time derivative of the pair variables

Through the repeated differentiated of the equation (2) with respect to the time, yields:

$$\begin{aligned}
\begin{pmatrix} 0 \\ \dot{X}_{1P}^i \\ \dot{Y}_{1P}^i \\ \dot{Z}_{1P}^i \end{pmatrix} &= \begin{pmatrix} \frac{\partial \mathbf{A}_1^i}{\partial \theta_1^i} \mathbf{A}_2^i \mathbf{A}_3^i \dot{\theta}_1^i + \mathbf{A}_1^i \frac{\partial \mathbf{A}_2^i}{\partial \theta_2^i} \mathbf{A}_3^i \dot{\theta}_2^i + \mathbf{A}_1^i \mathbf{A}_2^i \frac{\partial \mathbf{A}_3^i}{\partial \theta_3^i} \dot{\theta}_3^i \end{pmatrix} \begin{pmatrix} 1 \\ X_{4P}^i \\ Y_{4P}^i \\ Z_{4P}^i \end{pmatrix}, i=1,2, \\
\begin{pmatrix} 0 \\ \ddot{X}_{1P}^i \\ \ddot{Y}_{1P}^i \\ \ddot{Z}_{1P}^i \end{pmatrix} &= \begin{pmatrix} \frac{\partial^2 \mathbf{A}_1^i}{(\partial \theta_1^i)^2} \mathbf{A}_2^i \mathbf{A}_3^i (\dot{\theta}_1^i)^2 + \mathbf{A}_1^i \frac{\partial^2 \mathbf{A}_2^i}{(\partial \theta_2^i)^2} \mathbf{A}_3^i (\dot{\theta}_2^i)^2 + \mathbf{A}_1^i \mathbf{A}_2^i \frac{\partial^2 \mathbf{A}_3^i}{(\partial \theta_3^i)^2} (\dot{\theta}_3^i)^2 + \\ + 2 \frac{\partial \mathbf{A}_1^i}{\partial \theta_1^i} \frac{\partial \mathbf{A}_2^i}{\partial \theta_2^i} \mathbf{A}_3^i \dot{\theta}_1^i \dot{\theta}_2^i + 2 \frac{\partial \mathbf{A}_1^i}{\partial \theta_1^i} \mathbf{A}_2^i \frac{\partial \mathbf{A}_3^i}{\partial \theta_3^i} \dot{\theta}_1^i \dot{\theta}_3^i + 2 \mathbf{A}_1^i \frac{\partial \mathbf{A}_2^i}{\partial \theta_2^i} \frac{\partial \mathbf{A}_3^i}{\partial \theta_3^i} \dot{\theta}_2^i \dot{\theta}_3^i + \\ + \frac{\partial \mathbf{A}_1^i}{\partial \theta_1^i} \mathbf{A}_2^i \mathbf{A}_3^i \ddot{\theta}_1^i + \mathbf{A}_1^i \frac{\partial \mathbf{A}_2^i}{\partial \theta_2^i} \mathbf{A}_3^i \ddot{\theta}_2^i + \mathbf{A}_1^i \mathbf{A}_2^i \frac{\partial \mathbf{A}_3^i}{\partial \theta_3^i} \ddot{\theta}_3^i \end{pmatrix} \begin{pmatrix} 1 \\ X_{4P}^i \\ Y_{4P}^i \\ Z_{4P}^i \end{pmatrix}, i=1,2.
\end{aligned} \quad (4)$$

In the inverse analysis, these equations are solved with respect to the first and second – time derivative respectively of the pair variables.

The velocity and the acceleration components of the point  $P$  on the axes of the coordinate system attached to the platform (0) are considered as known.

In a similar manner are differentiated the equations (3).

In these matrix equations, the variables  $\theta_k^{2i}$ ,  $\dot{\theta}_k^{2i}$  and  $\ddot{\theta}_k^{2i}$ ,  $k=1,6$  of the pairs and their derivatives with respect to the time are imposed by the walking simulation program.

Because of these variables' particular values, the Denavit – Hartenberg transformation matrices have the following simpler particular forms:

$$A_1^{2i} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta_1^{2i} & -\sin \theta_1^{2i} & 0 \\ 0 & \sin \theta_1^{2i} & \cos \theta_1^{2i} & 0 \\ s_1^1 & 0 & 0 & 1 \end{bmatrix}$$

$$A_2^{2i} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_2^{2i} \cos \theta_2^{2i} & \cos \theta_2^{2i} & -\sin \theta_2^{2i} & 0 \\ a_2^{2i} \sin \theta_2^{2i} & \sin \theta_2^{2i} & \cos \theta_2^{2i} & 0 \\ s_2^1 & 0 & 0 & 1 \end{bmatrix}$$

$$A_3^{2i} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ a_3^{2i} \cos \theta_3^{2i} & \cos \theta_3^{2i} & -\sin \theta_3^{2i} & 0 \\ a_3^{2i} \sin \theta_3^{2i} & \sin \theta_3^{2i} & \cos \theta_3^{2i} & 0 \\ s_3^{2i} & 0 & 0 & 1 \end{bmatrix} \quad (5)$$

Keeping stable is a special problem that occurs while the robot walks, when one or more legs are in the transfer phase. When all the legs are in the support phase, it is obvious that the protection of the center of gravity is within the support polygon.

If one or more legs are in the transfer phase, the geometry of the support polygon changes and it occurs the risk that the protection of the center of gravity moves outside the support polygon.

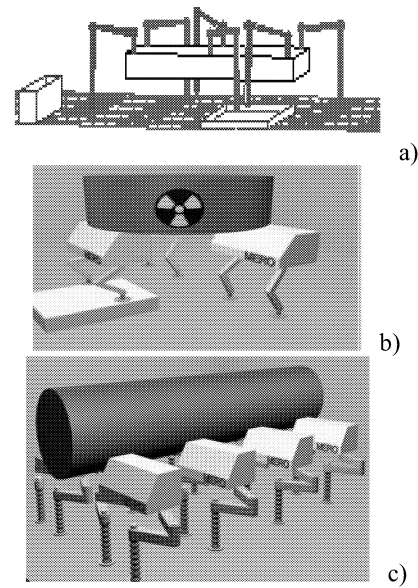
Solutions to such situations depend on how the modular walking robot are configured.

In fig. 6 a where are shown a sequence of the computer simulation of the successive gait of the modular walking robot, in C++ and in 6b, c simulation tripod and sucesive gait in Studio Max Animation

**Acknowledgements.** The authors wish to express their gratitude to the Research and Educational Ministry (MEC) for its support of the program of work reported herein. The work took place as part of the research project no. 263/2007-2010, ID 011/2007, in the framework of Grants of CNCSIS Program IDEAS, PN II.

## CONCLUSIONS

The movement simulation of the walking robots may be idealized into a mathematical model for the purpose of kinematic analysis. The techniques of idealization can play the decisive role in easiness, precision and time of calculus for the problem solving. The Denavit – Hartenberg method is numerically robust, the solutions are either exact in the sense that is possible to refine them up to an arbitrary accuracy. A modular walking robot could have one or more modules. The motions of the legs must be coordinated so that the conditions of the gait stability of the system to be ensured.



**Fig. 6 Computer graphics simulation of the gaits of the modular walking robot**

## REFERENCES

- [1] A.P. Bessonov, N.V. Umnov, (1973) "The Analysis of Gaits in six-legged Robots according to their Static Stability," Proc. Symp. Theory and Practice of Robots and Manipulators, Udine, Italy,
- [2] Denavit J., Hartenberg R.S., (1955), "A Kinematic Notation for Lower Pair Mechanisms Based on Matrices," Journal of Applied Mechanics, Tr. ASME, Vol. 77
- [3] Figliolini G., Rea P.; (2006) "Mechanics and Simulation of six -Legged Walking Robots", The 9<sup>th</sup> International Conference on Climbing and Walking Robots and Associated Technologies, Royal Military Academy, Brussels Belgium, Sept 11-14, 2006
- [4] I.Ion, I. Simionescu., A. Curaj, (2005) "MERO Modular Walking Robot Support of Technological equipments," The 8<sup>th</sup> International Conference on Climbing and Walking Robots, September 12-14, 2005 London, UK
- [5] I.Ion, I. Simionescu., A. Curaj, (2002) "Mobil Mechatronic System With Applications in Agriculture and Sylviculture." The 2<sup>th</sup> IFAC International Conference on Mechatronic Systems December 8-12, 2002-Berkeley –USA pp941-946.Pergamon Press
- [6] I.Ion, .A. Marin,. A. Curaj, L Vladareanu, (2008) "Design and Motion Synthesis of Modular Walking Robot MERO" Journal of Automation ,Mobile Robotics & Intelligent Systems ISSN 1897-8649 Vol@ No4 2008 pag.25 -30
- [7] S.M. Song, and J.K. Waldron, (1987) "An Analytical Approach for Gait Study and its Application on Wave Gait", International Journal of Robotics Research, Vol. 6, No. 2, 1987, pp. 60-71
- [8] J.K. Waldron, (1996) "Modeling and Simulation of Human and Walking Robots Locomotion" ,Advanced Scholl Udine ,Italy 1996

# MODELING, SIMULATION AND CONTROL OF ROTARY WING PLATFORMS IN A COMPUTER GENERATED FORCES TOOLKIT

Semuel Franko, Seniha Köksal, Mehmet Haklıdır  
Information Technologies Institute  
TUBITAK Marmara Research Center  
Gebze, Kocaeli  
Turkey

E-mail: [semuel.franko@bte.mam.gov.tr](mailto:semuel.franko@bte.mam.gov.tr)  
[seniha.koksal@bte.mam.gov.tr](mailto:seniha.koksal@bte.mam.gov.tr)  
[mehmet.haklidir@bte.mam.gov.tr](mailto:mehmet.haklidir@bte.mam.gov.tr)

## KEYWORDS

Computer Generated Forces (CGF), Military Training Simulation, LQR, control, Rotary Wing Platforms, HLA, RTI

## ABSTRACT

Platforms simulated by Commercial frameworks, supporting Computer Generated Forces (CGF) and control architecture, generally have simple motion equations with inadequate fidelity. Consequently, the commercial framework based CGF applications are not always able to meet the needs of training-critical military simulations that use motion models with high-fidelity. Therefore, the integration of platforms with high-fidelity motion models into the CGF applications based on commercial frameworks occurs as a critical problem.

In this paper, we present the modeling of rotary wing platforms with realistic 6 degrees of freedom motion equations, control of the platform via LQR based control methods and their integration into a commercial framework based CGF application within a high-fidelity virtual military training simulation.

In this study, VR-Forces has been chosen as the commercial framework to build the CGF and control architecture, and C++ programming language has been used to implement the software module for the proposed motion model. Furthermore this software module has been integrated to the VR-Forces back-end as a composite object. Finally two scenarios were formed to realize the simulations.

## INTRODUCTION

Distributed simulation systems are run independently in different environments; consequently, they do not have a common memory. A standard called HLA (High Level Architecture) is developed to support the communication and the reuse of these systems. HLA provides an architecture that reduces the cost and time for building a distributed simulation structure complex simulation applications require.

Training-critical military simulation applications generally utilize a distributed simulation approach since they mostly require fault tolerance, reduced execution time and geographical distribution. These applications provide

features for building scenarios, simulation of computer generated forces (CGF) and simulation of algorithmic behaviors.

VR-Forces is an HLA compliant and flexible software that meets the basic requirements of distributed simulation such as simulation management as it enables CGF on military, security or emergency intervention. There exist two types of VR-Forces applications which use Run-Time Infrastructure (RTI) for communication, front-end and back-end applications. The actual CGF simulation takes place in the back-end application and the remote front-hand GUI controls the back-ends.

Platforms, consisting of surface, air, ground and subsurface, simulated by VR-Forces have simple motion equations. Additionally, the environmental parameters that are used in these movement equations are not of appropriate fidelity to develop realistic movement models. Consequently, these movement models are insufficient for satisfying the requirements of high-fidelity military simulations used for training purposes.

In this study, algorithms and control methods of the motion module were developed in Matlab environment and they are implemented in C++ language. Then the entire model has been integrated into the component architecture of the back-end application of VR-Forces version 3.11 as a composite object. As the software module has been developed as an independent component, it is expected to easily be used with other applications like VR-Forces.

In this paper, firstly mathematical modelling of helicopter was discussed. Dynamic and kinematic equations of motion were represented according to the Padfield's (Padfield 2007) notation. The force and moments were explained that act mostly on the platform. Taylor's series expansion was used around trimming points of low and high velocity to obtain the linear motion equations and stability derivatives. State space form of the helicopter motion equations are formed by using related derivatives. Then the simulation architecture and integration of motion model was explained. At the end, simulation results were discussed with illustrations on two scenarios.

## MATHEMATICAL MODELLING OF HELICOPTER

### Coordinate Frames and Transformations

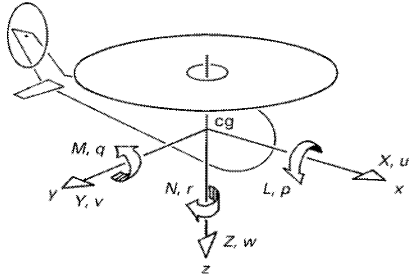
Basically, two frames are needed to demonstrate the motion of the helicopter, body fixed and earth fixed frames. In body frame, x shows longitudinal, y shows lateral, and z shows up/down movement. In the latter coordinate system, x points the north, y points east, and z points the center of the earth. To transform between body and earth frames, orthonormal rotation matrix R is used.

### Dynamic Equations of Motion

Translational motion, which is the motion of the center of gravity, can be defined by Newton's second law and Coriolis Effect. Linear accelerations (u,v,w) and angular accelerations (p,q,r) around x, y, and z axes can be defined as follows, where I defines moments of inertia and M defines inertia ratios:

$$\begin{aligned} \dot{u} &= vr - qw + \frac{F_x}{m} \quad (1) \\ \dot{v} &= pw - ur + \frac{F_y}{m} \quad (2) \\ \dot{w} &= uq - pv + \frac{F_z}{m} \quad (3) \end{aligned}$$

$$\begin{aligned} \dot{p} &= qr \frac{I_{yy} - I_{zz}}{I_{xx}} + \frac{M_x}{I_{xx}} \quad (4) \\ \dot{q} &= pr \frac{I_{zz} - I_{xx}}{I_{yy}} + \frac{M_y}{I_{yy}} \quad (5) \\ \dot{r} &= pq \frac{I_{xx} - I_{yy}}{I_{zz}} + \frac{M_z}{I_{zz}} \quad (6) \end{aligned}$$



Figures 1: Helicopter's 6 degrees of freedom movement

### Kinematic Equations

For translational kinematics, relation between body and earth fixed frame is as follows, where  $x_E, y_E, z_E$  identifies position of the helicopter with respect to earth-fixed frame.

$$\frac{dy}{dx} \begin{bmatrix} x_E \\ y_E \\ z_E \end{bmatrix} = R_{eb} \begin{bmatrix} u_B \\ v_B \\ w_B \end{bmatrix} \quad (7)$$

### Force and Moments Acting on Helicopter

Helicopter can be modeled by combining five subsystems: main-rotor, fuselage, empennage (consist of horizontal stabilizer and vertical fin), tail rotor and engine [Padfield 2007]. To define the force and moment (L,M,N) effects

originated from main rotor, tail rotor, gravity and drag on main rotor; mr, tr, g, and d subscripts are used respectively.

$$\begin{aligned} F_x &= X_{mr} + X_{tr} + X_g \quad (8) \\ F_y &= Y_{mr} + Y_{tr} + Y_g \quad (9) \\ F_z &= Z_{mr} + Z_{tr} + Z_g \quad (10) \\ L &= L_{mr} + L_{tr} + L_d \quad (11) \\ M &= M_{mr} + M_{tr} + M_d \quad (12) \\ N &= N_{mr} + N_{tr} + N_d \quad (13) \end{aligned}$$

### Trimming and Linearization

Nonlinear motion equations must be linearized about certain operating points. To increase the fidelity of the model 8 trim points have been used. First assuming that linear and angular accelerations are zero; setting the trimming forward/ side/ vertical velocity and heading rate to our desired values. By using Taylor's series expansion, external forces acting on platform become linear functions of perturbed states. Total force along x-axis, by the advantage of small perturbation theory ( $x = x_e + \Delta x$ ), can be written as follows:

$$\begin{bmatrix} F_x \\ F_y \\ F_z \\ L \\ M \\ N \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ L \\ M \\ N \end{bmatrix}_e + \frac{\partial X}{\partial u} \Delta u + \frac{\partial X}{\partial w} \Delta w + \dots = X_e + X_u \Delta u + X_w \Delta w + \dots \quad (14)$$

If we consider that the motion can be described nonlinearly as  $\dot{x} = F(x, u, t)$ , the linearized model can be defined as  $\dot{x} = Ax + Bu$ ,

where  $x = [u \quad w \quad q \quad \theta \quad v \quad p \quad \phi \quad r]$

and  $u = [\theta_{0mr} \quad a_1 \quad b_1 \quad \theta_{0tr}]$

The coefficients like  $X_u, X_w, \dots$  are called stability derivatives in flight dynamics. The result of these formulae can be found in appendix as A and B matrices.

### Obtaining the Stability Derivatives

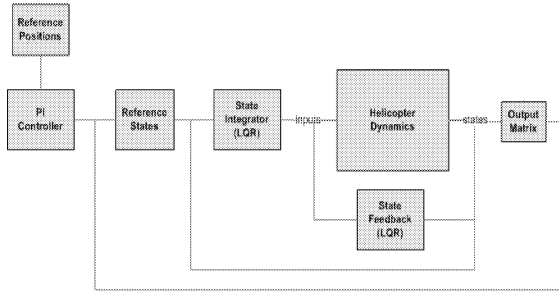
Stability derivatives can be calculated by using numerical and/or analytical methods. By using platform's main characteristics all derivatives can be figured. [Jensen and Nielsen 2005], [Munzinger 1998] For example  $X_u$  can be found analytically by the equations below:

$$\left[ \frac{\partial X_{mr}}{\partial u} = \frac{\partial (T a_1)}{\partial u} = \frac{\partial T}{\partial u} a_1 + T \frac{\partial a_1}{\partial u} \right] \quad (15)$$

Calculation of all derivatives was not mentioned in this paper. Further formulae can be found from [Prouty 2002]

### CONTROLLER DESIGN

Proposed LQR based controller consist of three subsystems. State feedback controller, state integrator and PI controller. Moreover, gain scheduling is used to reflect the change in platform dynamics with respect to the forward velocity.



Figures 2: Block diagram of controllers

Full-state feedback control algorithm tries to minimize the performance index (J), where  $x$ : states,  $u$ : inputs,  $Q$  and  $R$  are weighting matrices  $J = \frac{1}{2} \int_0^{\infty} (x^T Q x + u^T R u) dt$

After deciding appropriate weighting matrices according to set/rise time, overshoot and controlling effort; feedback gain (K) is calculated. By solving Riccati equation gains are obtained. Then the input becomes as  $u = -Kx$

### State Integrator

Full state feedback control gives adequate results. Nevertheless, controller will adjust the system to make states zero. For tracking control of the helicopter, the error between reference states and actual states must be taken into account. Therefore, the error term is defined as  $\dot{e} = \dot{x}_{desired} - \dot{x}_{actual}$ . After integrating  $\dot{e}$ ,  $e$  is obtained.

Then new control input becomes  $u = K_2 * e - K * \dot{x}$

### PI Controller

After controlling the states and setting that values according to the reference states, for position control, which has slower dynamics than attitude control a proportional-integral feedback controller is used. Position error is calculated as  $e_{pos} = [x, y, z]_{desired} - [x, y, z]_{actual}$ . By using classical  $K_p * e_{pos} + K_i \int e_{pos} dt$  formula, the trajectory control is realized.

## THE SIMULATION OF THE ROTARY WING PLATFORM

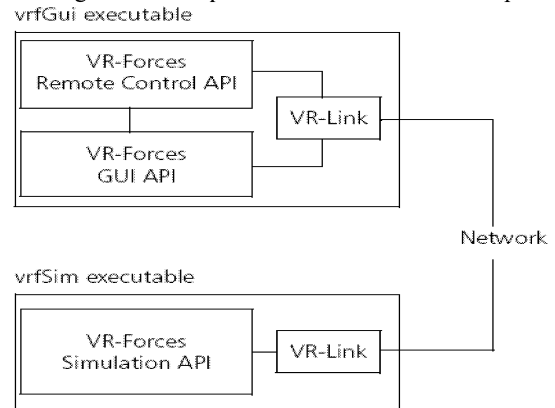
### Distributed Simulation Architecture

As Training-critical military applications mostly require integration of different simulation components that run on different hardware platforms, they utilize a distributed simulation approach. One of the main tasks that a simulation component can be devoted to is the simulation of Computer Generated Forces (CGF). Furthermore, simulation management and run control are the two other equally fateful tasks.

MÄK VR-Forces is a COTS CGF tool that is being used for a range of programs from operational analysis to embedded training [Lubetsky and Granowetter 2000]. Above RTI,

which provides network, services to VR-Forces application, there is two separate federates to provide interface for user operations and simulate the entities. (Figures 3) This implementation depends on the system architecture; also, these capabilities can be implemented in a single application as a single federate too.

User interface application provides the control services (start/stop simulation) over the whole system; furthermore, it allows user defined actions to be inserted to simulation within the content of a scenario. When one of these actions is made by user, related messages sent from the GUI application to the CGF application. According to these user inputs, corresponding actions are taken place into the CGF application, which actually contains a simulation engine that is handling all these requests within a simulation loop.



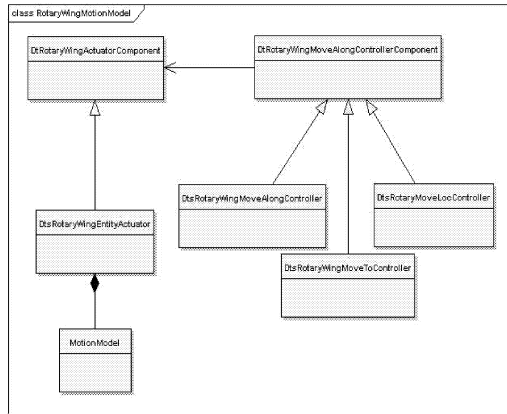
Figures 3: VR-Forces API implementation[Mak Technologies]

### Software Architecture

VR-Forces component architecture is used to realize the integration of the software module into the VR-Forces back-end application. In this architecture, there exist three types of components: namely controllers, actuators and sensors, which are used by component manager. Controllers process information about the tasks, which are assigned to entities by the user via the GUI, such as changing the characteristics of the entity or task assignment to the entity. There are different controllers each of which are specialized for a certain kind of task. Actuators obtain the necessary parameters from the controllers and realize the entity actions such as entity motion, etc. In other words, actuators update the entity's state information such as position, velocity, etc. in each simulation frame time.

### Model Integration to VR-Forces Architecture

For integrating the software module, which was developed for the movement model of the rotary wing platform, into the VR-Forces application, new controllers and actuator classes were derived from the original VR-Forces components. In this way new movement model was integrated into new actuator and new control parameters integrated into new controllers. (Figures 4)



Figures 4: Class Diagram of the Software Module

Since there are three different movement tasks that rotary wing can accomplish, three different controllers specified for controlling them. These are move along route, move to location and move to waypoint. In contrast, there is only one actuator, as it is responsible to release the same model for all tasks.

The operation of the motion module is started when the scenario runs and one of the movement tasks is assigned to the rotary wing platform. In order to accomplish assigned task, the user interaction values obtained by the command messages (such as the desired velocity of the platform, etc) are transferred to the actuator from the related controller and then motion module takes them via the actuator. Furthermore, the initial values used in the motion equations and characteristic parameters belongs to the platform are sent to the motion module owned by the entity's actuator.

Motion model of the platform calculates the new state information (position, velocity, etc.) of the rotary wing platform at the end of  $\Delta t$  from the time dependant velocity, acceleration and change of location values. The actuator class updates the state information of the platform by newly calculated state information. Updated information is broadcasted to other federates in the federation via RTI by the VR-Forces local network interface class which is created for the rotary wing platform.

## EXPERIMENTAL RESULTS

For testing the controllers' performance two following scenarios were formed. In primary scenario the user commands the helicopter to reach to four waypoints one by one. In latter, one route was established and commanded to the platform to follow.

### Movement Through Waypoints

For testing controller a waypoint scenario is formed. Initial point is set as  $[x \ z \ y] = [0 \ 0 \ 0]$  meters. Commanded composite velocity (of  $u$  and  $v$ ) is 20 m/sec. Four waypoints were selected as follows:

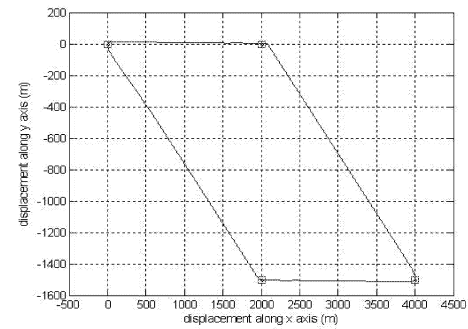
Waypoint 1= [2000 -300 -1500]

Waypoint 2= [4000 -300 -1500]

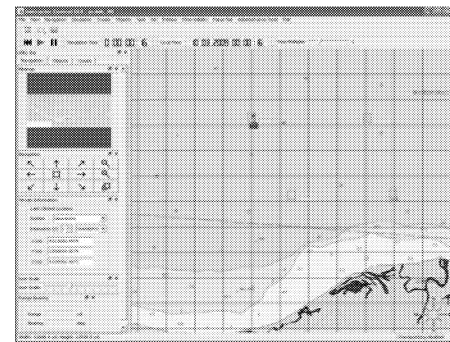
Waypoint 3= [2000 -300 0]

Waypoint 4= [0 -300 0]

In figure below, squares show waypoints, lines show the actual way of the helicopter.



Figures 5: 2D view of motion through waypoints



Figures 6: View of motion through waypoints in VR-Forces

### Movement Along Route

Initial point is set as  $[x \ z \ y] = [0 \ 0 \ 0]$  meters

A route consists of seven points was selected as follows:

Point 1= [-1.282482203782164 0 1.055818065669462]

Point 2= [-1.271867557262692 0 1.059178161341084]

Point 3= [-1.251208172955602 0 1.077556504674104]

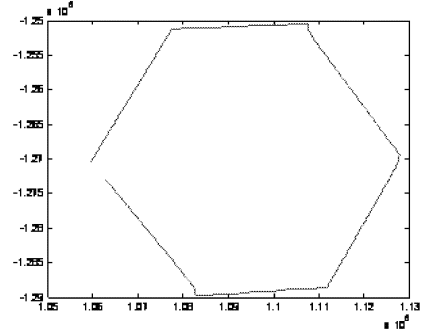
Point 4= [-1.250465372363537 0 1.107525134153801]

Point 5= [-1.288555974064607 0 1.111868795609908]

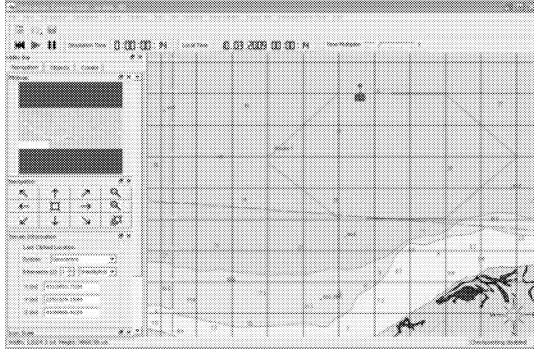
Point 6= [-1.289796838769078 0 1.082805280437650]

Point 7= [-1.272898446128637 0 1.062745052891792]

Composite velocity was the same with the preceding scenario. The result of the simulation can be seen in the following figure.



Figures 7: 2D view of motion through route



Figures 8: View of motion through route in VR-Forces

## CONCLUSIONS

In this study, a general 6 degrees of freedom motion model has been developed that includes force and moments acting on the helicopter. Proposed LQR based controller, consist of three subsystems-state feedback controller, state integrator and PI controller, has been implemented in this model. The model implemented as a software module has been integrated into the component architecture of the VR-Forces back-end. This integration has led the rotary wing platforms of the VR-Forces application to reach a high-fidelity level that is necessary for training-critical military simulations.

The performance of the controller has been illustrated through two different simulation case studies. The results are acceptable and satisfy for the needs of military simulation. Although we have designed our control to cover all influences, a more specified design can upgrade the performance in each different case. In addition, the simulation of environmental conditions and reflecting their effects to the corresponding platforms has enriched the simulation architecture and increased its usability for educational purposes.

As future work, we are planning to integrate torpedo object to our CGF platforms [Haklıdır et al. 2008] [Senyurek et al. 2008] to increase their strategic capability. The performance of the controller may be compared with an intelligent control technique or in different cooperation cases.

## APPENDIX

A matrix:

$$A = \begin{bmatrix} X_u & X_w & Q & X_q & -w & -g \cos(\theta) & X_r & R & X_p & 0 & X_v & V \\ Z_u & Z_w & Z_q & U & -g \cos(\phi) \sin(\theta) & Z_r & -P & Z_p & -V & -g \sin(\phi) \cos(\theta) & Z_v & \\ M_u & M_w & M_q & 0 & M_r & M_p & -2P \frac{I_{xz}}{I_{yy}} & M_r & -2R \frac{I_{xz}}{I_{yy}} & 0 & M_v & \\ 0 & 0 & \cos(\phi) & 0 & 0 & 0 & 0 & -\Omega_x \cos(\theta) & -\sin(\phi) & 0 & 0 & \\ Y_u & -R & Y_w & +P & Y_q & -g \sin(\phi) \sin(\theta) & Y_r & Y_p & +W & g \cos(\phi) \cos(\theta) & Y_v & -U \\ L'_u & L'_w & L'_q & +k_1 P - k_2 R & 0 & L'_r & L'_p & +k_1 Q & 0 & L'_v & -k_2 Q & \\ 0 & 0 & \sin(\phi) \tan(\theta) & \Omega_x \sec(\theta) & 0 & 0 & 1 & 0 & 0 & \cos(\phi) \tan(\theta) & 0 & \\ N'_u & N'_w & N'_q & -k_1 R - k_2 P & 0 & N'_r & N'_p & +k_2 Q & 0 & N'_v & -k_1 Q & \end{bmatrix}$$

B matrix:

$$B = \begin{bmatrix} X_{\theta_{0,ur}} & X_{a_1} & X_{b_1} & X_{\theta_{0,ir}} \\ Z_{\theta_{0,ur}} & Z_{a_1} & Z_{b_1} & Z_{\theta_{0,ir}} \\ M_{\theta_{0,ur}} & M_{a_1} & M_{b_1} & M_{\theta_{0,ir}} \\ 0 & 0 & 0 & 0 \\ Y_{\theta_{0,ur}} & Y_{a_1} & Y_{b_1} & Y_{\theta_{0,ir}} \\ L'_{\theta_{0,ur}} & L'_{a_1} & L'_{b_1} & L'_{\theta_{0,ir}} \\ 0 & 0 & 0 & 0 \\ N'_{\theta_{0,ur}} & N'_{a_1} & N'_{b_1} & N'_{\theta_{0,ir}} \end{bmatrix}$$

## REFERENCES

- Franko, S. 2009. Lqr Based Trajectory Control of Full Envelope Autonomous Helicopter, In *WCE 2009*, In process)
- Haklıdır, M.; Aldogan, D.; Tasdelen, I.; 2008. "High Fidelity Modeling and Simulation of Surface Platforms in a Computer Generated Forces Toolkit", In *SUMMERSIM 2008*, Edinburgh, Scotland.
- Jensen, R.; Nielsen, A. 2005. "Robust Control of an Autonomous Helicopter" Aalborg University & University at Buffalo
- Lubetsky, B.; Granowetter, L. 2000. "Implementing VR-Forces, a flexible CGF toolkit Mak Technologies." In *Spring Simulation Interoperability Workshop*
- Mak Technologies. VR-Forces Developer's Guide, 24.
- Munzinger, C. 1998. "Development of a Real-time Flight Simulator For an Experimental Model Helicopter" *Georgia Institute of Technology - School of Aerospace Engineering. M.S. Thesis*
- Prouty, R.W., 2002. "Helicopter Performance, Stability, and Control" *Krieger Pub Co*
- Padfield, G.D. 2007. "Helicopter Flight Dynamics: The Theory and Application of Flying Qualities and Simulation Modeling" *Blackwell Publishing*, Oxford, UK
- Senyurek, L.; Koksall, S.; Genc, H.M.; Aldogan, D.; Haklıdır, M.; 2008. "Implementation Of Fuzzy Control For Surface Platforms In A Computer Generated Forces Toolkit", In *International Conference on Computational Intelligence for Modelling, Control and Automation (CIMCA08)*, Vienna, Austria.

## BIOGRAPHY

**SEMUEL FRANKO** is a researcher in The Scientific and Technological Research Council of Turkey, Marmara Research Center (TUBITAK MRC). He is also a M.S. student in System Dynamics and Control Department of Istanbul Technical University. He received his B.S. degree in Mechanical Engineering from Istanbul University.

**SENİHA KÖKSAL** is a researcher in The Scientific and Technological Research Council of Turkey, Marmara Research Center (TUBITAK MRC). She is also a M.S. student in Bogazici University. She received her B.S. degree in Computer Engineering from Ege University.

**MEHMET HAKLIDIR** is a senior researcher in The Scientific and Technological Research Council of Turkey, Marmara Research Center (TUBITAK MRC). He is also a PhD student in Mechanical Engineering Department of Istanbul Technical University. He received his M.S. degree in Mechatronics Engineering from Istanbul Technical University and B.S. degree in Mechanical Engineering from Istanbul University.





# **SENSORS AND PROCESS CONTROL**



# MULTIVARIATE CONTROL STRUCTURE SYNTHESIS FOR FLUID CATALYTIC CRACKING UNIT USING RELATIVE GAIN ARRAY

Redah Alsabei, Zoltan K. Nagy\*, Vahid Nassehi

Department of Chemical Engineering, Loughborough University, Loughborough, Leicestershire,  
LE11 3TU, United Kingdom,

Emails: {r.alsabei, z.k.nagy, v.nassehi}@lboro.ac.uk

## KEYWORDS

Fluid Catalytic Cracking, Relative Gain Array, Five Lump Kinetics, Genetic Algorithm, Control Tuning.

## ABSTRACT

Fluid catalytic cracking (FCC) is one of the most important process in the petrochemical industries. The process is multivariable, strongly nonlinear, highly interactive and is subject to many operational, safety and environmental constraints, posing challenging control problems. A novel mode has been developed that captures the complex dynamics of the process, which combines together models of the reactor with a detailed five lumped kinetic model for the riser, as well as regenerator, and fractionator. Control studies of closed-loop simulations were carried out using an advanced proportional-integral-derivative (PID) algorithm which implemented the discrete position algorithm with backward approximation and anti-windup. The method of relative gain array (RGA) was used for the evaluation of potential control pairings for the multivariate system for which the PID controllers were implemented. The PID controllers were tuning using a genetic algorithm based optimization algorithm. The results indicate strong couplings among some of the control loops therefore there is a strong incentive for a more advanced multivariable controller such as nonlinear model predictive control (NMPC) based on economic objective.

## INTRODUCTION

The Fluid Catalytic Cracking (FCC) is one of the main processes in a refinery, because its product is high quality gasoline. In petroleum and petrochemicals, cracking is the process whereby complex organic molecules such as heavy hydrocarbons are broken down into simpler molecules (e.g. light hydrocarbons). This can be done with a thermic/steam or catalytic method. The fluid catalytic cracking unit (FCCU) has been selected as a model of this study because it produces more gasoline with a higher octane rating. It also

produces byproduct gases that are more olefinic, and hence more valuable, than those produced by thermal cracking. The productivity of the unit presented schematically in figure 1 increases with the cracking reactor operation temperature constrained only by its metallurgical limit.

Consequently, for a higher productivity, the system needs to be operated as close as possible to the metallurgical limits of temperature, with high flows of raw material; a condition has to be respected: that the control systems have to be able to avoid the violation of constraints resulting from safety or environmental considerations, even when the process is strongly disturbed. Cracking reactions are endothermic, therefore, for higher productivity the reactor should be operated at a temperature as high as possible. To assure the regeneration of the catalyst (the burn of the deposited coke) the regenerator, just as the reactor, must be operated as close as possible to its metallurgical limit of temperature. The restriction concerning the carbon monoxide concentration in the stack gas is very important from environmental considerations. Differential pressure between regenerator and reactor has strong influence upon the catalyst circulation, and because of that, it strongly influences the other important process variables. Therefore, its variations should be limited. The constraints regarding reactor and regenerator pressure, come from the mechanical resistance of the construction material. The appropriate control of the process must take into consideration all these restrictions, but it must allow at the same time a high feed flow processing for an increased productivity (Gould et al. 1970; Morari et al. 1993). This goal can be accomplished using two types of control schemes:

(I) including raw material flows (fresh feed flow and/or slurry recycle) into manipulated variables. These parameters have strong influence on all important process variables, therefore, they can be very efficient manipulated inputs. However, in this case we must avoid decrease of productivity that could be caused by the total feed flow variation.

(II) setting constant feed flows, which assure an expected productivity and choosing control structures

---

\* Corresponding author

without feed flows among manipulated inputs, capable of an appropriate control of the process.

A complex dynamic model has been developed for the study of possible control architectures for the FCCU. The model is developed in Matlab and Simulink interface on the basis of descriptions of the plant and operation data of the FCCU. Moreover the model has been used to study the dynamic behavior of the process, control the output variables and economic optimisation. The resulted model is described by a complex system of a high order differential algebraic equation (DAE) system, with 142 ordinary differential equations (ODEs) resulting from material and energy balances and more than 100 algebraic equations resulting from the discretization of the kinetic models. The dynamic simulator was used to analyze the process dynamics and to evaluate possible control pairings using the relative gain array (RGA) method (Agachi et al. 2006).

System interactions (Kalra and Georgakis 1994) suggest the use of an advanced control strategy (Goodwin et al. 2001). The purpose of the research was to find out schemes for a better control of the process and implement multivariate PID to the FCCU. An optimization based control tuning approach is also presented, which can be used for multivariate PID control tuning, using genetic algorithms. The control approach in this paper can be adapted to other cracking processes such as olefin cracking process.

## DYNAMIC MODELLING AND OPEN LOOP SIMULATION OF THE UOP FCCU

The fluid catalytic cracking typical UOP has been chosen as a model for this study based on the market demands, efficiency and wide application in many petroleum industries around the world. Typical UOP unit configuration is presented in Figure 1. Modelling and control studies for this unit have been done.

In current refineries, the FCC unit plays a prominent role, producing gasoline and diesel, as well as valuable gases such as ethylene, propylene and isobutylene, from feedstocks that comprise atmospheric gas oils, vacuum gas oils and hydrocracker bottoms. The significant economic role of the FCC unit in modern-day petroleum refining has attracted great interest in academia and industry in terms of developing and modelling control algorithms for efficient FCC application. The main parts of the FCC unit that have been modelled are: Feed system, preheat system, reactor, regenerator, air blower, wet gas compressor, catalyst circulation lines and main fractionator.

Accurately modelling industrial FCC units is very complicated due to the strong interactions between process variables, and the significant uncertainty surrounding the kinetics of the cracking reactions, coke

deposition leading to catalyst deactivation in the riser reactor and coke burning process in the regenerator.

The FCCU model has been developed on the basis of descriptions of the plant and operation data of the running FCCU. The model is sufficiently complex to capture the major dynamic effects that occur in an actual FCCU system. It is multivariable, strongly interacting and highly nonlinear. The developed simulator is able to depict the main dynamic characteristics of a typical commercial FCC process.

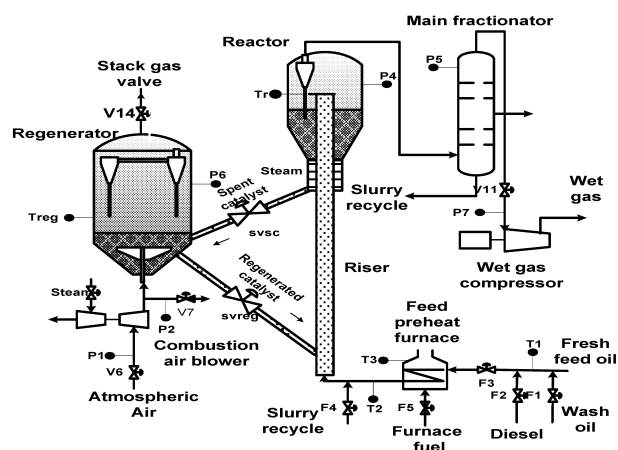


Figure 1: UOP type Fluid Catalytic Cracking Unit

The FCCU dynamic model includes the main systems: feed and preheat system, reactor (riser and stripper), regenerator, air blower, wet gas compressor catalyst circulation lines and the main fractionator. A yield model capable to predict the yields of valuable products and gasoline octane value is included in the model of the unit: the five-lump kinetic model for the riser section since it corresponds to the available literature data and it is sufficiently complex to describe the yields of valuable products and gasoline octane value.

The dynamic simulator was used to study the dynamic behaviour of the process. Set of FCCU dynamic simulations have been performed and studied based on nonlinear model. Different control loop was implemented to investigate the dynamic behaviour of FCCU in open loop and closed loop control system. The PID control algorithm was implemented and tested for the FCCU process to achieve closed loop stability.

An open-loop controller, also called a non-feedback controller, is a type of controller which computes its input into a system using only the current state and model of the system.

The model was implemented as Matlab universal oil products (UOP) FCCU, and simulations were carried out using the Simulink block diagram shown in (figure 2) to observe the dynamic behaviours for this model using different combinations of changes in set points and disturbances.

Different control structures were investigated for the most important disturbance that can appear frequently during the operation: feed composition change. In simulations a Matlab version of FCCU (UOP) model was used (McFarlane et al. 1993; Morari and Agachi 1992). Since feed composition is not modelled in detail, a feed composition change is simulated by altering the coking factor which governs coke yield (F3). The effect of this disturbance (5 % step change in F3 at  $t = 1.5$  h) on the uncontrolled process is shown in figure 3.

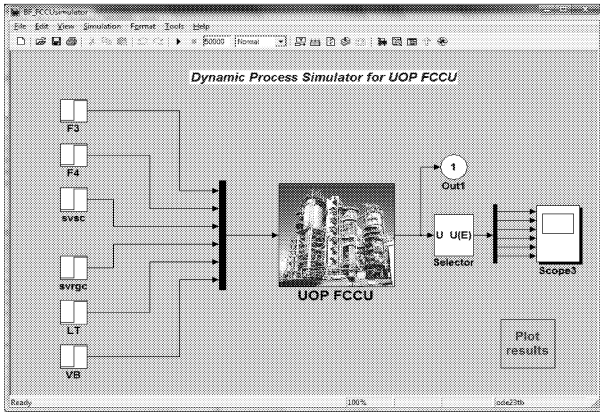


Figure 2: SIMULINK Block Diagram for Open Loop Simulation

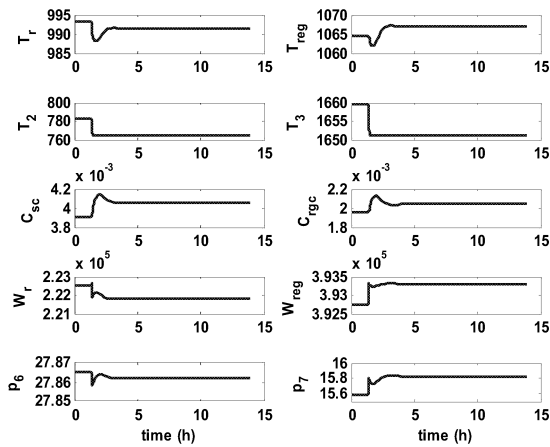


Figure 3: Simulation of FCCU Dynamic Behaviour in the Presence of the Feed Flow (+5%, increase)

### Control Structures Selection and Multivariable PID Control of the FCCU

The best control system will minimize intrusion on constraint boundaries and return the system to the nominal operating levels in minimum time. Different control structures from the above-mentioned types of control schemes were tried. The PID controller was studied. To determine the inherent control properties of the plant a controllability analysis was carried using the

relative gain array (RGA) method. The RGA is defined as the ratio of the open loop gain between an input  $j$  and an output  $i$  to the closed loop gain between the same input and output when for all the other loops perfect control is considered. The RGA is a useful measure of interaction between inputs and outputs for MIMO (multi-input-multi-output) systems. The RGA linearize systems of nonlinear matrix at different operating point may have different pairings.

For a system with  $m$  inputs,  $u = [u_1, \dots, u_m]$ ,  $n$  differential states  $x = [x_1, \dots, x_n]$  and  $k$  outputs,  $y = [y_1, \dots, y_k]$  the generic nonlinear model has the form,

$$\frac{dx}{dt} = f(x, u), \quad (1)$$

$$y = g(x, u). \quad (2)$$

Linearizing the nonlinear model equations (1)-(2) in the operating point given by  $x_0$  and  $u_0$  gives the linear state space model of the system:

$$\frac{dx}{dt} = Ax + Bu, \quad (3)$$

$$y = Cx + Du, \quad (4)$$

where the linearization matrices are given by:

$$A = \left[ \frac{\partial f}{\partial x} \right]_{x_0, u_0}, \quad B = \left[ \frac{\partial f}{\partial u} \right]_{x_0, u_0}$$

$$C = \left[ \frac{\partial g}{\partial x} \right]_{y_0, u_0}, \quad D = \left[ \frac{\partial g}{\partial u} \right]_{y_0, u_0}$$

The RGA matrix for a square plant can be calculated using the following equation:

$$RGA = (-CA^{-1} \cdot B + D) \times \left( (-CA^{-1} \cdot B + D)^{-1} \right)^T \quad (5)$$

where the  $\times$  symbol is the Hadamard product (element by element product) and  $A$ ,  $B$ ,  $C$  and  $D$  are coefficient matrixes from the linear state space model.

A relative gain between an input and an output close to 1 gives the best control pairs for independent control loops. In table 1, the RGA matrixes are shown for 6 of the studied control schemes.

Table 1 shows different control structures in which we implemented the PID controller. The best results were obtained, with the RGA matrix, with a control scheme (scheme 6 in table 1). The following controlled variables: temperature of fresh feed entering reactor riser ( $T_2$ ), reactor riser temperature ( $Tr$ ) and reactor

fractionator pressure ( $P_3$ ); the manipulated variables were furnace fuel flow (F5), slurry recycle flow rate (F4) and spent catalyst valve (svsc). The results obtained with the other schemes from table 1 were not

conclusive since there are strong couplings among some of the control loops specially when implement three or more control loops to the system.

Table 1: RGA Matrixes of Different Control Schemes

| 1    | F3 | F5 | svsc   | V14     | svrgc | 2    | svsc | svrgc |
|------|----|----|--------|---------|-------|------|------|-------|
| P6   | 1  | 0  | 0      | -0      | -0    | Tr   | -0   | 1     |
| T3   | -0 | 1  | 0      | 0       | 0     | Treg | 1    | -0    |
| Tr   | -0 | -0 | -0     | 0       | 1     |      |      |       |
| Treg | 0  | 0  | 1.0027 | -0.0027 | -0    |      |      |       |
| P5   | -0 | -0 | -0     | 1.0027  | -0    |      |      |       |

| 3    | F4 | V14     | svrgc   | svsc    | 4       | F3     | F4     | V11 | VB |
|------|----|---------|---------|---------|---------|--------|--------|-----|----|
| P6   | -0 | 0.0017  | -0.0044 | 1.0027  | P7      | 0      | -0     | 1   | 0  |
| Treg | 0  | -0.0017 | 1.0044  | -0.0027 | Tr      | 0.7911 | 0.2089 | -0  | 0  |
| Crgc | 0  | 1       | -0      | 0       | Xgas    | -0     | 0      | 0   | 1  |
| Csc  | 1  | -0      | 0       | -0      | Xslurry | 0.2089 | 0.7911 | -0  | -0 |

| 5    | svsc    | svrgc | V14     | 6  | F4 | F5 | svsc |
|------|---------|-------|---------|----|----|----|------|
| Tr   | -0      | 1     | 0       | P5 | 1  | 0  | -0   |
| Treg | 1.0027  | -0    | -0.0027 | Tr | -0 | 0  | 1    |
| Dpr  | -0.0027 | -0    | 1.0027  | T2 | 0  | 1  | 0    |

The MIMO system design (figure 4) presents the block diagrams implemented in Simulink environment to simulate the PID control with three independent control loops. In this study each controller was tested to control  $T_r$ ,  $T_2$  and  $P5$ . In order to provide a suitable environment without influencing other control variables, the best results were obtained by manipulating  $svsc$  to control  $T_r$ ,  $F5$  for  $T_2$  and  $F4$  for  $P5$ . The results are shown in (figure 5). The disturbances caused by increasing the flow of fresh feed  $F_3$  were used to study the performances of the PID controllers.

#### Constrained Optimization Based MIMO PID Tuning

The PID controller is in general a single input single output (SISO) controller which has a set of tuning parameters, which can be used to fine-tune the closed-loop response for good performance and stability.

Tuning the multiple interacting PID loops is generally very difficult due to the high level of coupling between the loops.

In practice often the loops appear to be tuned independently, however when they are activated simultaneously the process can exhibit poor control performance or can be destabilized. To take the coupling between the loops into account the controllers should be tuned simultaneously. The paper proposes a method for the tuning of the multiple PID control loops based on the a constrained optimization problem, which minimizes the overall integral square error ( $OISE$ ) of the control loops defined as,

$$OISE = \sum_{i=1}^{N_L} \int_0^{t_f} (e_i(t))^2 dt \quad (6)$$

where  $e_i(t)$  is the error function in the case of loop  $i$ ,  $N_L$  is the number of PID loops and  $t_f$  is the time

horizon on which the error is computed. From the objective function the control effort was omitted since the primary objective was to evaluate whether it is possible to find set of tuning parameters which would provide suitable control performance in a given input-output structure. However once a particular control architecture would be found, in a practical framework the control efforts could also be included in the objective function (6). The optimization problem for the tuning is expressed as follows

$$\min_{K_i, T_i} OISE \quad (7)$$

subject to:

$$K_{\min,i} \leq K_i \leq K_{\max,i} \quad (8)$$

$$T_{\min,i} \leq T_i \leq T_{\max,i} \quad i = 1, 2, \dots, N_L \quad (9)$$

where  $K_i$  and  $T_i$  are the proportional gain and integral time, respectively, for the PID controller of loop  $i$ , and the constraints (8) and (9) sets the minimum and maximum limits on the tuning parameters. The solution of the optimization problem (7)-(9) is difficult to obtain. In this paper a method is proposed based on the application of genetic algorithms (GA) for the solution of the optimization problem which is compared with sequential quadratic programming (SQP). The GA-based method has the advantage that it does not require the computation of the jacobian and hessian matrixes, which in the case of the high dimensional and stiff FCCU model often leads to ill-conditioned problems and hence difficulty in the optimization problem.

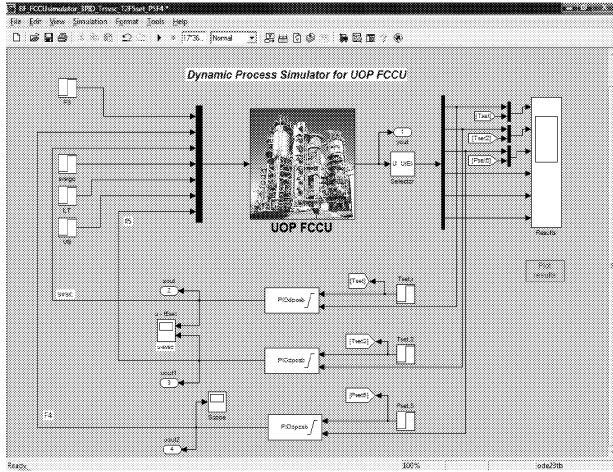


Figure 4: Simulink Block Diagram for Closed Loop Simulation

Table 2: Control tuning for advanced PID

| $K_1$  | $T_1$ | $K_2$ | $T_2$ | $K_3$ | $T_3$ | $OISE$ |
|--------|-------|-------|-------|-------|-------|--------|
| 0.0009 | 5.9   | 0.001 | 5.5   | 0.003 | 0.8   | 1.07E5 |

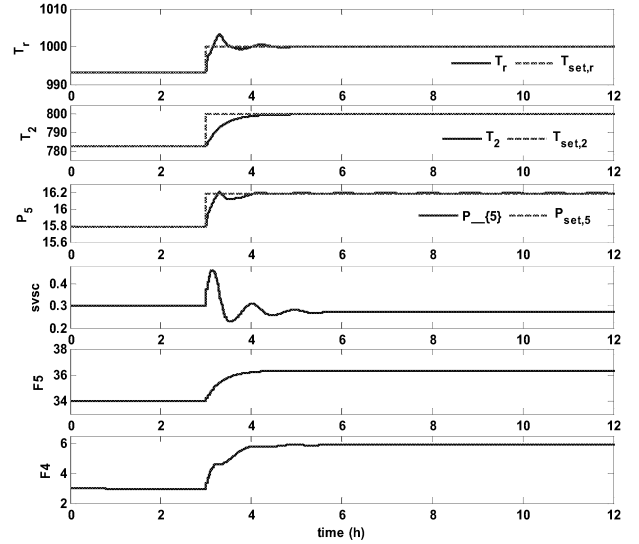


Figure 5: Performance of the Controller in the Presence of +20% Step Change ( $F_3=80 \rightarrow 96$  lb/s) at  $t = 3$  h

The simulation results show that a relatively small number of population used in the GA can converge quickly (within 20 generations) to the region of the global optimum. Table 2 contain sets of tuning parameters that have been obtained for each PID loops for which the control performance shown in figure 5 is achieved.

## CONCLUSIONS

The FCCU dynamic model has been simulated includes the main systems. The novelty of the model consists in that besides the complex dynamics of the reactor-regenerator system it also includes the dynamic model of the fractionator, as well as a model capable to predict the yields of valuable products and gasoline octane value: a kinetic model for the riser section.

The dynamic simulator was used to study the dynamic behaviour of the process. The simulations were conducted using different combinations of changes in set points and disturbances to exclude the possibility of cancellations of certain output dynamics. A simple RGA matrix based controllability analysis was performed in order to study the interdependence among the process variables. The PID algorithm was tried for different control structures. The results obtained strong couplings among some of the control loops, the PID controller parameters are very difficult to tune in order to satisfy all the outputs. The controllers are delicate and can easily drive the process into failure. Therefore, there is a strong incentive for a more advanced multivariable controller which is able to handle the strong couplings inherent in the process dynamics instead of implementing MIMO system with PID controller to achieve control objective.

## REFERENCES

- Agachi S., Nagy Z.K., Agachi, S. P., Nagy, Z. K., Cristea, M. V. and Imre-Lucaci, A.I., Model Based Control, Wiley-VCH Verlag GmbH & Co. KGaA, 2006.
- Garcia C.E., D.M. Prett, M. Morari (1989), *Automatica*, **25**, 335-348 .
- Goodwin, G. C., S. F. Graebe and M. E. Salgado (2001). *Control system design*, Prentice-Hall, Inc..
- Gould L.A., L.B. Evans and H. Kurihara (1970). Optimal Control of Fluid Catalytic Cracking Processes, *Pergamon Press*; 695-703
- Kalra L. and C. Georgakis (1994). *Ind. Eng. Chem. Res.*, **33**, 3063-3069.
- McFarlane R.C., R.C. Reineman, J.F. Bartee, C. Georgakis, *Comput. Chem. Eng.*, **17**, 1993, 275.
- Morari M., S. Agachi, I. Huq, J. Bomberger, B. Donno, A. Zheng (1993). Modifications to Model IV FCCUs to Improve Closed Loop Performance, Joint research project between Chevron Research and Technology Company and California Institute of Tehnology, Final Report.
- Morari M., S. Agachi (1992). Modeling and Control of a Model IV FCCU, Joint research project Chevron Research and Technology Company and California Institute of Technology, Quarterly Report.

## SYMBOLS

$A, B, C, D$  = continuous state space matrixes used in linear model

$C_{rge}$  = weight fraction of coke on regenerated catalyst

$C_{sc}$  = weight fraction of coke on spent catalyst

$D_{prrr}$  = differential pressure between the reactor and regenerator

$F3$  = total feed flow rate

$F4$  = slurry recycle flow rate

$F5$  = flow of fuel to furnace

$K_i$  = controller gain for parameter  $i$

$P_5$  = reactor fractionator pressure (psia)

$P_6$  = pressure in the regenerator (psia)

$P_7$  = wet gas compressor suction pressure (psia)

$svrgc$  = coefficient of the control valve on the regenerated catalyst pipe

$svsc$  = coefficient of the control valve on the spent catalyst pipe

$T$  = sampling time

$t$  = time

$T_2$  = temperature of fresh feed entering reactor riser

$T_3$  = furnace firebox temperature

$T_i$  = integral time for parameter  $i$

$Tr$  = temperature in the reactor roser

$Treg$  = temperature in the regenerator

$u$  = manipulated variable

$V11$  = wet gas compressor suction valve position (0-1)

$V14$  = stack gas valve position (0-1)

$VB$  = reboiler's liquid flow

$W_r$  = inventory of catalyst in reactor (lb)

$W_{reg}$  = inventory of catalyst in regenerator (lb)

$X$  = state variables

$x_{gas}$  = gas top of the column

$x_{slurry}$  = slurry bottom of the column

$y$  = output

### Subscripts

$l$  =  $l$ th sampling period

### Superscripts

$d$  = disturbance

$u$  = control actions

$k$  =  $k$ th sampling period

$T$  = matrix transpose operator



# SOFT-SENSOR FOR ON-LINE MASS OF CRYSTALS MEASUREMENT: APPLICATION TO CANE SUGAR CRYSTALLIZATION

Cédric Damour  
Michel Benne  
Brigitte Grondin-Perez  
Jean-Pierre Chabriat

EA 4079, Laboratory of Energetic, Electronics and Processes, University of La Reunion  
15 Av. René Cassin, BP 7151, 97715 Saint-Denis  
E-mail : {cdamour|benne|bgrondin|chabriat}@univ-reunion.fr

## KEYWORDS

Soft-Sensor, On-line Monitoring, Cane Sugar Crystallization

## ABSTRACT

In cane sugar industry, on-line monitoring and control of crystallization processes are performed using the electrical conductivity  $\kappa$  of sucrose solutions, which provides information on the supersaturation state. Nevertheless, previous studies point out the shortcomings of this approach:  $\kappa$  may be regarded as inappropriate to guarantee an accurate estimation of the supersaturation state in impure solutions. To improve the process control efficiency, additional information is necessary. The mass of crystals in the solution ( $m_c$ ) is relevant to complete information and to improve the accuracy and reliability of monitoring. Indeed,  $m_c$  inherently contains information about the mass balance and the crystal content ( $cc$ ). The main problem is that  $m_c$  is not available on-line. To overcome this technological problem, the development of a soft-sensor is proposed for  $m_c$  on-line measurement, which furthermore provides valuable information about the process efficiency. In this paper, a reliable soft-sensor is presented for a final crystallization stage (C sugar). Simulation results obtained on industrial data show the reliability of this approach,  $m_c$  and the crystal content ( $cc$ ) being estimated with a sufficient accuracy for achieving on-line monitoring in industry.

## INTRODUCTION

The crystallization process is the key stage in the production of sugar. Sugar is extracted from sucrose solutions and crystallization occurs in specific physicochemical conditions, defined by the supersaturation state. In order to improve exhaustion, crystallization is achieved grade wise, inside vacuum boiling pans, through three stages named A, B and C. As the purity of the solutions decreases from A to C stages, the efficiency of the C-grade crystallization is of high interest, because it is the last step where sugar could be extracted.

In cane sugar industry, this process monitoring and control is based on an electrical conductivity setpoint, the

most usual approach to control the supersaturation of the solutions and keep it in the appropriate range.

The observation of C crystallization operations reveals that the electrical conductivity  $\kappa$  of impure sucrose solutions is affected by the change of concentration, material's quality and other technical conditions such as the reduced pressure in the vacuum pan. In other words,  $\kappa$  may be regarded as inappropriate to guarantee an accurate estimation of the supersaturation state in impure solutions (Grondin-Perez et al. 2005). Experimental investigations point out the shortcomings of control strategies following a conductivity setpoint. Indeed, observation of growth phase operations in solutions containing a high level of impurities reveals  $m_c$  may not increase even if  $\kappa$  tracks the desired setpoint (Figure 1 & Figure 2).

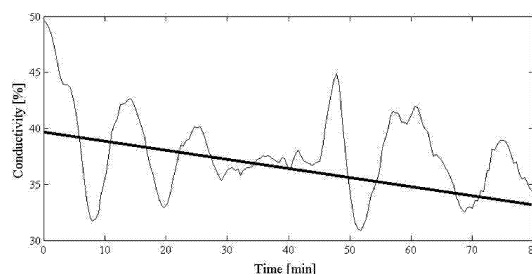


Figure 1: Conductivity Variations (setpoint in bold)

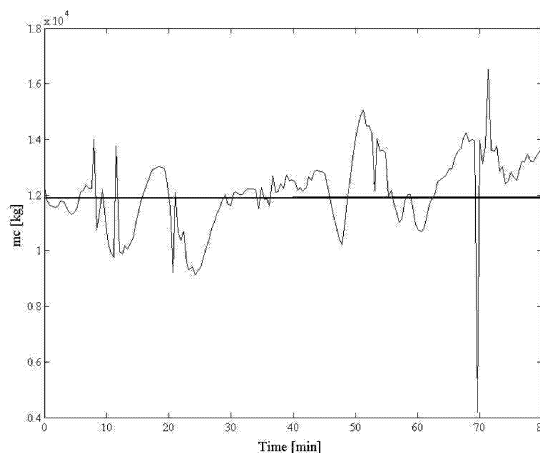


Figure 2: Mass of Crystal (rather constant)

Without mention of the effect of impurities on both measurements ( $\kappa$  and  $m_c$ , same figures) and performance, the usual control strategy based on a conductivity setpoint tracking has proved to be unsatisfactory, especially in terms of global productivity (time and energy loss).

As the crystallization performance is measured by the crystal growth, quantified by the increase in the mass of crystals,  $m_c$  appears to be a suitable performance criterion. Indeed,  $m_c$  inherently contains information about the mass balance and the crystal content ( $cc$ ). Better than the changes in  $\kappa$ , the variations in  $m_c$  tend to describe accurately the complex physical-chemical reactions occurring during the phase change. The problem is that  $m_c$  is not available on-line (Grondin-Perez et al. 2006).

These considerations lead to the concept of soft-sensors, first investigated a couple of years ago to control both supersaturation of liquors and crystal content ( $cc$ ) of massecuites, by measuring di-electric constant and electrical conductivity at radiofrequencies (Radford, 1985). Recently, on-line applications have been successfully implemented in cane sugar industry (Devogelaere et al. 2002; Bakir et al. 2005, Simoglou et al. 2005), which allows to consider alternative strategies for monitoring and process control of industrial crystallization processes.

The aim of this paper is to develop a reliable soft-sensor for on-line  $m_c$  measurement, applied to C crystallization. The rest of the paper is organized as follows. Section 2 presents the structure of the soft-sensor, based on 5 ordinary differential equations (ODE) to describe heat and mass transfer (HMT) with phase change. In section 3, validation is based upon industrial data. Simulation results show the reliability of this approach for achieving on-line monitoring of an industrial process.

## SOFT SENSOR FOR MASS OF CRYSTALS MEASUREMENT

This section considers the development of a soft-sensor of the mass of crystals  $m_c$ . A preliminary study dealt with the development of a tendency model, with a view to focus attention on the main step of the crystallization process: the growth phase (2.1). Ultimately, the model involves 5 ODE to describe the mass and energy balance (2.2).

### Focus on Crystal Growth from Impure Solutions

In cane sugar industry, to increase performance with minimum energy consumption, crystallization is achieved in three steps, named A-grade, B-grade and C-grade:

- crystallization of high purity liquors allows the extraction of A sugar from A molasses;
- crystallization of A molasses allows the extraction of B sugar from B molasses;
- and crystallization of B molasses allows the extraction of C sugar from C molasses, called blackstrap.

Since it is important to maximize the yield at each stage of the process, the fact remains that a key step in sugar recovery is C-grade crystallization (no more sugar can be extracted from blackstrap molasses in a cost-effective fashion). Consequently, global quality and productivity are very dependent on an efficient control of the C-grade crystallization.

The most widely approaches investigated to describe sugar extraction take into account three physical-chemical phenomena: nucleation, growth and dissolution. During the last 10 years, these investigations led to the development of models of A-grade processes (high purity liquors).

Little information is available about C-grade in the industrial context. There, low-grade sucrose solutions (impure solutions) contain a high proportion of impurities, which considerably changes their properties. It makes it difficult to adapt A-grade models to C-grade modeling. To overcome this difficulty, the most usual approach consists of considering nucleation and crystal agglomeration negligible in super-saturated conditions (Pautrat et al. 1997; Barth et al. 2006).

In other words, crystal growth is assumed to be preponderant, and this assumption underlies the development of a model dedicated to process control simulation.

### Nomenclature

#### Thermodynamic variables

|            |                                       |                    |
|------------|---------------------------------------|--------------------|
| $Bx^{\%}$  | Brix (mass fraction of dry substance) | %                  |
| $cc$       | crystal content                       | %                  |
| $Cp$       | specific heat capacity                | $J.kg^{-1}.K^{-1}$ |
| $F$        | flow rate                             | $m^3.s^{-1}$       |
| $h$        | specific enthalpy                     | $J.kg^{-1}$        |
| $L$        | specific latent heat                  | $J.kg^{-1}$        |
| $m$        | mass                                  | $kg$               |
| $\dot{m}$  | mass flow rate                        | $kg.s^{-1}$        |
| $Pte^{\%}$ | purity (mass fraction of sugar)       | %                  |
| $\dot{Q}$  | heating power                         | $W$                |
| $T$        | temperature                           | $^{\circ}C$        |
| $\dot{W}$  | stirring power                        | $W$                |
| $\rho$     | density                               | $kg.m^{-3}$        |
| $\alpha$   | adjusted parameter                    |                    |

#### Subscripts

|           |                                 |
|-----------|---------------------------------|
| $i$       | impurities                      |
| $c, Crys$ | crystal                         |
| $s$       | dissolved sucrose               |
| $w$       | water                           |
| $cw$      | condensate (from heating steam) |
| $f$       | feed                            |
| $hs$      | heating steam                   |
| $mc$      | mass of crystals                |
| $ml$      | mother liquor                   |
| $mg$      | magma: mother liquor & crystals |
| $vap$     | emitted vapor                   |

## Soft Sensor

Sugar crystallization consists in the exhaustion of a liquor to produce crystals. Extraction is induced by seeding the liquor in vacuum pans. During this operation, both the liquor and growing crystals are mixed together to obtain a homogeneous supersaturated magma. Super-saturation of the magma is obtained by vacuum evaporation. The three steps process is performed through a series of continuous and semi batch vacuum pans. At each step, several semi batch vacuum pans, operating as continuously stirred tank reactors (CSTR, Figure 3), are fed with different purity liquors.

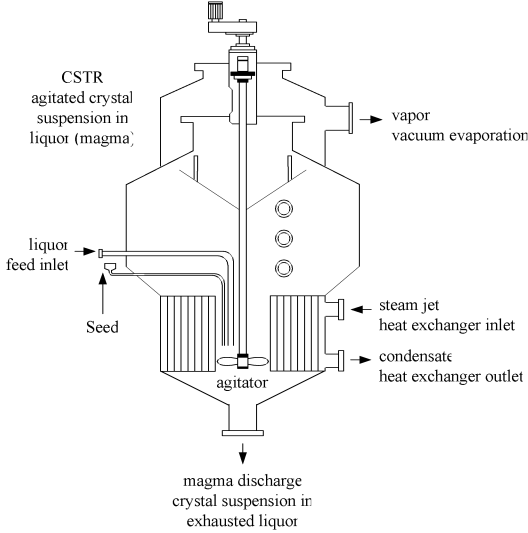


Figure 3: CSTR used for Sugar Crystallization

Because of the impact of many variables in terms of quality (liquors, molasses, heating steam, vacuum), it is important to develop a mathematical model that incorporates enough knowledge to describe the involved heat and mass transfer (HMT). A wide variety of mathematical models can be found in the scientific literature to describe these HMT, both steady state (Semlali Aouragh Hassani et al. 2001) and dynamic models (Feyo de Azevedo et al. 1993; Feyo de Azevedo et al. 1994). The main difficulty consists in the estimation of the crystallization rate, usually calculated by solving the population balance represented by moment equations (Hulburt and Katz 1964; Ramkrishna 1985).

Under the aforementioned assumptions, a dynamic model can be derived from five ordinary differential equations (Benne et al. 2008):

$$\frac{dm_w}{dt} = \rho_f F_f \left( 1 - \frac{Bx_f}{100} \right) + \rho_w F_w - \dot{m}_{vap} \quad (1)$$

$$\frac{dm_i}{dt} = \rho_f F_f \frac{Bx_f}{100} \left( 1 - \frac{Pte_f}{100} \right) \quad (2)$$

$$\frac{dm_s}{dt} = \rho_f F_f \frac{Bx_f}{100} \frac{Pte_f}{100} - \frac{dm_c}{dt} \quad (3)$$

$$\frac{dm_c}{dt} = cc \left( \rho_f F_f - \dot{m}_{vap} \right) + \alpha_{Crys} \quad (4)$$

$$\frac{dT_{mg}}{dt} = \frac{1}{m_{mg} Cp_{mg}} \left( \dot{W} + \dot{Q} + \rho_f F_f (h_f - h_{ml}) \dots \right. \\ \left. \dots + \rho_w F_w (h_w - h_{ml}) - \dot{m}_{vap} L_{vap} + \frac{dm_c}{dt} L_c \right) \quad (5)$$

with

$$\dot{m}_{vap} = \rho_{cw} F_{vap} : \text{mass flow rate of emitted vapor;}$$

$$\dot{Q} = \rho_{hs} F_{hs} L_{cw} : \text{heating power brought by heating steam condensation;}$$

and  $cc = \frac{m_c}{m_{mg}}$  : crystal content, ratio of  $m_c$  to the total mass.

## IDENTIFICATION AND VALIDATION

Taking into account the aforementioned hypothesis and further assumptions based on experimental considerations, two thermo-dynamic equilibrium relationships complete the description of HMT during the growth phase:

- the time delay between the heating steam ( $hs$ ) supply & the condensed water ( $cw$ ) outlet is inconsiderable (with regard to the process kinetics);  $\dot{Q}$  is calculated using the measured flow rate of the condensate:

$$F_{hs} = \alpha_Q F_{cw \leftarrow hs} \quad (6)$$

- to evaporate 1 kg of water about 1 kg of heating steam is needed; the evaporation capacity ( $F_{vap}$ ) is calculated using:

$$F_{vap} = \alpha_{vap} F_{cw \leftarrow hs} \quad (7)$$

The model identification has been performed using 4 databases randomly selected from one year industrial measurements. Three parameters had to be estimated to make the model fit the data:  $\alpha_{Crys}$ ,  $\alpha_Q$  &  $\alpha_{vap}$  (cf. equation (4),(6) and (7) above).

An iterative fitting procedure has been implemented in Matlab<sup>®</sup> environment. The algorithm is based on the minimization of a 3D quadratic criterion expressing the mean squared difference between simulated and experimental variables (lsqnonlin).

In the present approach, the optimization led to the following adjusted parameters:

$$\left[ \alpha_Q = 1.8177 ; \alpha_{vap} = 1.6455 ; \alpha_{Crys} = 0.9217 \right]$$

The validation step is performed using industrial data obtained using two on-line Brix measure-ments ( $Bx_{mg}^{\%}$  given by a microwave sensor and  $Bx_{ml}^{\%}$  given by a refractometer).

Prior to each run, the initial conditions for the state variables are determined from Brix, purity and level off-line measurements at time  $t = 0$  s. Figure 4 & Figure 5 show 2 examples of simulated versus measured results for  $m_c$  (database 1 & 2).

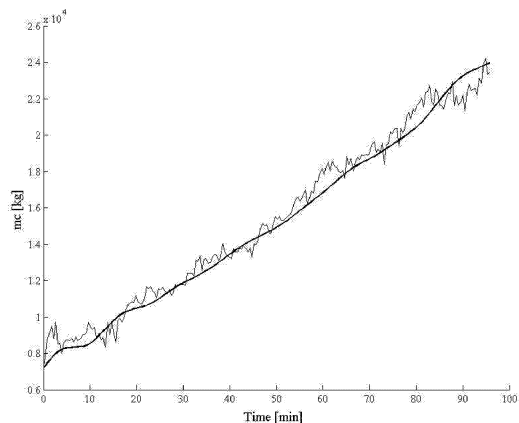


Figure 4: Simulated  $m_c$  (bold) compared with Industrial Measurements (database 1)

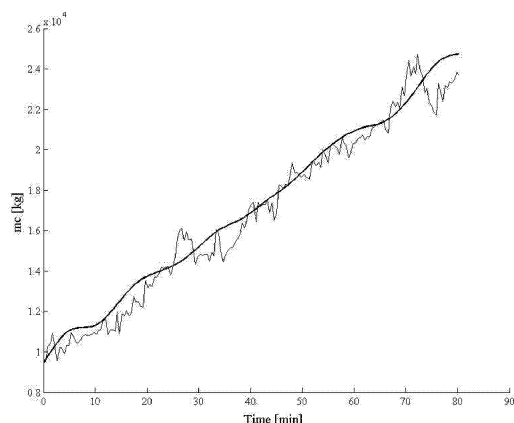


Figure 5: Simulated  $m_c$  (bold) compared with Industrial Measurements (database 2)

In comparison with industrial data, the simulator has proved to be able to calculate  $m_c$  with a relative error less than 4 %. The accuracy is sufficient for achieving soft-sensor objectives, without using on-line Brix measurements (given by microwave or refractometric sensors), the software being initialized with off-line measurements.

## CONCLUSION

In this paper, a soft-sensor for on-line measurement of the mass of crystals have been developed and applied to an industrial C crystallization process. Simulation results show the efficiency of this approach for achieving monitoring tasks. Associated with the widely used electrical conductivity  $\kappa$ ,  $m_c$  appears to be a suitable performance criterion, carrying additional information about the mass balance and the crystal content ( $cc$ ). In terms of process control, because of the large number of parameters and the multiple interactions between each other, the traditional strategies have proved to be unsatisfactory (Beyou 2008). These results and the relevant experimental observations allow drawing some perspectives about the use of  $m_c$  and  $cc$  for achieving alternative process control strategies.

## PROSPECT

In manufacturing industries equipped with a supervision and control system (SCS), computer-aided production management is carried out using a set of peripheral devices. In this study, the SCS is the MODUMAT 8000® (BAILEY-ABB®): the whole experimental data are carried round the production line through MEDIAMAT networks.

In 1997, a data server named I-MEDIA has been developed in the lab (Lorion 2001), so as to interface peripheral devices and a personal computer (PC) holding monitoring and control application tasks. Recently, an updated program has been developed to handle communication with Matlab® routines. The implementation of the model-based soft sensor *in situ*, using I-MEDIA, is planned for 2009.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the help of Mr. Jean-Claude Pony, Dir. of the BR sugar mill ([www.bois-rouge.fr](http://www.bois-rouge.fr)), and its technical staff, and Mr. Jean•Paul Dijoux, networks engineer ([www.cerf.fr](http://www.cerf.fr)).

## REFERENCES

- Bakir T., S. Othman, F. Puel & H. Hammouri, 2005, *Continuous-discrete observer for crystal size distribution of batch crystallization process*, 44<sup>th</sup> IEEE Conf. on Decision and Control 2005 & 2005 European Control Conf.), 12-15 Dec. 2005, pp. 6240-6244
- Barth S., 2006, *Utilization of FBRM in the Control of CSD in a Batch Cooled Crystallizer*, PhD thesis, Georgia Institute of Technology, 121 p.
- Beyou S., 2008, *Proposition d'un algorithme PID à paramètres variables. Pour l'amélioration de la conduite d'un procédé de cristallisation industriel*, PhD thesis, University of La Réunion (EA 4079), 181 p.
- Benne M., B. Grondin-Perez & J.-P. Chabriat, 2008, *Estimation of two parameters to fit a tendency model for dynamic simulation of an industrial crystallization process*, FoodSim 08, 26-28 June 2008, University College Dublin, Ireland
- Devogelaere D., M. Rijckaert, O.G. Leon & G.C. Lemus, 2002, *Application of feed forward neural networks for soft sensors in the sugar industry*, VII<sup>th</sup> Brazilian Symp. on Neural Networks, pp. 2-6
- Feyo de Azevedo S., J. Chorão, M. J. Gonçalves & L. Bento, 1993, *On-line monitoring of white sugar crystallization through software sensors. Part I*, International Sugar J., 95, pp. 483-488
- Feyo de Azevedo S., J. Chorão, M. J. Gonçalves & L. Bento, 1994, *On-line monitoring of white sugar crystallization through software sensors. Part II*, International Sugar J., 96, pp. 18-26
- Grondin-Perez B., M. Benne, C. Bonnetaze & J.-P. Chabriat, 2005, *Industrial multi-step forward predictor of mother liquor purity of the final stage of a cane sugar crystallization plant*, J. of Food Engineering, 66, pp. 361-367
- Grondin-Perez B., M. Benne & J.-P. Chabriat, 2006, *Supervision of C crystallization in Bois Rouge sugar mill using on-line crystal content estimation using*

- synchronous microwave and refractometric Brix measurements*, J. of Food Engineering, 76, pp. 639-645
- Hulburt H. M. & S. Katz, 1964, *Some problems in particle technology. A statistical mechanical formulation*, Chemical Engineering Science, Vol. 19, pp. 555-574
- Lorion R., L. Chane-Kuang-Sang, M. Benne & J.-P. Chabriat, 2001, *Industrial data server development to implement neural network software sensors*, MOSIM '01, Troyes, 25-27 avril 2001), ISBN 1-56555-212-1
- Pautrat C., J. Génotelle & M. Mathlouthi, 1997, *Sucrose crystal growth: effect of supersaturation, size and macromolecular impurities*, in Sucrose crystallization, science and technology, VanHook et al., Bartens, pp. 57-70
- Radford D. J., 1985, *EP0162580: massecuite supersaturation monitor*, TONGAAT HULETT GROUP LIMITED (ZA), Nov. 1985  
<http://www.freepatentsonline.com/EP0162580A1.html>
- Ramkrishna D., 1985, *The status of population balances*, Reviews in chemical engineering, 3 (1), pp. 49-95
- Semlali Aouragh Hassani N., K. Saidi & T. Bounahmidi, 2001, *Steady state modeling and simulation of an industrial sugar continuous crystallizer*, Computers & Chemical Engineering, Vol. 25, Issues 9-10, pp. 1351-1370
- Simoglou A., P. Georgieva, E. B. Martin, A. J. Morris & S. Feyo de Azevedo, 2005, *On-line monitoring of a sugar crystallization process*, Computers & Chemical Engineering, Vol. 29, Issue 6, pp. 1411-1422

# DESIGN GENERAL FUZZY CONTROLLER IMPLMENTED IN VHDL AND SYNTHESIZED USING FPGA

Rasha E. Majed  
Basil Sh. Mahmood  
Department of Computer and Information Engineering  
University of Mosul  
Mosul  
Iraq  
[rasha\\_elham@yahoo.com](mailto:rasha_elham@yahoo.com), [basil\\_mahmood@yahoo.com](mailto:basil_mahmood@yahoo.com)

**KEYWORDS:** Fuzzy Controller, FPGA, JTAG.

## ABSTRACT

In this paper, a fuzzy control system has been designed and implemented. It consists of three main parts: The first and last parts are represented by ADC and DAC respectively. The middle part represents a fuzzy controller which consists of many stages. The system utilizes FPGA chip of Spartan 3E kit. It can connect up to three inputs to control up to two independent outputs through a set of fuzzy rules that are stored in a Block RAM. The system can be programmed and set up by external computer connected to it through JTAG port. The system is based on Mamdani controller type which is adopted to implement various fuzzy controllers. The system software is divided into two parts: (1) the FPGA software that was written in VHDL to decode each component in fuzzy systems, then, synthesize it using Project Navigator. (2) PC software that was built using Visual Basic which supports GUI design. The PC software provides facilities like input-output MF design, fuzzy rule table decoder, and downloading fuzzy parameters to target hardware through JTAG port. The designed system was implemented and tested by considering a fuzzy controller which connects two inputs to control one output. Then, comparing the practical results which were obtained using ModelSim XE-III 6.0a with the theoretical results in addition to the results of the testing system are done. The response time of the designed fuzzy controller is estimated to be 0.48  $\mu$ s.

## INTRODUCTION

Fuzzy Logic was considered as one of the controlling methods on the complex systems, is based on linguistic variables and fuzzy sets which are described using membership functions (Xilan Tian et al. 2007). So, fuzzy

controller was represented as one of the applications of the fuzzy logic, its outputs were found using its input values depending on the fuzzy rules. Fuzzy logic controllers are proposed to control plants that are mathematically poorly understood are where experienced human operators are available for providing qualitative rules of control. Fuzzy controllers make effective use of this information and provide a linguistic description about the system and control instructions. This method is a model-free approach and provides nonlinear controllers to perform any nonlinear control actions (Hugang Han. 2008). Also from a practical point of view, a fuzzy control is easy to understand, inexpensive to develop and simple to implement. Now there are many methods to implement a fuzzy controller. The more usual implementation of fuzzy controllers is the simulation on a commercial microprocessor (Leong P.H.W and Tsoi K.H. 2005). This solution is inexpensive but in some real time problems is not useful. Many fuzzy VLSI chips have been developed to implement fuzzy control. Although due to cost reason most of them have used a serial implementation of rules and is not possible to use in high speed applications with low cost, in addition to no changes can be made once the chip is created. FPGAs (Field Programmable Gate Arrays) are general purpose computing devices which can be used to reduce board size and accelerate applications in almost any domain. As they continue to improve in density, performance, cost and ease of usage (Masmoudi N. et al. 1999). Fuzzy control system has been designed and implemented on an FPGA using VHDL. It consists of three main parts: The first is **ADC** and the third **DAC** respectively. And, the middle part is a **Fuzzy controller** which consists of set of pipelined stages **Fuzzification** stage transforms the input in degree of membership with linguistic values. The **Inference Engine** stage performs the inference operations on the **Rule Base** (Jantzen Jan. 1998). Finally, a **Defuzzification** stage transforms the fuzzy result of the inference process in a crisp output as shown in figure (1).

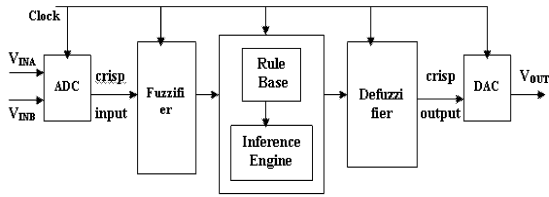


Figure 1: Block Diagram of Fuzzy Logic System Unit

The system utilizes FPGA chip at Spartan 3E kit. It can connect up to three inputs to control up to two independent outputs through set of fuzzy rules which are stored in a Block RAM (©2005Xilinx, Inc.). The system has JTAG port (©2004 Xilinx, Inc.s), which connects it with personal computer that is used for programming and setting up the system using Visual Basic that supports GUI design. The system was based on Mamdani controller type to implement various fuzzy controllers. This paper also presents the hardware implementation of FLC to control a simple plant. The output response of the closed-loop system is obtained from connecting sound card to FPGA, and then connects the latter to PC using Matlab.

## FUZZY LOGIC SYSTEM UNIT AND THE IMPLEMENTATION OF EACH UNIT

The input to the controller is scaled to the analog input values in the Spartan 3E ADC domain [0.4, 2.9]v and then crisp output of the defuzzification block is scaled back to [0, 3.3] v range of Spartan 3E DAC. The implementation specifications of the FLC in this paper are: no. of input variables=2, no. of output variable=1, no. of bits per input and output variable =12 (depending on the type of ADC and DAC), no. of bits for slope of membership function=12. The inputs to the FLC are crisp and for each input variable, the overlapping degree of its membership function is at most two. Membership functions in the input and output variable have the triangular shape.

### ADC

The first component in the fuzzy logic system is the Spartan 3E ADC that converts the analog voltage on  $V_{INA}$  or  $V_{INB}$  to 14-bit digital representation,  $D[13:0]$  as expressed by Equation (1)(© 2006 Xilinx, Inc. 2006):

$$D[13:0] = \text{gain} \times \frac{(V_{in} - 1.65 \text{ v})}{1.25 \text{ v}} \times 8192 \quad (1)$$

The gain is the current setting loaded into the programmable preamplifier. The reference voltage ( $V_{ref}$ ) for Spartan 3E ADC is 1.65v. Consequently,  $V_{ref}$  is subtracted from the input voltage on  $V_{INA}$ ,  $V_{INB}$ . The maximum range of the ADC is  $\pm 1.25 \text{ v}$ , centered on  $V_{ref}$ . Hence, 1.25v appears in the denominator to scale the analog input accordingly. Finally, the ADC presents a 14-bit two's complement digital

output. A 14-bit two's complement number represents values between  $-2^{13}$  and  $2^{13}-1$ . Therefore the quantity is scale by 8192 or  $2^{13}$ . Through applying Equation (1) above at the two input voltages, we obtain the digital values:

$$\begin{aligned} \text{Input}_1 (13:0) &= -1 \times ((1.63764 - 1.65/1.25) \times 8192) = 81_d = 0051_h \\ \text{Input}_2 (13:0) &= -1 \times ((1.651,65/1.25) \times 8192) = 0_d = 0000_h \end{aligned}$$

Then, subtract input2 from input1 to find the value of error, after delay subtract error from previous error to find the value of change of error as shown is Figure (2). The Figure (3), (4) are for the timing diagram of error and change of error.

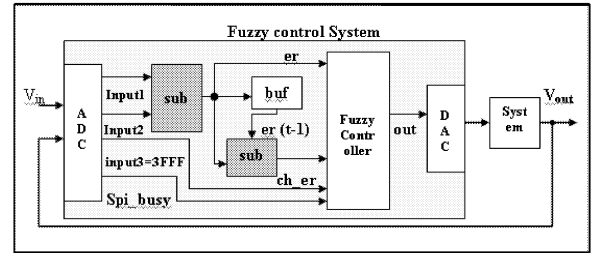


Figure 2: Block Diagram of Fuzzy System for finding error,  $ch\_error$

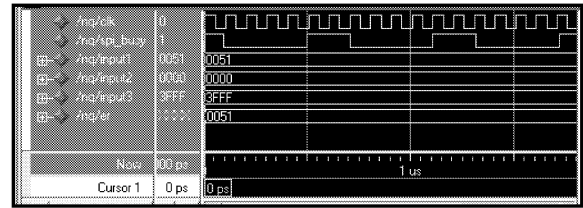


Figure 3: Timing diagram of error

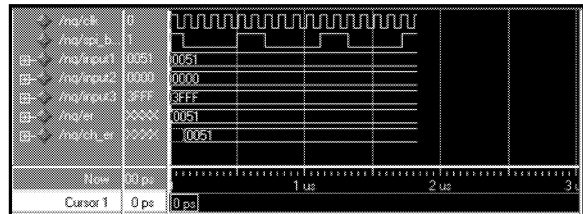


Figure 4: Timing diagram  $Ch\_error$

### Fuzzification

The first stage in the FLC is the fuzzifier that transforms crisp inputs into a set of membership values up the certain interval [1.625, 1.675] v in the corresponding fuzzy sets. The membership function shapes are typically triangular, trapezoidal or exponentially. In this paper, triangular shapes are used as shown in figure (5) for error and change of error with 12-bit interval  $[7D7, 829]_H$  and with 3-bit fuzzy labels Negative, Zero, and Positive for each inputs. **Input Multiplexer Module** is used to multiply inputs (After rounding ADC output to 12-bit and adding the result with  $2000_H$ ), no. of membership function for the input signal e, and its address in memory is selected. Then in the next step, Negative and Zero fuzzy subsets are activated and the

membership values for the two fuzzy subsets are formal within the region depending on the point at which the input line cut the two slopes. The 4 points of Negative and Zero subsets, and the positive slope of each one are stored in the Block RAM. The input is compared along with the points stored in the memory until it becomes less than the second point of the zero fuzzy subset. Then, the input is subtracted from the point of Negative fuzzy subset and multiplied with the positive slope in the **Fuzzification Module**. This gives the membership degree for zero subset and the degree for Negative fuzzy set is just the complement membership degree of Zero as the sum of all membership values is always equal 1. Then, the same operation on the signal ce is

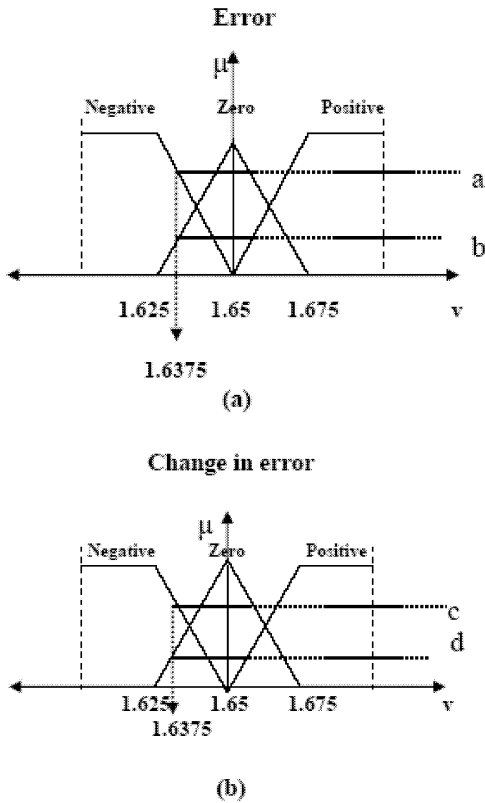


Figure 5: Triangular membership functions

repeated each after one clock pulse (Sequentially). The last step is the **Fuzzification Buffer** which is used to store the degrees and labels of the inputs in a buffer, and reads all of them at the same time. Figure (6) shows the block diagram of the fuzzification (Singh Sameep, et al. 2003). In this paper, we consider e, f equal to "FFFFFF"<sub>H</sub>, and their labels are equal to "111"<sub>B</sub> since input3 is not used. Utilizing pipelining, we used the handshaking signals (load, flag) to control on the system. Figure (7), (8), (9) show the state diagrams of Input Mux Module, Fuzzification, and fuzzifier buffer.

## Fuzzy Inference

The knowledge base of the fuzzy controller which is represented by the Block RAM consists of a data base of linguistic statements (rules), which states the relationship between the input domain fuzzy sets and output domain fuzzy sets. For a system with two inputs, error(e) and change of error(ce) and single output each having three fuzzy sets, the rules can be represented in tabular form as shown in table(1). In this paper, we stored these linguistic labels into 3-bit labels as shown in table (2). A maximum of 4 rules can be active at any time with triangular membership functions for e and ce ; these 4 rules will be repeated because the third input is not used. Figure (10) shows the components of the fuzzy inference. Linguistic labels and input degrees are assigned to the three membership degrees and labels on a clock coming from the fuzzifier for each input variable by **Linguistic Label Module** (depending on Fuzzy State Machine).

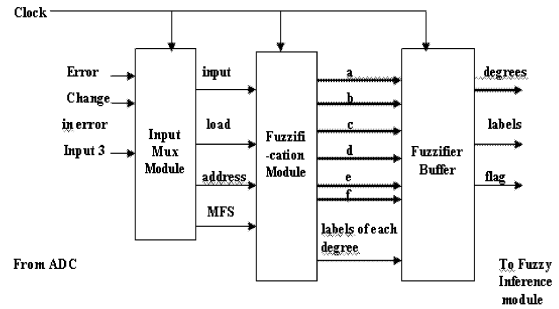


Figure 6: Block Diagram of Fuzzification

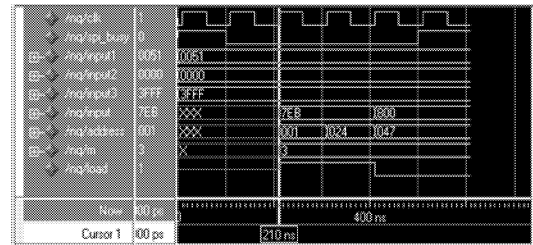


Figure 7: Timing diagram of Multiplexer Module

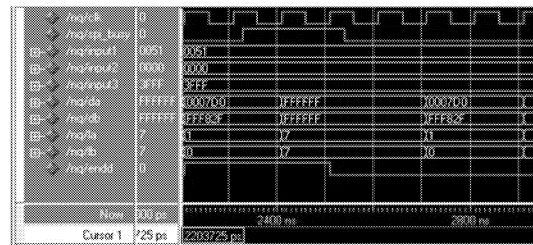
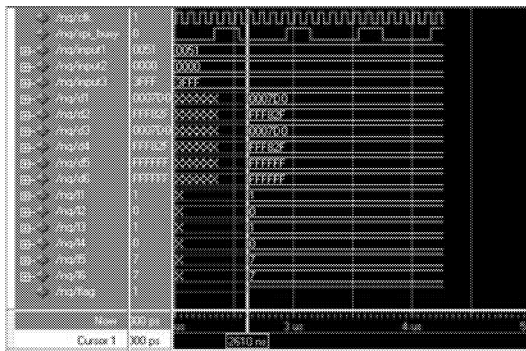


Figure 8: Timing diagram of Fuzzifier Module





The next step is the **Minimum Module** which compares the degrees of the antecedent of each active rule and outputs the minimum membership degree and the input labels on a clock. Then, finding the output labels depending on the input labels by comparing the input labels with all rules stored in the memory on a clock using **Rule Module**. The last step is to

Table 1: Fuzzy rule matrix

| Ch <sub>er</sub> \ er | Negative | Zero      | Positive |
|-----------------------|----------|-----------|----------|
| Negative              | Heating  | Cooling   | Cooling  |
| Zero                  | Heating  | No Change | Cooling  |
| Positive              | Heating  | Heating   | Cooling  |

Table 2: Binary fuzzy rule matrix

| er<br>Ch_er | "000" | "001" | "010" |
|-------------|-------|-------|-------|
| "000"       | "000" | "010" | "010" |
| "001"       | "000" | "001" | "010" |
| "010"       | "000" | "000" | "010" |

find the fuzzy output by using the minimum degree and the output labels at a clock. The **Fuzzy Output Module** divides the minimum degree by positive slope of output fuzzy set. Then, adds the results with the first point of the fuzzy set. The division is implemented using pipelined divider V8.0 (© 2004 Xilinx, Inc. 2005). This whole process is executed sequentially for all the repeated active rules depending on handshaking signals (as ready1, ready2). We used defuzzifier buffer to store the outputs of fuzzy inference and find acc, clear signals which are used in the next step. The figures (11-16) show the state diagrams of each component in the inference engine.

## Defuzzification

The last step is the defuzzification and the final output is determined by the Center Of Gravity method of all contributions of the output sets i.e. the output can be obtained by summing the product of fuzzy output with the degree from defuzzification buffer divided by the sum of all

the degrees of the previous step. The final crisp output is realized by the Equation (2) (Hugang Han. 2008):

$$W = \frac{(\text{Neg} \times x_1 + \text{Ze} \times x_2 + \text{Ze} \times x_3 + \text{Pos} \times x_4) \times 2}{(x_1 + x_2 + x_3 + x_4) \times 2} \quad (2)$$

Where  $x_1, x_2, x_3, x_4$  represent the minimum degrees for error and change in error and the terms Neg, Ze, Pos are the fuzzy outputs respectively. As shown in figure (17), the inputs to the defuzzifier are the degree, fuzzy output with acc and the clear signals from the previous step. As the **Multiplier** (© 2006 Xilinx, Inc. 2006) multiplier the degree with the fuzzy output, the **Accumulator** inside it accumulates the numerator and denominator as long as it stay active. The number of clock cycles the accumulator stays active is equal to the number of active rules. During this time it does the multiplication of the degree with the fuzzy output of rule along with the addition to produce the numerator. Also, for the denominator, the degree of each antecedent is added. Both the numerator and denominator are cleared to zero by the clear signal to start a new fuzzy logic process. The **Division** is implemented using the pipelined divider V8.0 (© 2004 Xilinx, Inc. 2005). Figures (18-21) show the state diagram of each component in the defuzzification module.

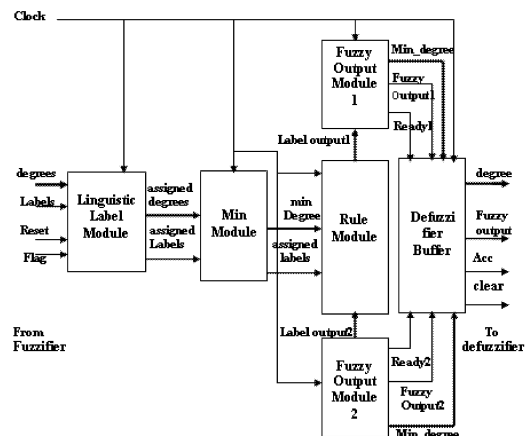


Figure 10: Block Diagram of Fuzzy Inference

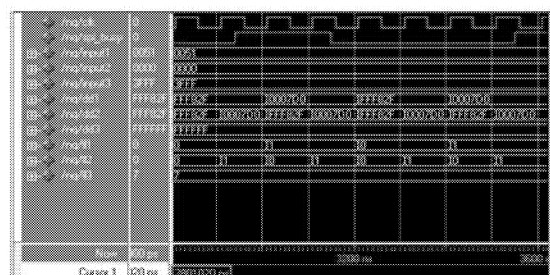


Figure 11: Timing diagram of Linguistic label Module

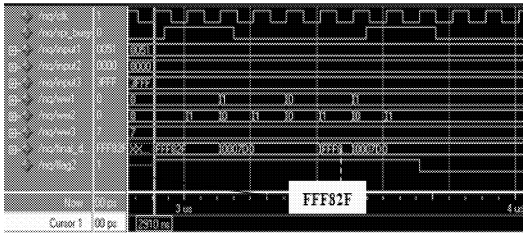


Figure 12: Timing diagram of Minimum Module

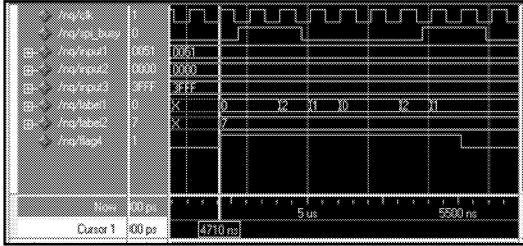


Figure 13: Timing diagram of Rule Module

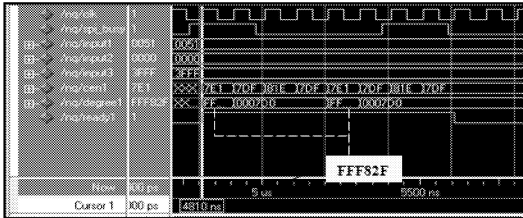


Figure 14: Timing diagram of Fuzzy Output Module 1

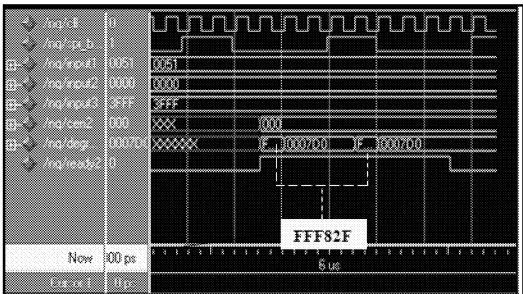


Figure 15: Timing diagram of Fuzzy Output Module 2

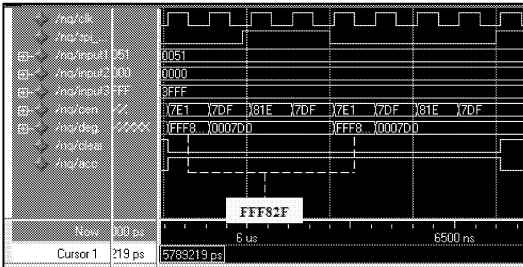


Figure 16: Timing diagram of Defuzzifier Buffer

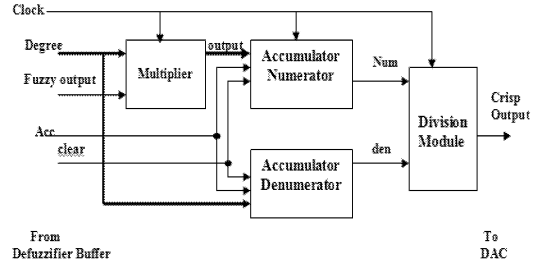


Figure 17: Block Diagram of Defuzzifier

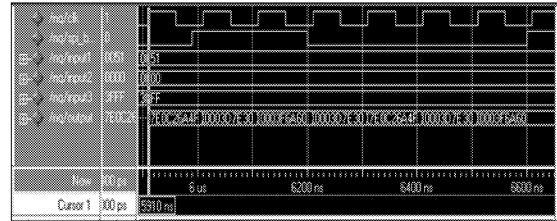


Figure 18: Timing diagram of Multiplier Module

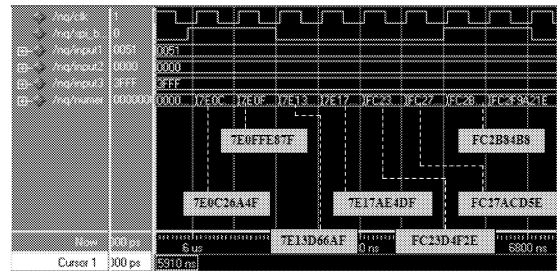


Figure 19: Timing diagram of Accumulator Numerator Module

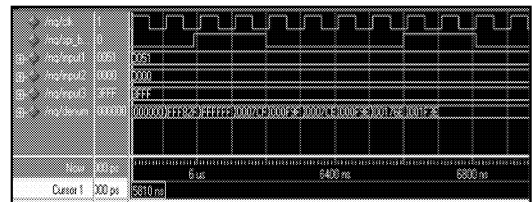


Figure 20: Timing diagram of Accumulator Denominator Module

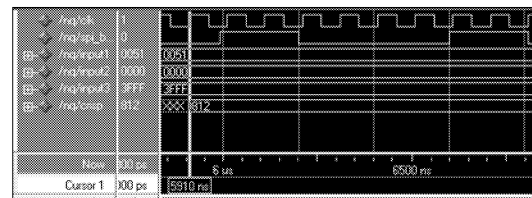


Figure 21: Crisp output of the defuzzifier

## DAC

The last component in the fuzzy logic system is the spartan 3E DAC that converts 12-bit digital output to analog output as expressed in Equation (3) (© 2006 Xilinx, Inc. 2006).

$$V_{out} = \frac{\text{Digital output}}{4096} \times V_{ref} \quad (3)$$

The reference voltage ( $V_{ref}$ ) for spartan 3E DAC is 3.3v. A 12-bit number represents values between 0 and  $2^{12}$ . Therefore the quantity is scaled by 4096 or  $2^{12}$ . Through applying Equation (3) above at digital output (convert output of defuzzifier 813<sub>H</sub> to 2067<sub>D</sub>), we obtain the analog value,

$$V_{out} = \frac{2067}{4096} \times 3.3 = 1.6653 \text{ v}$$

Practically, the results were checked on computer by simulating the controller on Matlab software, and these results were compared with the real values that are coming out from the controller as shown in table (3).

Table 3: Comparison between the results in Matlab and in FPGA

| index | Input1 | Input2 | Practical results | Theoretical results | Matlab results |
|-------|--------|--------|-------------------|---------------------|----------------|
| 1     | 1.65   | 1.625  | 1.679             | 1.67                | 1.68           |
| 2     | 1.65   | 1.65   | 1.66              | 1.65                | 1.65           |
| 3     | 1.675  | 1.675  | 1.709             | 1.7                 | 1.68           |
| 4     | 1.63   | 1.63   | 1.633             | 1.625               | 1.63           |
| 5     | 1.675  | 1.675  | 1.709             | 1.7                 | 1.68           |
| 6     | 1.675  | 1.65   | 1.71              | 1.701               | 1.68           |
| 7     | 1.662  | 1.662  | 1.673             | 1.665               | 1.65           |
| 8     | 1.637  | 1.637  | 1.673             | 1.6653              | 1.66           |
| 9     | 1.662  | 1.637  | 1.673             | 1.665               | 1.66           |
| 10    | 1.63   | 1.625  | 1.642             | 1.634               | 1.63           |
| 11    | 1.63   | 1.662  | 1.673             | 1.665               | 1.64           |

## TESTING SYSTEM

The FLC designed using VHDL is tested as a PFLC with equally spaced three fuzzy sets for a closed-loop control system. Since the maximum speed of the design was 63.680MHz, the clock speed of the FPGA was set at 50 MHz. The output scaling introduces a gain of 1 in the system as shown in figure (22).

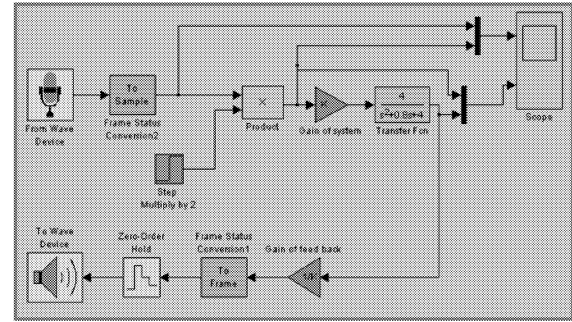


Figure 22: Closed Loop Control System

The figure (23) shows the input signal before and after multiplying it by 2 because the sound card uses the signals in the range (0-1) V. Figure (24) shows the output response of the system when  $W_n=2$ ,  $\zeta=0.2$ ,  $k=1$ . Figure (25) shows the output response of the system when  $W_n=20$ ,  $\zeta=0.2$ ,  $k=1$ . Figure (26) shows the output response of the system when  $W_n=2$ ,  $\zeta=0.2$ ,  $k=5$ . As shown in table (4), there is no largely differentiation between the practical properties of the response and its theoretical properties which were computed by equations.

## CONCLUSIONS

This paper presents an approach for the implementation of a fuzzy logic controller on an FPGA using VHDL. The FLC system is implemented on a Xilinx Spartan 3E FPGA and used to control a plant to demonstrate its validity and it can also be used for control purposes in other applications. The design of the FLC is highly flexible as the membership functions and rule base can be easily changed through setting programs. The characteristics of the design which were obtained from ISE 8.1 report are shown in table (5) (Rasha E. 2007). Through EQ (3), we computed executing time of the controller,

$$T_F = (P+N) \times T_{clk} \quad (3)$$

$$T_F = (15+9) \times (0.2 \mu) = 0.48 \mu\text{sec}.$$

Where N is no. of fuzzy rules and P is total no. of pulses to find the result of each component in the controller.

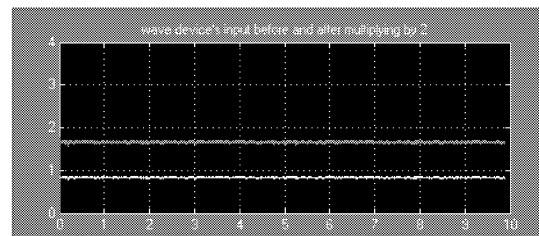


Figure 23: Step input signal

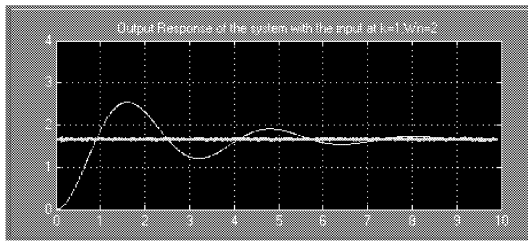


Figure 24: Output response at  $W_n=2$ ,  $k=1$

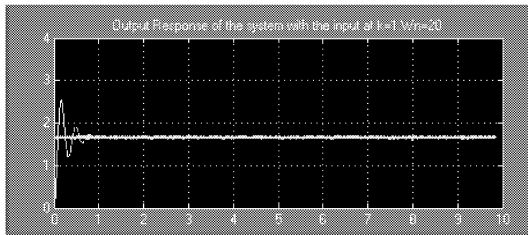


Figure 25: Output response at  $W_n=20$ ,  $k=1$

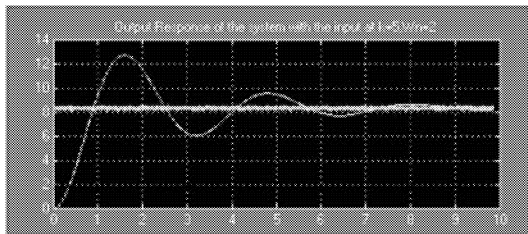


Figure 26: output response at  $W_n=2$ ,  $k=5$

Table 4: System Properties

## References

- Xilan Tian et al. 2007 "A Self-tuning Fuzzy Controller for Networked Control System" IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.1.
- Hugang Han. 2008 "FUZZY CONTROLLER DESIGN WITH INPUT SATURATION" International Journal of Innovative Computing, Information and Control

Table 4: System Properties

| System Properties | Theoretical Equations                       | Results                 |                    |                           |                    |
|-------------------|---|-------------------------|--------------------|---------------------------|--------------------|
|                   |   | First Result<br>$W_n=2$ |                    | Second Result<br>$W_n=20$ |                    |
|                   |   | Practical Result        | Theoretical Result | Practical Result          | Theoretical Result |
| Overshoot         | $V_{in} + e^{(-\pi\zeta/\sqrt{1-\zeta^2})}$ | 2.5<br>volt             | 2.19<br>volt       | 2.5<br>volt               | 2.19<br>volt       |
| Overshoot time    | $\frac{\pi}{(W_n\sqrt{1-\zeta^2})}$         | 1.6<br>Sec              | 1.602<br>Sec       | 0.1<br>Sec                | 0.160<br>Sec       |
| Settling time     | $\frac{4}{W_n\zeta}$                        | 9.5<br>Sec              | 10<br>Sec          | 0.95<br>Sec               | 1<br>Sec           |
| Rise time         | $\frac{0.8+2.5\zeta}{W_n}$                  | 0.6<br>Sec              | 0.64<br>Sec        | 0.07<br>Sec               | 0.064<br>Sec       |

Table 5: The practical properties of the designed Controller

|  |     |           |
|--|-----|-----------|
| Number of Slices                         | 94% |           |
| Number of 4 input LUTs                   | 53% |           |
| Number of bonded IOBs                    |     | 56        |
| Number of IOs                            | 24% |           |
| Number of GCLKs                          | 4%  |           |
| Maximum Frequency                        |     | 63.680MHz |
| Maximum output required time after clock |     | 4.364ns   |

ICIC International c°2008 ISSN 1349-4198 Volume 4, number 10, October pp. 2507.

- Leong P.H.W and Tsoi K.H. 2005 "Field Programmable Gate Array Technology for robotics Applications". IEEE International Conference on Robotics and Biomimetics.
- Masmoudi N. et al. 1999 "Hardware design of programmable Fuzzy Controller on FPGA". pp. 1675-1679; ISBN: 0-7803-5406-0, ©1999 IEEE.
- Jantzen Jan. 1998. "Design Of Fuzzy Controllers". Publisher Technical University of Denmark: Dept. of Automation, Publ. No98-E 864.
- ©2005Xilinx, Inc. 2005 "Using Block RAM in Spartan-3 Generation FPGAs". XAPP463 (v2.0), [www.xilinx.com](http://www.xilinx.com).
- © 2004 Xilinx, Inc. 2004 "JTAG Loader- Quick Guide-11". [http://www.Support.Xilinx.com/xlnx/xweb/xil-lx\\_home.jsp](http://www.Support.Xilinx.com/xlnx/xweb/xil-lx_home.jsp)
- © 2006 Xilinx, Inc. 2006 "Spartan 3E Starter Kit Board User Guide". UG230 (v1.0). [www.xilinx.com](http://www.xilinx.com).
- Singh Sameep, et al. 2003 "Implementation of a Fuzzy logic Controller on an FPGA using VHDL". 0-7803-7918-7/03 @ 2003 IEEE.
- © 2006 Xilinx, Inc. 2006 "Multiplier V8.0", DS255. [www.xilinx.com](http://www.xilinx.com)
- © 2004 Xilinx, Inc. 2005 "Pipelined Divider V3.0 "; DS305. Product specification 1-800-255-7778. [www.xilinx.com](http://www.xilinx.com)
- ©2005Xilinx, Inc. 2005 "Using Block RAM in Spartan-3 Generation FPGAs". XAPP463 (v2.0), [www.xilinx.com](http://www.xilinx.com).
- Rasha E. 2007 "Design General Fuzzy Controller, Implemented in VHDL and Synthesized using FPGA", M.Sc. thesis.

# Adjustment of conformity parameters of PID-type compensators using simulation by AMESim

Teodor Costinel Popescu  
Iulian Dutu  
INOE 2000-IHP,  
Bucharest, Romania,  
E-mail: popescu.ihp@fluidas.ro  
E-mail: dutu.ihp@fluidas.ro

Cătălin Vasiliu  
Marius Mitroi  
“POLITEHNICA” University  
of Bucharest, Romania,  
E-mail: vasiliu@fluid-power.pub.ro  
E-mail: mariusm@amcsit.ro

**KEYWORDS:** modeling, simulation, AMESim, PID compensator tuning, ground leveling machine

## ABSTRACT

The AMESim modeling and simulation environment makes it possible to adjust conformity parameters of PID-type (proportional, integrative, derivative) compensators which are characteristic for hydraulic servo mechanisms. This paper presents an example of tuning by simulation the following parameters:  $K_p$  (gain),  $K_i$  (integral time constant), and  $K_d$  (derivative time constant) which are characteristic of a PID compensator, the equivalent of an automatic regulator of a servo mechanism that adjusts position with laser feedback. This kind of servo mechanisms allows an automatic operating of the ground leveling machines, tracing an optical reference plane. The adjustment of the three parameters mentioned above is performed in order to optimize the automatic tracing system performances as precision and stability. The simulation model developed in AMESim contains two hydraulic servomotors for position adjustment: the first one for simulating the uneven ground, and the second one for tracing (similar to the one that is mounted on a ground leveling machine). Displacement in time of the first servo cylinder represents the profile of the uneven ground, while displacement of the second one stands for the vertical displacement of the blade of a ground leveling machine. The amount of displacement values of the two servo cylinders represents the time variation of the ground deviation from the optical reference plane.

## INTRODUCTION

The modern fluid control systems are using hybrid tuning algorithms as Fuzzy - PID error compensators. The high degree of nonlinearity of these systems leads to the wide use of modeling and simulation techniques for obtaining the tuning parameters by a virtual testing system. This testing manner offers a strong costs cut, and a useful reduction of the real experimental test. This paper is devoted to the use of a modern modeling and simulation engineering software created by Michel Lebrun and Claude Richards in 1989 for high speed solving a practical problem. After 20 years of intensive development of the symbolic libraries in different engineering fields, AMESim became an efficient tool for solving different applications of the fluid control systems. The case presented in this paper intends to offer a model of developing new applications of the electro hydraulic systems by this tool. The authors created both the laboratory model of the electro hydraulic control system, and the real system set up on a modern ground leveling machine. The comparison between the static and dynamic performances of the real system is found in good agreement.

## PROBLEM FORMULATION

To adjust a regulator is to determine the parameters of an imposed adjustment structure, of a settled degree, so that to achieve from the resulted system a behavior as close as possible to the desired one. In practice the most frequently used regulators are of type P, PI, PD and PID which calculate the  $u(t)$  command according to the following relations: (1), for a **P**: regulator: proportional; (2), for a **PI**: regulator proportional, integral; (3), for a **PD**: regulator proportional, derivative; (4), for a **PID** regulator proportional, integral, derivative, where:  $K_p$  – constant of the proportional part (gain),  $K_i$  – constant of the integral part,  $K_d$  – constant of the derivative part.

$$u(t) = K_p \cdot \varepsilon(t) \quad (1),$$

$$u(t) = K_p \cdot \varepsilon(t) + K_i \cdot \int \varepsilon(t) dt \quad (2),$$

$$u(t) = K_p \cdot \varepsilon(t) + K_d \frac{d\varepsilon(t)}{dt} \quad (3),$$

$$u(t) = K_p \cdot \varepsilon(t) + K_i \cdot \int \varepsilon(t) dt + K_d \frac{d\varepsilon(t)}{dt} \quad (4),$$

**PID** type regulators are used for the error signal in hydraulic rapid servomechanisms. Component **P** amplifies the error, develops a higher-speed system, but it can't cancel the stationary error; component **I** removes the stationary error, but it destabilizes the system, while component **D** stabilizes the system (VASILIU et al.).

The simulation model in AMESim (figure 1) represents a hydraulic servomechanism for position control with one external feedback by laser and two internal feedbacks, arising at the level of the two included servomechanisms, as follows. The upper servomechanism, that simulates the profile of the uneven land, and the lower servomechanism, a tracing one, that actuates the blade of the levelling machine in a vertical plane.

The hydraulic servo cylinder of the upper servomechanism has a mobile body, while the one of the lower servomechanism has a fix body. The first internal feedback loop arises between the displacement transducer of the cylinder with mobile body and the upper comparator of the simulation model. The second internal feedback loop arises between the displacement transducer of the cylinder with fix body and the internal comparator of the simulation model. The external feedback loop arises between the displacement transducer placed in the upper side of the model and the comparator placed in its lower side.

The above configuration can be a fair representation of the true system, included in the frame of the levelling machine.

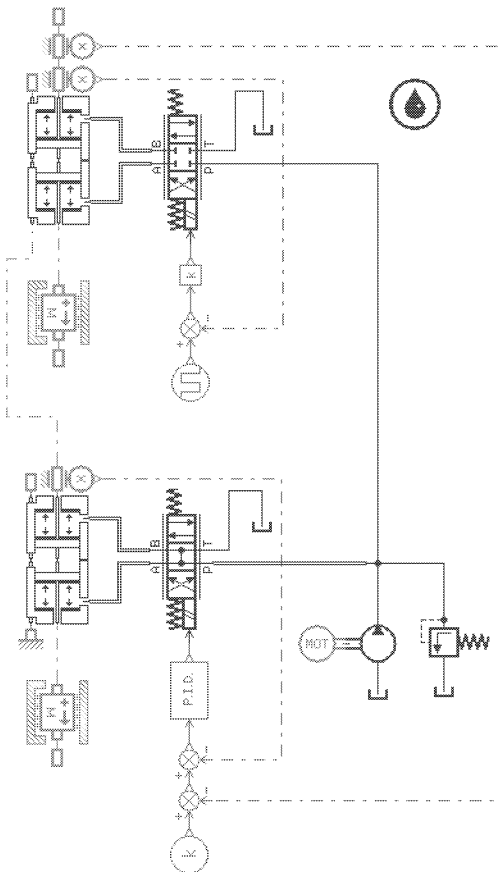


Figure 1. Simulation network in AMESim.

## ADJUSTING THE PARAMETERS OF A PID REGULATOR

The servomechanism that simulates the profile of the uneven land is excited by a rectangular signal with amplitude of 0.140 m and frequency of 0.025 Hz. In figure 2 three curves are setted: curve1 – variation of displacement over time of the servocylinder that simulates the profile of the uneven land; curve2 - variation of displacement over time of tracing the servocylinder, that actuates the blade of the navy machine in vertical plane; curve3 – the amount of the two displacement values, which is the variation over time of the deviation of the uneven land from the optical reference plane.

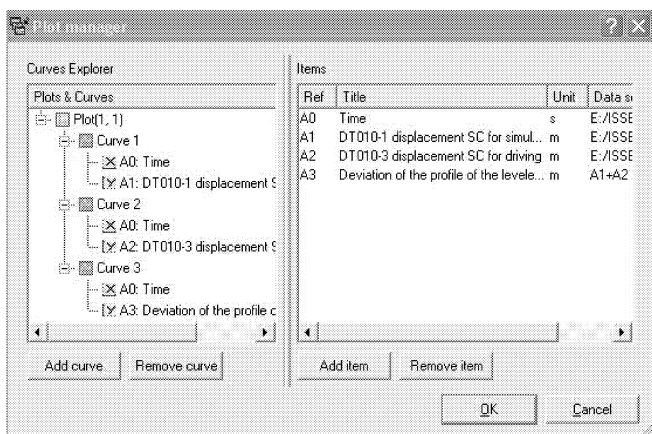


Figure 2. Setting the curves in "Plot manager".

The shape of the three curves can be seen in figure 3.

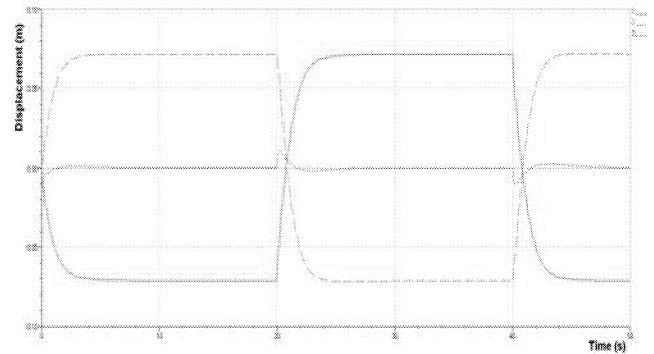


Figure 3. The reaction of the tracing servocylinder when the servocylinder that simulates the profile of the land is excited by a rectangular signal.

The curves in figure 3 resulted when setting the values of the parameters of a PID regulator inside "Change Parametres" box as in figure 4.

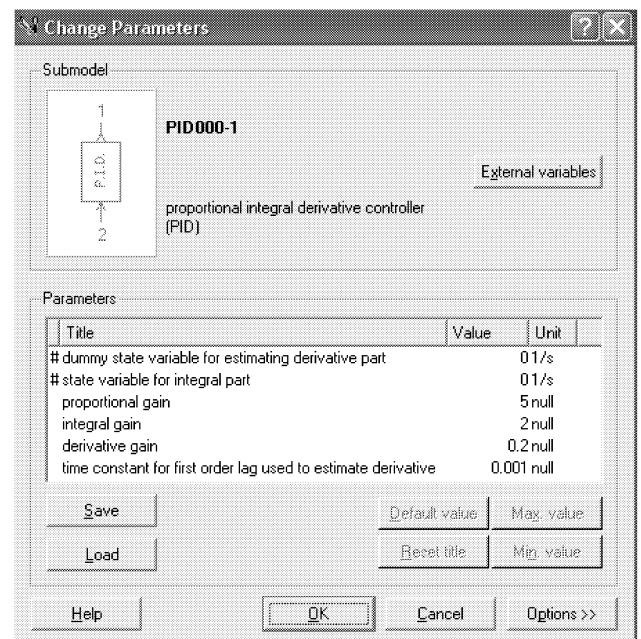


Figure 4. Random setting of the parameters of a PID regulator.

In these conditions the maximum value of the deviation of the leveled land from the optical reference plane is 0.01m (figure 5).

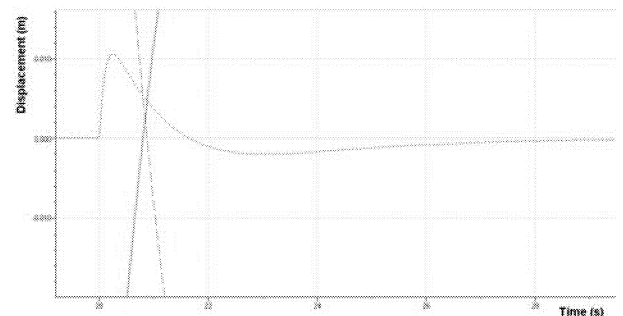


Figure 5. The maximum value of the deviation of the leveled land from the optical reference plane.

### Optimizing parameter $K_P$

Running the application in AMESim is repeated, this time canceling parameters  $K_I$  and  $K_D$  and selecting five values for parameter  $K_P$ , according to the settings in "Batch Control Parameter Setup" box (figure 6).

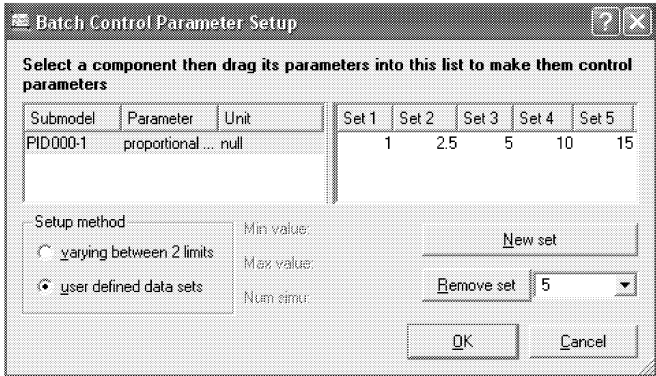


Figure 6. Setting values for parameter  $K_P$ .

In "Plot manager" box (figure 7.) there are shown the curves resulted when running the application in *Batch* mode, corresponding to five different values of parameter  $K_P$ . These curves represent: curve1...curve5 – variation over time of the displacement of the servocylinder that simulates the profile of the uneven land; curve6...curve10 - variation over time of the displacement of the servocylinder that actuates the blade of the navy machine; curve11...curve15 - variation over time of the deviation of the leveled land from the optical reference plane.

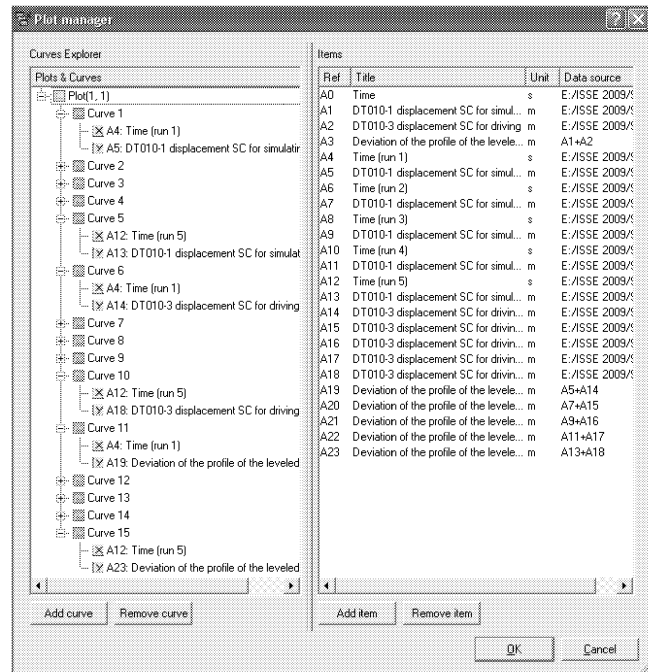


Figure 7. "Plot manager" box.

In figure 8 is shown the influence that the variation of the parameter  $K_P$  has upon the dynamics of the tracing servomechanism when exciting the servomechanism that simulates the profile of the uneven land by a rectangular signal with amplitude of 0.140 m and frequency of 0.025 Hz.

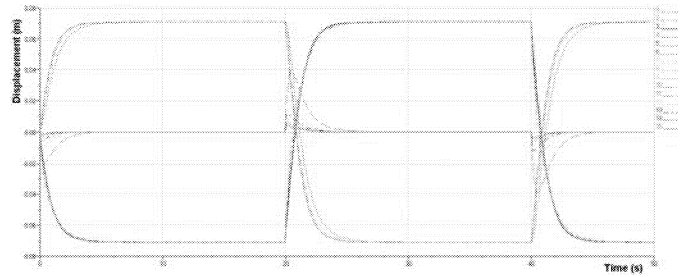


Figure 8. Influence of the variation of parameter  $K_P$  upon the dynamics of the tracing servomechanism.

In figure 9 is shown one detail of the variation over time of the amount of the displacement values of the two servocylinders, when applying the settings in figure 6. One can notice an increasing dynamics of the tracing servocylinder, in accordance with the increase of the value of parameter  $K_P$ .

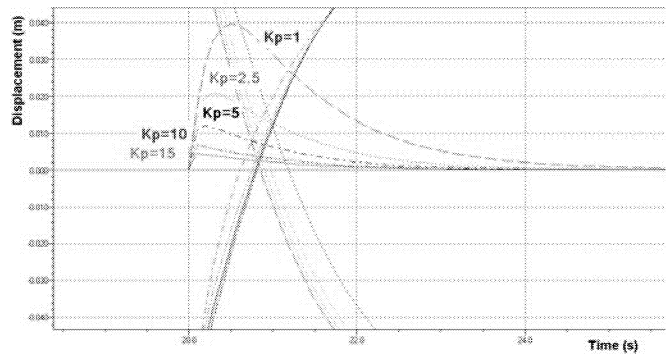


Figure 9. Variation in the deviation of the profile of leveled land from the reference plane, depending on variation of parameter  $K_P$ .

### Optimizing parameter $K_I$

Running the application in AMESim is repeated, this time canceling parameters  $K_P$  and  $K_D$  and selecting five values for parameter  $K_I$ , according to the settings in "Batch Control Parameter Setup" box (figure 10).

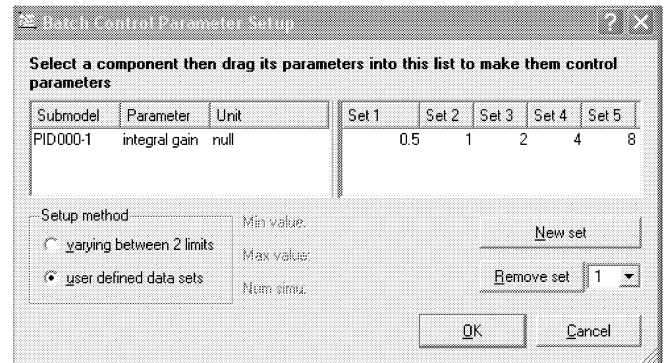
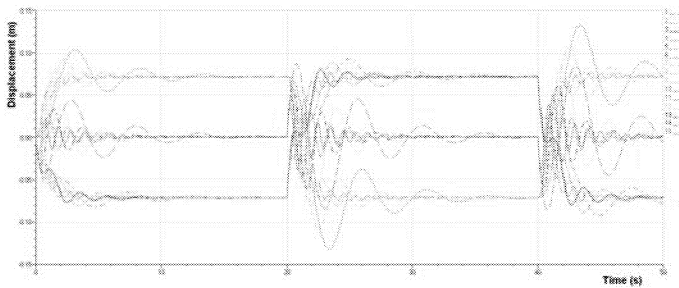


Figure 10. Setting values for parameter  $K_I$ .

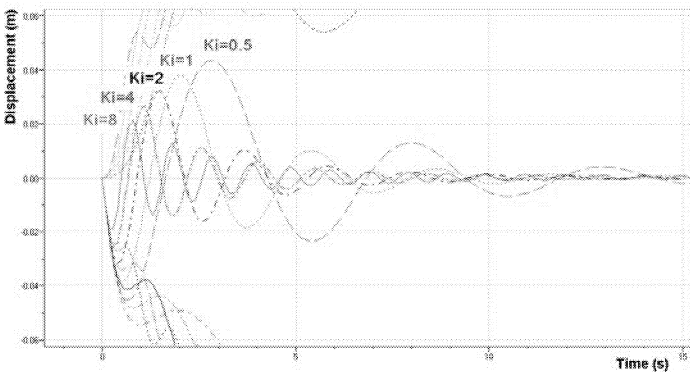


In figure 11 is shown the influence that the variation of parameter  $K_I$  has upon the dynamics of the tracing servomechanism when exciting the servomechanism that simulates the profile of the uneven land by a rectangular signal with amplitude of 0.140 m and frequency of 0.025 Hz.



**Figure 11. Influence of the variation of parameter  $K_I$  upon the dynamics of the tracing servomechanism.**

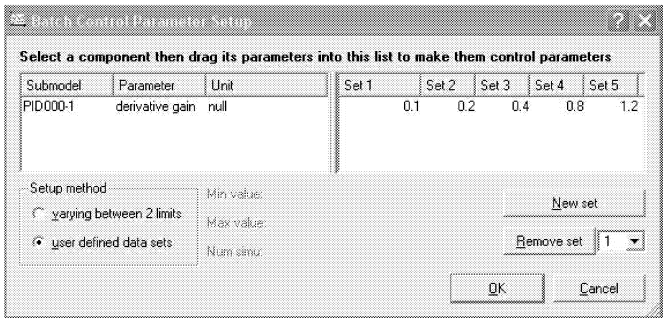
In figure 12 is shown one detail of the variation over time of the amount of the displacement values of the two servocylinders, when applying the settings in figure 10. One can notice that the stationary error in the tracing servomechanism is removed faster at a higher value of parameter  $K_I$ .



**Figure 12. Variation in the deviation of the profile of leveled land from the reference plane, depending on variation of parameter  $K_I$ .**

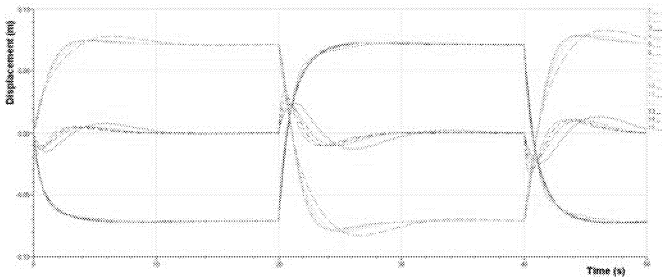
### Optimizing parameter $K_D$

Running the application in AMESim is repeated, this time setting parameters  $K_P=1$  ;  $K_I=0.5$  and selecting five values for parameter  $K_D$ , according to the settings in "Batch Control Parameter Setup" box (figure 13).

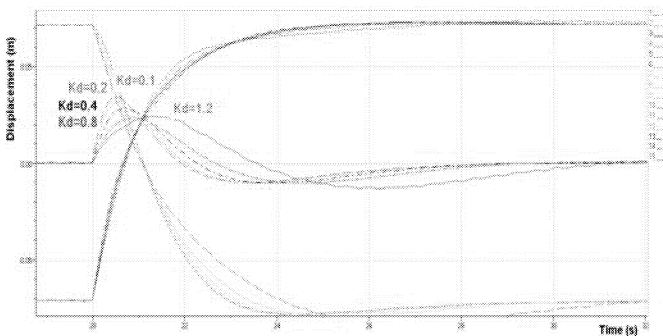


**Figure 13. Setting values for parameter  $K_D$ .**

In figure 14 is shown the influence that the variation of parameter  $K_D$  has upon the dynamics of the tracing servomechanism when exciting the servomechanism that simulates the profile of the uneven land by a rectangular signal with amplitude of 0.140 m and frequency of 0.025 Hz. In figure 15 is shown one detail of the variation over time of the amount of the displacement values of the two servocylinders, when applying the settings in figure 13. One can notice that the stabilization in the tracing servomechanism is attained faster at a lower value of parameter  $K_D$ .



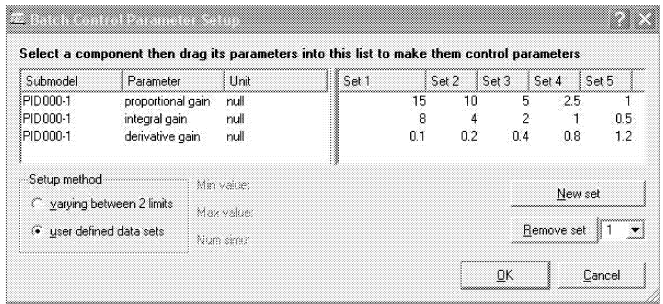
**Figure 14. Influence of the variation of parameter  $K_D$  upon the dynamics of the tracing servomechanism.**



**Figure 15. Variation in the deviation of the profile of leveled land from the reference plane, depending on variation of parameter  $K_D$ .**

### Optimizing parameter $K(K_P, K_I, K_D)$

Running the application in AMESim is repeated, this time selecting five pairs of values for parameters  $K_P$  ,  $K_I$  and  $K_D$  , according to the settings in "Batch Control Parameter Setup" box (figure 16).



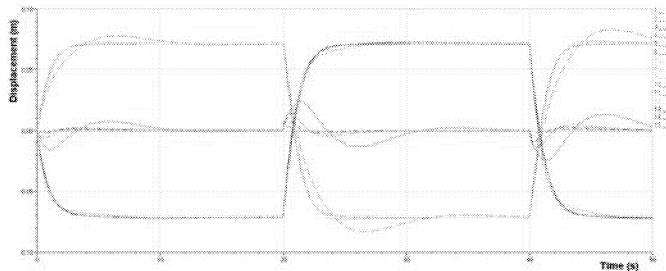
**Figure 16. Setting values for  $K(K_P, K_I, K_D)$ .**

In figure 17 is shown the influence that the variation of parameter  $K(K_P, K_I, K_D)$  has upon the dynamics of the tracing servomechanism when exciting the servomechanism

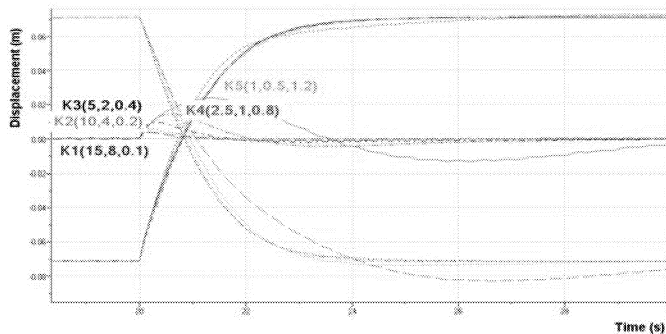


that simulates the profile of the uneven land by a rectangular signal with amplitude of 0.140 m and frequency of 0.025 Hz.

In figure 18 is shown one detail of the variation over time of the amount of the displacement values of the two servocylinders, when applying the settings in figure 16. One can notice that the optimal dynamics and stability of the tracing servomechanism is attained when PID regulator has the global parameter  $K(15,8,0.1)$ , where:  $K_P=15$ ,  $K_I=8$  and  $K_D=0.1$ .



**Figure 17. Influence of the variation of parameter  $K(K_P, K_I, K_D)$  upon the dynamics of the tracing servomechanism.**



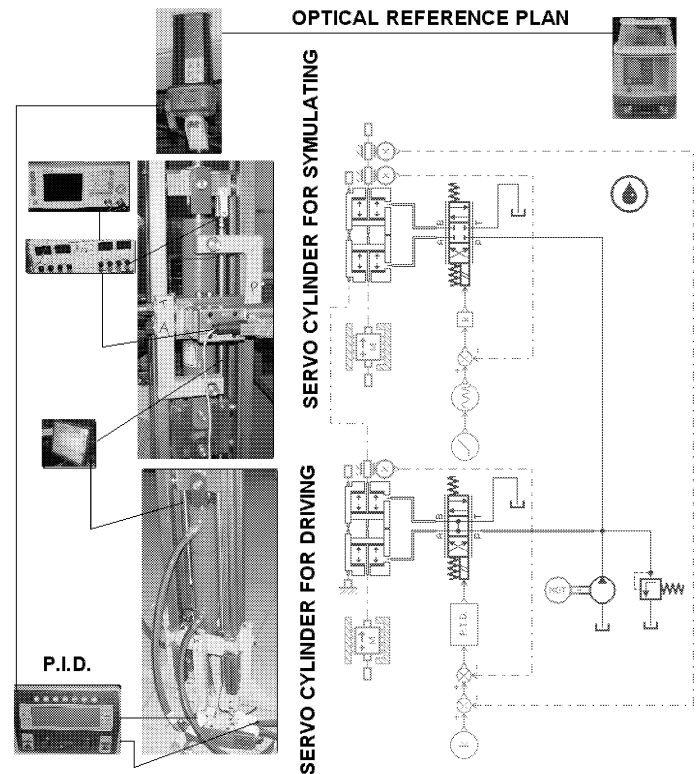
**Figure 18. Variation in the deviation of the profile of leveled land from the reference plane, depending on variation of parameter  $K(K_P, K_I, K_D)$ .**

## CONCLUSIONS

The paper develops a simulation model in AMESim for a hydraulic position adjustment servomechanism with response by laser (POPESCU et al. 2008), with a view to optimize the parameters of the PID regulator which is specific to this servomechanism.

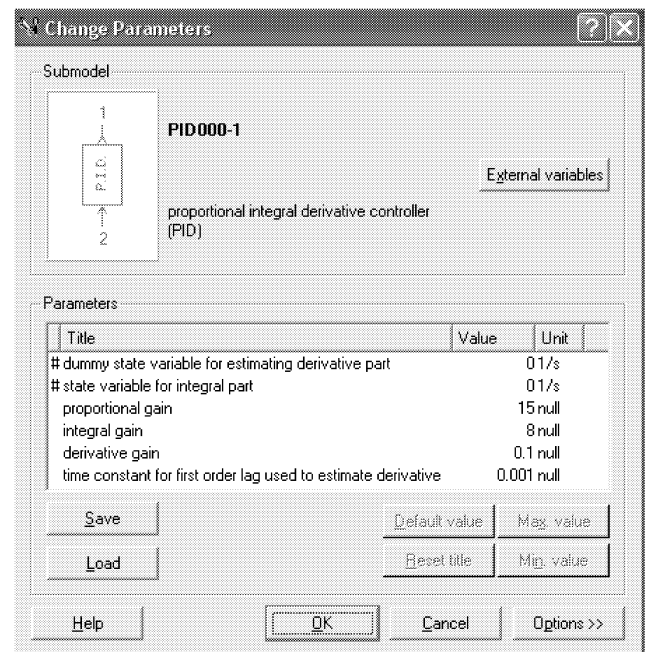
The hydraulic servomechanism (figure 19.) is equivalent to a test device for laser, hydraulic and electronic modules designed to equip those navy leveling machines with automatic regime of operation (POPESCU et al. 2009). Leveling is performed tracing an optical reference plane generated by a rotary laser emitter and received by a laser sensor. The laser emitter is mounted on a tripod, while the laser receiver is on the blade of the navy machinery. The model includes a simulator of the profile of the uneven land and a hydraulic servomechanism which traces the profile of the land and actuates the blade of the navy machine in vertical plane.

The simulation model was at the basis of developing the test device for laser, hydraulic and electronic modules designed to equip those navy leveling machines with automatic regime of operation.

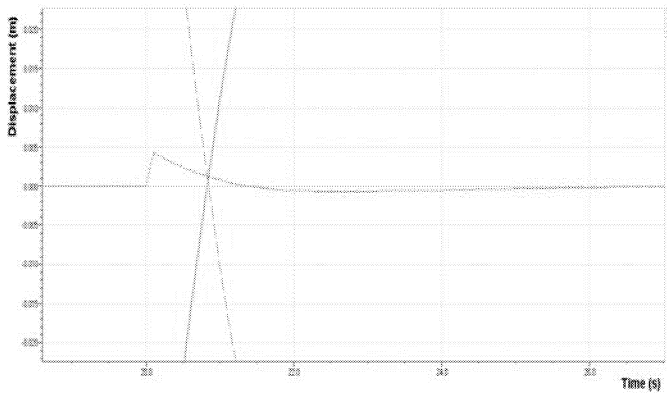


**Figure 19. Assimilating the simulation model in AMESim to a laboratory test device.**

By consecutively running the simulation model AMESim (AMESim Software Suit ) the optimal parameters of the PID regulator are attained at the values  $K_P=15$ ,  $K_I=8$  s and  $K_D=0.1$  s. When running the application with these optimal values of the parameters of the PID regulator, according to the settings in figure 20, one attains the minimum value of 0.004 m of the deviation of the leveled land from the optical reference plane (figure 21). This value is 2.5 times lower than the one resulted from the first running of the simulation model (figure 5).



**Figure 20. Setting optimal parameters for a PID compensator.**



**Figure 21. Maximum optimized value of the deviation of the leveled land from the optical reference plane.**

A typical way of using the above control system is shown in figure 22.



**Figure 22. Laser controlled electro hydraulic leveling machine (courtesy New Holland)**

## REFERENCES

- Michel LEBRUN, Claude RICHARDS., „How to create Good Models without Writing a Single Line of Code”, *Fifth Scandinavian International Conference on Fluid Power, Linköping*, (1997).
- \*\*\* IMAGINE SA, „Advanced Modelling And Simulation Environment, Release 4.2.1. User Manual”, Roanne, (2005).
- Michel LEBRUN, „EHA's Model Reduction Using Activity Indexes”. *Recent Advances in Aerospace Hydraulics, INSA Toulouse*, (2004).
- Nicolae VASILIU, Daniela VASILIU, „Electro Hydraulic Servomechanisms with Two Stages DDV for Heavy Load Simulators Controlled by ADWIN.” *Recent Advances in Aerospace Hydraulics, INSA Toulouse*, (2004).
- Nicolae VASILIU, Daniela VASILIU, „Fluid Power Systems, Vol.I. ”, *Technical Publishing House, Bucharest*, (2005)
- Nicolae VASILIU, Constantin CALINOIU, Constantin DRAGOI, „Experimental Identification of the Electro Hydraulic Servo Systems used in Speed Governors for Hydraulic Turbines”, *National Conference on Hydropower, Bucharest*, (2002).
- Teodor Costinel POPESCU, Andrei DRUMEA, Iulian DUȚU, „Numerical simulation and experimental identification of the laser controlled modular system purposefully created for equipping the terrace leveling installations”, *International Spring Seminar on Electronics Technology (ISSE) Budapest, Hungary; 7-11 May, 2008*, in: "Abstract Proceedings - Reliability and Life-time Prediction", ISBN:978-963-06-4915-5; pp.336-341, (2008).
- Teodor Costinel POPESCU, Marian BLEJAN, Gheorghe ȘOVĂIALĂ, „7297 2D-Experimental Research upon Accommodating the Functional Parameters of a Laser Controlled System Designed for a Grading Machinery with the Actual Operating Conditions of the Grading Machinery”, *ATOM-N 2008, 28...31 August 2008, Constanța, Romania, The 4<sup>th</sup> edition of the International Conference "Advanced Topics in Optoelectronics, Microelectronics, and Nanotechnologies"*, Proceedings of SPIE Vol. 7297 (SPIE, Bellingham, WA, 2009), Article CID Number 7297-86, ISSN: 0277-786X, ISBN: 9780819475596, 5pg., (07.01.2009).

## 5. BIOGRAPHY

**Teodor Costinel POPESCU**, PhD, graduated Institute of Civil Engineering – Faculty of Fluid Systems in 1978, and Polytechnic Institute of Bucharest – Faculty of Mechanical Engineering in 1989, Fluid Machines Faculty. Since 1983 he is working as senior scientific researcher in fluid power systems field at Fluid Power Systems Research Institute – INOE 2000 – IHP Bucharest, Romania.

**Iulian DUTU**, PhDc. graduated Mechanical and Mechatronic Engineering Faculty from University Politehnica of Bucharest in 2002. He works as scientific researcher in fluid power systems at Fluid Servo Systems and Power Electronics Division of INOE 2000 - IHP, Bucharest, Romania.

**Catalin VASILIU**, PhDc, graduated Electrical Engineering Faculty from University Politehnica of Bucharest in 2008. He works as scientific researcher in the Automotive Department of the Transport Faculty of the same university. He worked 6 month as research engineer in LMS Intl Company in the field of RTS of the hybrid systems by AMESim RT and LABView RT using PXI computers from National Instruments Corporation.

**Marius MITROI**, PhDc, graduated Power Engineering Faculty from University Politehnica of Bucharest in 2007. He works as IT Manager within Romanian Innovation Financing Agency AMCSIT - POLITEHNICA” from Bucharest, Romania; he is working also in the field of modeling and real time simulation of the fluid control systems.

# **FLUID FLOW SIMULATION IN ENGINEERING APPLICATIONS**



# NOTES ON THE IMPLEMENTATION OF BROWNIAN MOTION IN MESOSCOPIC FLUID-PARTICLE MODELS

Massimo Lai  
Dimitris Drikakis

FMaCS group, Department of Aerospace Sciences, Cranfield University  
Cranfield, Bedfordshire, MK43 0AL, United Kingdom  
email: m.lai@cranfield.ac.uk, d.drikakis@cranfield.ac.uk

## KEYWORDS

Physics, Simulators, Numerical Methods, Random Number Generators, Stochastic Models

## ABSTRACT

Mesoscopic simulation of nucleic acids transport can prove extremely beneficial to the growing field of microfluidic DNA biosensors. In microfluidic devices, unlike macroscopic ones, the diffusive transport becomes relevant as it is no longer overwhelmed by convection. Diffusion of colloid-sized particles is caused by their brownian motion in solution, whose modelling is therefore crucial for the correct description of diffusive effects. It is known that the random displacement and velocity perturbations follow a multivariate Gaussian distribution. We analyse the explicit form of the covariance matrix and show a decomposition that factorizes the probability density function into three pairs of bivariate distributions. A preliminary model is implemented and validated for a simple test case.

## INTRODUCTION

Modelling the brownian motion is an important part of a mesoscopic simulation of particles and polymer transport, because random agitation is the physical mechanism that triggers diffusive processes.

The explanation and description of Brownian motion was studied and formalized by the likes of Einstein, Langevin, Smoluchowski (Doi and Edwards 1986). Langevin's approach is to include explicitly the random force in the hamiltonian equations of motion of the colloid particle (assumed to be spherical) thus obtaining a stochastic ordinary differential equation which can be solved formally using the boundary condition  $\mathbf{v}(0) = \mathbf{v}_0$ ,

$$\frac{d\mathbf{v}}{dt} = -\beta\mathbf{v} + \mathbf{A}(t) \Rightarrow \mathbf{v} - \mathbf{v}_0 e^{-\beta t} = e^{-\beta t} \int_0^t e^{\beta\tau} \mathbf{A}(\tau) d\tau \quad (1)$$

where  $\beta = 6\pi R_h \eta / M$  is the friction coefficient,  $m$  is the mass,  $R_h$  the equivalent hydrodynamic radius of the particle,  $\eta$  is the newtonian viscosity of the solvent, and  $\mathbf{A}(t)$  is the stochastic acceleration, that needs to be char-

acterised for Eq. 1 to be of any practical use. Given a certain simulation timestep  $\Delta t$ , what we need to know is the *cumulative* effect of all the random accelerations caused by collisions occurred in the interval  $[t, t + \Delta t]$ , on the resulting position and velocity. When the “random flight” condition  $\Delta t \gg \beta^{-1}$  does *not* apply, the displacement over the timestep is no longer purely random, but it's affected by the initial velocity, which must be calculated as well. This can be often the case in mesoscopic fluid-particle simulations (Trebotich et al. 2005). If we call  $\mathbf{r}$  and  $\mathbf{r}_0$  the particle's position vector and initial position, under the hypotheses that

- the timescale of the random force variation is much faster than any other physical quantity in play;
- energy equipartition will hold for both solvent molecules and particles, so the particle velocity distribution will also converge to a Maxwellian curve

the joint probability distribution function (PDF)  $f(\mathbf{d}, \mathbf{u})$  for the vectors

$$\mathbf{d} = \mathbf{r} - \mathbf{r}_0 - \beta^{-1} \mathbf{v}_0 (1 - e^{-\beta \Delta t}) \quad (2)$$

$$\mathbf{u} = \mathbf{v} + \mathbf{v}_0 e^{-\beta \Delta t}, \quad (3)$$

(from which displacement and velocity can be readily obtained) over a timestep  $\Delta t$ , is given by (Chandrasekhar 1943):

$$\frac{1}{(2\pi)^3 (fg - h^2)^{3/2}} \exp \left[ -\frac{g|\mathbf{d}|^2 - 2h\mathbf{d} \cdot \mathbf{u} + f|\mathbf{u}|^2}{2(fg - h^2)} \right], \quad (4)$$

where the coefficients  $f, g, h$  are given by

$$f = \frac{1}{\beta^2} \frac{k_B T}{M} (2\beta \Delta t - 3 + 4e^{-\beta \Delta t} - e^{-2\beta \Delta t}) \quad (5)$$

$$g = \frac{k_B T}{M} (1 - e^{-2\beta \Delta t}) \quad (6)$$

$$h = \frac{1}{\beta} \frac{k_B T}{M} (1 - e^{-\beta \Delta t})^2 \quad (7)$$

where  $T$  is the absolute temperature and  $k_B$  the Boltzmann constant. Modelling brownian fluctuations involves sampling six random numbers  $d_x, d_y, d_z, u_x, u_y, u_z$ , from the joint PDF given by Eq. 4. In the following, a numerical implementation is described and tested.

## METHOD

For numerical implementation, Eq. 4 must first be recast in the canonical form of a Gaussian multivariate. When the vector of mean values is  $\mathbf{0}$  (as in brownian fluctuations) its general form is:

$$f(\mathbf{w}) = \frac{1}{(2\pi)^{n/2}|\mathbf{C}|^{1/2}} \exp \left[ -\frac{1}{2} \mathbf{w}^T \mathbf{C}^{-1} \mathbf{w} \right] \quad (8)$$

where  $\mathbf{w}$  is an  $n$ -dimensional row vector of random variables with zero mean, and  $\mathbf{C}$  is the  $n \times n$  covariance matrix, symmetric and positive definite. The off-diagonal terms of  $\mathbf{C}$  are responsible for the mutual correlations between the stochastic variables. If we write the vector  $\mathbf{w}$  as:

$$\mathbf{w} = [\mathbf{d} \ \mathbf{u}]^T = [d_x \ d_y \ d_z \ u_x \ u_y \ u_z]^T \quad (9)$$

the exponent in Eq. 4 becomes the quadratic form:

$$\mathbf{w}^T \mathbf{C}^{-1} \mathbf{w},$$

where the inverse matrix  $\mathbf{C}^{-1}$  and the covariance matrix  $\mathbf{C}$  are found to be

$$\mathbf{C}^{-1} = \frac{1}{fg - h^2} \begin{bmatrix} g & 0 & 0 & -h & 0 & 0 \\ 0 & g & 0 & 0 & -h & 0 \\ 0 & 0 & g & 0 & 0 & -h \\ -h & 0 & 0 & f & 0 & 0 \\ 0 & -h & 0 & 0 & f & 0 \\ 0 & 0 & -h & 0 & 0 & f \end{bmatrix} \quad (10)$$

and

$$\mathbf{C} = \begin{bmatrix} f & 0 & 0 & h & 0 & 0 \\ 0 & f & 0 & 0 & h & 0 \\ 0 & 0 & f & 0 & 0 & h \\ h & 0 & 0 & g & 0 & 0 \\ 0 & h & 0 & 0 & g & 0 \\ 0 & 0 & h & 0 & 0 & g \end{bmatrix} \quad (11)$$

The covariances (i.e. off-diagonal terms) are non-zero only between components of position and velocity along the same cartesian direction. Components along different directions are non correlated, and actually independent, their joint PDF being Gaussian (Baldi 1998). Consequently, the distribution in Eq. 4 can be factorized into three identical distributions:

$$\frac{1}{(2\pi)^2(fg - h^2)} \exp \left[ -\frac{gd_i^2 - 2hd_iu_i + fu_i^2}{2(fg - h^2)} \right], \quad i = x, y, z. \quad (12)$$

Now if we split  $\mathbf{w}$  into three vectors

$$\mathbf{w}_i = [d_i \ u_i]^T, \quad i = x, y, z, \quad (13)$$

the Eq. 12 can be rewritten as canonical Gaussian bivariate (consisting of a displacement and a velocity component each)

$$\frac{1}{(2\pi)^2|\mathbf{B}|} \exp \left[ -\frac{1}{2} \mathbf{w}_i^T \mathbf{B}^{-1} \mathbf{w}_i \right], \quad i = x, y, z, \quad (14)$$

where the covariance matrix and its inverse are:

$$\mathbf{B} = \begin{bmatrix} f & h \\ h & g \end{bmatrix}, \quad \mathbf{B}^{-1} = \frac{1}{fg - h^2} \begin{bmatrix} g & -h \\ -h & f \end{bmatrix}, \quad (15)$$

and their coefficients are already known. A common method for the implementation of Gaussian multivariates requires the Cholesky factorization of the covariance matrix (Allen and Tildesley 1987).

Namely, if  $\mathbf{x} = [x_1, x_2]^T$  is a vector of independent normal variables, and  $L$  a lower triangular matrix such that  $\mathbf{L}\mathbf{L}^T = \mathbf{B}$ , with  $\mathbf{B}$  positive definite, then the vector  $\mathbf{y} = \mathbf{L}\mathbf{x}$  follows a Gaussian multivariate with 0 mean vector and covariance matrix  $\mathbf{B}$  (Baldi 1998). The matrix  $\mathbf{L}$  is the so-called Cholesky factor of  $\mathbf{B}$ .

Our bidimensional case is quick to derive, and we get:

$$\mathbf{L} = \begin{bmatrix} \sqrt{f} & 0 \\ -\frac{h}{\sqrt{f}} & \sqrt{g - \frac{h^2}{f}} \end{bmatrix}, \quad (16)$$

that yields the recipe for each of the component of displacement and velocity, for  $i = x, y, z$ :

$$d_i = x_{i1} \sqrt{f} \quad (17)$$

$$u_i = x_{i1} \left( -\frac{h}{\sqrt{f}} \right) + x_{i2} \sqrt{g - \frac{h^2}{f}}, \quad (18)$$

where  $x_{i1}$  and  $x_{i2}$  are independent normal variables that can be generated for example by a Box-Muller or Marsaglia polar algorithm (Devroye 1986), provided that a good-quality generator of a uniform distribution in the interval (0,1) is available (Panneton et al. 2006). The definition of the two random vectors  $\mathbf{d}$ ,  $\mathbf{u}$ , and the knowledge of their PDF, provide the numerical scheme for the generation of a brownian trajectory. Be  $\Delta t$  the timestep length, and  $\mathbf{r}^t$  and  $\mathbf{v}^t$  position and velocity at the timestep  $t$ , rearranging the terms in the Eq. 2 we get:

$$\mathbf{r}^{t+\Delta t} = \mathbf{r}^t + \beta^{-1}(1 - e^{-\beta\Delta t})\mathbf{v}^t + \mathbf{d}^t \quad (19)$$

$$\mathbf{v}^{t+\Delta t} = (e^{-\beta\Delta t})\mathbf{v}^t + \mathbf{u}^t, \quad (20)$$

where  $\mathbf{d}^t$  and  $\mathbf{u}^t$  are correlated random vectors whose components are generated as described above.

## RESULTS

The model described in the previous section was implemented in a custom C code using the WELL random number generator (Panneton et al. 2006), and parametrised according to the data collected from an all-atom 40ns molecular dynamics (MD) simulation of a ssDNA polyadenine tetramer, in saline water at 1.0M ion concentration, at 298K, with timestep 1fs (unpublished results), in order to allow a comparison between the predicted diffusion coefficients. After correcting the MD results for solvent viscosity and finite-size effects,

the resulting model parameters, in SI units, were: hydrodynamic radius:  $R_h = 0.818 \cdot 10^{-9}m$ ; particle mass:  $M = 2.108 \cdot 10^{-24}Kg$ ; temperature:  $T = 298.0K$ ; friction coefficient:  $\beta = 7.095 \cdot 10^{-12}s^{-1}$ ; solvent viscosity at 298K and 1.0M salinity:  $\eta = 0.97 \cdot 10^{-3}Pa \cdot s$ . The resulting elements of the covariance matrix were  $f = 5.601 \cdot 10^3 pm^2$ ,  $g = 1.473 \cdot 10^3 pm^2 ps^{-2}$ , and  $h = 7.075 \cdot 10^1 pm^2 ps^{-1}$ .

The diffusion coefficient predicted by the all-atom simulation was  $D = (0.275 \pm 0.003) \cdot 10^{-9}m^2s^{-1}$ .

The decay time of the displacement vector autocorrelation function (ACF) was  $\tau = 0.122ps$ .

The decorrelation time and the diffusion coefficient from the MD simulation were used to validate the model.

First, the displacement (ACF) was computed for  $10^5$  steps of  $0.001ps$ . The resulting decay curve had a time constant of  $0.136ps$ , very close to the value given by the all-atom simulation, as shown in Fig. 1.

Secondly, the diffusion coefficient of the brownian particle was calculated from the slope of the window-averaged mean square displacement (MSD) of a trajectory of  $10^5$  steps, applying Einstein's relation:

$$D = \lim_{t \rightarrow \infty} \frac{\langle |\mathbf{r}(t) - \mathbf{r}_0|^2 \rangle}{6t}, \quad (21)$$

which gave  $D = 0.276 \cdot 10^{-9}m^2s^{-1}$ , again in good agreement with the all-atom simulation (Fig. 2).

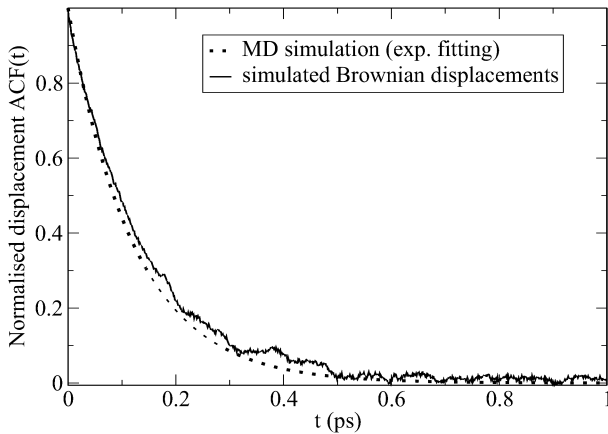


Figure 1: Computed ACF (solid line) and fitting of the ACF from MD simulation (dotted) for the brownian displacements calculated with a timestep of  $0.001ps = 1fs$ .

## CONCLUSIONS

In mesoscopic fluid-particle models, the representation of brownian perturbations requires the sampling of random numbers drawn from a 6-dimensional Gaussian multivariate with a given covariance matrix. We derived a factorization into 3 independent Gaussian

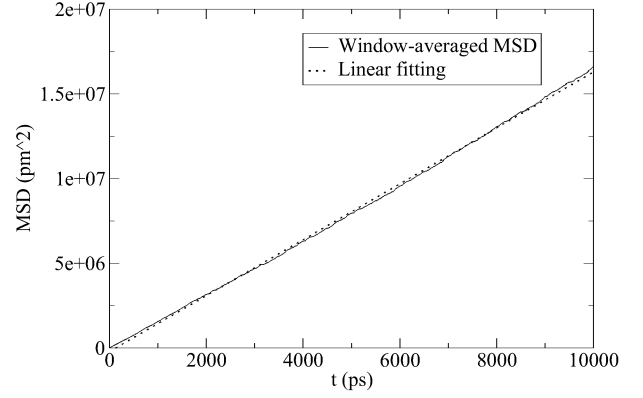


Figure 2: Window-averaged MSD (solid line) and linear fitting (dotted) for a trajectory of  $10^5$  timesteps of  $10ps$  each. The theoretical diffusion coefficient,  $0.275 \cdot 10^{-9}m^2/s$  is well reproduced.

bivariate, that allows straightforward implementation using simple numerical recipes. The model was implemented in a C code that was validated, for the simple case of an isolated diffusing ssDNA tetramer, against the values predicted by an all-atom MD simulation for displacement decorrelation time and diffusion coefficient. Future work will focus on the extension of the approach to diffusing chains.

This work has been supported by the European Commission under the 6th Framework Program (Project: DINAMICS, NMP4-CT-2007-026804).

## REFERENCES

- Allen M.P. and Tildesley D.J., 1987. *Computer simulation of liquids*. Oxford University Press, Oxford.
- Baldi P., 1998. *Calcolo delle Probabilit e Statistica, 2a edizione (in Italian)*. McGraw-Hill.
- Chandrasekhar S., 1943. *Stochastic Problems in Physics and Astronomy*. *Rev Mod Phys*, 15, no. 1, 1–89.
- Devroye L., 1986. *Nonuniform random variate generation*. Springer-Verlag, New York.
- Doi M. and Edwards S., 1986. *The theory of polymer Dynamics*. Oxford Science Publications, New York.
- Panneton F.; L'Ecuyer P.; and Matsumoto M., 2006. *Improved long-period generators based on linear recurrences modulo 2*. *ACM Trans Math Softw*, 32, no. 1, 1–16.
- Trebotich D.; Miller G.H.; Colella P.; Graves D.T.; Martin D.F.; and Schwartz P.O., 2005. *A tightly coupled particle-fluid model for DNA-laden flows in complex microscale geometries*. In *Computational Fluid and Solid Mechanics 2005*. 1018–1022.

# Estimation of Longitudinal Aerodynamics of a Transport Aircraft Using FDR Data Through Neural Networks

Ming-Hao Yang, Cjing-Shun, Ho,  
Fei-Bin, Hsiao  
Department of Aeronautics and  
Astronautics  
National Cheng Kung University  
No.1, University Road, Tainan City,  
Taiwan, ROC  
E-mail: [ming-hao@asc.gov.tw](mailto:ming-hao@asc.gov.tw),  
[csho,fbshiao@mail.ncku.edu.tw](mailto:csho,fbshiao@mail.ncku.edu.tw)

C. Edward Lan  
Department of Aerospace  
Engineering  
University of Kansas  
Lawrence, KS 66045, USA  
E-mail: [vortex@ku.edu](mailto:vortex@ku.edu)

Chien-Chun Kung  
Department of Mechtronics, Energy  
and Aerospace Engineering  
National Defense University  
No. 190, San Yuan 1 St., Tashi,  
Taoyuan, Taiwan, ROC  
E-mail: [cckung@ndu.edu.tw](mailto:cckung@ndu.edu.tw)

## KEYWORDS

Extended Kalman Filter, Flight Dynamics, Longitudinal  
Aerodynamic Derivatives, Neural Networks

## ABSTRACT

The main objective of this study is to estimate aircraft's longitudinal aerodynamic derivatives by modeling the flight data from flight data recorder, FDR, on a civil transport aircraft. The quality of recorded flight data is enhanced by Extended Kalman Filter to perform kinematic compatibility check and estimate unrecorded flight data at the same time. The thrust of the turbofan engines for the aircraft is estimated by a thrust model and the FDR data. The aerodynamic forces and moments are estimated with aircraft's dynamics equations. After obtaining the aerodynamic forces and moments, aerodynamic modeling based on multi-layer back propagation neural network and radial-based function neural network is performed to estimate the aircraft's longitudinal derivatives. The disadvantages of these two networks are also compared. It is shown that some longitudinal aerodynamic derivatives are degraded on the average with rapidly varying wind field when flying through atmospheric turbulence.

## INTRODUCTION

FDR provides useful information during aircraft accident investigation about the operation, motion and health conditions of aircraft. Because of the limitation of aircraft's navigation system and the recording media capability of FDR, there are inadequate data to allow investigators to perform a thorough investigation. Therefore the investigators sometimes have to utilize other tools, such as the training/engineering based simulators, for further aircraft performance and stability analysis (Aviation Safety Council 2005). The training /engineering simulators are based on the certificated aerodynamic characteristics from the static wind tunnel and flight test results. These data are not capable of describing the nonlinear dynamic flight conditions in accident flight or in adverse weather encounter.

In the present study, the recorded flight data from the FDR of a civil transport aircraft, including 3 normal flights at different flight phases and one cruise flight in atmospheric

turbulence, to evaluate the aircraft longitudinal aerodynamic characteristics, and the strength of atmospheric turbulence.

## AERODYNAMIC DERIVATIVES ESTIMATION

Generally, there are three methods to obtain these values:

### Approximate mathematical model (AMM)

AMM is the simplest and least accurate method to evaluate the flight stability in the aircraft preliminary design from mathematical formula, tables and charts. The most popular sources are the Engineering Sciences Data Unit, and the USAF DATCOM.

### Wind Tunnel Measurements (WTM)

For the civil transport aircraft, the airframe designers are not specifically required to undertake the dynamic testing. Therefore, when an aircraft at a highly dynamic situation, the nonlinear dynamic characteristics of the aircraft will be dominant, and the static wind tunnel test results will be unreliable to describe the dynamic situation. The rotary balance test for civil transport aircraft is typically conducted at very low Reynolds number, with the results used mainly in estimating the developed spin modes of general aviation and military aircraft. In addition, dynamic aerodynamic derivatives for a civil transport have also been acquired recently through the forced oscillation test at low speeds and low Reynolds numbers. However, how useful these dynamic data are in accident flight reconstruction and in describing the aircraft response in severe atmospheric turbulence encounter is yet to be verified. Therefore, for the time being it is imperative to estimate the aerodynamic derivatives from flight data for accident flight investigation.

### Flight Test Measurements (FTM)

In the late 1960s, FTM started being developed at NASA and some flight test research centers (Iliff et al. 1972). As the identification technology improving, the aerodynamic derivatives identification method also has evolved in complexity. The calculation method started from the least square to the Newton-Raphson and then evolved to the Maximum Likelihood (Iliff 1987). The system to be identified has progressed from a linear to a non-linear one. Today, the maximum likelihood has been widely used in this field (Hui et al. 2005). Afordmention methods are based on small disturbance from equilibrium trim flight condition. The aerodynamic derivatives from these methods represent



the quasi-steady values as the instantaneous slope. When an aircraft encounters adverse weather, such as severe atmospheric turbulence, or in a loss-of-control situation, the associated aerodynamics is expected to be highly nonlinear. Therefore, an appropriate identification method must be capable of handling more complex phenomena, such as artificial intelligent technique (Weng et al. 2006).

## METHODOLOGY

### Extended Kálmán Filter(EKF)

EKF is applied to the nonlinear optimal estimation. It uses the system's observations/measurements, system's dynamics, and measurement's dynamics, to estimate the noise and bias and estimate system parameters. EKF applies sequential and recursive processes by linear minimum variance theory (G.Minkler and J. Minkler, 1993).

### Atmospheric Turbulence Strength Identification

According to ICAO newest automatic pilot reporting system (APRS), the severity of atmospheric turbulence is based on turbulence index (TI), and is defined as instantaneous and RMS values of EDR, describing the dissipation of turbulence kinematic energy. The severity is identified as four levels, severe( $15 \leq TI \leq 27, \text{Peak EDR} > 0.5$ ), moderate( $6 \leq TI \leq 14, 0.3 < \text{Peak EDR} < 0.5$ ), light( $1 \leq TI \leq 5, 0.1 < \text{Peak EDR} < 0.3$ ) and No( $TI = 0, \text{Peak EDR} < 0.1$ ).

### Thrust Model for Turbofan Engine

The engine thrust is calculated from thrust models and the FDR data. The recorded data includes, engine pressure ratio, engine exhaust temperature, N1 and N2. The thrust force of turbofan engine is calculated by using the momentum principle with contributions from the core engine and fan section and can be shown to be (Jack 1996),

### Neural Network

The neural network (NN), as a simulator of linear and non-linear systems' dynamics, can learn through the input/output relations without physical mathematical description of dynamics (Suykens et al. 1996, Gupta et al. 2003). In the present study employs the multilayer backpropagation neural networks (MLBPNN) and radial based neural networks (RBFNN) to identify the aerodynamic characteristics of civil transport aircraft.

#### MLBPNN

The signal flow for MLBPNN consists of feedforward and backward processes. Feedforward signal maps the input signal to the output by the neuron cell's activity: weighting the input and mapping via the active function and delivering the signal to the next neuron. The difference between the network's output and the learning target is called as error signal, propagated backward. The present study adopts the steepest learning rule.

#### RBFNN

RBFNN is a type of feedforward NN. The structure of RBFNN consists input, hidden layer, and output layers. The radial based function, an Gaussian function, is located at hidden layer, and measures the distance between the input and weighted distance. The weighted distances are summed and compared to the target value through the output layer.

## RESULTS

The four segments are listed in Table 1:

Table 1 Flight data for this study

|          | Flight #1   | Flight #2  | Flight #3 | Flight #4  |
|----------|-------------|------------|-----------|------------|
| Alt (ft) | 32900~33100 | 2400~28000 | 0~11800   | 19900~3348 |
| Time     | 1~828       | 829~1842   | 1843~2128 | 2128~2727  |

### Turbulence Strength Identification

The instantaneous peak and RMS/Average of EDR are denoted by  $EDR_{INST}$  and  $EDR_{AVE}$  spatially. The results are shown in Figure 1. They show that only in Flight segment #1 turbulence was encountered at 240~270 seconds, and 780~829 seconds. The most severe turbulence at the first time was at 256 seconds. At that time, the  $EDR_{INST}$  was 0.557, the  $EDR_{AVE}$  was 0.436, and the TI was 19. Therefore, in flight segment #1 severe turbulence was encountered at the first time period. In the second time period, the most severe turbulence occurred at 788 seconds. The maximum value of  $EDR_{INST}$  and  $EDR_{AVE}$  are 0.16 and 0.1. The strength of turbulence is therefore classified as light.

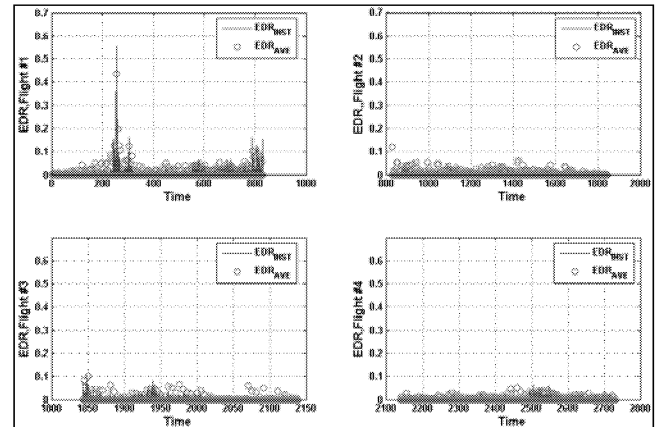


Figure 1: Turbulence strength estimated results

### Estimation of Flight Data and Thrust Force

To estimate the un-recorded flight data and remove the noise and bias from the recorded flight data, the EKF is performed as kinematic consistency. The thrust force is estimated by inserting the recorded flight data and the dimension of the engine into the thrust model. The calculated results of  $C_L$ ,  $C_D$ ,  $C_Y$ ,  $C_m$ ,  $C_l$ , and  $C_N$  are shown in Figure 2.

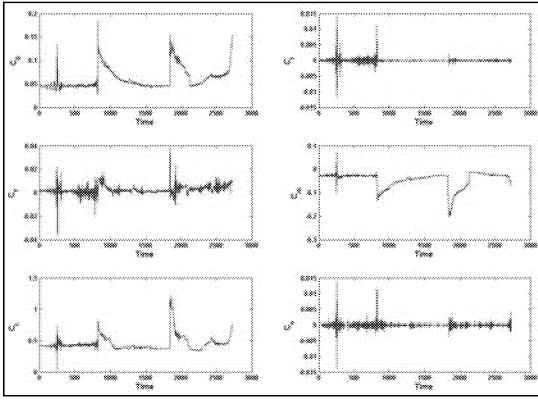


Figure 2: Aerodynamic coefficients variation for Flight segments #1 to #4

Figure 3 is the 3-dimensional wind field along the flight track. According to these calculated 3-dimensional wind field, the minimum variation of wind speed in the horizontal plane of navigation axis was  $-9.8$  knots.; while the maximum change was  $36$  knots.

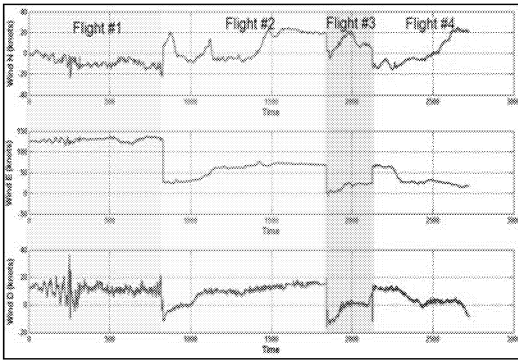


Figure 3: Three dimensional wind field results

### Aircraft Longitudinal Aerodynamic Derivatives

The aircraft longitudinal aerodynamic coefficients include the lift, drag, and pitching moment coefficients. In the system identification process, a quarter of the whole flight data will be used in the learning process, and the remaining flight data is used to validate whether the learning results could represent the aircraft flight dynamics. Furthermore, except for the identification of the lift coefficient related derivatives by MLPNN and pitching moment related derivatives, the elevator is assumed not to contribute to the lift and drag coefficients. Because the actual aerodynamic derivatives for the subject airplane are not available, the following results for the Convair 880 and Boeing 747 and listed in Table 2 (Nelson 1998) will be used for comparison.

Table 2 Some of the longitudinal aerodynamic derivatives of Convair 880 and Boeing 747

| Convair 880 |               |               |               |                     |          |                     |
|-------------|---------------|---------------|---------------|---------------------|----------|---------------------|
|             | $C_{L\alpha}$ | $C_{D\alpha}$ | $C_{m\alpha}$ | $C_{m\dot{\alpha}}$ | $C_{mq}$ | $C_{L\dot{\alpha}}$ |
| Ma=.8       | 4.8           | 0.15          | -0.65         | -4.5                | -4.5     | 0.19                |
| Boeing 747  |               |               |               |                     |          |                     |
| Ma=.9       | 5.5           | 0.47          | -1.6          | -1.2                | -25      | 0.3                 |

### Lift Coefficient Related Derivatives

The lift coefficient related derivatives are modeled by MLPNN and RBFNN. For the MLPNN method, the derivatives are assumed to be function of mach number, vertical wind's magnitude of time variation, rotational rates of body axes, elevator deflection, angle of attack and its rate, and sideslip angle and its rate, and the pitch trim adjustment, presented the operation of stabilizer(ds). The initial weighting for the MLPNN is randomly set, and it is a time consuming process to find suitable one to meet the non-linear flight dynamic for the subject airplane. After many trials and errors, the structure of MLPNN is chosen with 2 hidden layers. There are 10 neurons for the input to the first hidden layer, with linear active function. At the second layer, there are 3 neurons with tangential active function. The output layer has only one neuron with linear mapping from the second hidden layer. The correlation factor for the validation results to the target values is up to 98%. Some lift related derivatives, such as  $C_{L\alpha}$  (CL\_aoa)  $\cdot$   $C_{L\dot{\alpha}}$  (CL\_ds)  $\cdot$   $C_{L\ddot{\alpha}}$  (CL\_de), and  $C_{L\beta}$  (CL\_beta) are shown in Figure 4. The variation of the estimated results is larger than desirable, and the value of  $C_{L\dot{\alpha}}$  is negative, which are even physically incorrect.

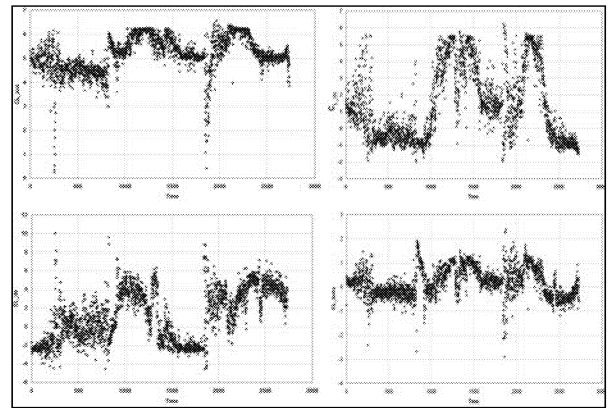


Figure 4: Some estimated results of lift coefficient related derivatives by MLPNN

Then, the study changes the structure of network from MLPNN to RBFNN, and including the time rates of  $p$ ,  $q$ ,  $r$  of body axes,  $(\dot{p}, \dot{q}, \dot{r})$ . For the optimization process of the RBFNN, the network has to update the centers of radial functions, and weightings. Similar to the MLPNN, these values also affect the learning results. The non-linearity of RBFNN is increasing as the numbers of center are increased, and the learning results will be mired in the local minimum of the learning error. Therefore, in this study, the initial value of centers of radial function and weightings are determined by the Fuzzy C-Means algorithm. The learning process of RBFNN has two stages, the first one is learning with one quarter of the whole flight data, and has 25 neurons. The second stage is focused on Flight segment #1, with 7 neurons. Some of the aerodynamic derivatives estimated results in the first stage are denoted as RBF-25, and those from the second stage are denoted as RBF-7. The above results are presented in Figure 5.

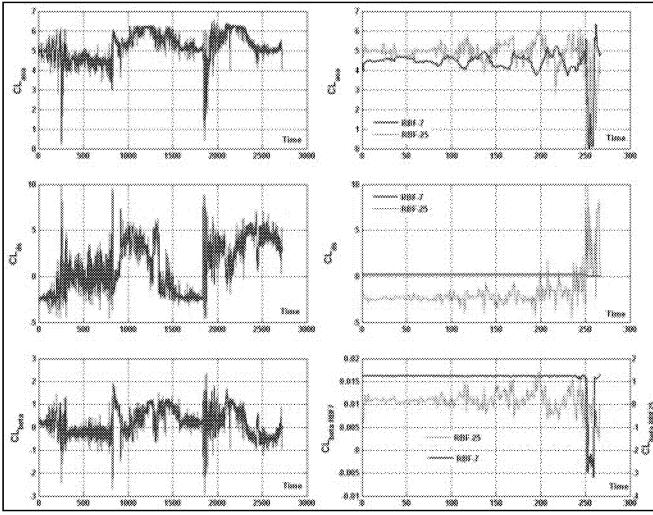


Figure 5: Some of the lift coefficient related aerodynamic derivatives estimated by RBFNN

In the period of Flight segment #1 when there was no turbulence encountered, the average values of  $C_{L\alpha}$  for RBF-25 and RBF-7 were 4.25 and 4.27 respectively. The average values of the  $C_{L\delta}$  at normal cruise condition for RBF-25 and RBF-7 were 0.222 and 0.189, and the average values of the  $C_{L\beta}$  for RBF-25 and RBF-7 were -0.017 and -0.016. The  $C_{L\alpha}$  for RBF-7 ranged from 5.17 to 0.41, but the value of RBF-25 varied from 10.15 to -1.31. The  $C_{L\delta}$  for RBF-7 varied from 0.189 to 0.001. The estimated results of RBF-7 are better than RBF-25, when the aircraft encountered turbulence. Compared to the results in Table 3, the values of  $C_{L\alpha}$  and  $C_{L\delta}$  for Convair 880 are 4.8 and 0.19 at Mach 0.8, and the values for Boeing 747 at Mach 0.9 are 5.5 and 0.33. Therefore, the estimated results are in the range for these two types of aircraft at cruise condition.

#### Drag Coefficient Related Derivatives

The input vector space, learning algorithm, initial setting and numbers of the center of radial based function and the weightings for drag coefficient related derivatives are the same as in the identification for lift coefficient related derivatives. The correlation factor for the target values and learning results are 89%. Some of the identification results for the drag coefficient related derivatives are presented in Figure 6. For Flight segment #1 before encountering turbulence, the average values of  $C_{D\alpha}$  for RBF-25,7 are 0.43 and 0.48. The average value of  $C_{D\delta}$  for RBF-25,7 are 0.09 and 0.001, and the average value of  $C_{D\beta}$  are -0.13 and 0.003. With the aircraft in atmospheric turbulence, the values of  $C_{D\alpha}$  varied in the range of 0.6 to 0.28 for RBF-7, and 2.85 to -3.49 for RBF-25. The value of  $C_{D\delta}$  and  $C_{D\beta}$  decrease as the turbulence is encountered. The value of  $C_{D\alpha}$  for convair 880 at Mach 0.8 and Boeing 747 at Mach 0.9 are 0.15 and 0.47 respectively. The estimated result for  $C_{D\alpha}$  for RBF-7 matches the Table 2, The drag coefficient contributing from pitch trim is not significant to the drag coefficient.

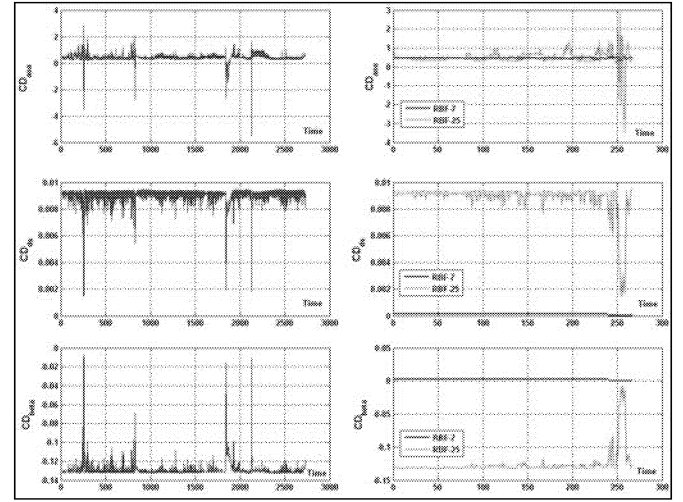


Figure 6: Some of the drag coefficient related aerodynamic derivatives estimated by RBFNN

#### Pitching Moment Coefficient Related Derivatives

The identification of the pitching moment related derivatives include the deflection of elevator. The network structure is RBFNN, but the initial setting for the center neurons and the weightings is changed to the orthogonal least square, instead of the Fuzzy C-means. The convergence for this learning scheme is slower, and need longer run time to model the pitching moment related derivatives. Figure 7 is some estimated pitching moment related derivatives. The average value of  $C_{m\alpha}$  was -0.337;  $C_{m\delta}$  was -2.79;  $C_{m\dot{\delta}}$  was -1.12, and  $C_{mq}$  was -5.56 at normal cruise condition. The value of  $C_{m\dot{\delta}}$  is -0.58 at Ma=0.8 for Convair 880, and the value for the Boeing 747 is -1.2. When the aircraft encountered turbulence, the values of  $C_{m\alpha}$ ,  $C_{m\delta}$ , and  $C_{m\dot{\delta}}$  ranged from -0.02 to -0.44, -2.92 to -0.05, and -1.64 to 0.11 respectively.

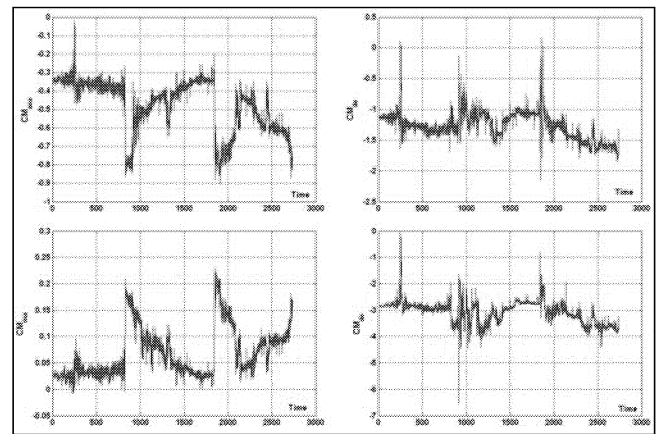


Figure 7: Some of the pitching moment coefficient related aerodynamic derivatives estimated by RBFNN

## CONCLUSIONS

In this study, the results indicated that the RBFNN is better than MLBPNN typically. For the RBFNN, the learning mechanism, based on weighting the distance between the

input and centers of radial based function to fit the desired output, is similar to the definition of quasi-steady aerodynamic derivatives, which are defined relative to the equilibrium flight condition. The centers of radial based function for various segments of flight data were similar to the equilibrium states at different flight phases or segments. From the above analysis, the aircraft encountered severe turbulence only at 240 to 270 sec. During this period of time, the worst degradation of  $C_{L\alpha}$ ,  $C_{m\alpha}$  and  $C_{L\delta}$  decreased up to 9.1%, 5.9% and 0.52%, compared to the steady cruise condition. The severe turbulence also reduced the pitching control effectiveness by up to 1.71%, and might even have positive  $C_{m\delta}$ . The positive pitching control effectiveness from elevator reminds the investigators about the possibility of the pilot induced oscillations when pilot encounter unexpected motion due to the elevator control. In addition, the degradation of the aerodynamic derivatives means that the performance and stability deviations from normal flight condition depicts the poor stability and handling quality when the aircraft encounter severe turbulence and we couldn't use the normal derivatives to do some performance analysis.

## ACKNOWLEDGEMENT

This study is supported by National Science Council, and the Aviation Safety Council, Taiwan. The project number is NSC97-3114-P-707-001-Y.

## REFERENCES

- Aviation Safety Council. 2005. *In-Flight Icing Encounter and Crash into the sea Transasia Airways Flight 791*, Aircraft Accident Report, Taipei, Taiwan (Apial).
- European Organization for Civil Aviation Equipment. 2003. *Minimum Operational Performance Specification for Crash Protected Airborne Recorder Systems ED112*. EUROCAE Working Groups 50, EUROCAE Office, Malakoff, France, (Jan).
- Iliff, Kenneth W. and Taylor Lawrence W, Jr. 1972. "Determination of Stability Derivatives from Flight Data Using a Newton-Raphson Minimization Technique", Technical report H-626, NASA Flight Research Center, C.A.(Mar).
- Iliff, Kenneth W. 1987. "Aircraft Parameter Estimation", Technical report H-1394, NASA Ames Research Center, Moffett, Fld. (Janu).
- Kenneth Hui, Lorenzo Auriti, and Joseph Ricciardi. 2005. "Advances in Real-Time Aerodynamic Model Identification", *Journal of Aircraft* Vol.42, No.1(Janu-Feb), 73-79 Journal.
- C-T, Weng, C-S. Ho, C.E. Lan, and M. Guan. 2006. "Aerodynamic Analysis of a Jet Transport in Windshear Encounter During Landing". *Journal of Aircraft*, Vol.43, No. 2(Mar-April), 419-427 Journal.
- Ming-Hao Yang, KC Chang, Michael Guan, Ching-Shun, Ho, and Fei-Bin, Hsiao. 2006. "Applied flight data to analyse the aircraft encounter turbulence". 2006 AASRC/CCAS Joint Conference(NCU, Zhongli City, Dec.10).
- G. Minkler and J.Minkler. 1993. Theory and Application of Kalman Filtering. Magellan Book Company. 160~260.
- Jack D. Mattingly, 1996. Elements of Gas Turbine Propulsion. McGraw-Hill, Inc. 213~337.
- Suykens, Johan A. K./Moor, and Bart L. R. de. 1996. *Artificial neural networks for modelling and control of non-linear systems*. Kluwer Academic Publishers.
- Gupta, Madan M., Jin, Liang., Homma, Noriyasu. 2003. *Static and dynamic neural networks :from fundamentals to advanced theory*. Wiley.
- Robert C. Nelson. 1998. *Flight Stability and Automatic Control*. McGraw-Hill, Singapore.

## BIOGRAPHY

**MING-HAO YANG** receives B.S. and M.S. degrees in Aeronautics and Astronautics Engineering from the National Cheng Kung University in Taiwan. Since graduating, he has joined the Investigation Laboratory of Aviation Safety Council as FDR and aircraft performance engineer. He has returned to the National Cheng Kung University to his Ph. D for his further study since 2006.

**CHING-SHUN HO** receives the Ph.D degree in aerospace engineering from the University of Texas at Austin, U.S.A.. Since 1990, he has been an associate professor of the department of aeronautic and astronautic engineering at the National Cheng Kung University, Taiwan. His main research interests are satellite navigation and flight safety analysis.

**CHUAN-TAU LAN** has taught Aerospace Engineering at the University of Kansas for 39 years. He has been appointed a Distinguished Professor since 1992. In 2007, he retired as a J. L. Constant Distinguished Professor Emeritus. Since 1970, he has been a Principal Investigation in NASA research projects in high angle-of-attack aerodynamics and flight dynamics. His current interests are flight safety issues and accident prevention related to aerial vehicles.

**FEI-BIN HSIAO** received Ph. D. degree in Aerospace Engineering from University of Southern California, USA. He joined the Institute of Aeronautics and Astronautics, National Cheng Kung University, in 1985 and was promoted to full Professor in 1990 and Distinguished Professor in 2004. He is a Fellow of American Institute of Aeronautics and Astronautics, corresponding member of International Academy of Astronautics (IAA), and Fellow of Aeronautical and Astronautical Society of Taiwan.

**CHIEN-CHUN KUNG** obtained his Ph. D. in Aeronautics and Astronautics engineering from the National Cheng Kung University in Taiwan. He is currently an assistant professor of aeronautical engineering at National Defense University, Chung Cheng Institute of Technology in Taiwan.

# **FLUID FLOW SIMULATION IN DESIGN**



# Numerical Simulation of Refrigerant Flow Through Adiabatic Capillary Tube: A Comparison between Homogeneous Equilibrium Model (HEM), Delayed Equilibrium Model (DEM) and Improved Delayed Equilibrium Model (IDEM)

Abdul Qaiyum Shaik

Henk Versteeg

Wolfson School of Mechanical and Manufacturing Engineering,

Loughborough University, Loughborough, LE11 3TU, UK

E-mail: A.Q.Shaik@lboro.ac.uk

## KEYWORDS

adiabatic capillary tubes, alternative refrigerants, metastable flow, numerical simulation, refrigeration and air-conditioning, two-phase flow.

## ABSTRACT

As the industries are moving towards the alternatives to HCFC-22, more accurate and general calculation tools are required to predict the capillary tube performance using the new refrigerants. In the present paper, three different methods: Homogeneous Equilibrium Model (HEM), Delayed Equilibrium Model (DEM) and Improved Delayed Equilibrium Model (IDEM) are used to study the behaviour of the new refrigerant mixture R410A through an adiabatic capillary tube. The numerical results were compared with the experimental results of Sanzovo and Mattos (2003). The comparison showed that good agreement exists between the experimental and numerical pressure and temperature profiles along the capillary tube. Also, it was noticed that IDEM predicts the mass flow rate better in compare with that of HEM and DEM.

## INTRODUCTION

A capillary tube is a simple tube of fixed length and smaller diameter, which is usually used as an expansion and refrigerant controlling device for small refrigerating systems and households refrigerators. Its simplicity, low initial cost and low starting torque of compressors are compelling reasons for its use. Many conventional refrigerants are being phased out from the industries as they are causing ozone depletion and global warming (Montreal Protocol, 1987). Because of this reason, the manufacturers have move to alternative refrigerants such as R-410A, a near-azeotropic mixture of HFC 32 and HFC 125 (50%/50% on mass basis) and R407C, a zeotropic mixture of HFC 32, HFC125 and HFC 134a (23%/25%/52% on mass basis). The thermodynamic characteristics of flows of CFC refrigerants through pipes and capillary tubes have been extensively studied. However, there is still a need for more accurate and general calculation tools to support the design of equipment using the new refrigerants.

A variety of theoretical and numerical models have been reported in the literature on the capillary tube behaviour with pure CFCs and HCFCs refrigerants such as CFC-12, HCFC-22 and R134A (Cooper et al., 1957; Mikol,1963; Koizumi

and Yokoyama, 1980; Li et al. 1990a, 1990b; Escanes et al., 1995; Bansal and Rupasinghe, 1998; Wong and Ooi, 1996a 1996b). A hunt for potential replacements for CFCs has led to the development of new alternatives such as pure HFCs, binary and ternary azeotropic as well as zeotropic HFC's refrigerant mixtures. Various models have been used to model the flow through capillary tubes using these new refrigerants. For example Sami and Tribes (1998) and Sami et al. (1998) developed a numerical model to predict the performance of capillary tube for zeotropic and azeotropic binary and ternary mixtures using Colebrook's friction factor. Numerical results showed that the proposed model fairly simulated their experimental data. Sanzovo and Mattos (2003) numerically studied the refrigerant mixture flow through adiabatic capillary tubes using two different models (homogeneous equilibrium model and separated flow model). They validated these models against their experimental results. All of the above models for the refrigerant mixtures neglected the metastable flow region. Li et al. (1990a) experimentally studied the metastable flow of refrigerant R12 through adiabatic capillary tube and reported that the flashing flow of refrigerant through capillary tube consists of for different regions: subcooled liquid region, metastable liquid region, two-phase metastable region and two phase equilibrium region (Figure 1). Garcia-Valladares (2004) numerically studied the flow of refrigerant mixtures using the separated flow model considering the metastable regions. The delayed equilibrium model proposed by Feburie

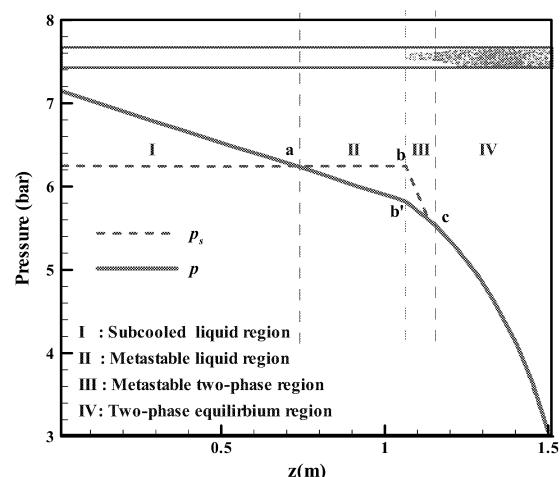


Figure 1 Typical pressure distribution along an adiabatic capillary tube

et al. (1993) was used to model the two-phase flow region. Recently, Vacek and Vins (2007) developed a numerical model to study the behaviour of capillary tubes in cooling circuits operating under either adiabatic or non-adiabatic conditions with saturated fluorocarbon refrigerants neglecting the metastability using two different models: homogeneous equilibrium and separated flow model. These refrigerants have several unique properties like high dielectric performance, chemical stability and radiation resistance. The reliability of the numerical was verified both with available data from the literature, for traditional refrigerants, and with their own experimental data using  $C_3F_8$ .

Although the DEM predicts the mass flow rate well with the experimental data for steam-water flows, it under predicts the mass flow rate for inlet conditions close to saturated liquid (Attou and Seynhaeve, 1999a). In the present paper, a new Improved Delayed Equilibrium Model (IDEM) outlined by Attou and Seynhaeve (1999a and 1999b) is used along with the other existing models (HEM and DEM) to predict the behaviour of refrigerant R410A through an adiabatic capillary tube. The numerical results are compared against the experimental data (Sanzovo and Mattos, 2003).

## MATHEMATICAL MODELLING

The basic governing equations are the one-dimensional conservation equations of mass, momentum and energy. The following assumptions are made in modelling the flow:

- One-dimensional adiabatic flow
- Straight horizontal, constant inner diameter and uniform surface roughness
- No slip between the phases ( $U_l = U_v = U$ )
- The effect of surface tension is neglected ( $p_l = p_v = p$ )
- Metastable region is included in the model
- Refrigerant is free from oil.

Metastability effects are included in the model using the Delayed Equilibrium Model (DEM) and Improved Delayed Equilibrium Method (IDEM) outlined by Feburie et al. (1993) and Attou and Seynhaeve (1999a and 1999b) respectively. The basics of the method can be best understood by considering the conceptual capillary tube sketched in Figure 2.

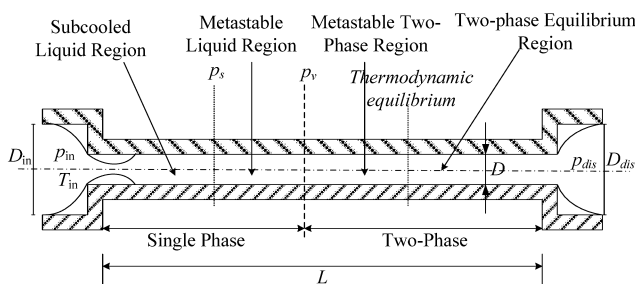


Figure 2 Schematic view of an adiabatic capillary tube

The fluid at the entrance of the capillary tube is in a known state (i.e. sub-cooled/ saturated / two-phase state). When the local pressure  $p < p_s$  (saturation pressure), the refrigerant enters into the metastable-liquid region. At  $p = p_v$

(vaporisation pressure), inception of vaporization takes place and the refrigerant enters into the metastable two-phase region. The under-pressure of vaporisation ( $p_v$ ) associated with the delay of vaporisation was evaluated using the correlation proposed by Chen et al. (1990) for sub-cooled liquid inlet and Lackme (1979) for saturated and two-phase inlet conditions respectively. In the metastable two-phase region the DEM and IDEM proposed by Feburie et al (1993) and Attou and Seynhaeve (1999a and 1999b) are used respectively. According to these models, evaporation does not happen instantaneously, but only a fraction  $y$  (the so-called vaporization index) of the refrigerant is transformed into saturated mixture, the other fraction  $(1-y)$  remains metastable liquid and is submitted to an isentropic evolution. The vaporisation index ' $y$ ' is a measure of the departure from local saturated state. If  $y=1$ , the fluid is saturated mixture and if  $y=0$ , the fluid is metastable liquid. In the flow direction ' $z$ ' a conversion of metastable fluid into saturated fluid will take place, which is evaluated using the following relaxation equations:

DEM:

$$\frac{dy}{dz} = 0.02 \frac{P}{A} (1-y) \left[ \frac{p_s - p}{p_c - p_s} \right]^{0.25} \quad 1$$

IDEM:

$$\frac{dy}{dz} = 0.01 \frac{P}{A} (1-y)^2 \left( \frac{U_{in}}{U} \right)^{1/10} \left[ \frac{p_s - p}{p_c - p_s} \right]^{1/4} \quad 2$$

where  $P$  is perimeter,  $A$  is area,  $U_{in}$  is superficial velocity at the inlet,  $p_s$  is saturated pressure and  $p_c$  is the critical pressure.

All the properties are evaluated using REFPROP v.7.0 (2002). The refrigerant properties for different regions are evaluated as follows:

*Subcooled liquid region* : fluid properties are estimated as a function of pressure and temperature

*Metastable liquid region* : fluid properties are estimated as a function of saturated pressure.

*Metastable two-phase region*: fluid properties are estimated as a function saturated pressure and metastable liquid properties are estimated as function of pressure and entropy

*Two-phase equilibrium region*: fluid properties are estimated as a function of saturated pressure.

HEM

For HEM, vaporisation is instantaneous. Vaporisation occurs at  $p = p_s$ . The phases are in thermal equilibrium. This model ignores slip effects and considers the two phases are being intimately mixed.

Pipe friction losses within the capillary tube are evaluated using Churchill's friction factor expression [cited by Lin et al., 1991] in conjunction with Dukler's viscosity model [cited by Zhou and Zhang, 2006]. The pressure drop associated with a sudden contraction for sub-cooled inlet is evaluated using equations given in ESDU report. (ESDU-05024). Where as the pressure drop associated with two-phase inlet is evaluated assuming frozen flow across the sudden contraction. As the liquid passes through the sudden



contraction some part of the saturated liquid gets converted into metastable liquid, which is evaluated using the mean enthalpy equation. The pressure recovery associated with a sudden expansion at the exit for two-phase is evaluated using equations given in ESDU report (ESDU-89012). The flow is critical when the refrigerant velocity reaches sonic velocity and the two phase mach number is unity.

## NUMERICAL SOLUTION

The numerical solution scheme subdivides the domain into 'N' control volumes (CV). Owing to high gradients produced at the end of the capillary tube, a non-uniform grid is generated according to the expression given by Escanes et al. (1995). For each CV, a set of algebraic equations is obtained by the discretization of the governing equations. The values of the flow variables at the CV outlet section are obtained by solution of this set of equations, from the known values at the CV inlet section and the "current" mass flow rate. The solution procedure marches forward in the flow direction from the entrance to the exit of the capillary tube. At this point, the calculations are terminated and the computed discharge pressure is compared with the actual value. If the difference is greater than a preset tolerance the mass flow rate is adjusted and the solution procedure is restarted at the entrance of the tube until the calculated value agrees with the actual exit pressure. The grid density was systematically varied to examine its effect on simulation results; grid independence was achieved with N (Total number of nodes) = 400 in all cases. Visual 'C' has been used to write the computer program.

## TEST CASES

The experimental data given by Sanzovo and Mattos (2003) with refrigerant R410A were used as test cases for the present study. The dimensions of the insulated capillary tube are : D (internal diameter) = 1.101 mm and L (length) = 1.5 m. The relative roughness ( $\epsilon/D$ ) =  $2.354 \times 10^{-4}$ . The internal diameter of the connecting tubes (to calculate the pressure drop in the inlet contraction and outlet expansion) is 10.0 mm. The cases were selected in such a way that they could cover a large range of inlet conditions (i.e. sub-cooled/saturated/two-phase). Case 1 in Table 1 and Table 2 was selected to compare the pressure and temperature profiles against the experimental results along the capillary tube for sub-cooled and two-phase inlet conditions, respectively. Cases 2a, 2b and 2c in Table 1 (sub-cooled temperature varying from 1.4 to 5.6) and Table 2 (quality varying from 0 to 0.04) are selected to compare the mass flow rate with the experimental data as the inlet condition changes from sub-cooled to two-phase flow.

## RESULTS AND DISCUSSION

Table 1 shows the comparison of mass flow rates using HEM, DEM and IDEM with the experimental mass flow rates of Sanzovo and Mattos (2003) for R410A with sub-cooled inlet conditions. From the table it is clear that IDEM predicts the mass flow rate better in comparison to that of

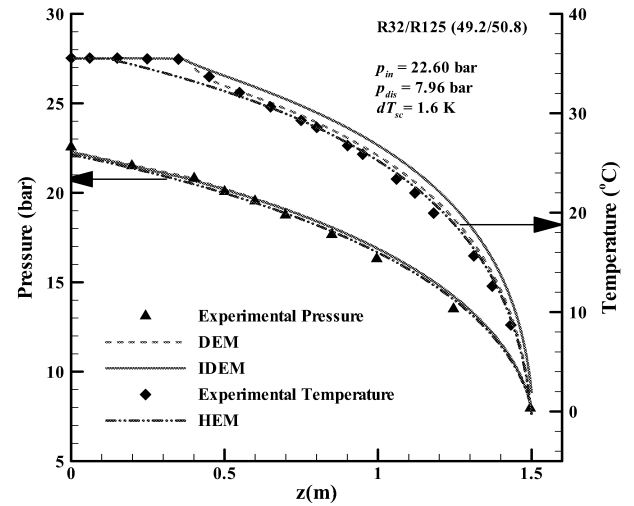


Figure 3 Comparison of pressure and temperature distribution along the capillary tube for R410A: subcooled inlet conditions

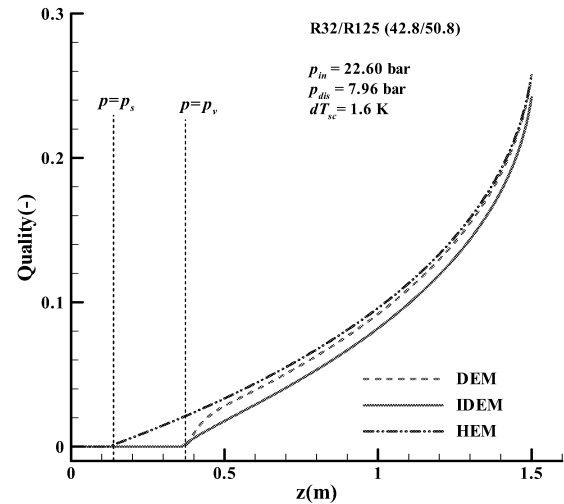


Figure 4 Comparison of quality distribution along the capillary tube for R410A: subcooled inlet conditions

Table 1 Comparison of mass flow rate for R-410A : Subcooled Inlet conditions (Sanzovo and Mattos, 2003) (D=1.101 $\times 10^{-3}$ m, L= 1.5m and  $\epsilon/D=2.354 \times 10^{-4}$ )

| Case | Refrigerant |             | $p_{in}$<br>(bar) | $\Delta T_{sub}$<br>(K) | $p_{dis}$<br>(bar) | $\dot{m}_{exp}$<br>(kg/h) | HEM                     |              | DEM                     |              | IDEM                     |              |
|------|-------------|-------------|-------------------|-------------------------|--------------------|---------------------------|-------------------------|--------------|-------------------------|--------------|--------------------------|--------------|
|      | R-32<br>(%) | R-25<br>(%) |                   |                         |                    |                           | $\dot{m}_{HEM}$<br>kg/h | Error<br>(%) | $\dot{m}_{DEM}$<br>kg/h | Error<br>(%) | $\dot{m}_{IDEM}$<br>kg/h | Error<br>(%) |
| 1    | 49.2        | 50.8        | 22.60             | 1.6                     | 7.94               | 22.93                     | 21.45                   | -6.45        | 21.89                   | -4.54        | 22.46                    | -2.04        |
| 2a   |             |             | 24.33             | 5.6                     | 796                | 26.07                     | 24.81                   | -4.82        | 25.27                   | -3.07        | 25.93                    | -0.53        |
| 2b   | 48.5        | 51.5        | 24.34             | 3.3                     | 796                | 24.97                     | 23.49                   | -5.92        | 23.95                   | -4.09        | 24.56                    | -1.65        |
| 2c   |             |             | 24.31             | 1.4                     | 797                | 24.05                     | 22.31                   | -7.23        | 22.79                   | -5.24        | 23.36                    | -2.85        |

HEM and DEM with a maximum error of -2.85%. Figure 3 shows the comparison of pressure and temperature profiles along the capillary tube. The evaluated pressure and temperature profiles are in good agreement with the experimental pressure and temperature profiles. The predicted pressure profiles are almost identical for all the models, but there are some differences in the predicted temperature profiles. Figure 4, shows the distribution of quality along the capillary tube for the three models. It can be noticed that for HEM vaporisation of refrigerant takes place at  $p = p_s$  whereas for DEM and IDEM, metastability causes the vaporisation to be delayed until  $p = p_v$ . As one can see from the figure once the vaporisation is started, the DEM predictions rapidly approach those of the HEM, indicating that the metastable model is predicted to reach equilibrium within a short distance of 0.4 m. This causes a noticeable kink in the predicted temperature and quality profiles for DEM. The IDEM predictions tend to the HEM curve less rapidly, indicating a more gradual return to equilibrium.

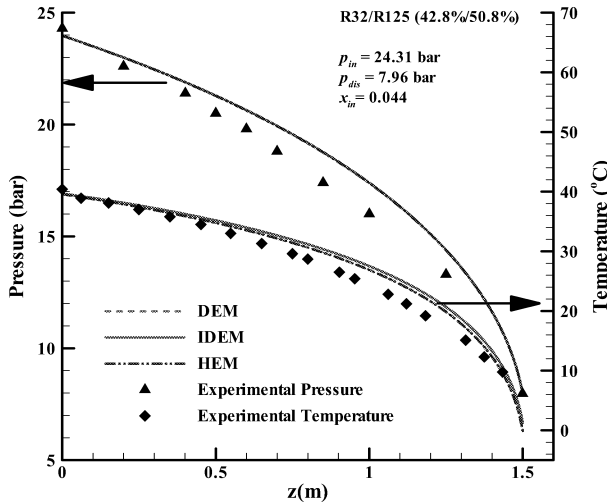


Figure 5 Comparison of pressure and temperature distribution along the capillary tube for R410A: Two-phase inlet condition

Table 2 shows the comparison of mass flow rate using the three different models for two-phase flow inlet condition. From the table, again it can be noticed that IDEM predicts the mass flow rate well compared to HEM and DEM, especially close to the saturated conditions. Figure 5 shows the distribution of pressure and temperature profiles along the capillary tube using the three different models. Again, a

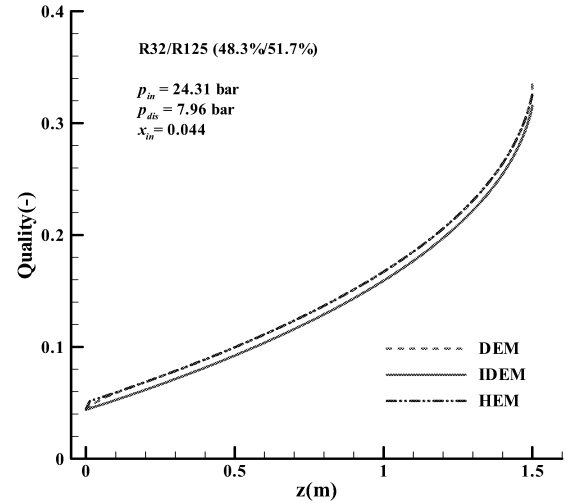


Figure 6 Comparison of quality distribution along the capillary tube for R410A: Two-phase inlet conditions

good agreement has been achieved with the pressure profiles. However, there are some discrepancies in temperature profiles. Figure 6 shows the distribution of quality along the capillary tube using the three different models. From the figures it can be observed that there is no significant difference between HEM and DEM as there is no delay in vaporisation because of the two-phase flow at the inlet.

In this set of test cases, as well as those with subcooled inlet conditions discussed earlier, the differences between DEM and IDEM predictions is due to the evaluation of vaporisation index. The coefficients in equation (1) and (2) dictates the rate of vaporisation along the z-direction. The constant coefficient of DEM equation (1) and the factor  $(1-y)$  are greater than the corresponding constant and the factor  $(1-y)^2$  in IDEM equation (2), hence the DEM predicts a more rapid return to the equilibrium state than IDEM. This is responsible for the kink in the temperature and quality profiles predicted by the DEM for subcooled inlet conditions (Figure 4). A much less noticeable kink also occurs in the results for two-phase inlet condition (Figures 5 and 6).

Table 2 Comparison of mass flow rate for R-410A : Two-phase Inlet conditions (Sanzovo and Mattos, 2003)  
( $D=1.101 \times 10^{-3}$  m,  $L= 1.5$  m and  $\varepsilon/D=2.354 \times 10^{-4}$ )

| Case | Refrigerant |             | $p_{in}$<br>(bar) | $x_{in}$<br>(K) | $p_{dis}$<br>(bar) | $\dot{m}_{exp}$<br>(kg/h) | HEM                     |              | DEM                       |              | IDEM                       |              |
|------|-------------|-------------|-------------------|-----------------|--------------------|---------------------------|-------------------------|--------------|---------------------------|--------------|----------------------------|--------------|
|      | R-32<br>(%) | R-25<br>(%) |                   |                 |                    |                           | $\dot{m}_{HEM}$<br>kg/h | Error<br>(%) | $\dot{m}_{DEM}$<br>(kg/h) | Error<br>(%) | $\dot{m}_{IDEM}$<br>(kg/h) | Error<br>(%) |
| 1    | 48.3        | 51.7        | 24.31             | 0.044           | 7.96               | 21.78                     | 19.94                   | -8.44        | 19.95                     | -8.41        | 20.27                      | -6.93        |
| 2a   |             |             | 26.07             | 0               | 796                | 23.94                     | 22.37                   | -6.55        | 22.89                     | -4.38        | 23.38                      | -2.35        |
| 2b   | 48.5        | 51.5        | 26.06             | 0.02            | 796                | 23.21                     | 21.97                   | -5.35        | 21.97                     | -5.35        | 22.35                      | -3.69        |
| 2c   |             |             | 26.07             | 0.04            | 797                | 22.85                     | 21.16                   | -7.40        | 21.16                     | -7.38        | 21.50                      | -5.90        |

## CLOSING REMARKS

Three different models (HEM, DEM and IDEM) were used to study the flow of alternative refrigerant R410A along an adiabatic capillary tube. All three models gave reasonable predictions, but our results showed that IDEM predicts the mass flow rate better compared with HEM and DEM, especially close to the saturated conditions at the entrance of the capillary tube. This result is similar to that of Attou and Seynhaeve (1999a and 1999b). The numerical predictions of pressure and temperature profiles for the sub-cooled inlet conditions using the three models were generally in good agreement with experimental results. The predictions were less accurate for two-phase inlet conditions, which may be attributable to limitations in the validity of the modelling assumptions (e.g. no slip between the phases) or to uncertainties in the experimental data. Overall, the findings of this study suggest that a good account of metastability is an important feature of accurate predictions of industrial refrigerant flows in the capillary tubes with the new refrigerant mixtures.

## ACKNOWLEDGEMENTS

Abdul Qaiyum Shaik gratefully acknowledges a scholarship from Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, UK for supporting his work.

## NOMENCLATURE

|           |  |
|-----------|--|
| $A$       | Cross sectional area ( $\text{m}^2$ )              |
| $D$       | Diameter (m)                                       |
| $L$       | Length (m)   |
| $\dot{m}$ | mass flux ( $\text{Kg/s}$ )                        |
| $N$       | Number of nodes                                    |
| $p$       | Pressure (Pa)                                      |
| $p_s$     | Saturation pressure (Pa)                           |
| $p_v$     | Pressure of vaporization (Pa)                      |
| $P$       | Perimeter (m)                                      |
| $T$       | Temperature (K)                                    |
| $U$       | Velocity (m/s)                                     |
| $x$       | Mass fraction ( Vapour quality)                    |
| $y$       | Mass ratio of total saturated phase to total phase |
| $z$       | Axial coordinate (m)                               |

### Greek letters

|                 |  |
|-----------------|--|
| $\varepsilon/D$ | Relative roughness of the capillary tube |
|-----------------|--|

### Subscripts

|     |            |
|-----|------------|
| c   | critical   |
| dis | discharge  |
| in  | inlet      |
| l   | liquid     |
| s   | saturation |
| v   | vapour     |

## REFERENCES

- Attou, A. and J.M. Seynhaeve. 1999a. "Steady-state critical two-phase flashing flow with possible multiple choking phenomenon Part 1: Physical modelling and numerical procedure". *Journal of Loss Prevention in the Process Industries*. Vol. 12, 335-345.
- Attou, A. and J.M. Seynhaeve. 1999b. "Steady-state critical two-phase flashing flow with possible multiple choking phenomenon Part 2: comparison with experimental results and physical interpretations". *Journal of Loss Prevention in the Process Industries*. Vol. 12, 347-359.
- Bansal, P.K., and A.S. Rupasinghe. 1998. "An homogeneous model for adiabatic capillary tubes". *Applied Thermal Engineering*. Vol. 18, Nos 3-4, 207-219.
- Chen, Z.H., R. Li, S. Lin, and Z.Y. Chen. 1990. "A correlation for metastable flow of refrigerant 12 through capillary tubes". *ASHRAE Transactions*. Vol. 96, No. 1, 550-554.
- Cooper L, C.K. Chu and W.R. Briskin. 1957. "Simple selection method for capillaries derived from physical flow conditions". *Refrigerating Engineering*. Vol. 65, 37-41.
- Engineering Sciences Data Unit (ESDU) Report 05024. "Flow through sudden contractions of duct area: pressure losses and flow characteristics".
- Engineering Sciences Data Unit (ESDU) Report 89012. 1989 "Two-phase flow pressure losses in pipeline fittings. Fluid Mechanics Internal Flow." Vol. 5b. ISBN 9780856796838.
- Escanes, F., D. Perez-Segarra, , and A. Oliva. 1995. "Numerical simulation of capillary-tube expansion devices". *International Journal of Refrigeration*. Vol. 18, No. 2, 113-122.
- Feburie, V., M. Giot, S. Granger and J.M. Seynhaeve. 1993. "A model for choked flow through cracks with inlet subcooling". *Int. J. Multiphase Flow*. Vol. 19, No. 4, 541-562.
- Garcia-Valladares, O. 2004. "Review of numerical simulation of capillary tube using refrigerant mixtures". *Applied Thermal Engineering*. Vol. 2004, 949-966.
- Koizumi, H. and K. Yokohama. 1980. "Characteristics of refrigerant flow in a capillary tube". *ASHRAE Transactions, Part 2*. Vol. 86, 19-27.
- Lackme, C. 1979. "Incompleteness of the flashing of a supersaturated liquid and sonic ejection of the produced phases". *Int. J. Multiphase Flow*. Vol. 5, 131-41.
- Li, R.Y., S. Lin, Z.Y. Chen and Z.H. Chen. 1990a. "Metastable flow of R-12 through capillary tubes". *International Journal of Refrigeration*. Vol. 13, No. 3, 181-186.
- Li, R.Y., S. Lin, and Z. H Chen. 1990b. "Numerical modelling of thermodynamic non-equilibrium flow of refrigerant through capillary tubes". *ASHRAE Trans*. Vol. 96, 542-549.
- Lin, S., C.C.K Kwok., R.Y Li, Z.H Chen and Z.Y Chen. 1991. "Local frictional pressure drop during vaporization of R-12 through capillary tubes". *Int. J. Multiphase Flow*. Vol. 17, No. 1, 95-102.
- Mikol, E. P. 1963. "Adiabatic single and two-phase flow in small bore tubes". *ASHRAE J*. Vol. 5, 75-86.

- Montreal Protocol. 1987. *Treaty Series*, 1990, No. 19 HMSO.
- REFPROP v. 7.0. 2002. "NIST Thermodynamic properties of refrigerants and refrigerant mixtures database". *Standard Reference Data Program*. Gaithersburg, Maryland 20899, USA, August 2002.
- Sami, S.M., and C. Tribes. 1998. "Numerical prediction of capillary tube behaviour with pure and binary alternative refrigerants". *Applied Thermal Engineering*. Vol. 18, No. 6, 491-502.
- Sami, S.M., B. Poirier and A.B. Dahamani. 1998. "Modelling of capillary tubes behaviour with HFC 22 Ternary Alternative Refrigerants". *International Journal of Energy and Research*. Vol. 22, 843-855.
- Sanzovo, F.A., and O. Mattos. 2003. "Refrigerant mixtures flow through capillary tubes: a comparison between homogeneous and separated-flow models". *HVAC & R Research*. Vol. 9, 33-53.
- Vacek, V and V. Vins. 2007. "A study of the flow through capillary tubes tuned for a cooling circuit with saturated fluorocarbon refrigerants". *International J of Thermophysics*. Vol. 28, 1490-1508.
- Wong, T. N., and K.T. Ooi. 1996a. "Evaluation of capillary tube performance for CFC-12 and HFC-134a". *International Communication in Heat and Mass Transfer*. Vol. 23, 993-1001.
- Wong, T. N., and K.T. Ooi. 1996b. "Adiabatic capillary tube expansion devices: A comparison of the homogeneous flow and the separated flow models". *Applied Thermal Engineering*. Vol. 16, No. 7, 625-634.
- Zhou, G and Y. Zheng. 2006. "Numerical and experimental investigations on the performance of coiled adiabatic capillary tubes." *Applied Thermal Engineering*. Vol. 26, No. 11-12, 1106-1114.

# SIMULATION BASED OPTIMIZATION OF MICRONEEDLE GEOMETRY TO IMPROVE DRUG PERMEABILITY IN SKIN

Ololade Olatunji<sup>1</sup>, Diganta B Das<sup>1</sup>, Barrak Al-Qallaf<sup>2</sup>

<sup>1</sup>Chemical Engineering Department, Loughborough University, Loughborough LE11 3TU, UK

<sup>2</sup>Department of Engineering Science, Oxford University, Oxford OX1 3PJ, UK

(Email: [O.Olatunji@lboro.ac.uk](mailto:O.Olatunji@lboro.ac.uk))

## ABSTRACT

Microneedles are a method of transdermal drug delivery which overcomes the barrier of the stratum corneum by piercing through with little or no pain. Microneedle have been shown to increase the skin permeability but the extent to which they do so is affected by the design of the microneedles due to certain parameters such as tip radius, base radius, pitch, number of microneedles on the patch and penetration depth. In this study a framework is developed to optimize these parameters and an optimum geometry for a microneedle shape is proposed. Simulations using two different softwares were then carried out to analyze the effect of these parameters on the diffusive flux and concentration of the drug in the blood. We found that the results of the numerical optimization and the results from the computer simulations led to the same conclusion about the optimum dimensions for the chosen geometry.

## INTRODUCTION

Microneedles have been introduced as a means of delivering drug into the body transdermally in an almost painless manner. These micron sized needles pierce the upper layer of the skin (i.e. stratum corneum) which acts as the main barrier to transdermal delivery of drugs. The needles should extend into the epidermis but are not expected to touch the nerves in the dermis or the subcutaneous tissues. The drug diffuses across the rest of the epidermis by diffusion and then into the dermis where it is absorbed by the blood vessels (Henry et al, 1998).

Microneedles can be either hollow or solid. Solid microneedles can either have the drugs coated on (Gill and Prausnitz, 2006) or the drug can be encapsulated in a polymer matrix that forms the needle to form dissolving microneedles (Lee et al, 2008) or the poke with patch approach where the needles are inserted and removed then followed by application of a patch on the surface (Martanto et al, 2004). Hollow microneedles have holes in them through which the drug travels to get to the dermis (Stoeber and Liepmann, 2005). As far as the author

knows the first published study on the fabrication and use of microneedles for transdermal drug delivery was by Henry et al (1998).

Gill and Prausnitz (2006) developed a simple, versatile and controlled microneedle coating process. The microneedles were fabricated using the laser cutting technique and coated using the novel dip coating method they developed. They were able to uniformly coat a range of substances such as calcien, vitamin B and plasmid DNA onto microneedles using this coating method. They later in 2007 used the same method to develop a rational basis for designing coating solution formulations for uniform and thick coatings (Gill and Prausnitz, 2007). They analysed the factors that affected the coating thickness, coating concentration and uniformity. Such factors include concentration of surfactant and viscosity enhancer, type of surfactant and viscosity enhancer and number of dips. McAllister et al (2003) used microneedles to deliver insulin to hairless rats. Microneedles were shown to deliver insulin at a rate good enough to have significant biological effect as the blood glucose concentration was reduced by up to 80% using microneedles.

Transdermal drug delivery by microneedles in general has many advantages over oral delivery and delivery using hypodermic needles. Drugs delivered transdermally avoid the harsh conditions in the gastro-intestinal tract implying that the drug has a higher bioavailability. Also delivering drugs transdermally provides a controlled release rate and maintains a therapeutic concentration of the drug in the blood over long periods. This is important for drugs such as insulin where the patient might require a basal rate throughout the day (Tojo, 2005)

The structure of the skin has been outlined by Henry et al (1998). In transdermal delivery of drugs especially those of high molecular weight such as insulin, the major barrier is the permeation across the stratum corneum of the upper skin layer. Microneedles pierce through this barrier and the

diffusion of the drug through the lower layers of the skin is much faster than in the stratum corneum (Henry et al, 1998). Also it has been shown that the drugs delivered in smaller quantities are more easily absorbed than the drugs delivered in large boluses (John and Gareth 1997). Therefore compared to delivering a relatively large amount at a time using hypodermic needles, using microneedles is expected to result in better absorption into the blood stream.

It was suggested that the needle design has an impact on how well the microneedles deliver drugs (Teo et al, 2006). Most of the previous studies on microneedles have been focused on developing fabrication methods and determining performance of the microneedles in vitro and in vivo. Not a lot of study has been done on the effects of the microneedle dimensions on permeation through the skin. This study is concerned with the microneedle design and the effect of this design on permeability, it aims at finding an optimum geometry for tapered microneedles with circular base and then analysing how a number of geometrical parameters affect the drug delivery. This is done by using numerical simulations to optimize a developed equation that relates the permeability to parameters such as radius and needle length. Then using computer simulations, the effect of these parameters on the diffusive flux and blood drug concentration profile was analysed.

Similar studies have been done by Al-Qallaf et al (2008). The distribution of the microneedles in a patch was optimized based on the top view of the microneedle when inserted in skin. However for effective design of these systems, one needs to optimise both the distribution and geometry of the microneedles. The focus of this paper is to optimise the geometry of the microneedles.

## METHODOLOGY

### Theoretical Model for Optimization

#### Governing equation

Envisage a cross section through the centre of a row of microneedle inserted into a portion of the skin as shown in figure 1.

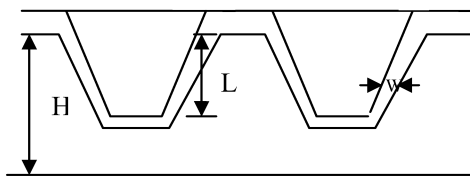


Figure 1. Diagram of array inserted in skin

H: epidermis thickness; L: Depth to which the needle penetrates into skin; W: annular gap width.

(Annular gap width at any point along the needle length) =  $\varepsilon \times$  (radius at that point)

$\varepsilon$  is a constant having a value of 0.025 McAllister (2003).

First the analogies on how the drug gets from the microneedle coating into the blood are explained. When the microneedle is inserted into the skin;

- The skin fluid (mostly water) fill up the annular space between the needle wall and the surrounding skin
- The drug coating on the microneedle surface dissolves into the surrounding fluid that occupies the annular gap
- The dissolved drug then diffuses across the epidermis, past the epidermal/dermal junction, into the dermis where it is absorbed by the blood microcirculation.
- The drug will dissolve into the surrounding skin fluid until the solubility limit is reached or until all the drug on the coating is diffused
- Knowing the annular gap width and the solubility of the coated drug, wastage can be avoided by coating up until the solubility limit of the drug so that no drug remains on the needle.
- So since the annular gap width is filled by fluid into which the drug dissolves and the total area of the annular gap width affects permeability this means geometrical parameters such as radius and needle length can be related to permeability.

These analogies are deduced from previous studies and publications on skin structure (Stalheim-Smith and Fitch, 1993), the presence of water in the epidermis (Machin et al 1984), microneedle application method (Henry et al, 1998) significance of the solubility limit of the permeant (Hadgraft, 2004) and McAllister (2003)'s explanation of the annular gap width.

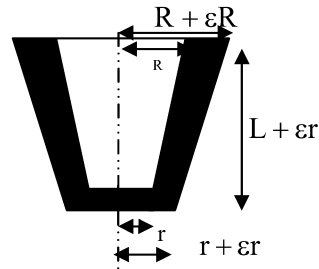


Figure 2. Side view of transverse section through one microneedle including annular gap width.

Where R is the base radius of microneedle, r is the tip radius and  $P_t$  is the pitch or spacing between needles.

Total area =

$$2 \left[ \frac{1}{2} ((R + \varepsilon R) - (r - \varepsilon r)) \times (L + \varepsilon L) \right] + 2(r + \varepsilon r)(L + \varepsilon L) \quad (1)$$

Area of unshaded region

$$= L(r + R) \quad (2)$$

Subtracting (2) from (1) gives area of annular space

$$= \varepsilon(LR + rL) + (\varepsilon + \varepsilon^2)(Rr + r^2) \quad (3)$$

This gives f, fractional skin area with microneedle inserted as

$$f = \frac{n(\varepsilon(LR + rL) + (\varepsilon + \varepsilon^2)(Rr + r^2))}{P_t n H} \quad (4)$$

The pitch, centre to centre spacing of microneedles, can be defined as

$$P_t = \frac{x}{n} \quad (5)$$

Where  $x$  is the length of one side of a square array and  $n$  is the number of microneedles on a row. Putting equation (4) into the permeability equation used by McAllister (2003) shown in equation (6).

$$K = f \frac{D}{H} \quad (6)$$

Where  $K$  is the permeability and  $D$  is diffusion coefficient. Then taking out the constants we have an optimization function given in (7).

$$\frac{KH^2}{\varepsilon D} = g = \frac{(LR + rL) + (\varepsilon + 1)(Rr + r^2)}{P_t} \quad (7)$$

The epidermis thickness can be included alongside a constraint;

$$L \leq H - z \quad (8)$$

Where  $z$  is an arbitrary value that ensures a distance between the needle tip and the dermis in this study  $z$  is taken as 0.005cm and  $H$  is 0.015cm. Therefore for a given skin thickness, we can also find the optimum penetration length. And the optimization function can now be given as,

$$g = \frac{n(R + r)(L + rc)}{xH^2} \quad (9)$$

Where  $c = \varepsilon + 1$

The other two constraints used are shown in equation (10) and (11);

$$x/n \geq \sigma R \quad (10)$$

This same constraint has been used by Al-Qallaf and Das (2008) in their optimization. This ensures that the needles do not overlap in the array.  $\sigma$  is the aspect ratio i.e. the ratio of pitch to radius and is taken to be 2.7 in this case as previous study already shows this to be the best aspect ratio.

$$r \leq R \quad (11)$$

This is to ensure that the tip radius is never bigger than the base radius.

The optimization function indicates the extent of permeability. A high value of  $g$  indicates a high permeability and a low value indicates a low permeability.

#### Method of solution

This equation is solved using a java optimization programme that has been reported by Al-Qallaf and Das (2008). The programme solves the problem by iterating through the whole sample spaces for each parameter before selecting combinations that gives the highest value of  $g$ .

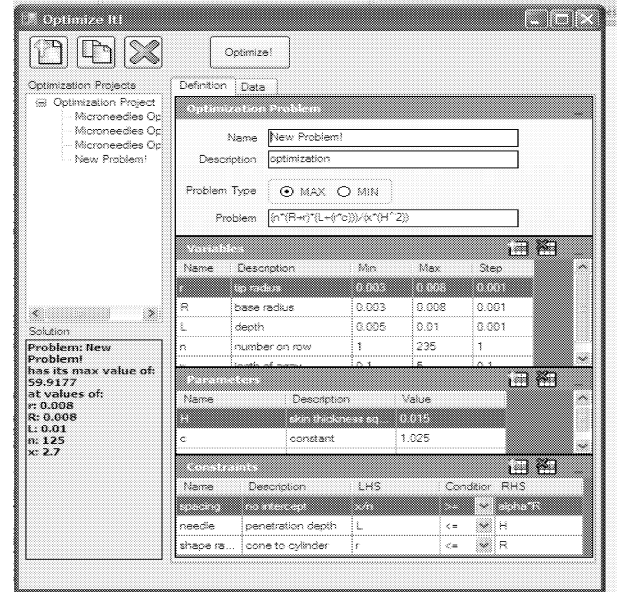


Figure 3. The Graphical User Interface (GUI) of Optimization programme

#### Simulations of Drug Transport in Skin

3D simulations in COMSOL® (Comsol, 2005) were used to solve the diffusion equation in order to obtain flux values. COMSOL uses the finite element method to solve the partial differential equation (Zimmermann, 2004). A constant concentration, on the needle coating and a sink condition in the dermis was assumed. With the imposed boundary conditions, the software was able to solve for values of diffusive flux at different times till the steady state diffusive flux was obtained.

The simulations were used to simulate a coated microneedle inserted in skin (figure 5) under the assumption that diffusion of the drug through the viable epidermis is the limiting process in this form of transdermal delivery (Davidson et al, 2008). The flux term is related to the permeability such that the steady state flux is simply a product of permeability and concentration (Moss et al, 2001)

$$J_{ss} = KC_v \quad (12)$$

The effective skin thickness was calculated by rearranging the Fick's first law equation for steady state diffusion.

$$J_{ss} = -D \frac{dC}{dx} \quad (13)$$

Where  $J_{ss}$  is the steady state flux of insulin across skin,  $D$  is the diffusion coefficient of insulin,  $C$  is the concentration of insulin and  $dC/dx$  is the concentration gradient. We define that at steady state

$$\frac{dC}{dx} = \text{constant} = \frac{C_0 - C_1}{h_{eff}} \quad (14)$$

Where  $C_0$  is insulin concentration at the coated area of microneedles,  $C_1$  is insulin concentration at the epidermal/dermal junction and  $h_{eff}$  is the skin thickness after microneedle has been inserted.

By combining equation (13) with equation (14), we obtain the following equation for steady state diffusive flux  $J_{ss}$ :

$$J_{ss} = \frac{-D(C_0 - C_1)}{h_{eff}} \quad (15)$$

When  $C_0$  is defined to be zero (i.e., sink condition), the effective skin thickness is determined as follows (Davidson et al., 2008):

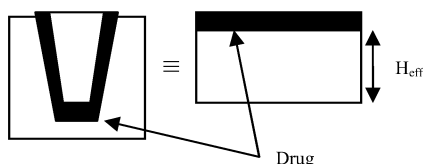
$$h_{eff} = \frac{DC_1}{J_{ss}} \quad (16)$$

The results from COMSOL simulations were then linked to SKIN-CAD<sup>®</sup> (Biocom system, 2006) results by using the effective skin thickness as the distance to the microcirculation. So the equivalent of a drug coated unto the microneedle surface in COMSOL is a patch of the same drug placed on skin stripped of the stratum corneum, with a viable epidermis thickness equal to the effective skin thickness.

### Simulations for Blood Concentration Profile

The software SKIN-CAD<sup>®</sup> is a simulator of skin pharmacokinetics that was used to obtain a transient profile for skin penetration and blood concentration (Tojo, 2005). The software was used to model the application of the drug as a patch unto skin stripped of the stratum corneum. An application time of 4 hours was entered; the effective skin thickness and application area depended on the dimensions of the microneedle that was being tested. Insulin was used as the model drug (MW 5734 (Kutsky et al, 1973)). The feasibility of microneedles for delivery of insulin has been studied by Ito et al (2006). The properties of insulin used in this study are listed in table 1. The mathematical framework of SKIN-CAD<sup>®</sup> has been avoided in this work since it has been explained previously (Al-Qallaf et al., 2007). The absorption kinetics of insulin is quite complicated and varies from patient to patient and from administration to administration for the same patient. This variability and effect of factors such as dose size, insulin concentration and insulin crystals has been studied by Soeborg et al (2008). For the purpose of this study the one-compartment pharmacokinetics model has been defined in the simulation and the insulin is taken to be rapid acting insulin analogue. Skin binding and drug metabolism in the skin were neglected and 100% absorption of the drug was assumed (Davidson et al, 2007).

The area of application in SKIN-CAD<sup>®</sup> was taken as the total drug area in contact with the epidermis. This is equivalent to the total surface area of the needle with the dissolved coating when it is inserted in the skin. This is illustrated in figure 4.



**Figure 4.** Area of application in SKIN-CAD<sup>®</sup> is equivalent to the area of drug in contact with epidermis when the needle is inserted in the skin.

SKIN-CAD<sup>®</sup> uses the finite difference method to calculate the diffusion through the viable epidermis. The one compartment pharmacokinetics model used to obtain the blood concentration profile is given in equation 17.

$$V_p \frac{dC_p}{dt} = \frac{dQ}{dt} S_a - K_e C_p V_p \quad (17)$$

$V_p$  is the volume of distribution in the blood,  $C_p$  is the concentration of the drug in the blood,  $t$  is time  $K_e$  is the elimination rate constant,  $dQ/dt$  is the rate of skin penetration,  $S_a$  is the area of the drug releasing surface of the delivery system (Tojo, 2005).

**Table 1.** Properties of insulin

| Properties                                     | value   | Reference           |
|--|---------|---------------------|
| Diffusion coefficient (m <sup>2</sup> /s)      | 1.0E-06 | Lv et al, 2006      |
| Elimination rate constant (min <sup>-1</sup> ) | 0.1     | Soeborg et al, 2008 |
| Volume of distribution (L)                     | 12      | Soeborg et al, 2008 |

## RESULTS AND DISCUSSION

The following sections present the results that were obtained. First of all we discuss the results of the numerical optimization. The effects of each parameter on the drug permeability are then analysed. Subsequently the simulations are used to determine the effect of the parameters on the steady state flux and the drug concentration in blood.

### Optimum Geometry of Microneedles

In this study, the optimization equation (9) along with the constrain equation (8, 10, 11) were implemented using java optimization programme. As mentioned previously, the highest value of the optimization function (g) which is directly proportional to the skin permeability has been determined for the input parameters as shown in Table 2.

The results suggest that the highest permeability for the parameters chosen will be obtained for a needle of cylindrical shape with a radius of 0.008 cm and a penetration depth of 0.01 cm (assuming an epidermis thickness of 0.015 cm), for an array 2.7 cm long containing 125 microneedles. The maximum value of g was 59.9177.

**Table 2.** Range of values used for optimization.

| Parameter | Range tested       |
|-----------|--------------------|
| n         | 1 – 231            |
| x         | 0.1 – 5 (cm)       |
| L         | 0.005 – 0.01 (cm)  |
| R         | 0.003 – 0.008 (cm) |
| r         | 0.003 – 0.008 (cm) |



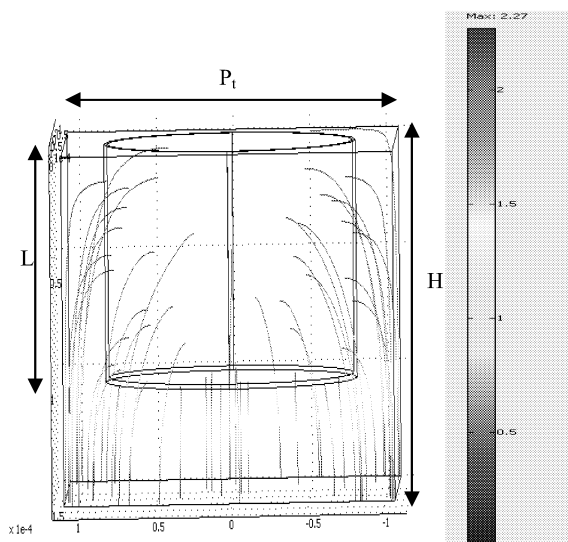


Figure 5. COMSOL simulation of optimum geometry

It is now required to observe the relationship between each parameter and the optimization function, and hence the effect on permeability.

#### Effect of Microneedle Base Radius

For a needle 0.01cm long, tip radius of 0.003 cm and array are 5cm long containing 125 needles, Figure 6 shows the effects of varying the base radius on the optimization function. The graph suggests that increasing the base radius will result in an increase in the permeability of the drug in the skin.

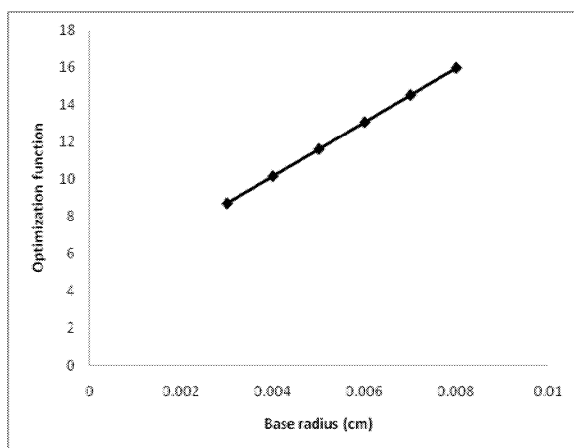


Figure 6. The effect of base radius (R) on optimization function (g)

#### Effect of Microneedle Tip Radius

Keeping the base radius as  $R=0.007\text{cm}$  the tip radius was varied so that the shape of the microneedle ranged from fully conical to cylindrical as shown in figure 7 the tip radius shows the same relationship to permeability as the base radius.

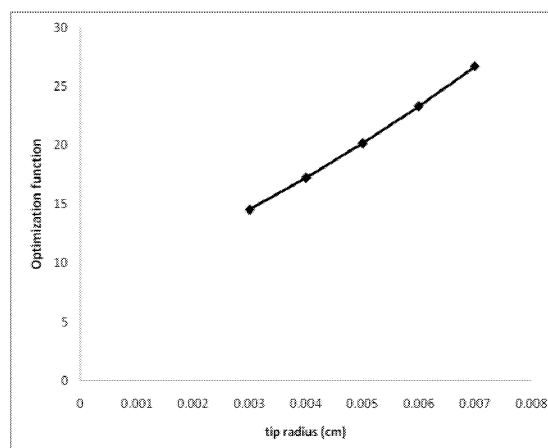


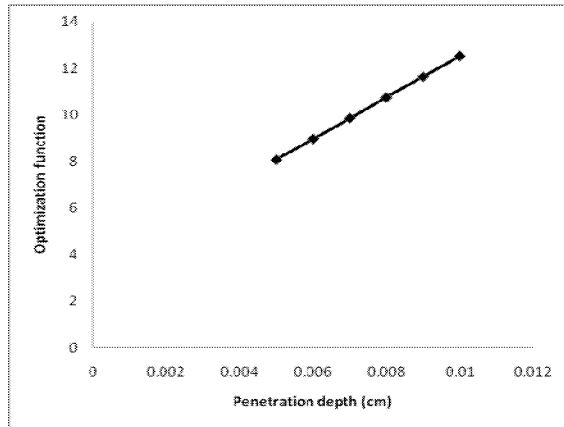
Figure 7. The effect tip radius (r) on the optimization function (g)

The permeability is the highest for a cylindrical tip than a sharper tip. However it might be desired to make the tips smaller to ease insertion. Davis et al (2006) developed a theoretical method for predicting needle insertion force and fracture force for microneedles. They concluded that the insertion force increases with increasing interfacial area between the needle tip and the skin surface. All the needles that were tested were in the range 0.003cm to 0.008cm in radius. The same range was used for the radius in our simulations as all the microneedles fabricated in this range where shown to insert into skin without fracture.

As expected the graph suggests that the optimization function will continue increasing to infinity as the size of the needle increases. This relationship is expected from equation 9. This is somewhat unrealistic, as it suggests that the biggest needle is better and these defeat the whole idea of “micro” needles. It has been said that delivering drugs in smaller quantity will result in better absorption (Allen et al, 2004) so there needs to be a biological factor so that up until a certain permeability will increase but beyond a certain value the permeability will start to decrease and the graph will show a different shape with a maximum point.

#### Effect of Penetration Depth of Microneedle

Due to the elastic nature of the skin and the presence of hair on the surface the microneedle will not penetrate its full length into the skin. Hence the penetration depth is considered instead of the total length of the microneedle. The values for  $g$  were plotted against changing penetration depth for a cylindrical needle of radius 0.004cm, for an array 5cm long containing 125 needles. The relationship between the penetration depth and the optimization function is shown in figure 8. As expected this indicates that for a given skin thickness, a longer needle will give a higher permeability.

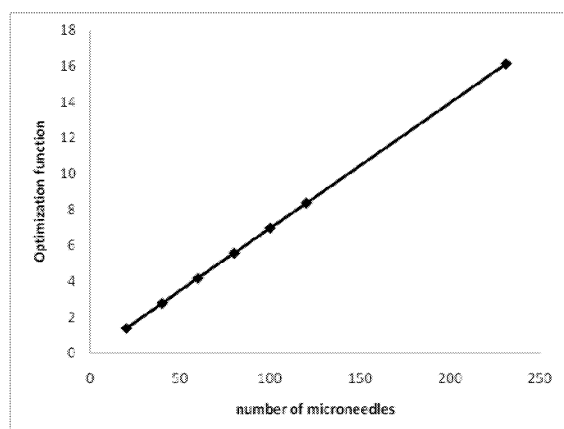


**Figure 8. The effect of penetration depth (L) on optimization function (g)**

This seems logical as the closer the needle gets to the dermis the lesser the distance the drug has to diffuse through. The constraint here is the fact that the needle has to be long enough to pierce the stratum corneum but short enough to avoid the nerves in the dermis. It is recommended that the needles should be longer than 0.001cm and shorter than 0.02cm (Teo et al, 2005).

#### Effect of Number of Microneedles

For a 5cm long array, different values of the number of microneedles were selected and the values plotted against the corresponding values of the optimization function. The microneedle spacing is dependent on the size of the array and the number of microneedles in the array. This relationship is represented by equation 5. Figure 9 therefore represents the relationship between the number of microneedles per row and the permeability, as well as the pitch and the permeability. The graphs shows that the more densely packed the needles are in the array, the higher the permeability of the skin. However the case may not be the same as what the results suggest in reality when it comes to the needles being inserted into the skin. If the array is too densely packed it may cause difficulty on insertion. It is necessary to have a constraint related to the mechanical property of the skin which gives a limit on the spacing.



**Figure 9. The effect of the number of microneedles per row in an array (n) and the optimization function (g)**

The results in figure 9 show that increasing the number of microneedles in a row of array will increase the permeability. This also indicates that reducing the pitch, in other words a denser array, will increase the permeability assuming that the needles are evenly spaced out.

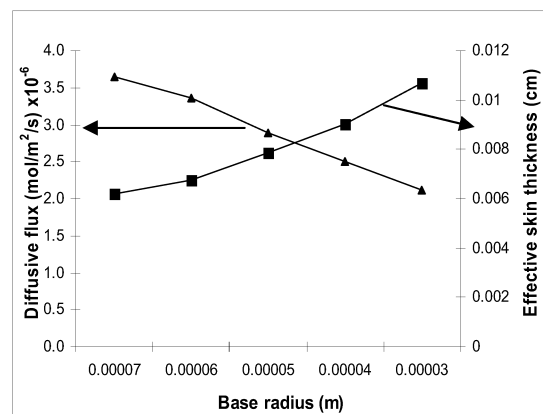
#### Drug Transport in Skin

Having had a two dimensional mathematical framework relating the permeability to the geometrical parameters, the effects of these parameters on the flux and effective thickness were obtained from 3D simulations in COMSOL. Typical results from these simulations are shown in the figures 10 to 13. The steady state diffusive flux increases for every parameter except the pitch where the steady state flux decreases as the pitch increases. These results correspond with the results from the numerical optimization as the flux and the permeability are related by equation (12) as mentioned earlier.

A number of simulations were also carried out keeping all other dimensions constant except for the dimension that was being tested.

#### Drug Concentration in Blood

Using SKIN-CAD<sup>®</sup> the effect of the important parameters on the blood drug concentration profile was analysed. The results are shown in figure 14 to 17. Figure 14 shows that a less spaced out array will result in a higher concentration of the drug in the blood, Figure 15 shows that the deeper the needle penetrates into the skin, the higher the optimum blood concentration, Figure 16 indicates that a wider needle will give a higher peak blood drug concentration while figure 17 shows that the wider the tip, the higher the peak blood concentration will be.



**Figure 10. The effect of base radius on the effective skin thickness and diffusive flux**

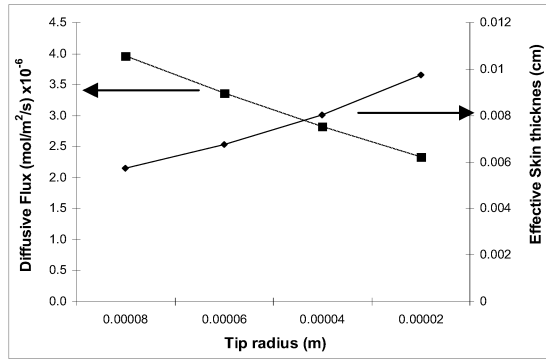


Figure 11. The effect of tip radius on the effective skin thickness and diffusive flux

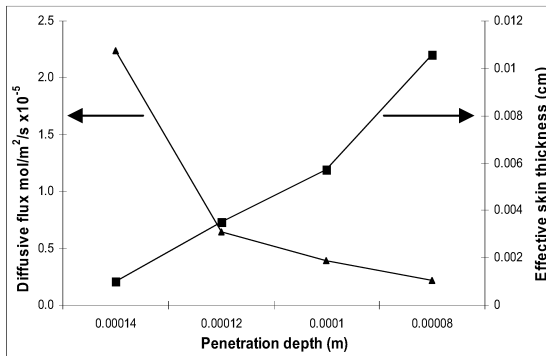


Figure 12. The effect of the penetration depth on the effective skin thickness and diffusive flux

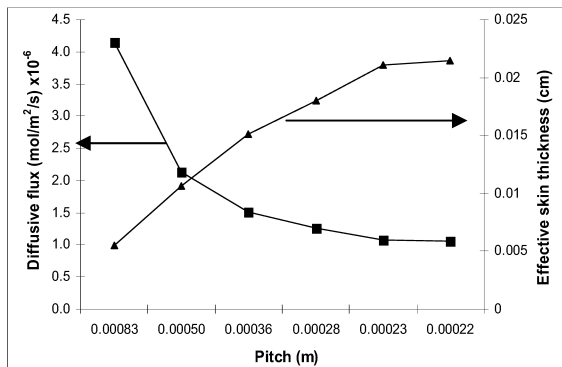


Figure 13. The effect of pitch on the effective skin thickness and the diffusive flux

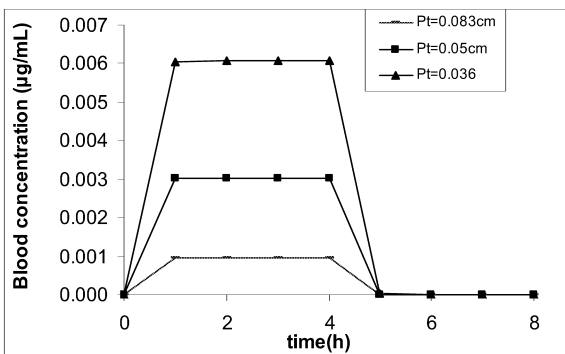


Figure 14. Effect of the pitch ( $P_t$ ) on the Insulin concentration in the blood

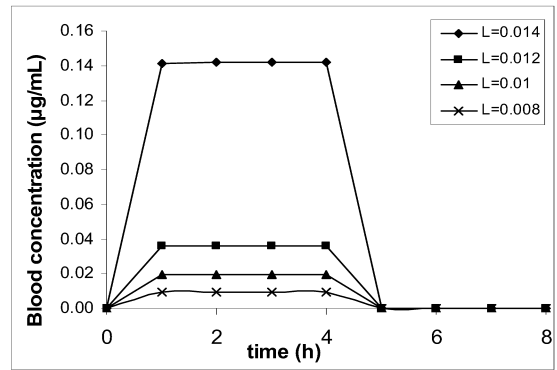


Figure 15. Effect of penetration depth ( $L$ ) on the Insulin concentration in the blood

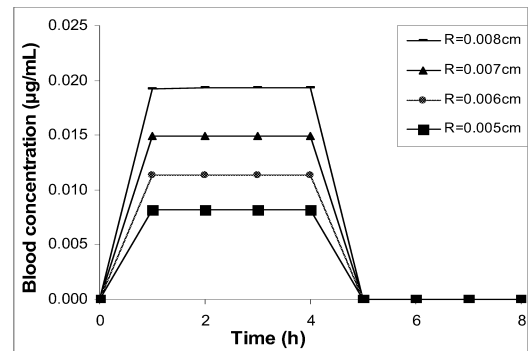


Figure 16. Effect of the base radius ( $R$ ) on the Insulin concentration in the blood

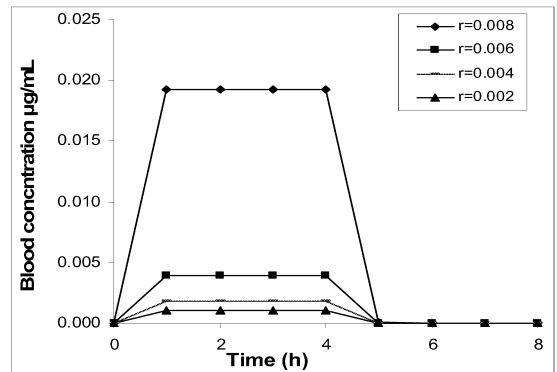


Figure 17. Effect of tip radius ( $r$ ) on the Insulin concentration in the blood.

Al-Qallaf and Das (2008) presented a framework for optimizing the dimensions of microneedles in a square array in order to increase drug permeability in skin. Their study looked at optimizing the microneedle dimensions which included microneedle radius, pitch, area of patch and number of microneedles for any given microneedle. The results showed that increasing the radius and number of microneedles will increase the optimization function, while increasing the area will cause a decrease in the optimization function. Both optimization equations related permeability to geometrical parameters and the results from the previous optimization agrees with the results obtained here. The framework developed in this paper is for optimizing the dimensions of a particular microneedle shape.

## CONCLUSION

This study was aimed at relating geometrical parameters to the performance of solid microneedles in delivering drugs by developing a mathematical framework that relates drug permeability to these parameters. A microneedle shape was selected and sensitivity analysis was performed to analyse what effect a change in length, base radius, pitch, number of microneedles per row and tip radius would have on the permeability. Our simulations helped to analyse the effects of these parameters on the flux and concentration of the drug in blood. The vital findings of this work are that they give significant evidence that the design of the microneedle has effect on the drug delivery efficiency and they must be considered in the development of a microneedle based drug delivery system. Possibility of insulin delivery using solid coated microneedles was also considered (numerically) as the pharmacokinetic properties (diffusion coefficient through skin, volume of distribution and elimination rate constant) used in the simulation were those of insulin.

Our results give the bases for follow up experiments or further optimization on microneedle designs and the distribution of the microneedles in the array.

## REFERENCE

- Al-Qallaf, B., Das, D.B., Mori, D., Cui, Z.F., 2007. Modelling transdermal drug delivery of high molecular weight drugs from microneedle systems. *Journal of philosophical transaction of the royal society Series A*, 365(2007), pp.2951-2967.
- Al-Qallaf, B., Das, D.B., 2008. Optimization of Square Microneedle Arrays For Increasing Drug Permeability in Skin. *Chemical Engineering Science*, 63(2008), pp.2523-2535.
- Allen, V L., Ansel, H C., Popovich N G., 8<sup>th</sup> ed. 2005. *Ansel's Pharmaceutical Dosage Forms and Drug Delivery Systems*, Philadelphia: Lippincott Williams & Wilkins.
- Biocom Systems, Inc. 2006 SKIN-CAD<sup>®</sup>. Simulator for skin pharmacokinetics. Fukuoka, Japan: Biocom Systems, Inc.
- Comsol, AB, 2005, FEMLAB<sup>®</sup> 3 multiphysics modelling, Documentation, Comsol AB, Stockholm, Sweden.
- Davidson, A., Al-Qallaf, B., Das D.B., 2008. Transdermal drug delivery by coated microneedles: Geometry effects on effective skin thickness and drug permeability. *Chemical Engineering Research and Design*, 86(2008), pp.1196-1206.
- Davis, S. P., Landis, B. J., Adams, Z. H., Allen, M. G. & Prausnitz, M. R., 2004. Insertion of microneedles into skin: measurement and prediction of insertion force and needle fracture force. *Journal of Biomechanics*, 37(2004), pp.1155-1163.
- Gill, S. H., Prausnitz, R. M., 2006. Coated microneedles for transdermal drug delivery. *Journal of controlled release*, 117(2007), pp.227-237.
- Gill, S H., Prausnitz, R M., 2007. Coating formulations for microneedles. *Pharmaceutical research*, 24(7), pp.1369-1380
- Hadgraft, J., 2004. Skin Deep. *European Journal of Pharmaceutical and Biopharmaceutics*, 58(2004), pp.291-299.
- Henry, S., McAllister, V D., Mark, G. A., Prausnitz, R. M., 1998. Microfabricated microneedles: A novel approach to transdermal drug delivery. *Journal of Pharmaceutical Sciences*, 87(8), pp.922-925.
- Ito, Y., Hagiwara, E., Saeki, A., Sugioka, N., Takada, K., 2006, Feasibility of microneedles for percutaneous absorption of insulin invitro using hairless mice. *European Journal of pharmaceutical sciences*, 29(2006), pp.82-88.
- Kutsky J. R., 1973. *Handbook of Vitamins and Hormones*. London: Litton Educational Publishing Inc.
- Lee W. J., Park J., Prausnitz R M., 2007. Dissolving microneedles for transdermal drug delivery. *Biomaterials*, 29(2008), pp.2113-2124.
- lv, Y. G., Liu, J., Gao, Y. H., Xu, B., 2006. Modelling of transdermal drug delivery with a microneedle array. *Journal of Micromechanics and Microengineering*, 16(2006), pp. 2492-2501.
- Machin J, Lampert G.J, O'Donnell M.J, 1985. Component Permeabilities and water Contents in Periplaneta Integument: Role of the Epidermis Re-examined. *Company of Biologists Limited*, 117(1985), pp.155-169.
- Martanto, W., Davis, P S., Holiday, R N., Wang, J., Gill, S. H., Prausnitz, R. M., 2004. Transdermal delivery of insulin using microneedles in vivo. *Pharmaceutical Research*. 21(6), pp.947-951.
- McAllister, D.V., Wang, P. M., Davis, S. P., Park, J. H., Canatella, P. J., Allen, M. G. Prausnitz, M. R. 2003. Microfabricated needles for transdermal delivery of macromolecules and nanoparticles: fabrication methods and transport studies. *PNAS*. 100, pp.13755-13760.
- Moss, G.P., Dearden J.C., H., M.T.D Patel. Cronin, 2001, Quantitative structure permeability relationships (QSPRs) for percutaneous absorption. *Toxicology in Vitro*, 16(2002), pp.299-317
- Stalheim-Smith, A., Fitch, K G., 1993. *Understanding Human Anatomy and Physiology*. St. Paul: West publication Company Inc.
- Stoeber, B., Liepmann, D., 2005. Arrays of Hollow Out-of-plane Microneedles for Drug Delivery. *Journal of Microelectromechanical systems*, 14(3), pp.472-479.
- Teo A L., Shearwood C., Kian C. N., Jai L., Shabbir, M., 2005. Transdermal Microneedles for Drug Delivery Application. *Materials Science and Engineering B*, 132(2006), pp.151-154.
- Tojo K., 2005. *Mathematical Models of Transdermal and topical drug delivery*, 2<sup>nd</sup> ed. Japan: Biocom systems Inc.
- Søeberg T., Rasmussen C H., Mosekilde E., Colding-Jorgensen M., 2008. Absorption Kinetics of insulin after subcutaneous administration. *European Journal of Pharmaceutical Sciences*, 36(2009), pp.79-90.
- Zimmerman B.J., 2004. *Process modelling and simulation with Finite element methods*, Series A Vol 15, Singapore: World scientific publishing Co. Pte. Ltd.

## Acknowledgement

One of the authors (O.O.) would like to thank Loughborough University for a PhD studentship given to her to work on the subject of this paper.

# Modelling of Flashing Propellant Flow in Pressurised Metered Dose Inhalers

Henk Versteeg and Abdul Qaiyum Shaik  
Wolfson School of Mechanical and Manufacturing Engineering,  
Loughborough University, Loughborough, LE11 3TU, UK  
E-mail: H.K.Versteeg@Lboro.ac.uk

## KEYWORDS

Metered-dose inhalers, two-phase flow, flashing, propellant, metastability, delayed equilibrium model, homogeneous equilibrium model.

## ABSTRACT

The thermodynamics and fluid mechanics of propellant flow in pressurised metered-dose inhalers (pMDIs) is complex and poorly understood. Previous work by Fletcher (1975) and Clark (1991) has shown that metastability may play a significant role in determining the flow conditions inside pMDIs. This paper presents a study of the thermodynamics and one-dimensional multiphase fluid dynamics of propellant flows in pMDIs under quasi-steady state conditions. We report the findings of the homogeneous equilibrium model (HEM) with those of a delayed equilibrium model (DEM) reported recently by Shaik & Versteeg (2008) for a range of different quasi-steady inlet conditions. These results are compared against the experimental and numerical predictions reported in Clark (1991). On the basis of the findings the role played by metastability is discussed.

## INTRODUCTION

Pressurised metered-dose inhalers (pMDIs) are the most common devices for therapeutic aerosol delivery to the lungs in the treatment of asthma, bronchitis, and other pulmonary diseases. Figure 1 shows a schematic of a typical pMDI consisting of a canister and actuator. Drug is stored in the canister as suspension or in solution in propellant (previously R12/R114 mixtures, post 1990's R134a or R227 are used).

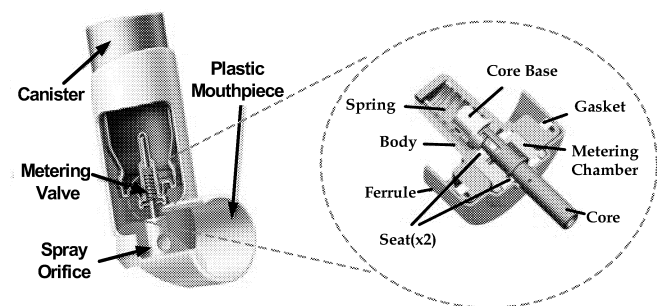


Figure 1 Schematic of a pressurised metered dose inhaler

The canister is equipped with a metering chamber and a spring-loaded valve. The metering chamber is filled with a precisely known volume of drug/propellant mixture. The drug is dispensed by depressing the actuator, which causes the metering chamber contents to flow through an orifice into the valve stem and actuator sump (together called

expansion chamber). The pressure drop across this valve orifice causes partial flashing of the propellant. A two-phase mixture enters the spray orifice where further propellant expansion takes place. The drug leaves the pMDI as part of a fine propellant aerosol with typical droplet size of 1-5  $\mu\text{m}$ .

To predict spray formation from pMDIs it is essential to have accurate knowledge of thermodynamic state, void fraction and velocity of the fluid at the spray orifice exit. Fletcher (1975) and Clark (1991) developed semi-empirical models of flashing propellant flows through a pMDI based on assumptions of thermodynamic equilibrium in the metering and expansion chambers and homogeneous frozen flow in the valve and spray orifices. The results of their models were in broad agreement with experimental trends except for some puzzling discrepancies. Spray exit velocities were overpredicted and propellant temperatures in the expansion chamber were higher than the saturation temperature at the computed pressures, suggesting that metastability may play a role in the flashing propellant flow.

Recently, Shaik & Versteeg (2008) presented the findings of a numerical model for the prediction of *steady*, metastable propellant flow through a two-orifice system - valve orifice, expansion chamber and spray orifice - based on the Delayed Equilibrium Model (DEM). The model was successfully validated against Fletcher's and Clark's experimental results for *continuous* propellant flows with saturated liquid at inlet. Metered discharge is a *transient* phenomenon where the mass of propellant reduces in the course of the propellant discharge event. This has two consequences:

- Inlet pressure and temperature exhibit a monotonic fall as the metering chamber empties
- Fluid is a vapour-liquid mixture at the valve orifice inlet

Clark's (1991) experimental results indicated that mass flow rates into and out of the expansion chamber were substantially equal during a sizeable portion of metered discharge events and that, except during the initial and final stages of the discharge event, the flow is quasi-steady. In this paper, we use our steady-state model of metastable flow to simulate this quasi-steady phase of propellant flashing. We compare the performance of our delayed equilibrium model (DEM) with experimental data from Clark (1991) and the well-established homogeneous equilibrium model (HEM).

## MATHEMATICAL MODEL

The basics of the method can be best understood by considering the conceptual two-orifice system sketched in Figure 2. The valve orifice, expansion chamber and spray orifice are represented as a series of straight cylindrical ducts of known diameter, length and hydraulic roughness.

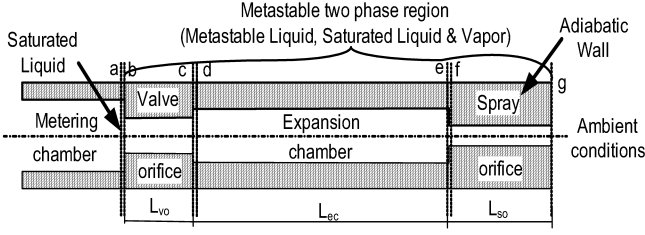


Figure 2 Schematic view of two-orifice system

The governing equations are the steady, one-dimensional conservation equations of mass, momentum and energy for the vapour-liquid propellant mixture. The following assumptions are made in modelling the flow:

- One-dimensional, quasi-steady, adiabatic flow.
- No slip between the liquid and vapour phases: ( $U_l = U_v$ ).
- The effect of surface tension is neglected: ( $p_l = p_v = p$ ).

In the homogeneous equilibrium model (HEM), vaporisation is instantaneous and propellant liquid and vapour are everywhere saturated. In our simulations of metastability effects we have used the Delayed Equilibrium Model (DEM) outlined in Feburie et al. (1993). As the fluid enters the valve orifice, initially saturated liquid is converted to metastable liquid due to the sudden pressure drop. The propellant does not immediately return to the saturated state at its new pressure, because evaporation does not take place instantaneously. In the DEM, only a fraction  $y$  (the so-called vaporization index) of the propellant is assumed to be transformed into saturated mixture, the other fraction  $(1-y)$  remains metastable liquid, which undergoes an isentropic evolution. The specific volume  $v_m$  and enthalpy  $h_m$  of the propellant are given by the following constitutive relations:

$$v_m = \frac{1}{\rho_m} = (1-y)v_{lm} + (y-x)v_l + xv_v \quad 1$$

$$h_m = (1-y)h_{lm} + (y-x)h_l + xh_v \quad 2$$

where  $v$  = specific volume,  $h$  = specific enthalpy,  $x$  = vapour mass fraction,  $y$  = vaporization index; subscripts: m = mixture, l = liquid, v = vapour, lm = metastable liquid.

The vaporization index ' $y$ ' is a measure of the departure from local saturated state:  $y=1$  is saturated,  $y=0$  is metastable. In the DEM, the conversion of metastable into saturated fluid in the flow direction ' $z$ ' is evaluated using the following relaxation equation (Feburie, 1993):

$$\frac{dy}{dz} = 0.02 \frac{P}{A} (1-y) \left[ \frac{p_{sat} - p}{p_c - p_s} \right]^{0.25} \quad 3$$

where  $P$  is duct perimeter,  $A$  is cross-sectional area,  $p_{sat}$  is saturated pressure and  $p_c$  is the critical pressure.

Inlet conditions at the valve orifice must be defined from knowledge of the propellant state in the metering chamber. Experimental values of metering chamber pressure are available. In order to specify the problem we also require temperature  $T$ , quality  $x$  and the vaporization index  $y$  of the mixture in the metering chamber. In absence of measurements of these variables we have opted to neglect metastability in the metering chamber and assumed that the fluid will be at thermodynamic equilibrium in this space. This fixes  $T = T_{sat}(p)$  and  $y = 1$  in the metering chamber. The

propellant flow regime in the metering chamber is also unknown. To specify quality  $x$  at the valve orifice inlet we explore two possibilities:

- Two-Phase Flow (TPF): vapour and liquid are homogeneously dispersed in the metering chamber and the quality  $x$  is obtained from energy equation (4).

$$x = [h_{l,0} - h_{l,sat}(p)] / [h_{v,sat}(p) - h_{l,sat}(p)] \quad 5$$

where  $h_{l,0}$  is liquid enthalpy at ambient temperature and subscript sat indicates saturated fluid at pressure  $p$ .

- Liquid-Only Flow (LOF): propellant is stratified in the metering chamber with vapour at the top and saturated liquid at the bottom. The valve orifice is located near the bottom of the metering chamber, so the fluid entering the valve orifice will be liquid-only, thus  $x = 0$ .

In reality, the metering chamber fluid would be metastable and the flow regime would be somewhere between the above two scenarios. Improved definition of the flow regime in the metering chamber is left for future study. Propellant properties are obtained from REFPROP v.7.0 (2002). Saturated properties of the propellant in the two-orifice system are estimated as a function of local pressure; where the liquid is metastable, the properties are estimated using its pressure and temperature.

## NUMERICAL SOLUTION

The governing equations are solved by marching in the flow direction through the region of interest, which is subdivided into  $N = 700$  control volumes (CV). Owing to the high gradients at the exit of the valve and spray orifice, a non-uniform grid concentrated at the exit sections (Figure 3) is generated using the distributions proposed by Escanes *et. al* (1995) in conjunction with grid expansion factor  $k = 3.5$ :

*Valve orifice*

$$\Delta z_i = \frac{L_{vo}}{\tanh(k)} \left[ \tanh\left(k \frac{i}{N_{vo}}\right) - \tanh\left(k \frac{(i-1)}{N_{vo}}\right) \right] \quad 6$$

*Expansion Chamber*

$$\Delta z_i = \frac{L_{ec}}{N_{ec}} \quad 7$$

*Spray orifice*

$$\Delta z_i = \frac{L_{so}}{\tanh(k)} \left[ \tanh\left(k \frac{i}{N_{so}}\right) - \tanh\left(k \frac{(i-1)}{N_{so}}\right) \right] \quad 8$$

where  $N_{vo}(=300)$ ,  $N_{ec}(=100)$ ,  $N_{so}(=300)$ , are the number of CVs in the valve orifice, expansion chamber and spray orifice for a grid-independent solution.

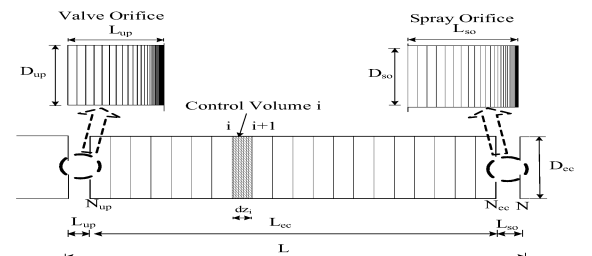


Figure 3 Schematic view of two-orifice system

Starting at the valve orifice inlet, values of the flow variables at each CV outlet are obtained from the known values at the

CV inlet and the “current” mass flow rate. When the calculation reaches the spray orifice the computed discharge pressure is compared with the actual value. If the difference is greater than a preset tolerance the mass flow rate is adjusted and the solution procedure is restarted at the entrance of the tube until calculated and actual discharge pressures agree.

## TEST CASES

Clark (1991) carried out an extensive programme of measurements on metered discharges of propellant R12 for a wide range of orifice, metering chamber and expansion chamber dimensions geometries. The test case used for the present study has the following geometry and operating conditions:

- Metering chamber volume:  $100 \mu\text{L} = 10^{-7} \text{ m}^3$
- Valve orifice:  $D_{vo} = 0.26 \text{ mm}$ ,  $L_{vo} = 0.5425 \text{ mm}$
- Expansion chamber:  $D_{ec} = 3.8 \text{ mm}$ ,  $L_{ec} = 11 \text{ mm}$
- Spray orifice:  $D_{so} = 0.26 \text{ mm}$ ,  $L_{so} = 1 \text{ mm}$
- Hydraulic roughness of all surfaces:  $\varepsilon = 1.5 \mu\text{m}$
- Ambient pressure 1.013 bar, which is the actual discharge pressure.
- Ambient temperature  $20^\circ\text{C}$ , which fixes the conditions of the metering chamber fluid prior to discharge
- Other inlet conditions are given in Table 1.

Table 1 Inlet conditions for numerical simulations

| Case | $t \text{ (ms)}$ | $p_{in} \text{ (bar)}$ | $x_{in}$        |           |
|------|------------------|------------------------|-----------------|-----------|
|      |                  |                        | HEM & DEM (LOF) | DEM (TPF) |
| 1    | 0                | 5.35                   | 0.0             | 0.0       |
| 2    | 25               | 4.85                   | 0.0             | 0.023     |
| 3    | 50               | 4.80                   | 0.0             | 0.025     |
| 4    | 75               | 4.73                   | 0.0             | 0.029     |
| 5    | 100              | 4.50                   | 0.0             | 0.039     |
| 6    | 150              | 4.09                   | 0.0             | 0.059     |
| 7    | 200              | 3.68                   | 0.0             | 0.080     |

(LOF) indicates liquid-only (stratified) inlet conditions  
(TPF) indicates two-phase (homogeneous) inlet conditions

## RESULTS AND DISCUSSION

The predicted distributions of pressure, temperature, void fraction and velocity for the metered discharge of propellant R12 at  $t = 25 \text{ ms}$  (Case 2 in Table 1) are shown in Figure 4 to Figure 6. These give the profiles of flow variables along the length of the two-orifice system using DEM (LOF), DEM (TPF) and HEM. The trends are representative of those predicted at other times. The sudden pressure drop in Figure 4 at the inlet ( $z = 0$ ) is due to flow acceleration at the abrupt entry to the valve orifice. Thereafter the pressure remains almost constant inside the expansion chamber. The next large pressure drop at the entrance of the spray orifice ( $z = 0.0115 \text{ m}$ ) is due to rapid fluid acceleration in the abrupt entry to the spray orifice. The propellant pressure drops along the spray orifice due to the frictional and acceleration effects. The large pressure drop at the exit plane of the spray orifice indicates that the flow is choked. The predicted pressure profiles are quite similar for the three different

modelling assumptions: the HEM yields slightly higher predictions, whereas the DEM with liquid-only inlet condition (LOF) gives the lowest pressure profile.

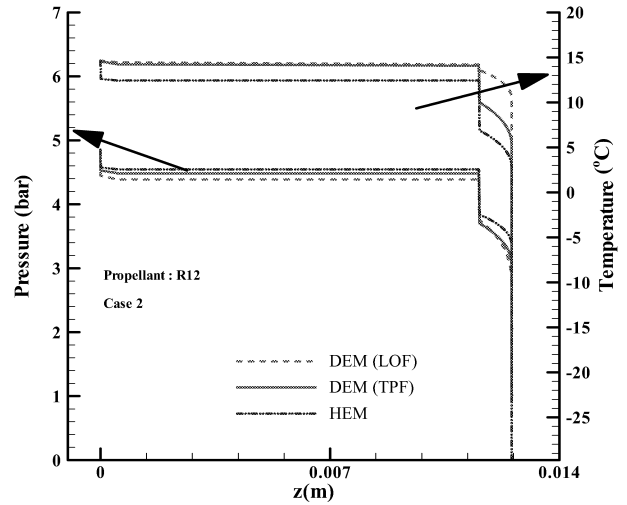


Figure 4 Predicted pressure and temperature profiles for case 2 with DEM (LOF), DEM (TPF) and HEM

Following Zhou and Zheng (2006) the average propellant temperature is computed using the superheated liquid temperature ( $T_{lm}$ ) and the saturation liquid temperature ( $T_l$ ):

$$T = (1 - y)T_{lm} + yT_l \quad 9$$

Figure 4 shows that the temperature distribution is similar to that of pressure, but considerable differences are found in the spray orifice. The DEM (LOF) modelling assumption yields more metastable propellant and hence higher temperatures. The HEM involves saturated propellant and hence predicts the lowest temperatures profiles.

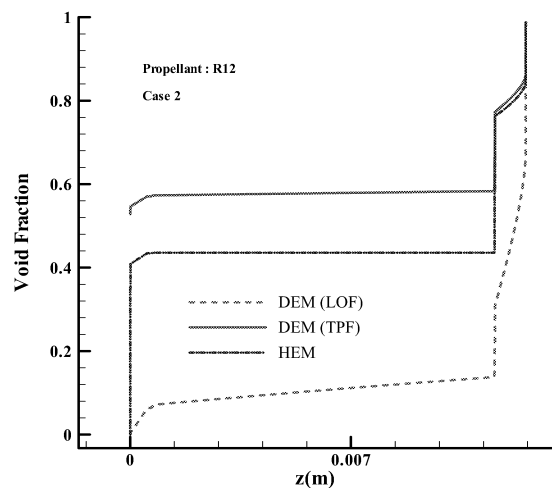


Figure 5 Predicted void fraction distribution for case 2 with DEM (LOF), DEM (TPF) and HEM

Figure 5 gives the distribution of void fraction along the two-orifice system. The void fraction increases as pressure decreases - rapidly where the pressure reduction is abrupt. Figure 6 shows the predicted velocity profiles in the two-

orifice system. The velocity is found to increase due to (i) area reduction, e.g. in abrupt contractions and (ii) void fraction increase. The DEM (TPF) has the highest void fractions and velocity, whereas the DEM (LOF) gives the lowest values as the level of metastability are highest for this modelling assumption. The differences are particularly large for the predicted void fraction, which will have a substantial impact on spray formation and highlights the importance of correct prediction of the evaporation rate.

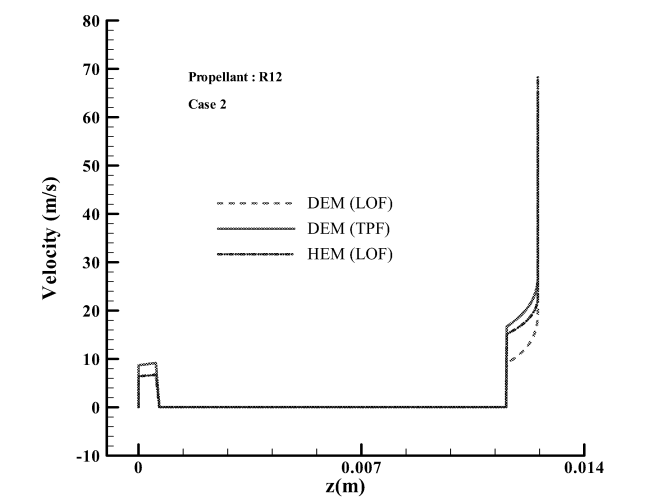


Figure 6 Predicted velocity distribution along the twin orifice system for case 2 with DEM (LOF and TPF) and HEM

Table 2 shows the numerically predicted quasi-steady mass flow rates at the start of the discharge event and at six later instants for DEM (LOF and TPF) and HEM. The DEM liquid-only flow (LOF) scenario predicts the largest mass flow rate compared with the DEM two-phase flow (TPF) scenario and the HEM, which yield similar results. Using the metering chamber volume of 100  $\mu\text{L}$ , liquid density of 1304  $\text{kg/m}^3$  and approximate discharge event duration of 300 ms (Clark, 1991), we can make an estimate of the average mass flow rate of  $4.35 \times 10^{-4} \text{ kg/s} = 1.565 \text{ kg/h}$ . Table 2 shows that the prediction with the DEM (LOF) scenario is very close to this estimate.

| Table 2 Predicted mass flow rates for different models |          |   |                 |                 |
|--|----------|---|-----------------|-----------------|
| Case   | $t$ (ms) | Numerical Mass Flow Rate $\dot{m}$ (kg/h) |                 |                 |
|  |          | HEM                                       | DEM liquid-only | DEM homogeneous |
| 1  | 0        | 1.064                                     | 1.744           | 1.744           |
| 2  | 25       | 0.995                                     | 1.652           | 1.040           |
| 3  | 50       | 0.987                                     | 1.642           | 1.010           |
| 4  | 75       | 0.976                                     | 1.627           | 0.971           |
| 5  | 100      | 0.944                                     | 1.583           | 0.872           |
| 6  | 150      | 0.883                                     | 1.501           | 0.731           |
| 7  | 200      | 0.819                                     | 1.413           | 0.613           |
| Average $\dot{m}$                                      |          | 0.936                                     | 1.573           | 0.910           |

Figure 7 shows the comparison of predicted expansion chamber pressure at different instants during the metered

discharge of R12 with experimental results from Clark (1991). All models underpredict the chamber pressure: the two DEM models (LOF) and (TPF) are almost identical, whereas the HEM is slightly closer to the measured expansion chamber pressure.

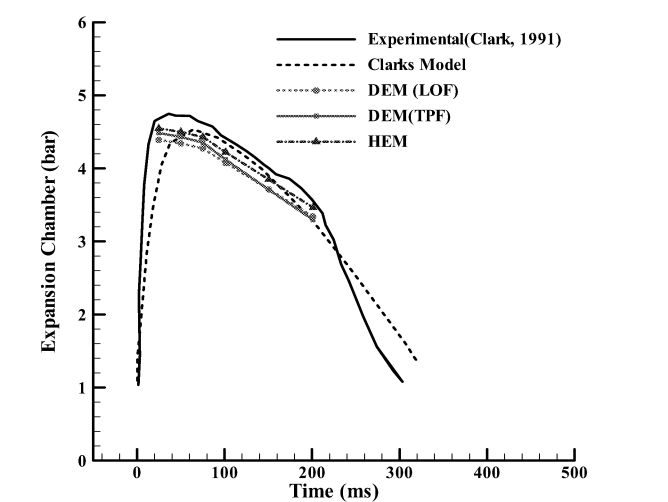


Figure 7 Comparison of expansion chamber pressure against experimental results from Clark (1991)

Figure 8 shows the variation of expansion chamber temperature against the Clark's (1991) experimental results. The predictions of the DEM models lie close together and yield expansion chamber temperatures that are higher and, hence, closer to the experimental results than the HEM.

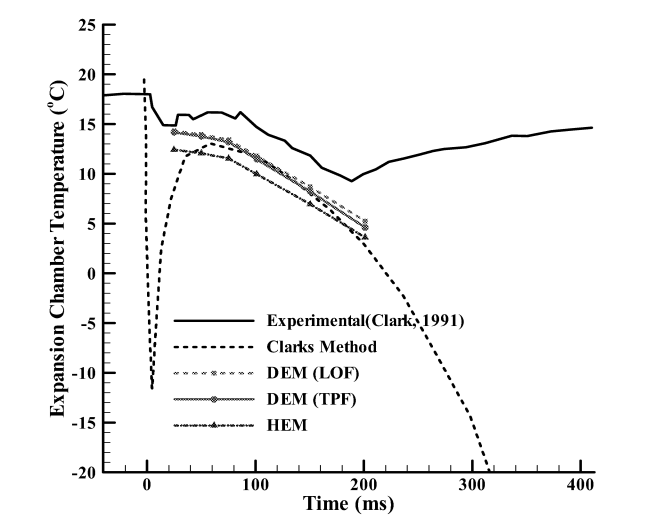


Figure 8 Comparison of expansion chamber temperature against experimental results from Clark (1991)

The fact that the expansion chamber pressure is predicted well with HEM, whereas the expansion chamber temperature is predicted most closely with DEM (LOF and TPF), indicate a modest level of metastability. Our new model improves the prediction of the expansion chamber pressure at 25 ms compared with Clark's model. Moreover, Table 2 shows that calculations based on homogeneous metering chamber conditions (HEM and DEM-TPF) lead to underprediction of the mass flow rate. Taken together, these pieces of evidence



suggest that Clark's model assumption of homogeneous metering and expansion chamber conditions causes underprediction of the propellant flow rate and subsequent failure to capture the correct initial transient pressure rise in the expansion chamber.

As highlighted above, metastability will cause large differences in the void fraction in the expansion chamber and spray orifice. Recent flow visualisations by Versteeg and Hargrave (2006) in a transparent prototype actuator suggested a slug flow regime in the spray orifice. The vapour void fraction will dictate the thickness of liquid films on the walls and of liquid bridges between the vapour slugs. Thus, metastability may have a strong influence on the conditions for spray formation at the spray orifice exit, suggesting that its accurate prediction is very important. The outcome of the DEM relaxation equation 3 depends on the value of its constant. Following Feburie (1993) it was set equal to 0.02, but it should be noted that this value originates from experiments on steam-water systems. A larger coefficient would increase evaporation, whereas a smaller coefficient would encourage metastability, inhibit evaporation and hence reduce the predicted void fraction. Further work is required to evaluate if 0.02 is the most appropriate value for propellant flows in pMDIs. Finally, it was noted earlier that the fluid in the metering chamber was measured by Clark to be metastable. No attempt was made to represent this in our model. This inaccuracy affects the initial conditions of our model and has made it difficult to improve upon Clark's predictions of expansion chamber pressure and temperature. Work is currently in progress to establish the best possible account of flow conditions and metastability in the metering chamber from the available experimental data.

## CLOSING REMARKS

In the present paper, metered propellant discharge from a pMDI was modelled as a two-orifice system with different inlet conditions representing different instants during the quasi-steady phase of a pMDI discharge event. To gain an insight into the role played by metastability steady-state simulations with different modelling assumptions - DEM (LOF), DEM (TPF) and HEM - were compared with Clark's (1991) experimental results and his model predictions. Detailed examination of the results suggested metastability strongly affects the void fraction in the expansion chamber and subsequent void fraction in the spray orifice. Large differences were found between predicted mass flow rates and spray orifice exit conditions for the two DEM models. Moreover, all models underpredicted the expansion chamber pressure and temperature. Our results drew attention to the need to examine whether changes to the coefficient of the relaxation equation for vaporisation index  $y$  can improve the predictive capability. The importance of accurate initial conditions was also highlighted, suggesting that improved predictions of pMDI discharge events will also benefit from further study of the flow regime and thermodynamic state of the propellant in the metering chamber.

## ACKNOWLEDGEMENTS

Abdul Qaiyum Shaik gratefully acknowledges a scholarship from Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, UK.

## NOMENCLATURE

|                      |  |
|----------------------|--|
| $A$                  | Cross sectional area ( $m^2$ )                       |
| $D$                  | Diameter (m)   |
| $f$                  | friction factor                                      |
| $G$                  | mass flux ( $kg/sm^2$ )                              |
| $L$                  | Length (m)   |
| $h_m$                | mean specific enthalpy (J/kg)                        |
| $i$                  | $i^{th}$ Control Volume                              |
| $k$                  | concentration factor ( $k=3.5$ )                     |
| $\dot{m}$            | mass flow rate (Kg /h)                               |
| $N$                  | Number of nodes                                      |
| $p$                  | Pressure (Pa)  |
| $p_s$                | Saturation pressure (Pa)                             |
| $p_c$                | critical pressure (Pa)                               |
| $P$                  | Perimeter (m)  |
| $q$                  | wall heat flux                                       |
| $T$                  | Temperature ( K )                                    |
| $U$                  | Velocity (m/s)                                       |
| $x$                  | Mass fraction (Vapour quality)                       |
| $y$                  | Vaporisation index                                   |
| $z$                  | Axial coordinate (m)                                 |
| <i>Greek letters</i> |  |
| $\varepsilon/D$      | Relative roughness of the capillary tube             |
| $\tau$               | wall shear stress [ $\tau = (f/4)(G^2/2\rho)$ ] (Pa) |
| $\rho$               | mean density ( $kg/m^3$ )                            |
| $v$                  | specific volume ( $m^3/kg$ )                         |
| <i>Subscripts</i>    |  |
| c                    | critical   |
| dis                  | discharge  |
| ec                   | expansion chamber                                    |
| in                   | inlet  |
| l                    | liquid   |
| lm                   | metastable   |
| m                    | mean   |
| sat                  | saturation   |
| so                   | spray orifice  |
| v                    | vapour   |
| vo                   | valve orifice  |

## REFERENCES

- Clark, A.R. 1991. "Metered atomization for respiratory drug delivery". PhD Thesis, Loughborough University of Technology.
- Escanes, F., D. Perez-Segarra, , and A. Oliva. 1995. "Numerical simulation of capillary-tube expansion devices". *International Journal of Refrigeration*. Vol. 18, No. 2, 113-122.
- Feburie, V., M. Giot, S. Granger and J. M. Seynhaeve. 1993. "A model for choked flow through cracks with inlet

- subcooling". *Int. J. Multiphase Flow*, Vol. 19, No. 4, 541-562.
- Fletcher, G.E. 1975. "Factors affecting the atomisation of saturated liquids". PhD Thesis, Loughborough University of Technology, Loughborough, UK.
- Hardy, Ph. And P. Mali. 1983. "Validation And Development of a Model Describing Subcooled Critical Flow Through Long Tubes". *Energie Premaire*, Vol 18, pp 5-23.
- Lin, S., C.C.K Kwok,. R.Y Li, Z.H Chen and Z.Y Chen. 1991. "Local frictional pressure drop during vaporization of R-12 through capillary tubes". *Int. J. Multiphase Flow*. Vol. 17, No. 1, 95-102.
- REFPROP v. 7.0. 2002. NIST Thermodynamic properties of refrigerants and refrigerant mixtures database, Standard Reference Data Program, Gaithersburg, Maryland 20899, USA.
- Shaik, A. Q. and H. K. Versteeg. 2008, "Model for the prediction of Internal Flow Conditions in Pressurised Metered Dose Inhalers(pMDIs)." *Drug Delivery to Lungs 19*, Edinburgh, UK.
- Versteeg HK and Hargrave GK. 2006, "Fundamentals and resilience of the original MDI actuator design", *Proceedings of Respiratory Drug Delivery IX*, Boca Raton, FL, USA, Vol. 1, pp. 91-100, April 23-27, 2006.
- Zhou, G and Y. Zheng. 2006. "Numerical and experimental investigations on the performance of coiled adiabatic capillary tubes." *Applied Thermal Engineering*. Vol. 26, No. 11-12, 1106-1114.

## BIOGRAPHY

**ABDUL QAIYUM SHAIK** is a PhD student in the Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University since 2005. His research area is numerical modelling of multi phase flows.

**HENK VERSTEEG** is Senior Lecturer at the Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University. His research interests are respiratory drug delivery.

# **POROUS MEDIA AND CHEMICAL ENGINEERING SIMULATION**



# Modeling of three dimensional Darcy flow through soils

Bahareh Kaveh-Baghbaderani, Vahid Nassehi, Abhijeet Kulkarni

Chemical Engineering Department  
Loughborough University  
Loughborough LE11 3TU  
Email: v.nassehi@lboro.ac.uk

## KEYWORDS

Underground flow, Hydrodynamics, Darcy flow, Finite element modelling, Three dimensional porous flow .

## ABSTRACT

The present work deals with the impact of surface flow on hydrodynamic conditions in saturated underground domains. A three dimensional finite element analysis of water flow has been used to obtain the required simulations. The results clearly show the effects of the surface flow on the hydrodynamic conditions of the subsurface porous regions. This analysis is an important prerequisite for the prediction of contaminant mobility in soils and hence provides a convenient tool for the prediction of environmentally important subsurface flow processes. For low permeability cases, considered here, governing equations consist of water continuity and Darcy equations. These equations are solved using a robust and reliable finite element procedure.

## INTRODUCTION

Seepage in soils is an essential topic of study in many civil engineering and environmental protection processes. For example, in the design of earth dams and retaining structures, that require quantification of drainage, amount of seepage need to be identified. Similarly seepage is the determining factor in the contaminants mobility as leachates in subsurface domains. Seepage flow models have been developed by many researchers (e.g. see Cedergren, 1994; Reddi, 2003). These investigations have shown that seepage regimes often have complicated characteristics mainly because of the heterogeneity of soils through which water flows. Heterogeneity of soil media arises for various reasons such as the presence of staggered layers of soil with different porosity. Therefore reliable mathematical description and modelling of seepage flow requires formulation of realistic features of the problem and the boundary conditions affecting the flow. Transport equations describing flow through porous media depend on the properties of the fluid and factors such as permeability and porosity of media (Bear, *et al.* 1991). In this regard it is important to distinguish between saturated and unsaturated domains. Seepage flow in unsaturated lands has been the subject of many studies (e.g. Trolborg, 2009; Uromeihi,

2007). However, unlike the unsaturated situations, the influence of surface flow on saturated domains has not been widely studied. This may be due to the common belief that surface flow makes only small contribution to the changes of hydrodynamic conditions under the ground in saturated domains. However as the results presented in this paper show, there is a significant link between surface flow and subsurface conditions in saturated lands. Simulations obtained in this work can therefore be considered as quantitative analysis of the link between surface and subsurface flows for saturated cases.

The flow through porous soils is affected by hydraulic gradient and the coefficient of soil permeability. Permeability is a function of the range of grain size and shape, stratification, consolidation and cementation of the material. The rate of flow is commonly assumed to be directly proportional to the hydraulic gradient, however this is not always true under realistic conditions. Theoretically, due to the increasing load, permeability of soils decreases with increasing depths. Therefore a layered heterogeneous strata is the common feature of flow domain in most environmental studies. In table 1 a typical range of soil permeability used in the present study is shown.

| Degree of permeability | Range of coefficient of Permeability( $m^2$ ) | Soil type  |
|------------------------|---|--|
| High                   | $10^{-6}$                                     | Medium and coarse gravel   |
| Medium                 | $10^{-6}$ to $10^{-8}$                        | Fine gravel; coarse; medium and fine sand; dune sand; clean sand – gravel mixtures |
| low                    | $10^{-8}$ to $10^{-10}$                       | Very fine sand, silty sand, loose silt, loess, well fissured clays                 |
| Very low               | $10^{-10}$ to $10^{-12}$                      | Dense silt, dense loess clayed silt, poorly fissured clays                         |
| Impermeable            | $10^{-12}$                                    | Unfissured clays   |

Table 1 Values of soil permeability

Porous flow under the ground is governed by hydraulic head. Hydraulic head consists of the velocity, pressure and elevation heads. The velocity head in soils is usually negligible in comparison with the other heads, therefore here the main driving force is taken to be combined elevation and pressure heads, which for simplicity is represented by a single pressure term. Saturated flow in a porous medium should be modelled using either Darcy

(Darcy, 1856); or Brinkman (Brinkman, 1947) equations depending on the permeability/porosity of the medium and flow Reynolds number (Wakeman and Tarleton, 2005). For very low Reynolds number (creeping flow) and porosity less than 0.6, the most suitable equation is Darcy's equation. This is the dominant situation in most types of seepage flow of water in soils. In a previous paper we have considered the validation of the Darcy's law for isotropic, homogenous, incompressible, saturated and isothermal porous media (Kulkarni, *et al.* 2008). The Darcy equation inherently implies perfect slip conditions at domain boundary walls and does not include any wall effects (Ishizawa and Hori, 1966). Therefore an accurate solution scheme for this equation should be capable of yielding a slip velocity on the porous domains. The finite element scheme used in the present work can very effectively cope with such boundary conditions. This technique can also take into account irregular geometries. However, considering the large scale of lands where environmental phenomena needs to be studied, any irregularity of the domain walls can be ignored. In this work we have used a block as a representative section of a typical underground flow domain.

## MODEL EQUATIONS

The governing model equations for the seepage flow of an incompressible Newtonian fluid such as water are represented as:

### A. EQUATION OF CONTINUITY

The continuity equation (i.e. expression of conservation of mass) for an incompressible fluid is represented (using vector notations) as:

$$\nabla \cdot \vec{u} = 0 \quad (1)$$

Where  $\vec{u}$  is the velocity vector.

### B. EQUATION OF FLOW

As mentioned earlier in this work we have selected the Darcy equation to represent the flow equation (i.e. expression of conservation of momentum). Using vector notation this equation (Nield and Bejan, 1992) is written as:

$$\rho \frac{\partial \vec{u}}{\partial t} + \nabla p + \frac{\mu}{K} \cdot \vec{u} = 0 \quad (2)$$

Where  $\rho$  is fluid density,  $p$  is the pressure,  $\mu$  is viscosity of the fluid and  $K$  is the permeability of the porous medium. In its most general description  $K$  should be regarded as a second order tensor which is represented, in a matrix form, as:

$$K = \begin{bmatrix} K_{xx} & 0 & 0 \\ 0 & K_{yy} & 0 \\ 0 & 0 & K_{zz} \end{bmatrix} \quad (3)$$

Where  $K_{xx}$ ,  $K_{yy}$  and  $K_{zz}$  are the principle components of the permeability tensor along the  $x$ ,  $y$  and  $z$  directions of a Cartesian coordinate system. Any anisotropy in a porous medium can hence be taken into account by assigning appropriate values to the components of the permeability tensor.

Conjunctive solution of equations (1) and (2) poses a mathematical problem as the first equation does not include a pressure term. Full mathematical analysis of the problem is somewhat obscure and requires lengthy explanations. However, it has been shown that a stable and accurate solution for these equations can be obtained provided that the solution scheme satisfies a condition known as the LBB condition (Reddy, 1986). A convenient way of satisfying this condition is to replace equation (1), which is the expression of incompressibility, with a modified form (Zienkiewicz and Wu 1991) as:

$$\frac{1}{\rho c^2} \frac{\partial P}{\partial t} + \nabla \cdot \vec{u} = 0 \quad (4)$$

Where  $c$  is the speed of sound in the fluid. Equation (4) represents conservation of mass for a slightly compressible fluid.

## BOUNDARY CONDITIONS

In the present work simulations of underground flow of water are obtained by the conjunctive finite element solution of equations (2) and (4) in a three-dimensional domain subject to the following boundary conditions:

1. Top surface: Constant velocity tangent to the surface.
2. Side walls: Perfect slip conditions.
3. Bottom surface: Perfect slip conditions.
4. Downstream surface: Zero pressure set as an arbitrary datum.

## SOLUTION ALGORITHM

Numerical solution of equations (2) and (4) via the finite element method starts with the domain discretization into a computational mesh. Selection of a particular type of element is of importance in generating reliable results. In this work we have used 8 noded (brick elements) which generate tri-linear approximations for both pressure and velocity fields. Application of the Galerkin finite element technique (Nassehi, 2002) generates a set of working equations which can be used according to the algorithm shown in Fig. 1 to obtain the required numerical results.

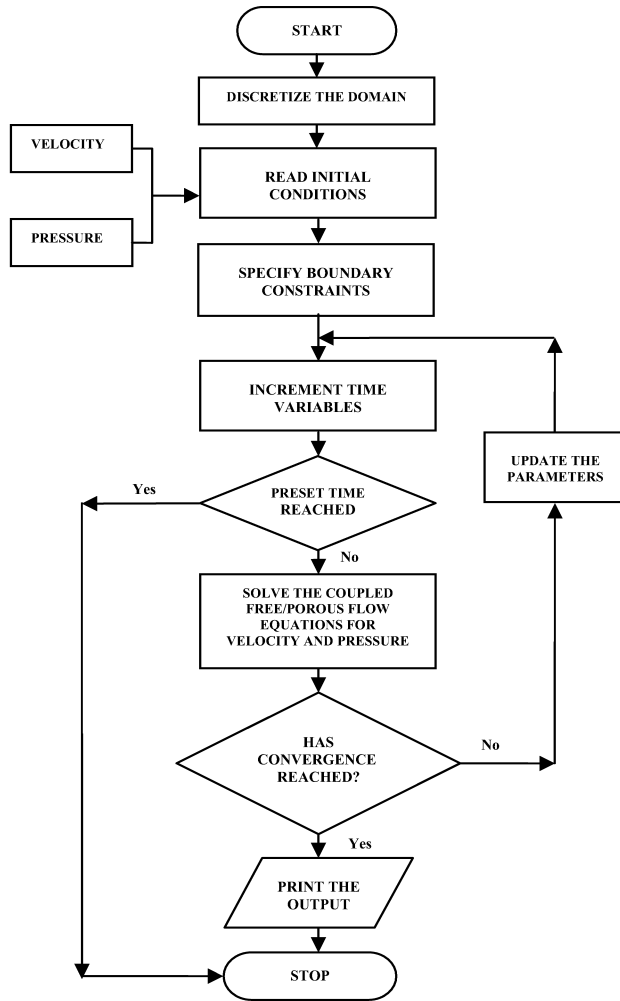


Figure 1: The solution algorithm implemented using the in-house developed program.

## COMPUTATIONAL RESULTS AND DISCUSSION

In this section the results obtained for a block domain (prism) subject to varying surface slope are presented and discussed.

In these simulations, the fluid under consideration is water with properties at  $20^{\circ}\text{C}$ , as viscosity  $= 0.001\text{ Kg m}^{-1}\text{s}^{-1}$  and density  $= 1000\text{ Kg m}^{-3}$ . The velocity of sound in water is taken to be approximately as  $10000\text{ ms}^{-1}$ . Depending upon the property of the permeable medium (which can be isotropic or anisotropic), appropriate values of permeability are used to cover a range of realistic situations. The time increment ( $\Delta t$ ) used in the solution scheme is 20 seconds.

### Block domain consisting of isotropic permeable medium

A block domain of 60 m width (W) x 60 m height (H) along x, y in one side of the domain and 60 m (W) x 30 m (H) along x, y in the other side of the domain and 200m length (L) along z axis is modelled. Therefore initially a surface slope of  $8.53^{\circ}$  is considered. The magnitude of surface velocity is  $0.2\text{ ms}^{-1}$ . The computational grid used to simulate porous flow in this domain comprises of 5760

eight noded brick elements and 6929 nodes, as shown in figure 2. The permeability coefficient for this homogeneous isotropic porous domain is taken as: ( $K_{xx} = K_{yy} = K_{zz} = 10^{-9}\text{ m}^2$ )

The schematic representations of the boundary conditions imposed over the different sides of this domain are shown in figures 3 and 4.

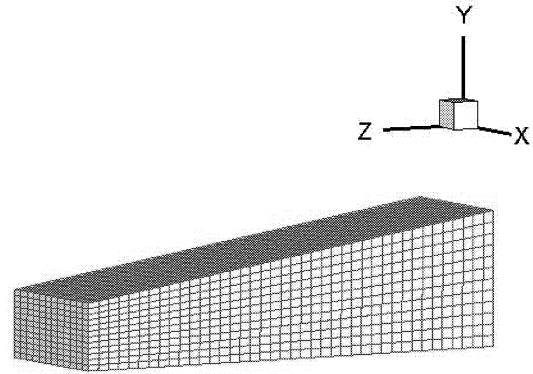


Figure 2: Finite element mesh for the homogeneous domain.

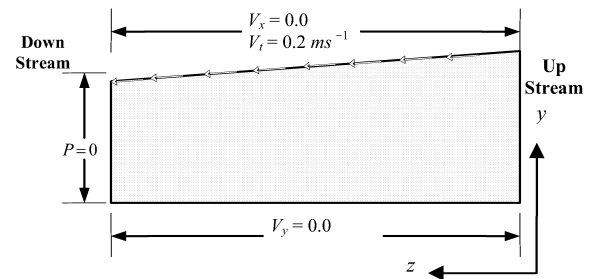


Figure 3: Schematic representation of the boundary condition in the yz plane

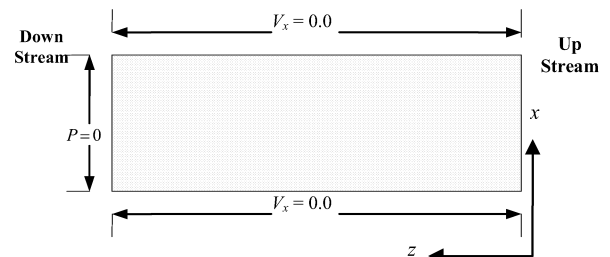


Figure 4: Schematic representation of the boundary condition in the xz plane

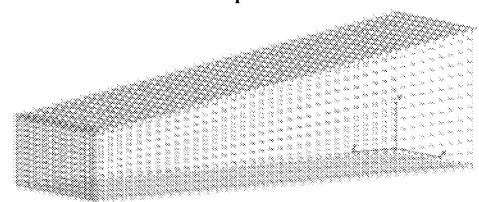


Figure 5: Three dimensional schematic representation of the boundary points at different faces of the domain.

Figure 5 shows the position of the boundary nodes on the faces of the permeable domain under consideration. In addition to the top surface flow velocity ( $0.2\text{ ms}^{-1}$ ) at the

downstream exit a zero pressure datum is prescribed. Slip wall boundary conditions are imposed on the sides of the domain. This means that velocity components vertical to each wall is set to be zero whilst the other components are left to be free.

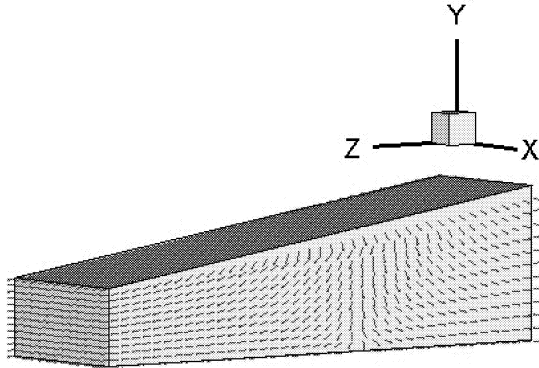


Figure 6: Velocity vector plot over the domain.

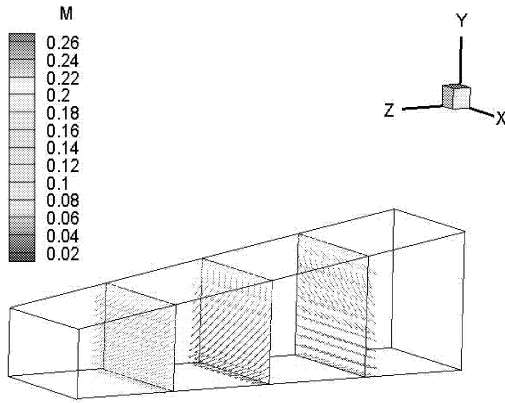


Figure 7: Velocity vector plot on three sample cross sections

To make the representation of the velocity field clearer, only the computed velocity vectors on three sample cross sections are shown in figure 7. The corresponding nodal pressures and magnitude of velocity are shown in figures 8 and 9, respectively. This result shows that, as intended, viscous stress associated with the bulk matrix of the fluid is transferred to the solid porous matrix.

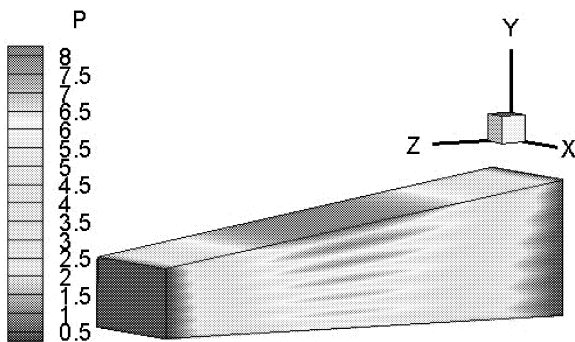


Figure 8: Pressure plot in the domain.

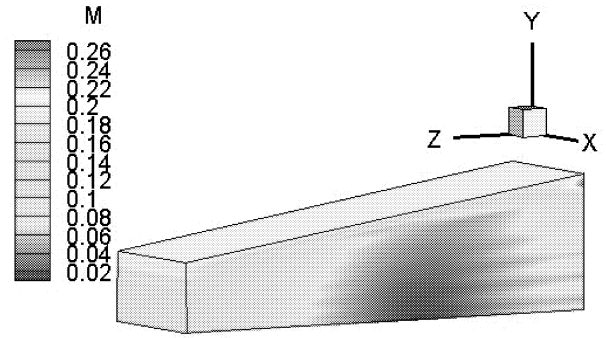


Figure 9: Velocity contour in the domain.

In figure 10 the simulated velocity in the middle section of domain parallel to zy plane is shown.

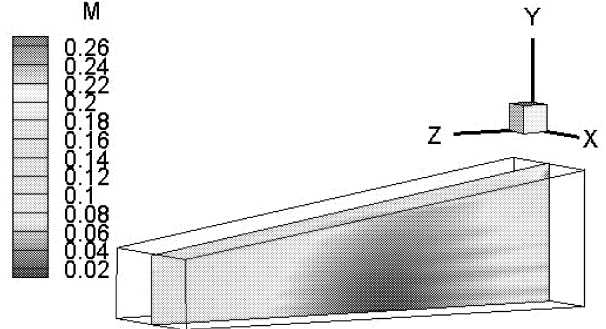


Figure 10: Velocity contour in the section parallel to the zy plane at  $x=30$  m.

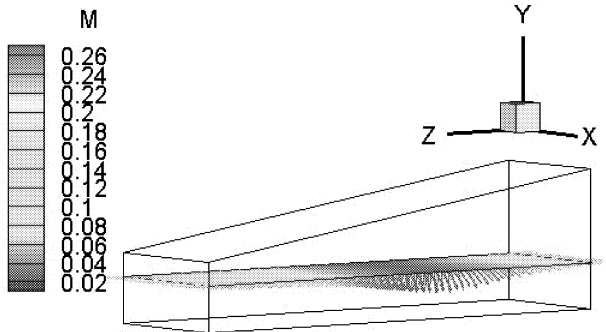


Figure 11: Velocity contour and vector plot in the surface parallel to the xz plane at  $y=20$  m.

Figures 12, 13 and 14 clearly show that despite the ground being saturated significant flow under the ground is generated by the imposed surface flow and, therefore, causing distribution of any existing contaminants in all directions.

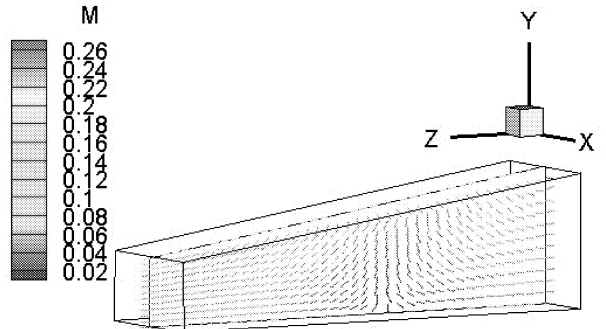


Figure 12: Velocity vector in the section parallel to the zy plane at  $x=30$  m in the domain.



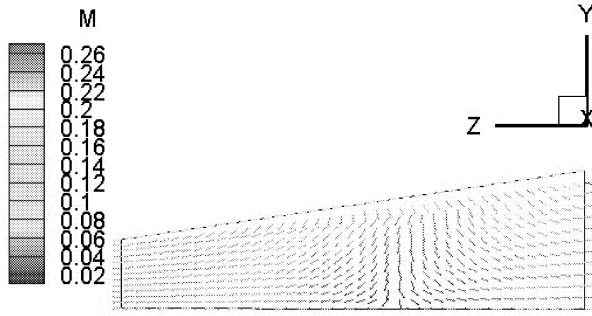


Figure 13: Velocity vector plot in the surface parallel to the yz plane at  $x=30$  m.

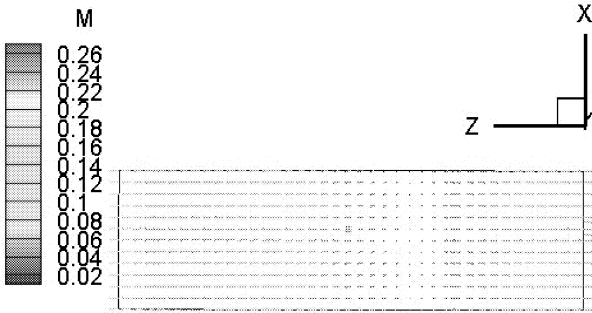


Figure 14: Velocity vector plot in the surface parallel to the xz plane at  $y=20$  m.

To study the influence of the surface slope on the intensity of underground circulation by the surface flow, a second domain which has a different surface slope of  $12.68^\circ$  is also simulated. All other boundary conditions are kept to be similar to the previous case. Simulation results shown in figures 15 to 19 clearly indicate the strong relation between the surface slope and the underground circulations caused by the surface flow. For example despite showing a pattern of circulation similar to the previous case, velocities have increased by as much as 38%.

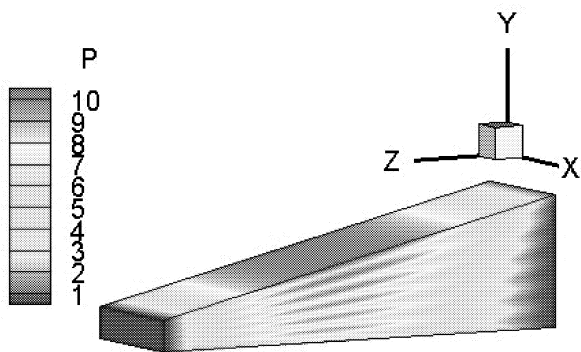


Figure 15: Pressure plot in the second domain.

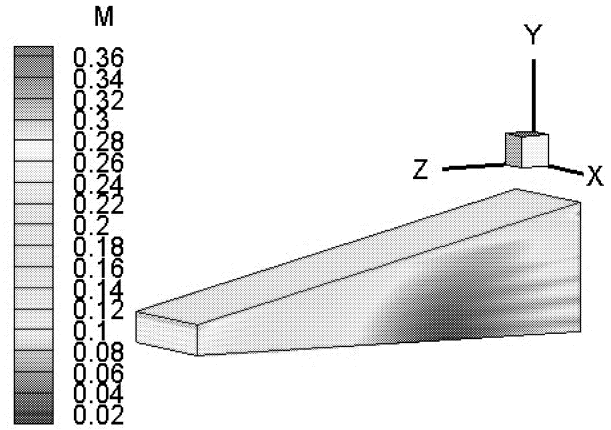


Figure 16: Velocity contour in the second domain.

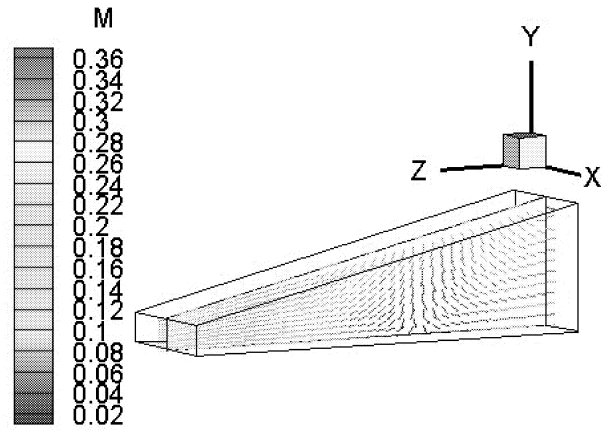


Figure 17: Velocity vector in the section parallel to the zy plane at  $x=30$  m in the second domain.

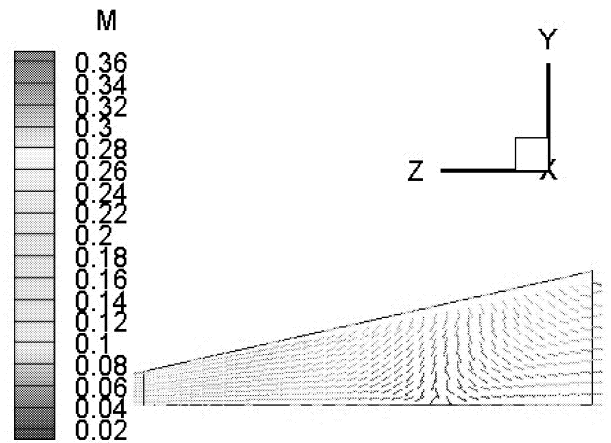


Figure 18: Velocity vector plot in the surface parallel to the yz plane at  $x=30$  m

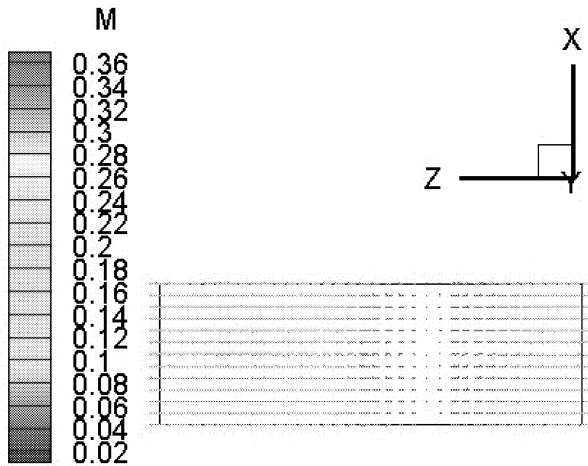


Figure 19: Velocity vector plot in the surface parallel to the xz plane at  $y = 20$  m.

The results shown in this section are self consistent and can be described logically and their patterns indicate a logical behaviour. The simulations do not show any sign of instability or any other numerical problems.

## CONCLUSIONS

The described simulation is based on the finite element solution of three dimensional Darcy and continuity equations. This provides a powerful means for the investigation of seepage flow of water in subsurface regions. The accuracy of the model is verified by its ability to preserve the agreement between the predicted velocity and pressure results with the theoretical expectations. Our simulations have shown that despite the land being saturated, rain water flow over the ground results in significant disturbance of the hydrodynamic equilibrium under the ground and hence causes movement of contaminants. It is also shown that water circulation under the ground is significantly affected by the surface flow velocity and the surface slope. As shown by these results the direction of flow under the ground does not exclusively correspond to the direction of water flow over the ground and can be in both upstream and downstream directions.

There is no theoretical difficulty that may prevent the extension of the methodology described in this paper to heterogeneous lands.

## REFERENCES

Bear, J., Buchlin, J., Vonkarman institute for fluid dynamics, (1991). "Modelling and applications of transport phenomena in porous media". Dordrecht, Boston, Kluwer Academic Publishers.

Brinkman, H.C., (1947). "On the permeability of media consisting of closely packed porous particles". Appl. Sci. Res. A1, pp. 81-86.

Cedertgren, H.R., (1989). "Seepage, Drainage, and Flow Nets", John Wiley and Sons, New York

Darcy, H.P.G., (1856). "Les Fontaines Publiques de la Ville de Dijon, Exposition et Application des Principes à Suivre et des Formules à Employer dans les Questions de Distribution d'Eau." Victor Dalmont, .

Das, D.B., Nassehi, V. and Wakeman, R.J., (2002). "A finite volume model for the hydrodynamics of combined free and porous flow in sub-surface regions". *Advances in Environmental Research*, 7(1), pp. 35-58.

Ishizawa, S. and Hori, E., (1966). "The flow of viscous fluid through a porous wall into a narrow gap" ( a consideration of the slip of fluid on porous wall surface). *Japanese Society of Mechanical Engineers*, 9, pp. 719.

Kulkarni, A., et al, (2008). "finite element modelling of under ground flow in low permeability land". ISC Lyon, France.

Nassehi, V. and Das, D.B., (2007). "Computational methods in the management of hydro-environmental systems". London, UK, IWA Publishing.

Nassehi, V. (2002). "Practical Aspects of Finite Element Modelling of Polymer Processing". John Wiley & Sons, Inc., West Sussex.

Nield, D.A. and Bejan, A., (1992). "Convection in porous media". New York, Springer -Verlag.

Reddy, J. N.(Junuthula Narasimha) 1984. "An Introduction to the Finite Element Method", McGraw - Hill Book Company, New York.

Reddi, L.N. (2003), "Seepage in soils, principles and applications" John Wiley, Hoboken, N.J.; Great Britain.

Troldborg, M., et al, (2009). "Unsaturated zone leaching models for assessing risk to groundwater of contaminated sites". *Journal of contaminant hydrology*, 105(1), pp. 28-37.

Uromeihy, A., (2007). "Evaluation and treatment of seepage problems at Chapar-Abad Dam", Iran

Wakeman, R.J. and Tarleton, E.S.,(2005). "Solid/liquid separation - Principles of industrial filtration", Oxford, Elsevier.

Zienkiewicz, O.C. and Wu, J., (1991). "Incompressibility without tears", How to avoid restrictions of mixed formulation. *International Journal for Numerical Methods in Engineering*, 32(6), pp. 1189-1203.

# Mathematical Model of Biofilm Growth

Yi-Ping Lo<sup>ab</sup>, John Ward<sup>a</sup>, Felipe Iza<sup>b</sup>, Rob Seager<sup>b</sup>, Michael Kong<sup>b</sup>.

Schools of <sup>a</sup>Mathematics and <sup>b</sup>Electronic and Electrical Engineering, Loughborough University

## Abstract

The aim of the paper is to derive a mathematical model of bacterial biofilm growth in two-dimensions using an Individual-Based model (IbM) to describe cell growth behaviour in prescribed environments. The approach extends the ideas of Xavier *et. al.* [1] and Drasdo and Hohme [2]. Bacteria are treated as volume occupying particles that respond in various ways to the concentrations of substrates present, whereby new cellular growth requires can lead to the repositioning of neighbouring cells in order to accommodate the changes in local volume. A Metropolis algorithm is applied to handle the cellular positioning adjustment and the coupled reaction-diffusion equations for each of the substrates are solved using finite-difference methods. The advantage of an IbM approach is that it can easily be extended to include multiple bacteria species with varying traits responding to multiple diffusible substrates. As an example, results are presented for a biofilm consisting of two species, which have differing tolerances to low nutrient environments.

## Introduction

Biofilms are multicellular bacterial cell colonies enclosed in a bacteria produced matrix of extracellular polymeric substrates (EPSs), irreversibly attached to a surface or interface [3]. Biofilm can adhere to surfaces of most material, abiotic or biotic, such as teeth, plants and food [4]. Once bacterial cells adhere on a surface, cells start growing and producing EPS to form a biofilm. The constituents of a biofilm can vary considerably and dynamically, consisting mostly of water, bacteria and EPS. The role of EPS not only encloses the bacteria, but stabilises the cell-cell interaction within the biofilm and limits the diffusion of substrates through it [4]. Figure 1 illustrates the initiation, development and maturation of a biofilm. Biofilm are the cause of many medical and industrial problems. Bacteremia and endocarditis may occur due to biofilm developing in the blood vessels [5]. In industrial settings, microbial contamination can induce corrosion of metals due to biofilms growing on metal surfaces [6]. Current techniques to remove or reduce biofilms include the application of mechanical force, such as vortexing and sonication [7]. The mathematical modelling approach that is currently being undertaken is to formulate a framework to simulate growth and treatment of biofilms, for example to investigate novel treatments such as the use of plasma beams in controlling biofilm growth.

Biofilm growth depends on a range of environmental factors, such as nutrient levels (oxygen, carbon and

nitrogen, [9, 3, 10]). When nutrient is at sufficient levels, most of bacterial cells produce EPS and an energy storage substrate poly- $\beta$ -hydroxybutyrate (PHB) [11]. When starved of carbon and other energy sources, cells take their energy from PHB in order to survive [11].

In this paper, we derive a hybrid continuum-discrete mathematical model of biofilm and development, often referred to as an individual-based model (IbM). In short, discrete objects (e.g. bacteria) and its attributes are simulated against a backdrop of factors described using continuum models. In comparison to continuum models, IbMs are generally easier to simulate in higher dimensions and easier to simulate the mechanics of cell contact interactions (basically applying simple rigid body mechanics). There are two main issues of the IbM model. Firstly, the diffusion and consumption of the substrate in the biofilm, e.g. nutrient, oxygen, carbon. etc. Secondly, cells are growing and dividing which leads cell contact and movement.

In the remainder of the paper we will give an outline of the model development and solution procedure, and then present results of a particular case study.

## 2-D Individual-Based Modelling

The general modelling approach will enable us to consider a biofilm consisting of multiple species responding to numerous substrates. Here, we will focus as an example inspired from [1], of a two species biofilm consisting of an EPS producing strain and PHB producing strain. The former strain will be the most active at high nutrient levels, but more vulnerable to death in low nutrient environments. The nutrient is sourced at the top of the domain and infiltrates the domain via diffusion where it is consumed by the cells. To simplify the modelling, we assume cells are cylindrical discs, with radius  $r$  that varies as the cell grows. In this particular investigation,

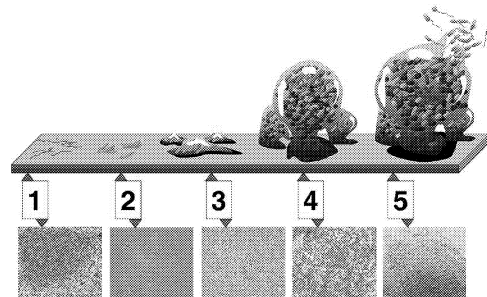


Figure 1: Schematic and photomicrographs of the development of a biofilm from initial colonisation to maturation. [8]

EPS only acts as secondary food source and is not involved in the biofilm structure. The PHB produced by the PHB producing strain is only available to the individual cells that has produced it. The modelling domain is the rectangular region  $0 \leq x \leq x_0$  and  $0 \leq y \leq y_0$ , where the base on which bacteria initially settle is at  $y = 0$  and nutrients infiltrate from the top  $y = y_0$ .

## Cells division and movement

The model assumes that cell particles cannot move actively, but can move by being pushed by neighbouring cells or through cell division. When a particle grows to reach a critical size,  $R_D$ , the particle divide to produce two identically sized particles (see Figure 2 (a)). The centre of one of these particles is maintained in the original position, whilst the other is attached but positioned randomly. Cell contacts that results from growth will lead to, on first iteration, to overlap of neighbours. To eliminate we apply the so called Metropolis Algorithm [12] to force cells apart in an appropriate manner. We have modified the particle response formula to reflect the role of particle inertia in particle contact interaction. For a particular (primary) particle, the direction of movement resulting from an overlap is given by the vector sum  $\mathbf{R}_0$  of the overlap radii, namely

$$\mathbf{R}_0 = \sum_{n=1}^N \frac{kr_0 + r_n - d}{1 + \left(\frac{r_0}{r_n}\right)^2} \mathbf{f}_n, \quad (1)$$

where  $N$  is the number of neighbouring particles (indexed using  $n$ ),  $r_0$  is the radius of the primary particle,  $r_n$  is the radius of the neighbouring particle  $n$ ,  $d$  is the distance between the centre of this two particles,  $k$  is a factor to adjust average minimum distance between particles to realistic value, which is taken to be 1.3 [13] and  $\mathbf{f}_n$  is the unit vector in the direction of the line connecting the centres of the primary and neighbouring particle  $n$  (see Figure 2(b)). Equation 1 ensures that a smaller particles moves further than a larger ones when

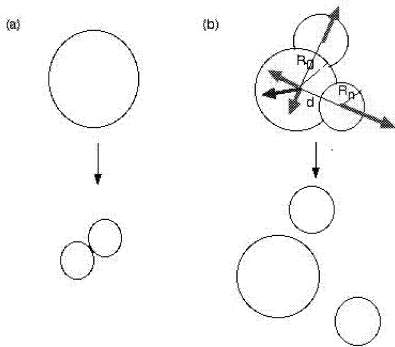


Figure 2: (a) Schematic of the division. One cell maintain at the same location, the other just attaches on it. (b) Schematic of cells movement. When cells overlap, cells push each other.

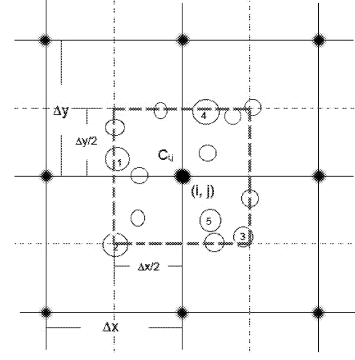


Figure 3: Schematic of the the mesh lattices of the domain. All of the cell particles are smaller than lattice.  $C_{i,j}$  is the nutrient concentration in the bold-dashed border  $(i, j)$ .

they contact, which is not guaranteed in related studies [12, 13].

## Nutrient diffusion

The nutrient can be viewed as a carbon source. The nutrient diffuses into the domain from the top and is consumed by the bacteria, described by the quasi-steady diffusion equation

$$D \left( \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \right) - q \frac{C}{C + K} + p = 0, \quad (2)$$

where  $C$  is the nutrient concentration,  $D$  the nutrient diffusion coefficient and  $K$  is a Michaelis-Menten constant. The functions  $q(B_E, B_P)$  describe consumption of nutrients and  $p(B_E, B_P)$  is a nutrient source from dead cells, where  $B_E$  and  $B_P$  are the volume fractions of the EPS and PHB producing bacteria respectively. Zero flux is assumed on all boundaries except  $C = C^*$  on  $y = y_0$ .

For numerical purposes, we assume the domain is divided up into a uniform mesh, in which the lattice size is larger than a bacterial particle radius (see Figure 3). Finite difference methods and Broyden's method [14] was used to solve (2). The volume fractions of bacterial particles used to evaluate functions  $p$  and  $q$  were calculated from the volume occupied by the cells in a box surrounding each mesh point (see Figure 3).

## Numerical results

For the simulations below, we started by seeding randomly 10 cells into the domain, 5 each of the EPS and PHB producing strains. The initial radius of each particle was set randomly, but smaller than  $R_D$ . The result shows a simulation of 13 hours biofilm growth (see Figure 4). Figure 5 shows that nutrients have been significantly depleted in the region of where the biofilm is forming. After 13 hours growth, the total number of EPS producer particles is significantly greater than

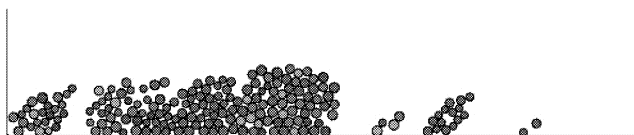


Figure 4: The distribution of the bacteria strains after a simulated time of 13 hours. The red cells are PHB producer particles and blue cells are EPS producer particles. The red cell intensity is uniform throughout, indicating that these particles are mostly alive. The intensity of blue for the EPS producer particles is variable with the darker the blue, the greater the proportion of living and active cells present.

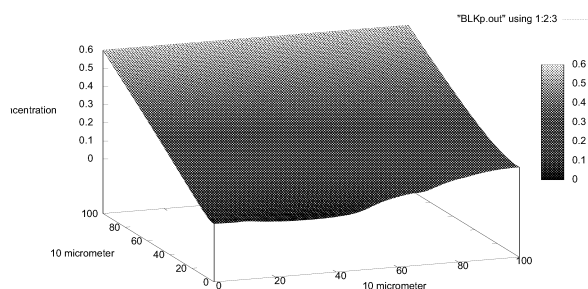


Figure 5: The distribution of nutrient concentration in the domain at simulated time of 13 hours.

PHB producer particles. This reflects their greater activity in the nutrient rich environment that existed in the early stages, however, they have grown to a depth in which nutrient diffusion limitations is evident, resulting in cell death. In contrast, the PHB cells that were originally less active are successfully surviving in the low nutrient environments due to the PHB reserves accrued in the early stages of relatively high nutrient availability. The simulation highlights the possibilities of significant heterogeneity in strains and the live/dead cell distribution that can develop within a biofilm. The blue intensity seems to be positioned randomly (light blue indicating greater cell death), these being due to the EPS cells dividing in a random direction being able to infiltrate deeper into the biofilm, moving any dead particles present elsewhere. The living cells deeper into the biofilm will eventually die, but the mingling of the dead particles in the living cell regions is an interesting prediction and may influence how cells behave within the biofilm; a feature that will be explored in the future. This simulation serves to demonstrate what the model in its current state can achieve and the type of results it can generate.

## Summary

In this paper, an individual-based model of biofilm growth was presented. The intention is to construct a

modelling framework that will give a realistic description of biofilm growth that can straightforwardly be adapted and extended to model a wide variety of situations. We have combined and extended the modelling approaches of earlier biofilm studies (Kreft *et al.* [13]) and models of tumour growth (Drasdo and Hohme [2]). We are currently extending these ideas further to account, for example, the important role of the EPS in biofilm development and cell movement; by assuming EPS as a compressible, but not too compressible, entity our approach will further depart from the earlier works and will, we feel, result with a more accurate model. As a case study of the current model, we looked at the effect of nutrient limitation on a biofilm consisting of EPS and PHB producing bacteria, where simulations showed the heterogeneous distribution of bacteria strains that can form. Such non-uniform distributions can have consequences in the applications of treatment, in which the issue of penetration and phenotypic variation can impact on a treatment's success. This will be explored in the future with our investigations into the effect of plasma beam therapy on biofilms.

## References

- [1] J.B. Xavier, C. Picioreanu, and M. van Loosdrecht. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environmental Microbiology*, 7(8):1085–1103, 2005.
- [2] D. Drasdo and S. Hohme. Individual-based approaches to birth and death in avascular tumors. *Mathematical and Computer Modelling*, 37:1163–1175, 2003.
- [3] R.M. Donlan. Biofilms: Microbial life on surfaces. *Emerging Infectious Diseases*, 8(9):881–889, September 2002.
- [4] K-P. Katharine and E. Karatan. Biofilm development in bacteria. *Advances in Applied Microbiology*, 57:97–111, 2005.
- [5] K. Okuda, T. Kato, and K. Ishihara. Involvement of periodontopathic biofilm in vascular diseases. *Oral Dis.*, 10:5–12, 2004.
- [6] I.B. Beech and J. Sunner. Biocorrosion: Towards understanding interactions between biofilms and metals. *Curr. Opin. Biotechnol.*, 15:181–186, 2004.
- [7] R.M. Donlan and J.W. Costerton. Reviews biofilms: Survival mechanisms of clinically relevant microorganisms. *Clinical Microbiology Reviews*, 15(2):167–193, April 2004.
- [8] Davis. 5 stages of biofilm development. 2007. accessed on [wikipedia] url: <http://en.wikipedia.org/wiki/file:biofilm.jpg>.
- [9] S. Hogg. *Essential Microbiology*, chapter 4. Microbial Nutrition and Cultivation. Chapter 5. Microbial Growth, pages 79–107. John Wiley & Sons. Ltd, 2005.
- [10] R. E. McKinney. *Environmental Pollution Control Microbiology*, chapter 3. Bacteria growth, pages 52–85. Marcel Dekker Inc., New York, 2004.
- [11] E. A. Dawes and D. W. Ribbons. Some aspects of the endogenous metabolism of bacteria. *American Society for Microbiology*, 28(2):126–149, 1964.
- [12] D. Drasdo, R. Kree, and J.S. McCaskill. Monte carlo approach to tissue-cell populations. *Physical Review E*, 52(6):6635–6657, 1995.
- [13] J-U. Kreft, C. Picioreanu, J.W.T. Wimpenny, and van Loosdrecht C.M. Individual-based modelling of biofilms. *Microbiology*, 147:2897–2912, 2001.
- [14] Y. Chen and D. Cai. Inexact overlapped block broyden methods for solving nonlinear equations. *Applied Mathematics and Computation*, 136:215–228, 2003.

# MODELLING, SIMULATION AND CONTROL OF A SEMI-BATCH INDUSTRIAL POLYMERIZATION REACTOR

Nadja Hvala\*, Teodora Miteva\*, Dolores Kukanja\*\*

\*J. Stefan Institute, Department of Computer Automation & Control

Jamova cesta 39, SI-1000, Ljubljana, Slovenia

E-mail: [nadja.hvala@ijs.si](mailto:nadja.hvala@ijs.si), [teodora.miteva@ijs.si](mailto:teodora.miteva@ijs.si)

\*\*Mitol d.d., Partizanska cesta 78,

SI-6210 Sežana, Slovenia

E-mail: [dolores.kukanja@mitol.si](mailto:dolores.kukanja@mitol.si)

## KEYWORDS

Semi-batch polymerization, modelling, energy balance model, gPROMS, simulation, optimization.

## ABSTRACT

This paper presents a detailed temperature model for a polymerization process and its performance on real plant data. The model is based on energy balance, and together with the already designed reaction model enables the design and test of possible control strategies for shortening the batch time and increasing the polymer production. The model was used to design a control algorithm that adjusts the initiator and monomer flow during the batch. The simulation results suggest that shortening of a batch cycle for about 1 hour could be expected, which means around 6% production increase.

## INTRODUCTION

The research work in this paper considers the polymerization process in MITOL, a chemical factory in Sežana, Slovenia. This is the leading glue manufacturer in the country with the production rate more than 90% of its capacity. The challenge we are facing now is to design optimal control strategies in order to increase the production rate and decrease the batch time, while keeping the quality parameters at the desired levels.

The production process is a semi-batch polymerization process with initiator and monomer added during the batch. The large quantities of raw materials and equipment resources prevent us from performing experiments for batch optimization directly on the plant. Therefore we need a good predictive model of the process. With such a model, the study of possible control approaches to plant optimization is done by simulation, and only the most promising approaches are then tested and verified on the real plant.

While modelling of polymerization processes is well addressed in scientific literature and also supported by standard chemical engineering software packages (e.g. Aspen Polymers), the model was designed from scratch because of specific process characteristics (semi-batch process controlled by initiator and monomer dosing during the batch

and not by usually used coolant feed temperature). For that purpose a model consisting of differential and algebraic equations (DAE) has already been constructed (Aller et al., 2007). The model predicts four outputs, i.e. conversion, particle size diameter, solids content and viscosity. They are computed from the initial and during-the-batch dosing of batch components. The model is, however, not able to predict the changes in the temperature profile as a result of different batch operating conditions. Therefore, the model was extended by additional differential equations to estimate the reactor temperature as a function of the concentrations of the reacting chemicals.

In this paper we present the modelling of the polymerization reactor temperature using the energy balance model. The model was validated on a set of industrial batches and then used in simulation experiments to optimise the initiator and monomer dosing during the batch.

## PROCESS DESCRIPTION

The process considered is polymerization of vinyl acetate using potassium persulfate (KPS) as initiator and polyvinyl alcohol as protective colloid. The production process is semi-batch polymerization. In the beginning initial amounts of monomer and initiator are added into the reactor as well as the whole amount of polyvinyl alcohol and water. The reactor is closed and the heating starts. The heating of the reactor is done by pumping hot water in the heating jacket around the reactor. The hot water is pumped with a temperature of 90°C. When the temperature in the reactor reaches approximately 65-70°C, the operator stops the hot water. The exothermic reaction and the heat of the remaining water filling the heating coat continue to rise the temperature. After the reactor temperature reaches a certain level, the remaining monomer starts to be pumped into the reactor with a continuous flow. The monomer flow is adjusted to control the mass of the monomer dosing tank. The mass set-point is on-line computed and linearly decreased to get a constant monomer flow. The opening of the monomer dosing valve is determined by a P (proportional) controller. The temperature of the added monomer is the outdoors temperature.

At the beginning of the reaction, the temperature can reach up to 82°C, but afterwards, for the product quality reasons, it is necessary to keep the temperature between 75°-80°C.

Temperature control is performed by manually adding a small amount of initiator every time the temperature decreases. When all the monomer is added into the reactor, a larger amount of initiator is added in order to terminate the reaction. The temperature is then allowed to reach 90°C. The reaction is considered as finished when the temperature starts to decrease again.

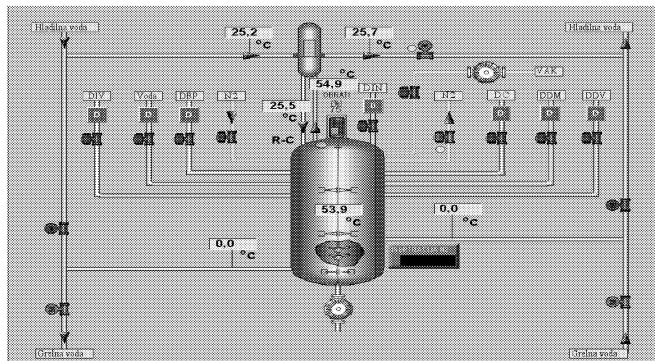


Figure 1: Scheme of the polymerization reactor

The main variables affecting the duration of the reaction and possibly used in the batch optimization are: reactor temperature, flow rate of the monomer and the addition of the initiator.

## PROCESS MODEL

## Reaction model

The main reactor ingredients are water, vinyl acetate as monomer, potassium persulfate (KPS) as initiator, polyvinyl alcohol as stabiliser and converted monomer.

The moles of *initiator* in the reactor  $I$  are increased because of initiator flow in the reactor  $I_e$ , and decreased because of initiator thermal decomposition that is modelled as a first order rate equation (Finch, 1973, Pramojaney, 1982)

$$\frac{dI}{dt} = I_e - k_d I \quad (1)$$

To account for higher rate of decomposition of KPS at higher temperature the initiator decomposition rate  $k_d$  was determined as

$$k_d = 8.4 \cdot \left( \frac{T - T_{start}}{40} \right)^2 + 1.6 \cdot \left( \frac{T - T_{start}}{40} \right) \quad (2)$$

where  $T$  is reactor temperature and  $T_{start}$  is the temperature when an appreciable decomposition is observed.

The moles of *monomer* in the reactor  $M_m$  are increased because of the monomer flow in the reactor ( $q_m$ ) that is present in a substantial part of the batch, and decreased because of the polymerization reaction

$$\frac{dM_m}{dt} = q_m - r_{pol} \quad (3)$$

where polymerization reaction rate  $r_{pol}$  takes into account the consumption of monomer in the water phase due to initiation, propagation and chain transfer reactions, and, above all, in the particle phase, which is the main locus of polymerization. It is given as

$$r_{pol} = M_w(k_{iwm}[R_\cdot] + k'_{iwm}[M_1\cdot] + k_{pw}([R_w\cdot] - [M_1\cdot]) + k_{rm}([R_w\cdot] - [M_1\cdot])) + k_{pp}[M_p\cdot] \bar{n} \sum_{i=1}^{NoC} PSD_i \quad (4)$$

where  $M_w$  is the mass of monomer dissolved in the water,  $[M_p]$  is the concentration of monomer in the particle phase (monomer swelling the particles),  $[R\cdot]$  is the concentration of primary radicals from initiator,  $[M_1\cdot]$  is the concentration of monomeric radicals of length 1,  $[R_w\cdot]$  is the concentration of oligomeric initiator and monomeric radicals. Rate constants  $k_{iwm}$ ,  $k_{iwm}'$ ,  $k_{pw}$ ,  $k_{pp}$  and  $k_{irm}$  stand for propagation of (unitary) initiator radicals, propagation of (unitary) monomeric radicals, propagation of initiator and monomeric radicals in the water phase, propagation in the polymer phase and chain transfer to monomer reaction, respectively. The amount of radicals in the particle phase is expressed by the average number of radicals per polymer particle  $\bar{n}$ , and the amount of particles. The particles are modelled according to their size and are presented with a preset number of groups (*NoC*) of different particle size.  $PSD_i$  is the mass of particles in group  $i$ . In our case *NoC* was equal to 5. More detailed information on radical and particle modelling can be found in (Aller et al., 2007; Gilmore et al., 1993a; Gilmore et al., 1993b).

The *polyvinyl alcohol* covers the particles, providing steric stabilization. The partition of the PVOH between the water phase, the droplets and the particles is modelled with the Langmuir model (Immanuel et al., 2002), based on adsorption isotherms

$$V_w[G_w] + K_{gd}[G_w]V_d + \frac{A_{\infty}K_{ad}[G_w]A_{ps}}{1 + K_{ad}[G_w]} = \frac{G_0}{MW_{PVOH}} \quad (5)$$

where  $[G_w]$  is the concentration of PVOH in the water phase.  $K_{gd}$  is the partition coefficient for surfactant between droplets and water,  $V_d$  is the volume of droplets,  $A_{c\infty}$  and  $K_{ad}$  are Langmuir adsorption constants,  $A_{ps}$  is the total surface area of the swollen particles,  $G_\theta$  is the total mass of PVOH and  $MW_{PVOH}$  its molecular weight.

## Temperature model

Temperature in the reactor is determined from the energy balance model (Zeaiter et al., 2002; Freire and Giudici, 2003) that takes into account the reaction heat capacity, and the energy produced and consumed during the reaction.

The reaction heat capacity  $K_R$  is calculated from the heat capacities of reactor ingredients, i.e. monomer in the reactor  $M_m$ , the monomer converted into polymer  $M_{conv}$ , water and polyvinyl alcohol  $G_\theta$

$$K_R = M_m MW_m Cp_{mon} + M_{conv} MW_m Cp_{pol} + \rho V_w Cp_{water} + G_0 Cp_{pol} \quad (6)$$

where  $Cp$  denotes specific heat capacities of different reactor ingredients.

Energy is consumed or released because of the following:

- heating of the reactor through the heating jacket,
- heating the incoming monomer,
- heat produced in the exothermic reaction,
- cooling the reactor by the reflux through the condenser,
- heat losses to the surroundings.

The heating of the reactor through the heating jacket  $\Delta H_{jacket}$  is given by

$$\Delta H_{jacket} = UA(T_{jacket} - T) \quad (7)$$

where  $T$  is the temperature in the reactor,  $T_{jacket}$  is the temperature in the reactor jacket,  $U$  is the heat transfer coefficient and  $A$  is the heat transfer area. The temperature in the jacket is modelled as follows

$$dT_{jacket}/dt = -\frac{UA}{m_w Cp_{water}}(T_{jacket} - T) \quad (8)$$

where  $m_w$  is the mass of the water in the reactor jacket and  $Cp_{water}$  is specific heat capacity of water.  $UA$  was estimated at low temperatures as  $0.00018 \cdot K_R$  as described in Zeaiter et al. (2002).

The energy needed to heat the incoming monomer  $Q_{mon}$  depends on the monomer feed rate  $q_m$ , the monomer specific heat capacity  $Cp_{mon}$  and its temperature  $T_{mon}$

$$Q_{mon} = q_m MW_m Cp_{mon} T_{mon} \quad (9)$$

The heat produced in the exothermic polymerization reaction  $Q_{pol}$  is proportional to polymerization reaction rate  $r_{pol}$  and the heat of polymerization  $\Delta H_r$

$$Q_{pol} = -\Delta H_r r_{pol} \quad (10)$$

where  $r_{pol}$  is determined in (4).

The cooling of the reactor by the reflux through the condenser  $Q_{cond}$  is also taken into account. It is determined from the estimated flow through the condenser  $q_{FC}$ , its specific heat capacity  $Cp_{Cout}$ , temperature difference of the incoming and outgoing flow and the vaporization heat  $\lambda$

$$Q_{cond} = q_{FC} Cp_{Cout} (T_{Cout} - T_{Cin}) + q_{FC} \lambda \quad (11)$$

where  $\lambda$  is calculated by the following equation

$$\lambda = -0.00281 T_{Cout}^2 - 0.32738 T_{Cout} + 408.178 \quad (12)$$

The heat losses  $Q_{loss}$  to the surroundings are modelled as

$$Q_{loss} = C_{loss} (T - T_{ext}) \quad (13)$$

where  $C_{loss}$  is an adjustable parameter and  $T_{ext}$  is the outside temperature.

Temperature model is divided into three parts depending on the stage of the process and the reaction (Figure 2).

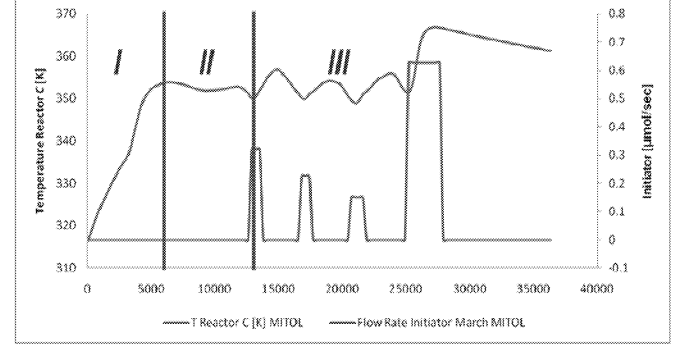


Figure 2: Division of temperature model in three phases

The first part starts when we close the reactor and start heating it up with hot water through the heating jacket. It ends when the pumping of hot water is stopped and the temperature reaches around 80°C. In that first part the heat losses are negligible and there is no flow through the condenser. The temperature in the reactor is determined from the following equation

$$K_R \frac{dT}{dt} + T \frac{dK_R}{dt} = q_m MW_m Cp_{mon} T_{mon} - C_{pol,1} \Delta H_r r_{pol} - \Delta H_{jacket} \quad (14)$$

The left side of the equation accounts for the changes in the internal energy of the reactor, which is due to the change of the reactor temperature (first term) and the change of its heat capacity (second term).

The second part of the modelled temperature lasts from around 80°C until the first initiator addition during the batch. In that part the reactor is no more heated through the heating jacket, while the heat losses and the cooling by the reflux are still negligible. The temperature model includes only monomer enthalpy and the reaction heat. The equation is the following

$$K_R \frac{dT}{dt} + T \frac{dK_R}{dt} = q_m MW_m Cp_{mon} T_{mon} - C_{pol,2} \Delta H_r r_{pol} \quad (15)$$

In the last part, the model for reactor temperature takes into account also the reflux through the condenser and the heat losses to the surroundings. The equation has the following form

$$K_R \frac{dT}{dt} + T \frac{dK_R}{dt} = q_m MW_m Cp_{mon} T_{mon} - C_{pol,3} \Delta H_r r_{pol} - Q_{cond} - C_{loss,1} Q_{loss} \quad (16)$$

Close to the end of the batch, the values of model parameters  $C_{pol,3}$  and  $C_{loss,1}$  are changed to  $C_{pol,4}$  and  $C_{loss,2}$ , respectively.



## Process output model

Based on simulated process components the output quality parameters, i.e. conversion, particle size diameter, solids content and viscosity are also determined. The conversion  $x$  is given by

$$x = 100 \frac{M_{conv}}{M_t} \quad (17)$$

where  $M_t$  is the total mass of monomer added into reactor. The particle size diameter  $D_{pu}$  is calculated as

$$D_{pu} = \left( \frac{6000 M_{conv} MW_m}{\pi \rho_p N_p} \right)^{1/3} \quad (18)$$

where  $N_p$  is the total number of particles and  $\rho_p$  is the polymer density. Solids content  $SC$  is expressed as the sum of the weights of the polymer and the PVOH divided by the total weight of the latex

$$SC = \frac{M_{conv} MW_m + G_0}{M_t MW_m + G_0 + \rho V_w} \quad (19)$$

The viscosity is estimated from the following expression

$$v = \frac{v_0}{\left( 1 - \frac{SC}{SC_{ref}} \right)^2} \quad (20)$$

where  $v_0$  is the viscosity of the water and  $SC_{ref}$  is a solids content reference value estimated from real plant measurements of the solids content and viscosity.

The overall model has thirteen differential equations, sixty-three algebraic equations and nine adjustable parameters.

## MODEL SIMULATION RESULTS

### Temperature model validation

The presented model was simulated using gPROMS modeling tool (PSE, 2004). Model simulations were performed based on real plant data. The temperature model parameters (Table 1) were taken from the literature (Berger and Meyerhoff, 1989) or estimated ( $UA$ ,  $C_{pol,i}$ ,  $C_{loss,i}$ ) so that a satisfactory agreement with real plant temperature data is obtained. Figure 3 shows the simulated reactor temperature as a comparison to real plant measurements. The inputs to the model were real plant monomer and initiator flows that were calculated from the weight changes of the monomer and initiator dosing tanks.

With the presented model we are able to get relatively good estimates of the reactor temperature. From Figure 3 it can be seen that the estimated profile follows the dynamics of the real one. Not so good estimates are obtained at the beginning and at the end of the batch. This may be attributed to still insufficient tuning of the corresponding model parameters.

At the beginning of the batch, the most influential part of the model is the heating of the reactor through the heating jacket, while at the end of the batch the heat losses are most important. So, the corresponding model parameters still need to be estimated more precisely or even on-line adjusted during the batch. Namely, Clarke-Pringle and MacGregor (1997) report that heat transfer coefficient decreases during the batch because of increased viscosity, and from batch to batch because of increased fouling of the reactor walls between cleaning periods. In addition, the gel effect may cause an increase in the reaction rate during the batch. This is well in accordance with our experience, where increasing values of reaction rate parameter  $C_{pol,i}$  were used during the batch (parameters  $C_{pol,1}$  to  $C_{pol,4}$ ), while the values of the heat losses parameter  $C_{loss,i}$  were adjusted for each batch and decreased during the batch (parameters  $C_{loss,1}$  and  $C_{loss,2}$ ). Some inspection into this problem is still needed.

Table 1: The values of temperature model parameters

| Parameter       | Value                                  |
|-----------------|--|
| $C_{p_{Cout}}$  | $1.17 \text{ J g}^{-1} \text{ K}^{-1}$ |
| $C_{p_{mon}}$   | $1.17 \text{ J g}^{-1} \text{ K}^{-1}$ |
| $C_{p_{pol}}$   | $1.77 \text{ J g}^{-1} \text{ K}^{-1}$ |
| $C_{p_{water}}$ | $4.18 \text{ J g}^{-1} \text{ K}^{-1}$ |
| $\Delta H_r$    | $-87.5 \text{ kJ mol}^{-1}$            |
| $MW_m$          | $86.09 \text{ g mol}^{-1}$             |
| $T_{Cin}$       | $287.15 \text{ K}$                     |
| $T_{Cout}$      | $358.15 \text{ K}$                     |
| $T_{ext}$       | $283.15 \text{ K}$                     |
| $T_{mon}$       | $283.15 \text{ K}$                     |
| $UA$            | $0.00018 \cdot K_R$                    |
| $C_{pol,1}$     | 0.3                                    |
| $C_{pol,2}$     | 0.45                                   |
| $C_{pol,3}$     | 0.6                                    |
| $C_{pol,4}$     | 0.85                                   |
| $C_{loss,1}$    | $0.9 \text{ J s}^{-1} \text{ K}^{-1}$  |
| $C_{loss,2}$    | $0.5 \text{ J s}^{-1} \text{ K}^{-1}$  |

### Overall model validation

The overall model was used to calculate the product quality parameters. Table 2 shows the validation of the model for five batches, where batch quality parameters were measured after each batch, while in normal production they are determined after mixing several batches. The production requirements are to produce batches with conversion over 99%, solids content between 44-46%, and viscosity between 27000–34000mPas.

From the table it can be seen that the model output variables are in good agreement with real plant data. The modelled conversion is for some batches a little lower than measured, which can be attributed to lower simulated temperature in the last part of the batch. Besides, some bigger deviations can be noticed for batch 1214 for all output variables. The RMSE (root mean square error) for conversion is 2.5%, for solids content 1.5% and for viscosity 13% (excluding the batch 1214 it is below 10%).

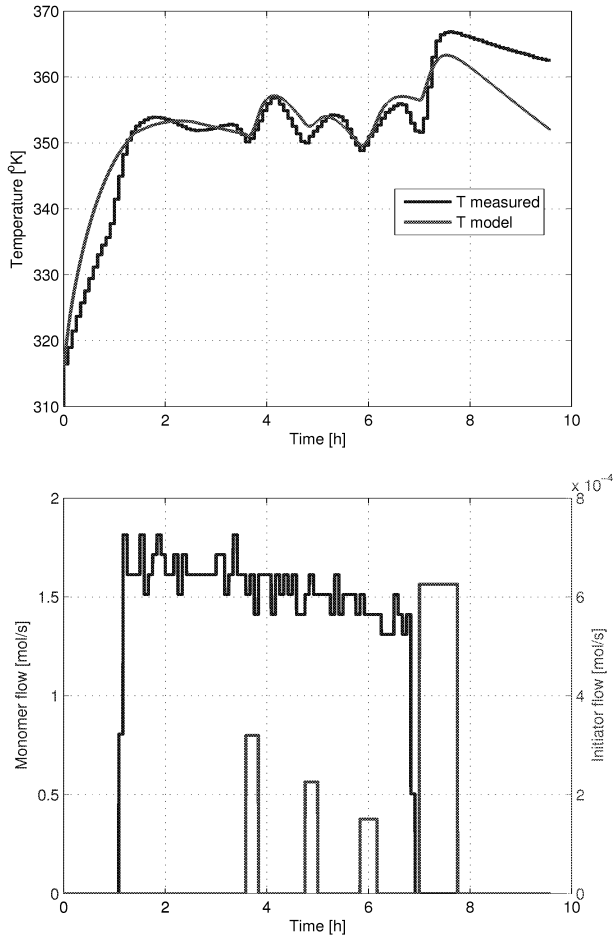


Figure 3: Simulated and real plant reactor temperature (upper diagram) and real plant monomer and initiator flows (lower diagram).

Table 2: Real plant and model output results

| Batch No. | Time [s] | Conversion [%] |       | Solids content [%] |       | Viscosity [mPas] |       |
|-----------|----------|----------------|-------|--------------------|-------|------------------|-------|
|           |          | Process        | Model | Process            | Model | Process          | Model |
| 1192      | 34500    | 99.95          | 99.78 | 46.4               | 47.15 | 37520            | 33618 |
| 1203      | 31800    | 99.90          | 99.15 | 45.9               | 46.91 | 31200            | 35419 |
| 1214      | 33600    | 99.89          | 94.92 | 45.4               | 44.59 | 25040            | 31879 |
| 1253      | 35700    | 99.13          | 98.87 | 46.9               | 47.06 | 34000            | 35911 |
| 1256      | 35100    | 99.37          | 97.07 | 46.4               | 45.87 | 30160            | 31706 |

## INITIATOR AND MONOMER DOSING CONTROL

The model was used for the design and test of control algorithms to shorten the batch time. Proper temperature control is one of the main concerns, and different advanced and model-based approaches are reported (Beyer et al., 2008; Zeaiter et al., 2002; Crowley and Choi, 1996; Pringle and MacGregor, 1997). In our case, the main idea was to maximize the monomer flow during the batch, which may cause, however, an undesired increase of reactor temperature, and consequently, lead to the synthesis of off-spec polymer at elevated temperatures. For that purpose the control algorithm was designed as a combined monomer and initiator dosing control based on reactor temperature (Figure 5).

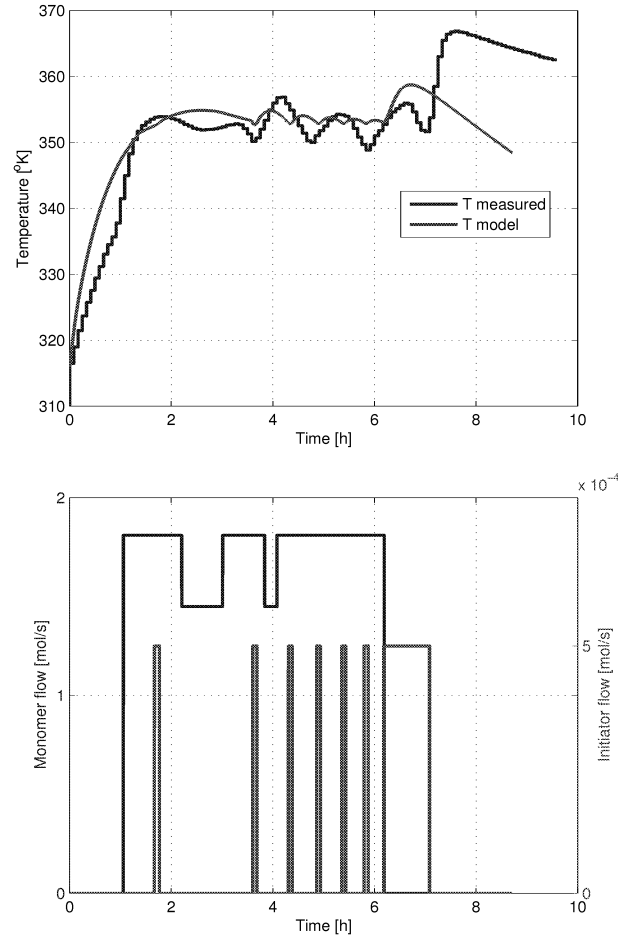


Figure 4: Simulated temperature in the reactor with applied temperature control in comparison with real plant temperature (upper diagram), and the corresponding monomer and initiator flows (lower diagram)

The initiator dosing control resembles the operator control strategy and is adding initiator if the reactor temperature is below the low temperature limit  $T_L$ . On the contrary, the monomer dosing reduces the monomer flow if the reactor temperature exceeds the high temperature limit  $T_H$ . When all the monomer is added into the reactor, a larger amount of initiator is added into the reactor in a similar way as in the real plant.

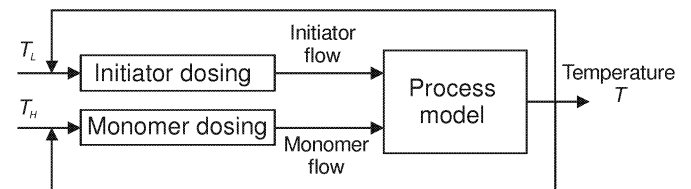


Figure 5: Temperature control acting on initiator and monomer flow

The simulation results of the proposed control strategy are shown in Figure 4. The values of  $T_L$  and  $T_H$  were chosen as 353K and 354.5K, respectively. The initiator flow rate was 0.0005mol/s, the monomer flow rate was 1.81mol/s, and the reduced monomer flow was 1.45mol/s.

From the figure it can be seen that with the applied control, the temperature during the monomer addition is kept in a narrower region compared to the real plant temperature. This is achieved despite the higher monomer flow (compare the real plant results in Figure 3 and simulation results with the applied control in Figure 4). Tighter temperature control is mainly achieved by the initiator dosing that prevents very high temperature peaks by small initiator additions. At the same time, by more frequent additions it does not allow the temperature to drop off and so preserves high reaction rate. In this way the reduction of the monomer flow at high temperatures is rarely needed, which contributes to shorter batch time.

Table 3 shows the output results obtained by simulating the five batches with the proposed initiator dosing control algorithm and keeping the monomer flow at 1.81 mol/s (560 kg/h). It can be seen that the output results are similar to that when simulating the real plant behaviour, only the conversion is on average a little higher, which can be attributed to higher reactor temperature due to tighter temperature control. It can be also seen that the reaction batch time was on average reduced for about 1 hour, which means about 6% production increase at a total batch time of 16 hours. While the simulation results of the proposed strategy are quite promising, they should be validated with experiments on the real plant.

Table 4 also presents the average values and standard deviations for the output parameters of five batches as measured on the plant, simulated with the model and simulated together with the control. From the table it can be seen that the simulation results are quite in accordance with the plant measurements, only the viscosity average value as predicted by the model is higher than measured, while the viscosity standard deviation is smaller than that obtained in the plant.

Table 3: Simulation results with the applied control strategy

| Batch No. | Time [s] | $\Delta$ Time [s] | Conversion [%] | Solids content [%] | Viscosity [mPas] |
|-----------|----------|-------------------|----------------|--------------------|------------------|
| 1192      | 30130    | 4370              | 99.55          | 47.04              | 33653            |
| 1203      | 29750    | 2050              | 98.80          | 46.72              | 35304            |
| 1214      | 30790    | 2810              | 99.19          | 46.89              | 35496            |
| 1253      | 30640    | 5060              | 99.06          | 46.83              | 36451            |
| 1256      | 30560    | 4540              | 98.67          | 46.66              | 34293            |

Table 4. Comparison of average output values and their standard deviation for the real process, the simulation model and the simulation with the applied control

| Output parameter   |          | Process | Model | Control |
|--------------------|----------|---------|-------|---------|
| Conversion [%]     | Average  | 99.65   | 97.96 | 99.05   |
|                    | St. dev. | 0.33    | 1.77  | 0.31    |
| Solids content [%] | Average  | 46.2    | 46.32 | 46.83   |
|                    | St. dev. | 0.51    | 0.98  | 0.13    |
| Viscosity [mPas]   | Average  | 31584   | 33707 | 35039   |
|                    | St. dev. | 4149    | 1740  | 975     |

## CONCLUSIONS

This paper presents a detailed temperature model for a polymerization process and its performance on real plant data. The model is based on energy balance, and together with the already designed reaction model enables the design and test of possible control strategies for shortening the batch time and increasing the polymer production. The model was used to design a control algorithm that adjusts the initiator and monomer flow during the batch in such a way to keep high monomer addition while controlling the reactor temperature within tight temperature limits. The simulation results based on real plant data show good model agreement with real plant data with respect to reactor temperature as well as output product parameters. The simulation results of the control algorithm suggest that shortening of a batch cycle for about 1 hour could be expected, which means around 6% production increase. The proposed control measures will be applied on a real plant, while more sophisticated modelling and control approaches, i.e. on-line adjustment of model parameters and predictive control will be tested in the future research work.

## ACKNOWLEDGMENT

The support of the European Commission in the context of the 6th Framework Programme (PRISM, Contract No. MRTN-CT-2004-512233) is greatly acknowledged.

## NOMENCLATURE

|                     |   |
|---------------------|---|
| $A$                 | Heat transfer area [m <sup>2</sup> ]  |
| $A_{cso}$           | Langmuir adsorption constants [mol m <sup>-2</sup> ]  |
| $A_{ps}$            | Total surface area of the swollen particles [m <sup>2</sup> ]   |
| $C_{pCout}$         | Specific heat capacity of reflux flow [J g <sup>-1</sup> K <sup>-1</sup> ]  |
| $C_{pmon}$          | Specific heat capacity of monomer [J g <sup>-1</sup> K <sup>-1</sup> ]  |
| $C_{pol}$           | Specific heat capacity of polymer [J g <sup>-1</sup> K <sup>-1</sup> ]  |
| $C_{pwater}$        | Specific heat capacity of water [J g <sup>-1</sup> K <sup>-1</sup> ]  |
| $D_{pu}$            | Diameter of polymer particles [cm]  |
| $G_0$               | Total mass of PVOH [g]  |
| $[G_w]$             | Conc. of PVOH in the water phase [mol l <sup>-1</sup> ]   |
| $\Delta H_{jacket}$ | The heat from the heating jacket [J s <sup>-1</sup> ]   |
| $\Delta H_r$        | Heat from the polymerization [J mol <sup>-1</sup> ]   |
| $I$                 | Moles of initiator in the reactor [mol]   |
| $I_e$               | Initiator feed rate [mol s <sup>-1</sup> ]  |
| $K_{ad}$            | Langmuir adsorption constant [l mol <sup>-1</sup> ]   |
| $K_d$               | Initiator decomposition rate constant [s <sup>-1</sup> ]  |
| $K_{gd}$            | partition coefficient for surfactant between droplets and water [-]   |
| $K_R$               | reaction heat capacity [J K <sup>-1</sup> ]   |
| $k_{iwm}$           | Rate constant for propagation of (unitary) initiator radicals [l mol <sup>-1</sup> s <sup>-1</sup> ]                        |
| $k_{iwm}'$          | Rate constant for propagation of (unitary) monomeric radicals [l mol <sup>-1</sup> s <sup>-1</sup> ]                        |
| $k_{pp}$            | Rate constant for propagation in the polymer phase [l mol <sup>-1</sup> s <sup>-1</sup> ]                                   |
| $k_{pw}$            | Rate constant for propagation of initiator and monomeric radicals in the water phase [l mol <sup>-1</sup> s <sup>-1</sup> ] |
| $k_{trm}$           | Rate constant for chain transfer to monomer reaction [l mol <sup>-1</sup> s <sup>-1</sup> ]                                 |
| $M_m$               | Moles of monomer in the reactor [mol]   |

|              |  |
|--------------|--|
| $M_t$        | Total mass of monomer added in the reactor [mol]                             |
| $M_{conv}$   | Moles of monomer converted into polymer [mol]                                |
| $M_w$        | Moles of monomer dissolved in the water [mol]                                |
| $m_w$        | Mass of water in the reactor jacket  |
| $[M_p]$      | Conc. of monomer in the particle phase [mol l <sup>-1</sup> ]                |
| $[M_f]$      | Conc. of monomeric radicals of length 1 [mol l <sup>-1</sup> ]               |
| $MW_{PVOH}$  | PVOH molecular weight [g mol <sup>-1</sup> ]                                 |
| $MW_m$       | Monomer molecular weight [g mol <sup>-1</sup> ]                              |
| $\bar{n}$    | Average number of radicals per polymer particle [-]                          |
| $N_p$        | Total number of particles [-]  |
| $NoC$        | Number of groups of particles of different size [-]                          |
| $PSD_i$      | Moles of particles of group $i$ [mol]  |
| $SC$         | Solids content [%]   |
| $SC_{ref}$   | Solids content reference value estimated from plant measurement [%]          |
| $q_m$        | Monomer feed flow rate [mol s <sup>-1</sup> ]                                |
| $q_{FC}$     | Flow through the condenser [g s <sup>-1</sup> ]                              |
| $Q_{cond}$   | Heat removed by the reflux [J s <sup>-1</sup> ]                              |
| $Q_{loss}$   | Reactor losses [J s <sup>-1</sup> ]  |
| $Q_{mon}$    | Energy needed to heat the feeding monomer [J s <sup>-1</sup> ]               |
| $Q_{pol}$    | Heat produced in the polymerization reaction [J s <sup>-1</sup> ]            |
| $[R\cdot]$   | Conc. of primary radicals from initiator [mol l <sup>-1</sup> ]              |
| $[R_w\cdot]$ | Conc. of initiator and monomeric radicals [mol l <sup>-1</sup> ]             |
| $[R_w\cdot]$ | Conc. of initiator and monomeric radicals [mol l <sup>-1</sup> ]             |
| $r_{pol}$    | Polymerization reaction rate [mol s <sup>-1</sup> ]                          |
| $T_{Cin}$    | Temperature in the reflux into the react [K]                                 |
| $T_{Cout}$   | Temperature in the reflux of the reactor [K]                                 |
| $T_{ext}$    | Outside temperature [K]  |
| $T_{jacket}$ | Temperature in the reactor jacket [K]  |
| $T_{mon}$    | Temperature of the feeding monomer [K]                                       |
| $U$          | Heat transfer coefficient [W m <sup>-2</sup> K <sup>-1</sup> ]               |
| $V_d$        | Volume of droplets [l]   |
| $V_w$        | Water volume [l]   |
| $x$          | Conversion [%]   |
| $\lambda$    | Vaporization heat [J g <sup>-1</sup> ]                                       |
| $\rho$       | Water density [g l <sup>-1</sup> ]   |
| $\rho_p$     | Polymer density [g l <sup>-1</sup> ]   |
| $\nu$        | Viscosity of the latex [Poise]   |
| $\nu_0$      | Viscosity of the water [Poise]   |
| $C_{pol,i}$  | Adjustable temperature model parameters [-]                                  |
| $C_{loss,i}$ | Adjustable temperature model parameters [J s <sup>-1</sup> K <sup>-1</sup> ] |

## REFERENCES

Aller, F., Kandare, G., Blázquez, L. F., Kukanja, D., Jovan, V., and Georgiadis, M. C. (2007). Model-based optimal control of the production of polyvinyl acetate. European

Congress of Chemical Engineering ECCE-6, Copenhagen, 16-20 September, 2007.

Berger K.C., and Meyerhoff G. (1989). *Polymer handbook* (3th edition). Wiley, New York.

Beyer, M. A., Grote, W., Reining G. (2008). Adaptive exact linearization control of batch polymerization reactors using a Sigma-Point Kalman Filter. *Journal of Process Control*, **18**, 663-675.

Clarke-Pringle, T., and MacGregor J. F. (1997). Nonlinear adaptive temperature control of multi-product, semi-polymerisation reactors. *Computers and Chemical Engineering*, **21**, 1395-1409.

Crowley, T. J., and Choi K. Y. (1996). On-line monitoring and control of a batch polymerization reactor. *Journal of Process Control*, **6**, 119-127.

Finch A. (ed.) (1973). *Polyvinyl Alcohol: Properties and Applications*. John Wiley and Sons Ltd, 1973.

Freire, F. B., and Giudici R. (2003). Temperature Oscillation Calorimetry by Means of a Kalman-like Observer: The joint estimation of  $Q_r$  and  $U_A$  in a Stirred Tank Polymerization Reactor. In *Polymer Reaction Engineering V* (ed. João B. P. Soares), Quebec, Canada, Macromolecular Symposia, Wiley.

Gilmore, C. M., Poehlein, G. W., and Schork, F. J. (1993a). Modeling Poly(Vinyl Alcohol) – Stabilized Vinyl Acetate Emulsion Polymerization. I. Theory. *Journal of Applied Polymer Science*, **48**, 1449-1460.

Gilmore, C. M., Poehlein, G. W., and Schork, F. J. (1993b). Modeling Poly(Vinyl Alcohol) – Stabilized Vinyl Acetate Emulsion Polymerization. II. Comparison With Experiment. *Journal of Applied Polymer Science*, **48**, 1461-1473.

Immanuel C. D., Cordeiro C. F., Sundaram, S. S., Meadows E. S., Crowley T. J., Doyle F. J. III (2002). Modeling of particle size distribution in emulsion co-polymerization: comparison with experimental data and parametric sensitivity studies. *Computers and Chemical Engineering*, **26**, 1133-1152.

Process Systems Enterprise Ltd. (2004). *gPROMS Introductory User Guide*. London, UK.

Pramojaney, N. (1982). *A Mathematical Model for an Emulsion Polymerization Reactor*. Ph.D. thesis, Lehigh University.

Zeaiter, J., Gomes, V. G., Romagnoli, J. A., and Barton, G. W. (2002). Inferential conversion monitoring and control in emulsion polymerisation through calorimetric measurements. *Chemical Engineering Journal*, **89**, 37-45.

# SOLVING CHALLENGING PROBLEMS IN THE OIL INDUSTRY USING ARTIFICIAL INTELLIGENCE BASED TOOLS

Ana Vila Fernandes  
Department of Products Exchange  
Petrogal – Galp Energia  
Sines  
Portugal  
E-mail: anavilafernandes@gmail.com

Henrique Vicente  
Department of Chemistry and  
Chemistry Centre of Évora  
The University of Évora  
Évora, Portugal  
E-mail: hvicente@uevora.pt

José Neves  
Department of Informatics  
The University of Minho  
Braga  
Portugal  
E-mail: jneves@di.uminho.pt

## KEYWORDS

Data Mining; Knowledge Discovery from Databases; Decision Support; Effluent Quality; Decision Trees.

## ABSTRACT

Predictive modelling is a process used in predictive analytics to create a statistical model of future behaviour. Predictive analytics is the area of data mining concerned with forecasting probabilities and trends. On the other hand, Artificial Intelligence (AI) concerns itself with intelligent behaviour, i.e. the things that make us seem intelligent. Following this process of thinking, in this work we have as the main goal the assessment of the impact of using AI based tools for the development of *intelligent* predictive models, in particular those that may be used to classify industrial waste, such as the residual waters in a refinery, based on the type of the mixtures of crude oil that arrive into the site to be processed.

Indeed, these models will enable the prediction of the quality classes of the effluents that will be disposed, in order to assure that Industrial Residual Water does not affect negatively the ecology of the receptors or the Staff Health and Safety and obeys the current legislation. The software employed was Clementine 11.1 and C5.0 Algorithm was used to induce decisions trees. The data in the database was collected from 2006 to 2007, and includes production data and effluent analysis data.

## INTRODUCTION

New technological breakthroughs provided new ways of creating and storing information. Indeed, different organizations daily accumulate great amounts of information referring to different processes, based on the assumption that large volumes of data may be a source of knowledge which can be used to improve their behaviour, by discovering trends and specificities, and accelerating the course of efficient decision making.

However, the conventional tools for data analysis have a great number of drawbacks, since they do not allow the detection of singularities inside such a data chaos. Undeniably, having in mind a response to a given number of difficulties (e.g. those resulting from the use of great amounts of data, multiple sources of data or several

application domains) a new area of Knowledge Discovery from DataBases (KDDB) was brought to life and its tools and techniques to problem solving have been since then enforced. The designation KDDB was formally adopted in 1989 and refers to a process that involves the identification and recognition of patterns in a Database, in an automatic process, i.e. obtaining relevant, unknown information, that can be useful in a decision making process, without a previous formulation of hypothesis (Fayyad et al. 1996; Thuraishingham 1999).

The interest in ecological mining has been growing in the last decades. Undeniably, several Data Mining techniques have been used to find patterns in water quality databases, such as Decision Trees (DTs), Artificial Neural Networks (ANNs) and Genetic Algorithms (GAs) (Dzeroski 2002; Santos et al. 2005; Chau 2006; Kuo et al. 2007). Although DTs have not been extensively used in ecological modeling, they have the advantage of expressing regularities explicitly and thus are easy to inspect for ecological validity.

In this paper, an approach to make the assessment of effluent quality by Data Mining models is exploited. Currently, the assessment of effluent quality is done through analytical methods, which configures a very restricted approach due to the distances, the number of parameters to be considered and the financial resources spent to obtain their values. Moreover, in the present setting, it would be added the latency times between the sampling moment and the outcome of the laboratory analyses. Due to these constraints, the development of Data Mining based models in conjunction with the development of a Decision Support System is a better alternative for the quality management of effluent resources (Turban et al. 2004). In reality, these models have shown to be very helpful to predict the classification of industrial residual water in five different classes, according to the current legislation and in three groups, which are Satisfactory discharge (S), UnSatisfactory discharge (US) and Avoidable discharges (A).

The present study took place in *Petróleos de Portugal - Petrogal, S.A. (Petrogal or the Company)* located 2,5 Km east from the coastal town of Sines, Portugal (Figure 1). *Petrogal* is the *Galp Energia* group company that operates the refining and oil products distribution business. The *Company* has operations across the whole domestic territory and is also present, through associate companies, in *Spain, Africa, Venezuela* and *Brazil*. In *Portugal*, *Petrogal* is the

only refiner with a daily refining capacity of 310,000 barrels and *The Company's* main activities consist of the research and exploration of crude oil and natural gas, the refining, transport, distribution and sale of crude oil and derivative products and natural gas. *Petrogal* has majority control over the distribution of refined products with the trademark *GALP*, which it owns (Galp 2006).

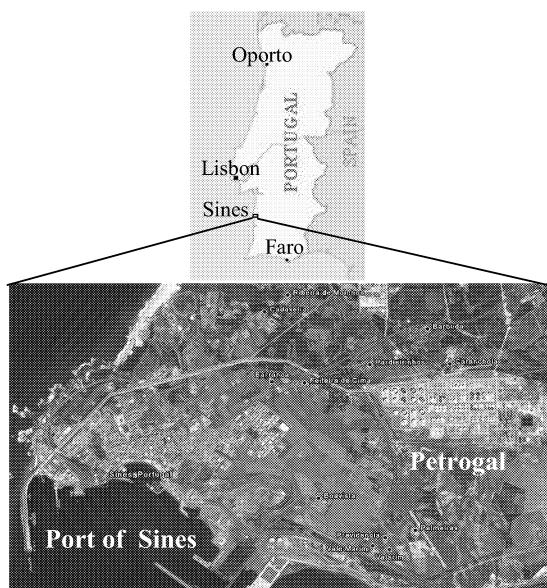


Figure 1: Petrogal location. Source - Google Earth

The paper is organized as follows: firstly, the ecological data is presented and explained and the decision trees are introduced. The experiments are then performed and described, being the results analysed in terms of different criteria. Finally, the conclusions are drawn.

## MATERIALS AND METHODS

### Sample Collection and Preservation

Separated effluent samples were collected directly to glass bottles and preserved. For chemical oxygen demand, analysis of the samples were collected in wide-mouth dark glass bottles of 500 cm<sup>3</sup> and preserved with concentrated sulphuric acid, pH ≤ 2; for oil and grease analysis the samples were collected in wide-mouth glass bottles of 1000 cm<sup>3</sup> and preserved with hydrochloric acid 1:1, pH ≤ 2; for phenolic compounds analysis the samples were collected in wide-mouth glass bottles of 1000 cm<sup>3</sup> and preserved with 2 cm<sup>3</sup> of concentrated sulphuric acid and for sulphide analysis the samples were collected in wide-mouth dark glass bottles of 500 cm<sup>3</sup> and preserved with 1 cm<sup>3</sup> of zinc acetate 1 mol dm<sup>-3</sup> and 0,5 cm<sup>3</sup> of sodium hydroxide 6 mol dm<sup>-3</sup>.

The effluent samples were collected from a location situated before the wastewater collector of the wastewater treatment plant of Ribeira de Moinhos (Figure 2) where, according to Santo André Wastewater Plant Regulation, the sampling frequency is two times per week (RARISA 2007).

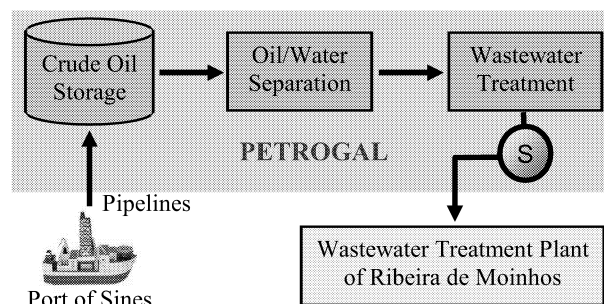


Figure 2: A simplified diagram of the Petrogal plant, showing sampling site S

### Analysis

The determination of pH was executed by the electrometric method, according to SMEWW 4500-H<sup>+</sup>. In this method the basic principle is potentiometric measurement, using a Metrohm 780 pH meter; the Total Suspended Solids (TSS) were evaluated according to SMEWW 2540 D. In this method the TSS are dried at 103 to 105°C. The oil and grease were determined according to SMEWW 5520 C. Measurements were carried out on a ATI MATTSON, Génesis Séries FTIR; the Phenolic Compounds were determined by molecular absorption spectrometry for use at a wavelength of 500 nm according to SMEWW 5530 D; the sulphides were determined by molecular absorption spectrometry for use at a wavelength of 664 nm according to SMEWW 4500-S<sup>2</sup>D. The molecular absorption spectrometry measurements were carried out on a UNICAM UV 300 spectrophotometer. Finally, the Chemical Oxygen Demand (COD) was determined according to SMEWW 5220 B (Eaton et al. 2005).

### Database

The data used in this study was collected from 2006 to 2007, containing a total of 153 records with 52 fields. Table 1 shows a synopsis of the relevant fields of the dataset, and Figure 3 presents a dataset fragment. The main fields in the dataset were the characteristics of the refining process, the crude oil predominant and the effluent features. For the refining process characterization, the indicators chosen were Process Time: it is the duration of the production process, in days; Load: it is the crude oil load processed in the refinery in m<sup>3</sup>/h, which is in the interval 732...1430 m<sup>3</sup>/h; API Density: density is a required measurement of the value of crude oil - one oil of a density inferior to 0,83 g cm<sup>-3</sup> is an oil of high economical value (Gomes and Alves 2007); TAN (Total Acid Number): is defined as the amount, in mg of KOH, necessary to titrate one gram of crude oil; (%) of each Crude Oil Type: as an indicator the % of crude oil that constitutes each one of the mixtures that get into the refinery to be processed. Each mixture contains different types of crudes that come from different places in the world.

Table 1: The Main Fields in the Dataset

| Refining Process                         | Crude Oil Predominant               | Effluent Features     |
|--|-------------------------------------|-----------------------|
| 1. Process Time /days                    | 1. Name                             | 1. pH                 |
| 2. Load / m <sup>3</sup> h <sup>-1</sup> | 2. API Density                      | 2. TSS                |
| 3. API Density                           | 3. Sulphur / wt %                   | 3. COD                |
| 4. TAN (Total Acid Number)               | 4. Acidity / mg KOH g <sup>-1</sup> | 4. Oils and Grease    |
| 5. % of each Crude Oil Type              | 5. Country                          | 5. Phenolic Compounds |
|  | 6. Region                           | 6. Sulphides          |

|   | A                   | B                                     | C           | D    | E                 |     | AS              | AT  | AU  | AV  |
|---|---------------------|---------------------------------------|-------------|------|-------------------|-----|-----------------|-----|-----|-----|
| 1 | Process Time / days | Load / m <sup>3</sup> h <sup>-1</sup> | API Density | TAN  | % Alba Condensate | ... | Region of World | pH  | TSS | COD |
| 2 | 4.2                 | 35900                                 | 0.8745      | 0.15 | 0.0               |     | Middle East     | 8.1 | 64  | 272 |
| 3 | 5                   | 35900                                 | 0.8549      | 0.68 | 10.0              |     | Latin America   | 8.0 | 98  | 294 |
| 4 | 5                   | 35900                                 | 0.8549      | 0.68 | 10.0              |     | Latin America   | 8.0 | 63  | 208 |
| 5 | 3                   | 35900                                 | 0.8631      | 0.67 | 20.0              |     | Africa          | 8.0 | 65  | 264 |

Figure 3: A Dataset fragment

According to the Santo André Wastewater Plant Regulation (RARISA, 2007), the classification of the quality of the industrial effluent will be classified into five classes (Table 2), in a non linear scale, I, II, ... V, where I denotes a slightly polluted industrial effluent and V denotes

an extremely polluted industrial effluent, which represents serious risks in terms of public and environmental health and implies the payment of high taxes by Petrogal (Figure 4). If any parameter exceeds the maximum for class V the discharge is forbidden and the classification is FD.

Table 2: Classification of the Quality of the Industrial Effluent (RARISA, 2007)

| Parameters         | Unit   | Class I  | Class II   | Class III  | Class IV   | Class V     |
|--------------------|--|----------|------------|------------|------------|-------------|
| pH                 | Sörensen scale   | $\geq 6$ | $\geq 6$   | $\geq 6$   | $\geq 6$   | $\geq 4,5$  |
|                    |  | $\leq 9$ | $\leq 9$   | $\leq 9$   | $\leq 9$   | $\leq 10$   |
| COD                | mg (O <sub>2</sub> ) dm <sup>-3</sup>                  | $< 150$  | $\geq 150$ | $\geq 300$ | $\geq 600$ | $\geq 1000$ |
|                    |  |          | $< 300$    | $< 600$    | $< 1000$   | $< 2000$    |
| TSS                | mg dm <sup>-3</sup>                                    | $< 100$  | $\geq 100$ | $\geq 200$ | $\geq 300$ | $\geq 500$  |
|                    |  |          | $< 200$    | $< 300$    | $< 500$    | $< 1000$    |
| Oils and Grease    | mg dm <sup>-3</sup>                                    | $< 5$    | $\geq 5$   | $\geq 20$  | $\geq 35$  | $\geq 50$   |
|                    |  |          | $< 20$     | $< 35$     | $< 50$     | $< 100$     |
| Sulphides          | mg dm <sup>-3</sup>                                    | $< 2$    | $\geq 2$   | $\geq 4$   | $\geq 7$   | $\geq 10$   |
|                    |  |          | $< 4$      | $< 7$      | $< 10$     | $< 20$      |
| Phenolic Compounds | mg (C <sub>6</sub> H <sub>5</sub> OH) dm <sup>-3</sup> | $< 5$    | $\geq 5$   | $\geq 10$  | $\geq 15$  | $\geq 20$   |
|                    |  |          | $< 10$     | $< 15$     | $< 20$     | $< 40$      |

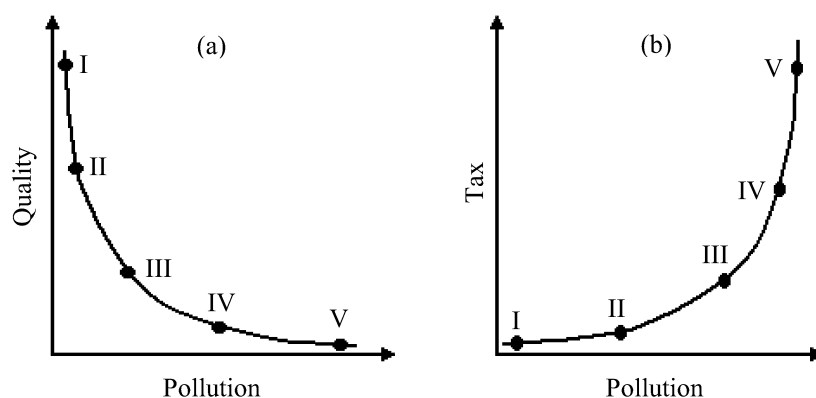


Figure 4 - The Industrial Effluent Quality Classes vs. Pollution (a). The Tax applied vs. Pollution (b)

The original dataset presented biased distributions: in 37.8% of the observations the industrial effluent is Class II, 31.8% is Class III, 14.1% is Class V, 7.4% is Class IV, 5.9% is Class FD and 3.0% is Class I.

Before attempting the DM modelling, the data was pre-processed. The original data set contained attributes with missing values. In particular, the chemical parameters presented some blank values. Since it was not possible to obtain the correct values, the blank registers were discarded, remaining a total of 135 examples. In addition, in order to enhance the Decision Tree learning process, the industrial effluent quality classes were reduced from the domain {I, II, III, IV, V, FD} to the domain {S, US, A} (Pyle 1999). The group S includes the classes I and II and denotes a Satisfactory industrial discharge, the group US includes the classes III and IV and denotes an UnSatisfactory industrial discharge and, finally, the group A, that includes the classes V and FD, and denotes an Avoidable industrial discharge.

## Decision Trees

The prediction of effluent quality, according to the criteria followed by RARISA (RARISA 2007), was defined as a classification problem. The Decision Tree (DT) is one of the most efficient and popular DM classification algorithms. It adopts a branching structure of nodes and leaves, where the knowledge is hierarchically organized. Each node tests the value of a feature, while each leaf it is assigned to a class label.

In this study, the C5.0 Algorithm (Quinlan 1996) was used to induce the DTs, under the SPSS Clementine

System. In fact, the use of DTs enables the automatic extraction of production rules that can be incorporated in a Decision Support System, using, for example, either the Structured Query Language (SQL) commands or the Predictive Model Markup Language (PMML), which is an XML-based language, developed by the Data Mining Group (DMG), which provides a way for applications to define statistical and data mining models and to share models between PMML compliant applications. (The Data Mining Group (DMG) is an independent, vendor led group which develops data mining standards, such as the Predictive Model Markup Language (PMML). DMG's XML related standards can be found at Source Forge).

## RESULTS

### Framework

Attending to the fact that the effluent class is only known after the conclusion of the whole process, it was decided to develop models to forecast the effluent type (i.e. S, US, A) based on the characteristics of the refining process and on the characteristics of the input crudes, being the experimental work done with two different strategies. Under the former approach (Model 1), it was used the predominant crude in mixture, its characteristics, the percentage of the other crudes and the characteristics of the refining process, while in the second one (Model 2), a DT was built using the percentage of each crude in the crude oil mixtures and the characteristics of the refining process (Figure 5).

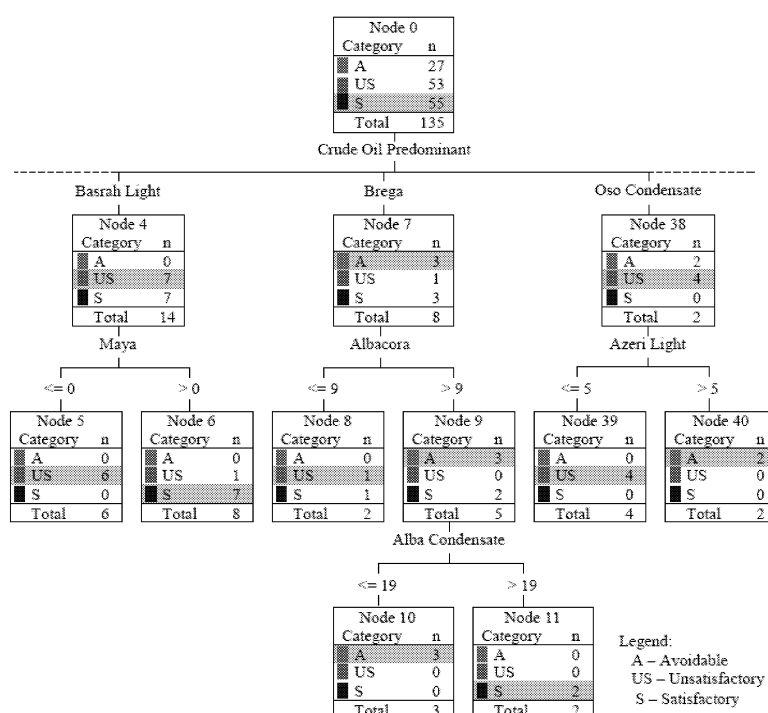


Figure 5: An extract of the Decision Tree for Model 1



These two approaches will be compared under the same framework, in order to verify which one of them is the more efficient to forecast the industrial effluent type.

## Tests

The classification models for industrial effluent quality were developed using the C5.0 algorithm. To ensure statistical significance of the attained results, 10 (ten) runs were applied in all tests, being the accuracy estimates achieved using the Holdout method (Souza et al., 2002). In each simulation, the available data is randomly divided into two mutually exclusive partitions: the training set, with 2/3 of the available data and used during the modelling phase, and the test set, with the remaining 1/3 examples, being used after training, in order to compute the accuracy values.

A common tool for classification analysis is the coincidence matrix (Kohavi and Provost, 1998), a matrix of size  $L \times L$ , where  $L$  denotes the number of possible classes. This matrix is created by matching the predicted (test result) and actual (industrial effluent quality real condition) values.  $L$  was set to 3 (three) in the present case.

In preliminary experiments, several feature selections were tested for Model 2. Figure 6 displays the specification of Model 1, according to the PMML v.2.1 standard, while Table 3 shows examples of the decision rules obtained for each strategy. It should be noted that the algorithm for induction of DTs did not use the data related to the refining process, despite being available, having chosen the predominant crude in mixture and the composition (in percentage) of each crude in the crude oil mixtures.

```
<?xml version="1.0" encoding="UTF-8" ?>
<PMML xmlns="http://www.dmg.org/PMML-3_1"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="3.1">
  <Header copyright="Copyright (c) Integral Solutions Ltd., 1994 - 2007. All rights reserved.">
    <Application name="Clementine" version="11.1" />
    <Annotation>
      <Extension extender="spss" name="Annotation" value="Exported with PMML extensions for use with SPSS SmartScore" />
    </Annotation>
  </Header>
  <DataDictionary numberOffFields="15">
    ...
    <SimpleRule confidence="1.0" id="23" nbCorrect="6.0" recordCount="6" score="S" weight="1.0">
      <CompoundPredicate booleanOperator="and">
        <SimpleSetPredicate booleanOperator="isIn" field="Crude Oil Predominant">
          <Array n="1" type="string">"Marlim"</Array>
        </SimpleSetPredicate>
        <SimplePredicate field="Albacora" operator="lessOrEqual" value="0" />
        <SimplePredicate field="Brass River" operator="lessOrEqual" value="3" />
        <SimplePredicate field="Marlim" operator="greaterThan" value="42" />
        <SimplePredicate field="Alba Condensate" operator="lessOrEqual" value="27" />
        <SimplePredicate field="Mellitah" operator="lessOrEqual" value="4" />
        <SimplePredicate field="Yoho" operator="lessOrEqual" value="0" />
      </CompoundPredicate>
    </SimpleRule>
    ...
  </RuleSetModel>
</PMML>
```

Figure 6: An extract of PMML specification for Model 1

Table 3: A snapshot of the Decision Rules for each Model

| Model 1  | Model 2   |
|--|---|
| ...<br>if Predominant = Marlim<br>and Yoho <= 0<br>and Brass River <= 3<br>and Mellitah <= 4<br>and Albacora <= 0<br>and A.Condensate <= 27<br>and Marlim > 42<br>Then → S<br><br>n = 6<br>confidence = 1,0<br>... | ...<br>if Iranian Heavy <= 0<br>and Slops <= 6<br>and Sirtica <= 6<br>and Forties <= 0<br>and Basrah Light > 10<br>and Brass River <= 0<br>Then → S<br><br>n = 15<br>confidence= 0,867<br>... |

## Discussion

Table 4 presents the coincidence matrixes for each approach. These values denote the average of the 10 (ten) runs. The results reveal that the second model is more accurate in predicting unsatisfactory industrial discharge cases (i.e. US group). Model 2 presents a value of approximately 78% for the training set and 76% for the test set, while Model 1 presents approximately 77% for the training set and 75% for the test set. Although both models provide relatively good results, Model 2, which does not focus on the predominant crude oil in the mixture, produces better results.

Table 4: The Coincidence Matrix for each Model

| Group | Model 1      |    |    |          |    |   |
|-------|--------------|----|----|----------|----|---|
|       | Training Set |    |    | Test Set |    |   |
|       | S            | US | A  | S        | US | A |
| S     | 29           | 11 | 1  | 9        | 3  | 2 |
| US    | 4            | 33 | 1  | 2        | 12 | 1 |
| A     | 3            | 3  | 14 | 0        | 1  | 6 |

| Group | Model 2      |    |    |          |    |    |
|-------|--------------|----|----|----------|----|----|
|       | Training Set |    |    | Test Set |    |    |
|       | S            | US | A  | S        | US | A  |
| S     | 29           | 7  | 2  | 12       | 2  | 3  |
| US    | 3            | 30 | 5  | 4        | 10 | 1  |
| A     | 3            | 0  | 11 | 1        | 0  | 12 |

Although in this comparison all 3 (three) classes were considered, the test sets and the training sets only contain a very limited number of avoidable (A) examples in the two models. Therefore, and being aware that the inner characteristics of the tools used in this work may bring some noise into the classification process, it still makes sense to perform an analysis considering only S and US groups (which is quite good), taking into account that the situations S and US correspond to the most favourable situations, namely Satisfactory industrial discharge and

UnSatisfactory industrial discharge. Clearly, we want to search only the most promising search paths into the population of possible solutions. However, Table 4 shows that both models could misclassify UnSatisfactory situations (US) as Satisfactory ones (S), though in small numbers. In these cases if the result from the systems were followed, then some effluents might be disposed wrongly, which could cause major problems. Indeed, all false negatives will need to be removed before such systems may be trusted.

In order to fulfil this goal (to be understood here as the work that will stem from the present one), and with respect to the computational paradigm it will be considered extended logic programs with two kinds of negation, classical negation,  $\neg$ , and default negation, *not*. Intuitively, *not*  $p$  is *true* whenever there is no reason to believe  $p$ , whereas  $\neg p$  requires a proof of the negated literal. An extended logic program (program, for short) is a finite collection of rules and integrity constraints, standing for all their ground instances, and is given in the form:

$p \leftarrow p_1 \wedge \dots \wedge p_n \wedge \text{not } q_1 \wedge \dots \wedge \text{not } q_m$ ; and  
 $? p_1 \wedge \dots \wedge p_n \wedge \text{not } q_1 \wedge \dots \wedge \text{not } q_m, (n, m \geq 0)$

where  $\wedge$  stands for the logical operator *and*,  $?$  is a domain atom denoting *falsity*, the  $p_i$ ,  $q_j$ , and  $p$  are classical ground literals, i.e. either positive atoms or atoms preceded by the classical negation sign  $\neg$  (Neves 1984). Every program is associated with a set of abducibles. Abducibles may be seen as hypotheses that provide possible solutions or explanations of given queries, being given here in the form of exceptions to the extensions of the predicates that make the program. Indeed, in our approach to problem solving, we will not get a solution to a particular problem, but rather a logic representation (or program) of the universe of discourse.

On the other hand, logic programming enables an evolving program to predict in advance its possible future states and to make a preference. This computational paradigm is, therefore, particularly advantageous since it can be used to predict a program evolution employing the methodologies for problem solving that benefit from abducibles, in order to make and preserve abductive hypotheses. It is on the preservation of the abductive hypotheses that our approach will be based.

Designing such a selection regime presents, still, unique challenges. Most evolutionary computation problems are well defined, and quantitative comparisons of performance among the competing individuals are straightforward. By contrast, in selecting an abstract and general logical representation or program, performance metrics are clearly more difficult to devise. Programs must be tested on their ability to adapt to a changing environment, to make deductions and draw inferences, and to choose the most appropriate course of action from a wide range of alternatives. Above all they must learn how to do these things on their own, not by implementing specific instructions given to them by a programmer, but by continuously responding to positive and negative environmental feedback. In order to accomplish such goal, i.e. to model the universe of discourse in a changing

environment, the breeding and executable computer programs will be ordered in terms of the quality-of-information that stems out of them, when subject to a process of conceptual blending (Neves et al. 2007). In blending, the structure or extension of two or more predicates is projected to a separate blended space, which inherits a partial structure from the inputs, and has an emergent structure of its own. Meaning is not compositional in the usual sense, and blending operates to produce understandings of composite functions or predicates, the conceptual domain, i.e. a conceptual domain has a basic structure of entities and relations at a high level of generality (e.g. the conceptual domain for journey has roles for traveller, path, origin, destination). In our work we will follow the normal view of conceptual metaphor, i.e. metaphor will carry structure from one conceptual domain (the source) to another (the target) directly.

## CONCLUSION AND FUTURE WORK

The use of DM techniques can solve complex problems in environmental applications, such as the real-time diagnosis of effluent quality in industrial processes. The use of these techniques allows the reduction of costs in chemical analysis and real-time intervention, which will preserve the environment and reduce the costs of the fines issued by Sistema de Santo Andre, through the prediction of the effluent quality from the crude oils chosen to be refined, taking into account their origins and their characteristics, which may influence the quality of the effluents that has to be treated.

In this work, two classification models were tested, using DTs. The results of the models were quite satisfactory and quite similar, regarding their outcome: S (i.e. Satisfactory industrial discharge) on Model 1 and US (i.e. UnSatisfactory industrial discharge) on Model 2 are the most favourable situations, which allow us to conclude that the application of one of the two models will be useful in the Oil Industry, not only minimizing fines issued to the company, but also protecting it from the inherent devastating environmental consequences.

Therefore, and according to the objectives of Petrogal, not only in establishing a compromise of integrating Health, Safety and Environment on the strategy and activities of the company, but also close by its continuous improvement on its performance, giving a decisive contribute to achieve a sustainable development and excellence in business, it is necessary to act on a pro-active way. Consequently, the use of predicting methods that have inherent processes of decision making, and that come from the scientific area of AI, may contribute to solve or minimize the problems referred to above. The proposed approach for further work opens room for the development of automatic tools for environmental decision support, which are expected to enhance the ecological response, being an abstract definition of such systems given above.

## ACKNOWLEDGEMENTS

The authors would like to thank The University of Minho for providing the software facilities for implementing the present solution, and to Petrogal, namely to Mr. Fernandes and Mr. Letras, in making available the data collected, which was fundamental to conduct this work.

## REFERENCES

- Chau, K.-W. 2006. "A review on integration of artificial intelligence into water quality modelling". *Marine Pollution Bulletin* 52, No.7, 726-733.
- Dzeroski, S. 2002. "Environmental Sciences". In *Handbook of Data Mining and Knowledge Discovery 2002*, W. Klösgen and J. Zytkow (Eds.). Oxford University Press, Oxford, 817-830.
- Eaton, A., Clesceri, L., Rice, E. and Greenberg, A. 2005. *Standard Methods for the Examination of Water and Wastewater*. 21st Edition, American Public Health Association, USA.
- Fayyad, U., Piatetsky-Shapiro, G., Smith, P. and Uthurusamy, R. 1996. *Advances in Knowledge Discovery and Data Mining*. AAAI Press/MIT Press, Massachusetts, USA.
- Galp Energia. 2006. "Annual Report". Petróleos de Portugal, Petrogal, SA, Portugal. (In Portuguese)
- Gomes, J. S. and Alves, F. B. 2007. *O Universo da Indústria Petrolífera*. Fundação Calouste Gulbenkian, Lisbon, Portugal. (In Portuguese)
- Kohavi, R. and Provost, F. 1998. "Glossary of Terms". *Machine Learning* 30, No.2/3, 271-274.
- Kuo, J.-T., Hsieh, M.-H., Lung, W.-S. and She, N. 2007. "Using artificial neural network for reservoir eutrophication prediction". *Ecological Modelling* 200, 171-177.
- Neves, J. 1984. "A Logic Interpreter to Handle Time and Negation in Logic Data Bases". In *Proceedings of ACM 1984 Annual Conference*, San Francisco, October 24-27, USA.
- Neves, J., Machado, J., Analide, C., Abelha, A., and Brito, L. 2007. "The Halt Condition in Genetic Programming". In *Progress in Artificial Intelligence, EPIA 2007 (LNAI 4874)*, Guimarães, Portugal.
- Pyle, D. 1999. *Data preparation for Data Mining*. Morgan Kaufmann, San Francisco, USA.
- Quinlan, J. R. 1996. "Bagging, boosting and C4.5". In *Proceedings of AAAI'96 National Conference on Artificial Intelligence*, AAAI Press, Menlo Park, USA, 725-730.
- RARISA 2007. "Regulamento de Recolha e Tratamento de Água Residual Industrial do Sistema de Santo André". Águas de Santo André - Grupo Águas de Portugal. (In Portuguese)
- Santos, M. F., Cortez, P., Quintela, H., Neves, J., Vicente, H. and Arteiro, J. 2005. "Ecological Mining - A Case Study on Dam Water Quality". In *Data Mining VI - Data Mining, Text Mining and their Business Applications 2005*, A. Zanasi, C.A. Brebbia and N.F. Ebecken (Eds.), WIT Press, Southampton, UK, 523-531.
- Souza, J., Matwin, S. and Japkowicz, N. 2002. "Evaluating Data Mining Models: A Pattern Language". In *Proceedings of 9th Conference on Pattern Language of Programs*, Illinois, USA.
- Thuraisingham, B. 1999. *Data Mining Technologies, Techniques, Tools and Trend*. CRC Press LLC, USA.
- Turban, E., Aronson, J. and Liang, T.-P. 2004. *Decision Support Systems and Intelligent Systems*. 7th Edition, Prentice Hall, New Jersey, USA.

## AUTHOR BIOGRAPHY

**ANA FERNANDES** was born in Évora, Portugal. At the University of Évora she studied Physics and Chemistry, obtained her degrees in 2005 and her master's degree in Environmental Science in 2007. She is now manager of the Quality Management System of the Municipal Water Laboratory of Santiago do Cacém, Portugal, and teaches at the Department of Environment of the Piaget Institute in Santo André, Portugal. Her current research interests include Water Quality Control, Environment, Health and Safety Labor, Quality, Data Mining and Knowledge Discovery from Databases and Intelligent Information Systems.

E-mail: [anavilafernandes@gmail.com](mailto:anavilafernandes@gmail.com)

**HENRIQUE VICENTE** was born in S. Martinho do Porto, Portugal and went to the University of Lisbon, where he studied Chemistry and obtained his degrees in 1988. He joined the University of Évora in 1989 and received his PhD in Chemistry in 2005. He is now Auxiliary Professor at the Department of Chemistry at the University of Évora. He is a researcher at the Chemistry Center of Évora and his current interests include Water Quality Control, Lakes and Reservoirs Management, Data Mining and Knowledge Discovery from Databases and Intelligent Information Systems.

E-mail: [hvicente@uevora.pt](mailto:hvicente@uevora.pt)

**JOSÉ NEVES** was born in Vila do Conde, Oporto, Portugal. He went to the Heriot Watt University, in Edinburgh, Scotland, where he obtained his degrees in MSc (1980) and PhD (1983). He is now Full Professor of Computer Science at the Department of Informatics at the University of Minho. His current research activities span the fields of Non-monotonic Logics, Knowledge Representation and Reasoning Systems and Evolutionary Intelligence.

E-mail: [jneves@di.uminho.pt](mailto:jneves@di.uminho.pt)

# MULTI PROJECT SCHEDULING IN THE CHEMICAL INDUSTRY USING A GENETIC ALGORITHM

Sven Tackenberg, Sebastian Schneider, Christopher M. Schlick

*Chair and Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Bergdriesch 27, Aachen, Germany  
s.tackenberg@iaw.rwth-aachen.de*

**Keywords:** Project Management, Resource Constraint Multi Project Scheduling Problem (RCMPSP), Genetic Algorithm (GA), C3 Model, Task Scheduling.

**Abstract:** Due to impatient customers and competitive threats, it has become increasingly important for the chemical industry to shorten the lead time of development projects. Furthermore, these organizations are faced with the challenge of planning and managing the simultaneous execution of multiple dependent projects under tight time and resource constraints. Within that kind of business environment, effective project management and task scheduling is crucial to organizational performance. A genetic algorithm approach with a novel genotype and GP mapping operation is presented to minimize the overall project duration and budget of multiple projects for a resource constrained multi project scheduling problem (RCMPSP) without violating inter-project resource constraints or intra-project precedence constraints. The C3 representation was used to model the processes due to its compact, generic and easily quantifiable nature. The implemented system is capable of calculating a number of project performance metrics that are focused towards determining the effect of stochastic rework of tasks, variable assignment of actors and the stochastic makespan for a specific task. The proposed Genetic Algorithm is tested on scheduling problems with and without stochastic feedback. This GA demonstrates to provide a quick convergence to a global optimal solution regarding the multi-criteria objectives.

## INTRODUCTION

Product and production system design is recognized as a crucial challenge by the companies of the chemical industry. Therefore the ability to quickly introduce new products into the market is a key factor for determining sustainable profitability (Clark and Fujimoto, 1991). As a result, novel methods for identifying, analyzing and optimizing the main performance shaping factors of development projects

as well as their interaction regarding complexity and coherence are necessary (Schlick et al., 2007).

Our vision is a novel approach to reduce the risk of multi-project management by using optimization methods for multi-project planning to support project managers' decision making. This concept should enable project managers to model, simulate and optimize a work organization regarding their multi objective target system (cost, lead time, utilization etc.) at each point of time during a development project. However, as a consequence of the inherent complexity of development projects (Hölldt-Otto and Magee, 2006), scientific methods for a multi-criteria optimization of valid development project models are missing. Based on results of the latest research, first results of a research project are presented to close the identified gap between work process modeling and optimization methods in order to continuously improve the performance of an organization's project portfolio.

The rest of the paper is organized as follows. In the next section, we review the C3 methodology to describe various elements of a project organization and the sequencing complexity of a C3 model. Section 3 provides a short summary of task scheduling and selected approaches for the systematical improvement of project organizations by optimization algorithms. The section 4 presents the developed genetic algorithm. It comprises the GA structure as well as the chromosome representation and initialization, the developed genetic operators and especially the transformation of a chromosome representation into a specific project organization model. To investigate the performance of the GA we discuss the results of computational experiments for a project portfolio of an enterprise of the chemical industry. The paper concludes with a brief summary of the work completed, a critical acclaim and possible extension in future.

## PROJECT ORGANIZATION

Designing and managing a project organization

demand a wide range of decision making and the extensive execution of activities that are necessary to ensure the achievement of project specific objectives.

## Project Organization Modeling

A semi-formal graphical representation for a project organization model is an intuitive and commonly used method of showing the interdependencies between tasks and responsibilities of actors within a project. The C3 modeling method for weakly structured and domain-spanning development processes was developed by our institute based on the Unified Modeling Language (UML) and the identified requirements of project managers (Killich et al., 1999). It was conceptualized especially for the recording and presentation of weakly structured cooperative work processes and has been successfully used for several years in research and industrial projects. The term “C3” is derived from the initial letters in the words cooperation, coordination and communication – the core elements of distributed work processes. Due to its semi-formal semantics, C3 is particularly suited for the analysis of creative processes that are also associated with uncertainties and iterations.

Figure 1 shows a sample C3 model where specific actors and resources to execute the tasks are undetermined. To complete a task, an activity has to be executed by at least one actor. Therefore it must be considered that one specific task can lead to different activities. The maximum allowable overlapping of activities is given by a value, assigned to the control flow. There is a feedback loop from task 7 to task 5. That is, the C3 model shows a 40% probability that task 5 will be repeated after task 7. The new effort for the achievement of task 5 is 30% of the original estimated effort. After the rework of activity 5 the probability for another feedback of task 7 to task 5 is reduced by 90%. The elements for the characterization of a task, such as requirements of actors, resources and information, estimated effort, are essential sub-elements of a C3 model. These elements are arranged in “swimlanes” according to their organizational affiliation. The tasks are supplemented by special elements that support the description of creative tasks as well as tasks with a certain degree of uncertainty regarding predecessor relationships (Kausch et al. 2007).

The semi-formal, independent C3 models of a multi project environment show the contextual relationships between tasks, yet only indicate an exact temporal sequence where it is truly necessary. All other temporal relationships, e.g., the assignment

of actors and resources (facilities and tools) to tasks, are not committed in the C3 model due to the fact that the specification is based on the current availability and the achieved intermediate results during the (simulated) course of a project.

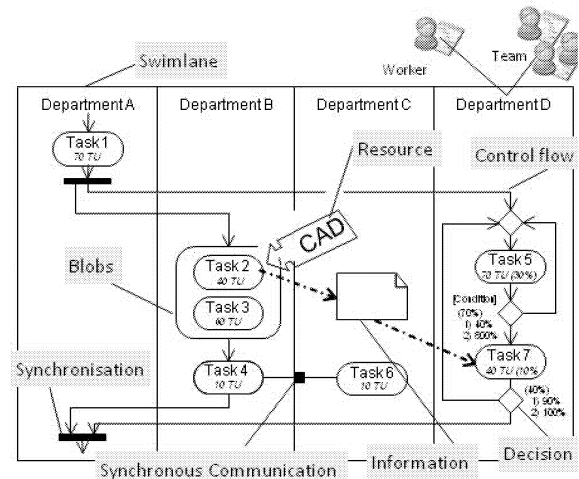


Figure 1 C3 model of a development project

Therefore the objective of a project manager is the prioritization of the precedence-constrained tasks of a project and the assignment of actors and resources to optimize a multi-criteria objective function such as minimizing project duration or project costs.

## Problem Complexity

The scheduling problem for a multi project environment of a company –  $m$  concurrent projects  $P_1...P_m$ , with a set of tasks  $TA_n = \{ta(1)...ta(n)\}$ , where  $n$  specifies the total number of tasks in project  $P_n$  – is known as the resource-constrained multi project scheduling problem (RCMPSP). The scheduling problem can be very complex as the number of projects, tasks, actors and resources increases. It was shown by Lenstra and Rinnooy (1978) that the scheduling of tasks under consideration of precedence and resource constraints is NP-hard (Garey and Johnson, 1979). Due to the fact that a C3 model can be transformed into a Design Structure Matrix – DSM (Browning and Eppinger, 2002) the sequencing and assigning process can be formulated as a *Quadratic Assignment Problem* (QAP). The QAP is well known as a NP-hard combinatorial optimization problem (Zhuang and Yassine, 2004). Therefore at present no algorithm can be found to solve real-world C3 models of arbitrary sizes for a multi project environment to optimality with an adequate performance. A benchmark study done by Hartmann (1998) demonstrated that a project with as few as 60 activities has not been solved to the global

optimum by computational experiments. Furthermore additional projects can extremely enlarge the number of feasible schedules. When considering the set of feasible project schedules for a RCMPSP  $\theta = \theta_T \cap \theta_R$ , where  $\theta_T$  denotes the set of precedent-feasible schedules and  $\theta_R$  denotes the set of resource-feasible schedules, there exist many possible  $\theta$  and many potential objectives for choosing between them. Therefore it is important to note, that the global optimum of  $\theta$  is not compulsively based on the local optima of  $\theta_T$  and  $\theta_R$ .

## LITERATURE REVIEW

Project scheduling is of great practical significance, and generalized models can be applied in product development, production planning as well as a variety of scheduling applications. Early attempts at project scheduling were focused on reducing the total project lead time (makespan) assuming unlimited resources. Well known techniques include the Critical Path Method (CPM) (Kelley, 1961) and the Project Evaluation and Review Technique, PERT (Malcolm, 1959). Scheduling problems have been extensively studied for many years by attempting to establish precise solutions using methods from the field of operations research (Kolisch and Padman, 2001).

It was shown that the general scheduling problem concerning precedence and resource constraints is NP-hard (Lenstra and Rinnooy, 1978). Therefore, exact optimization methods are too time consuming and ineffective for solving large organizational problems found in real enterprises. Yang et al. (2001) and Kolisch and Padman (2001) surveyed the most common methods that were developed for the resource-constrained project scheduling problem (RCPSP), such as dynamic programming, zero-one programming and implicit enumeration with branch and bound.

Surveys of heuristic and metaheuristic approaches which solve intractable scheduling problems quickly, efficiently and fairly satisfactorily can be found in Grünert and Irnich (2005) and Kolisch and Hartmann (1998). Nonobe and Ibaraki (2001) developed a technique to solve the RCPSP based on local search. Alternative approaches based on tabu search and genetic algorithms were presented by Shouman et al. (2006). The RCMPSP as a specific characteristic of the RCPSP deals with the scheduling of more than one project. Goncalves et al. (2008) solved a multi-project problem consisting of up to 50 single projects and 120 activities with a genetic algorithm (GA). The generation of solutions of the RCMPSP

based on GAs is also analyzed by Yassine et al. (2007a, 2007b). Kolisch (2000), however, proposed a list scheduling algorithm and KHosraviani et al. (2004) developed an optimization approach based on a Genetic tree. Furthermore the work of Linyi et al. (2007) has shown that metaheuristics are a promising approach to project scheduling problems due to fast convergence and easy implementation.

## DESIGN AND IMPLEMENTATION

This section introduces a novel GA to solve the RCMPSP given C3 models as input. The probabilities of an iterative execution of tasks during the project as well as stochastic values for the duration of an activity are integrated in problem encoding.

### GA Structure

As shown in Figure 2, the GA starts by creating an initial population of chromosomes through several runs of a simulation model (Tackenberg et al. 2008) or an algorithm. Each *chromosome* consists of a collection of *genes*. Genes are placed at different locations or *loci* of the chromosome and have values which are called *alleles*. The characteristic of each gene of one chromosome is thereby represented by an allele. The combination of genes (defined by loci and alleles) refers to the specific genetic makeup of an individual, termed as *genotype*. While the genotype corresponds to the structure of a GA, the term *phenotype* represents the decoded structure for the RCMPSP – a specific project organization model which can be regarded as one point in the search space. Classic genetic operations and functions (Goldberg, 1989) as fitness function, selection, crossover and mutation were adapted to C3 models to find optimal task sequences and assignments of actors or tools for the predefined objectives.

The *fitness function* is used to evaluate a chromosome how good the underlying project organization fulfils the multi criteria target system of a project manager. Next, the function *selection* chooses chromosomes that will be passed on to the next generation. To map the random process a *crossover* function is used to produce a new offspring chromosome from minimum two parent chromosomes according to a user-defined probability  $p_c$ . If an offspring takes the best parts from each of its parents, the result will likely be a better solution (Yassine et al., 2007a). Modified as well as unmodified chromosomes can be further mutated according to a user defined probability  $p_m$ . The *mutation* leads to a variability of the alleles regarding

the characteristic of the project organization. A new generation of chromosomes replaces the previous one, and the fitness of the new generation is evaluated. The cycle of functions is repeated until a

termination condition is met, –number of generations or fitness convergence in the population.

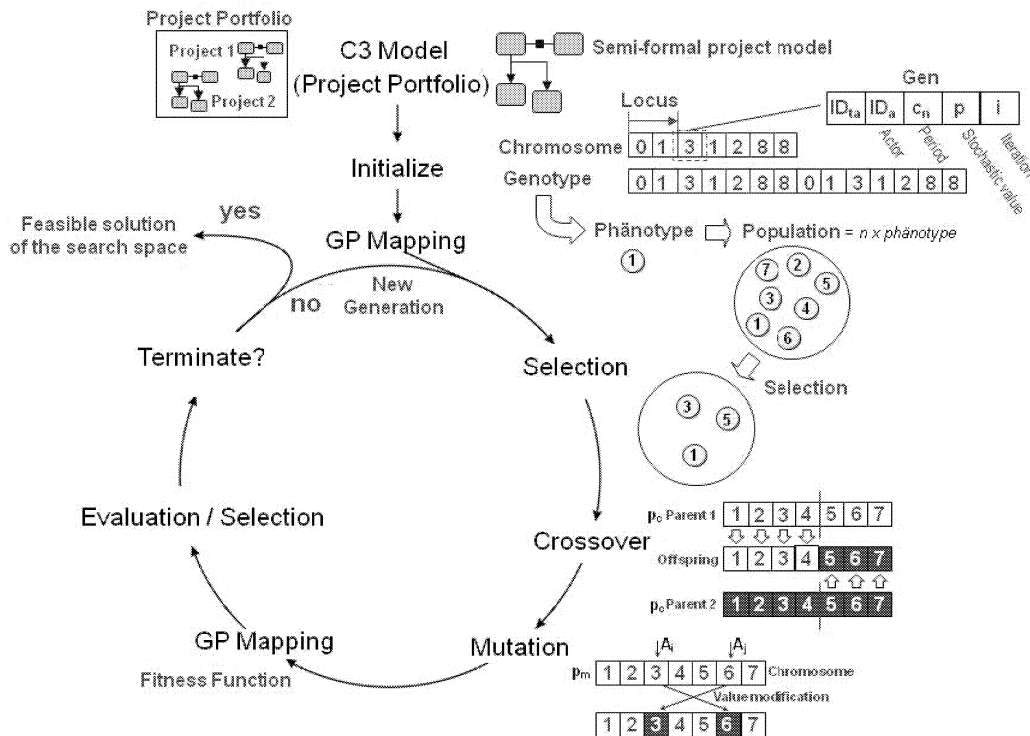


Figure 2 GA Structure for RCMPSP

The novelty of the presented approach is the characteristic of a chromosome, the evaluation of the fitness as well as the transformation of the modified chromosomes into a detailed, feasible project organization model. Therefore we will focus on these three aspects.

## Data Structure

The sequence of tasks as well as the assignment of actors (workers, teams) and resources (tools, machines, facilities) must be represented in a chromosome, which describes the project organization model. Various representation models for encoding a project as chromosomes in GAs exist, but the most common is the natural encoding by integer numbers. Yassine et al. (2007a) and Zhuang et al. (2004) introduce an encoding where a specific element or sub-element of a project is assigned exactly once to a locus in the permutation. Therefore each activity of a project is represented once in the chromosome. From our point of view this encoding

does not permit an extensive and efficient permutation of a realistic complex project organization. Especially uncertainties regarding the makespan of tasks, execution of iterations and restrictions regarding the assignment of actors cannot be easily integrated in the encoding of a project organization.

As shown in Figure 2 our approach introduces a novel representation of a project organization as a chromosome to fulfill the requirements of an adequate project representation. Each task of a project or multi project portfolio is given an unique identification number (ID) and each gene represents a task. A gene therefore contains information about the *Task ID*, *ID of Actors* and *Resources*, *Period* relating to the ending time of the predecessor tasks, *Stochastic factor* for the makespan of an activity and *Occurrence of iteration*. As shown in Figure 2 the structure of the gene is mandatory. A representation technique is used where the location of each gene in a chromosome is fixed and cannot be modified by

genetic operators. The information regarding the task and execution of activities is linked via pair representation  $\langle locus, allele \rangle$  – the locus of a gene is determined by the value of the corresponding task ID. Due to the composition of a multi project organization based on genes the chromosome length is set to the total task number of all considered projects. Chromosomes and genes are linked to a central database which contains the static values of the project portfolio specific actors, resources and iterations. Especially the information about the occurrence and the effect of iterations in development projects is complex. Therefore we use adapted DSM according to Gärtner et al. (2008):

- The *probability matrix* shows the probability of an iteration occurring.
- The *rework matrix* defines the work (makespan) performed during an iteration. Due to a worker's or organization's learning effect an exponential reduction of makespan according to the number of processed iterations can be considered.
- The *changes of probabilities matrix* describes the increase or decrease of a probability for an iteratively execution of tasks according to the number of processed iteration.

## Objective Function

Every optimization method must be able to assign a measure of quality to generated results in the search space to distinguish good and bad results (Yassine et al., 2007b). For this purpose a fitness function is used for GAs to assign each individual chromosome a fitness value.

To break down the RCMPSP it can be decomposed into a genotype-phenotype mapping  $f_{gp}$  and a phenotype-fitness mapping  $f_{pf}$  (Liepins and Vose, 1990; Yassine et al., 2007b). Therefore a genotypic search space  $\Phi_g$  as well as a phenotypic search space  $\Phi_p$  exist which can be either discrete or continuous. The genotypic search space  $\Phi_g$  covers all permutations of chromosomes and genes. According to the characteristic of the crossover and mutation operators  $\Phi_g$  includes valid as well as invalid solutions in  $\Phi_g$ . To transform only valid genotypes into a phenotype a function  $f_g$  is used, which assigns each element in  $\Phi_g$  a value as follows:  $f(x): \Phi_g \rightarrow \{0,1\}$ . According to the introduced decomposition, the genotype-phenotype mapping occurs first, where feasible genotype elements (value = 1) are mapped to elements in the phenotypic search space  $\Phi_g: f_g(x_g) : \Phi_g \rightarrow \Phi_p$  (see Sec. GP Mapping). The result of the mapping is a feasible representation of a detailed project organization (project plan). Subsequently, the

fitness of a phenotype in  $\Phi_p$  is calculated by:  $f_p(x_p): \Phi_p \rightarrow \mathbb{R}$ . Thus, the fitness of an element is a result of both mappings:  $f = f_p \circ f_g = f_p(f_g(x_g))$  (see also Yassine et al., 2007b).

For the definition of  $f_p$ , essential objectives for the scheduling problem must be considered. Based on the comprehensive survey by Kolisch and Padman (2001) we identified critical success factors for chemical engineering projects including traditional ones such as project duration and cost minimization but also more recent ones like the qualification of actors (Kausch et al., 2007). We use total project duration  $T$ , project cost  $C$  as the RCMPSP multi criteria performance measure to be minimized and degree of capacity utilization of actors  $U$  to be maximized. So we use the following fitness function:

$$\text{Minimize } T = \{\sum d_i + p_i \mid i = 1, \dots, n\}^* \quad \text{Notation (1)} \\ \text{ignore the overlapping of activities}$$

$$\text{Minimize } C = \{\sum c_i \mid i = 1, 2, \dots, n\}$$

$$\text{Maximize } U = \{\sum u_i \mid i = 1, 2, \dots, n\}$$

where

$d_i$  is the duration of task  $i$

$p_i$  is the maximum time period between task  $i$  and its predecessors

$c_i$  is the cost of task / activity  $i$

$u_i$  is the utilization of actor  $i$

The stochastic elements of a project model (makespan, occurrence of an iteration e.g.) have an immediate effect on the fitness of an individual. An exclusive evaluation based on the fitness value without considering the impact of stochastic elements may cause optimal scheduling solutions with a low probability of occurrence in reality. Therefore we subclassify the individuals of  $\Phi_p$  according to the occurrence of iterations. To ensure the consideration of project progressions with and without iterations, the evaluation/selection operator chose a minimum number of individuals of each iteration loop type. The quantity is determined according to the appearance probability of this specific project progression type.

The uncertainties regarding the makespan of activities to solve project specific tasks are also considered by the developed fitness function. Kausch et al. (2007) calculate the *grade of parallelism*  $Pa$  for a specific project to describe the quality of a scheduling solution:



$$Pa = \{\sum d_i / T \mid i = 1, \dots, n\} \quad (2)$$

In addition to the *grade of parallelism* a value for the probability is calculated with which a generated course of a project can be realized – based on the makespan probability distribution of all executed activities and iterations. These factors are considered by the determination of the fitness value.

## Constraints

The solution of the RCMPSP is subjected to the predecessor relationships between tasks, described in the C3 model. Due to the **precedence constraints** of each project, each task needs to be checked if its immediate predecessors have been sufficiently executed before being performed. Thereby a complete fulfillment of a predecessor task is not mandatory. To integrate aspects of Simultaneous Engineering (SE) an overlapping of coupled activities has been considered in the GA. Therefore the precedence relationships are described by the value “*minimum percentage of completion*”  $e_i$  in the C3 model. The precedence constraint can be formulated as:

$$\min t_i + (d_i \times e_i) \leq t_{i+1}, 0 < e_i < 1 \mid i = 1, \dots, n \quad (3)$$

where  $e_i$  represents the minimum percentage amount of work for task  $i$  to fulfill the requirements for an execution of task  $i+1$ .

Although projects and tasks may be unrelated by precedence constraints, they depend on a common pool of actors and resources. Due to the **resource constraints** two actors or resource conflicting tasks  $t_{ai}$ ,  $t_{ai+1}$  cannot be executed at the same time  $t$ :

$$t_i + d_i \leq t_{i+1} \text{ or } t_{i+1} + d_{i+1} \leq t_i, \mid i = 1, \dots, n \quad (4)$$

$$\text{if } a_i(t_{ai}) = a_i(t_{ai+1}) \text{ or } r_i(t_{ai}) = r_i(t_{ai+1}),$$

$t_{ai}$  is task  $i$

$a_i$  is the actor  $i$  with a specific characteristic

$r_i$  is the resource  $i$  with specific functions

$a_i(t_{ai})$  is actor  $i$  assigned to task  $i$ .

$r_i(t_{ai})$  is resource  $i$  assigned to task  $i$ .

Based on the task specific requirements at least one actor (worker, team) must be assigned to a task. We assume that an actor and a resource must be devoted to an activity until it is completed. An abort of an activity to start another activity is not allowed. In contrast to the approaches of Zhuang et al. (2004),

KHosraviani (2005) and Yassine et al. (2007a, 2007b) the processing time  $d_i$  and the actors  $a_i$  and resources  $r_i$  required for any task  $t_{ai}$ :  $a_i(t_{ai})$ ;  $r_i(t_{ai})$  are not fixed.

## Initialization

Due to stochastic elements of the project model and several concurrent projects there is only a small feasible search space in relation to  $\Phi$ . Therefore a random generation of alleles could result in the generation of a large number of infeasible solutions (Zhuang et al., 2004). So a simulation model was developed and is used to generate feasible genotypes for the initial population (Tackenberg et al., 2008).

Although an automatic mapping of a C3 model to a simulation model has been prototypically developed, the effort to generate an initial population is very time consuming. Therefore a permutation algorithm is used to generate an initial population of precedence feasible individuals. This algorithm proceeds as follows:

**Step 1:** A task from one of the considered projects is randomly chosen. The task is mapped to a gene:

- The gene is placed; task ID of the C3 model represents the locus.
- Values for the start time of an activity, the duration and the occurrence of an iteration are calculated based on the database entries and the corresponding probability distributions.
- Randomly an actor who fulfills the requirements (qualification, competence) is chosen, and it is checked if the actor is already selected for the given period. If the actor was selected before, continue this random selection until an adequate actor is found. The assignment of resources is analogue.

**Step 2:** Repeat step 1 until the set of unselected tasks of all considered projects is empty, which generates a chromosome that consists of all tasks.

**Step 3:** Repeat 1 and 2 until all chromosomes of a population size are generated.

The generated genes of the population are saved in a LinkedHashMap as objects of the class *Chromosome*. The chromosomes are objects of the class *Population* and are stored in a second LinkedHashMap. We use LinkedHashMaps – implemented in Java – due to the advantage of a dynamic generation and storage of objects.

The population size is determined by the project manager in consideration of the problem complexity.

To have an indication of an adequate population size Thierens defined an equation necessary for a successful GA (1995). The equation was used by us to get a first impression of the population size.

## GP Mapping

A chromosome is a blueprint for a project organization. The genotype-phenotype mapping  $f_g$  – GP-mapping – is used to generate a feasible, detailed project organization for each individual of a population. It results in a description of a project organization which covers information about the characteristics and interactions of tasks, the assignment of actors and resources to tasks as well as specific characteristics of a detailed project scenario, i.e., concrete information regarding starting time, makespans of activities, responsibilities etc. Attention should be paid to the fact that a specific definition of a task sequence and responsibilities of actors does not generate only one course of a project. Due to uncertainties regarding the required effort to solve a task and the exact starting and ending time of an activity there exist several courses of a project.

The GP Mapping operator acts to generate a project organization based on the information of the varied or unvaried chromosomes and the data base entries. The information in the data base is only generated once during the transformation of a C3 model to a chromosome. The steps for the GP mapping process are as follows:

**Step 1:** A task is randomly chosen and it is checked if its immediate predecessors have been sufficiently executed before set in the project plan. If not, step 1 is continued until a task is found.

**Step 2:** The makespan for the execution of a chosen task  $i$  is calculated, based on:

- Basic effort  $d_{e(i)}$  of task  $i$  estimated by project managers without consideration of actor's qualification, (constant value for each specific task, stored in database *Activity*)
- Qualification of assigned actor (constant value  $q$ , stored in database *Actor*)
- Random value  $h$  (variable value, stored in *gene of the chromosome*), based on a task specific probability distribution, e.g., Gaussian, right- or left-skewed  $\beta$ -distribution (constant parameters of the distribution are stored in database *Activity*).

The makespan of task  $i$  is calculated as follows:

$$d_i = d_{e(i)} \times q \times h \quad (5)$$

and is saved in the corresponding object of the class Chromosome.

**Step 3:** Calculation of the absolute starting time of activity  $i$ , based on:

- Period related to the ending time of the predecessor activities (variable factor  $z_i$ , saved in *gene of the chromosome*)
- Starting time and duration of the predecessor(s) saved in a LinkedHashMap (temporary data base entry).

The absolute starting time of activity  $i$  is calculated as follows:

$$t_i = t_{i-1} + (z_i \times d_{i-1}) \quad (6)$$

The variance of  $z_i$  is determined by the project manager,  $z_i > 0$ . For activities with more than one predecessor:

$$t_i = \max \{ t_i + (z_i \times d_i) \mid i = 2, \dots, n-1 \} \quad (7)$$

As an example, consider a chromosome with the two tasks  $ta1$  and  $ta2$ . Task  $ta1$  is the predecessor of  $ta2$  and starts at the starting time 0 Time Units (TU) with a duration  $d1$  of 10 TU. The value for  $z2$  of task  $ta2$  (characteristic of the allele) is 0.4. Therefore 6 TU before finishing  $ta1$  is the earliest time to execute  $ta2$  (starting time: 4 TU).

**Step 4:** Assignment of an actor to a task, based on:

- Actor ID (variable value), stored in gene of the chromosome.
- Status of the actor (employed, unemployed) for the considered time period:

$$p_i = t_i + d_i \quad (8)$$

The verification if an actor fulfills the minimum requirements of a task is previously done during the mutation of a gene.

**Step 5:** Checking if the assigned actor executes another activity at any particular time of the period  $p_i$ . If the actor is unemployed the assignment leads to a change of status (*employed*) of this actor for the period  $p_i$ . If an actor is employed, the starting time of the activity  $ta_i$  is modified under consideration of predecessor conditions and the earliest starting time, until a feasible solution is generated. In such a case the equations 6 or 7 and 8 are re-executed until a valid solution is found.

**Step 6:** Task Freeze – the parameters of the

considered task/activity are saved for the current chromosome and cannot be modified again during this generation. The activity is “placed” in the project plan.

**Step 7:** Repeat Steps 1 to 7 until all tasks of a chromosome have been placed to generate a whole project organization.

It is important to note that different chromosomes may potentially have the same fitness value and essentially represent the same project organization after the GP mapping. Although uncertainties of development projects have been considered in the GA model, due to the novel genotype and GP Mapping the mapping of a specific genotype into a phenotype produces always an identical project plan. Thereby the fitness value of the generated chromosomes is a significant performance indicator.

## Selection

Two popular types of selection approaches can be distinguished in literature: fitness-proportionate selection schemes and ordinal-based selection schemes. Fitness-proportionate schemes may often fail to provide adequate Selection Pressure (SP) when fitness variance in the population is very high or very low (Yassine et al., 2007a). SP is defined as the number of expected individuals (chromosomes) in the next generation and determines the performance of a selection operator (Bäck, 1994).

Therefore an ordinal-based selection scheme – Tournament Selection (Brindle, 1981) – is employed in this algorithm because of its ability to ensure an adequate SP independent of a specific fitness structure within the population. In tournament selection, a certain number of chromosomes is randomly selected, depending on the tournament size  $s$ . The best chromosome wins the tournament with probability  $p$  and overcomes the selection phase. We favor a binary tournament selection. It picks two individuals from a population of chromosomes and selects the better one. Therefore a chromosome’s fitness rank within a population is crucial rather than the value of its fitness.

## Crossover

Several crossover operators have been developed which enable a global exploration of the search space. The results of the different crossover operators are very heterogeneous. Therefore Whitfield et al. (2003) compared several crossover strategies for DSM sequencing. Due to the feasibility of mapping a

C3 model into a DSM the results of Whitfield et al. give a conclusion about their performance regarding the considered C3 modeled RCMPSP. A crossover procedure based on a version of one point crossover is used that works as follows (Murata and Ishibuchi, 1994; Yassine et al. 2007a):

**Step 1:** Two chromosomes that passed the selection phase are chosen randomly from the population (probability  $p_c$ ). One of them is randomly designated as the “primary” parent (*Parent 1*). These two chromosomes *Parent 1* and *Parent 2* undergo crossover according to the crossover probability  $p_c$ .

**Step 2:** If these both chromosomes undergo crossover a position (locus) along both parents is chosen by random. The position of *Parent 1* corresponded with locus of *Parent 2*.

**Step 3:** Select and place the genes of the first part of *Parent 1* into the positions at the beginning of *Offspring 1*. The second part of *Parent 2* is set into the loci right of the cutting position to complete *Offspring 1*. Due to the fixed assignment  $\langle \text{locus}, \text{Task ID} \rangle$  for all chromosomes of a generation the crossover operator only generates valid solutions.

**Step 4:** The generation of the second offspring (*Offspring 2*) is taken place analogously.

Figure 2 provides a graphical example of this process. The offspring first inherits four genes from parent 1 at loci (1,2,3,4) and then the remaining genes from parent 2 at loci (5,6,7).

## Mutation

Mutation is able to produce new chromosomes and can be helpful when the effects of crossover diminish, diversity slowly disappears, and the GA begins to converge (Yassine et al., 2007a). Due to empirical results of Whitfield (2003) we favor a modification of a two operation swap (Murata and Ishibuchi, 1994). A chromosome that passed the selection phase is chosen randomly from the population for mutation (probability  $p_m$ ). Two genes ( $A_i, A_j$ ) or a multiple of two are then accidentally selected. The first gen at loci  $A_i$  partly exchanges values (alleles) with the gen of loci  $A_j$  under consideration of predecessor and resource constraints. In particular the assignment of actors and resources are swapped but also the starting time of the activity (*period* relating to the ending time of the predecessor tasks), stochastic value to determine makespan or the occurrence of iteration.

## COMPUTATIONAL RESULTS

We now present test results for the GA and the Petri-Net simulation model. We performed tests for C3 models with different project specific characteristics. The results give a first impression of the performance of the GA. Further studies are currently in progression.

### Case Study

The performance of the developed GA is tested on a RCMPSP – three a posteriori modeled development projects in an enterprise of the chemical engineering industry. The projects for the development of three large scale chemical engineering plants have respectively 62 different tasks (see Appendix 1) with project specific characteristics. While the projects are unrelated, the execution of the 186 tasks depends on the common pool of actors and resources of the involved organization units as well as precedence relationships of tasks within a project. But there exist no precedence relationships between the three projects. Tasks durations range from 1 to 60 time units. Every project is characterized by two iterations which are combined to a cascade. For the real and complex development processes an ideal sequence and assignment of actors is not known, and finding the global optimum may be difficult with respect to the problem size and constraint.

### C3 related results

We expected the population size and crossover rate to be a problem: the larger the crossover rate and the population, the greater is the chance that the best individuals of a generation are not continuously improved. As the population size increases, the best fitness value for each population improves. Figure 3 provides us an insight into the change of the best fitness value over population size and how crossover probability  $p_c$  impacts the performance. When population size reaches the task string size (number of tasks: 186), the optimal makespans become stable. These results are consistent with the findings of Zhuang et al. (2004). For such kind of scheduling problem, a decent solution is expected when population size is the number of genes. With large enough population, the initialization ensures that good schemas appear. When crossover rate,  $p_c = 0.85$ , the GA generated much better fitness values than those at  $p_c = 0.05$ . It indicates that the implemented crossover operator dictate the evolution.

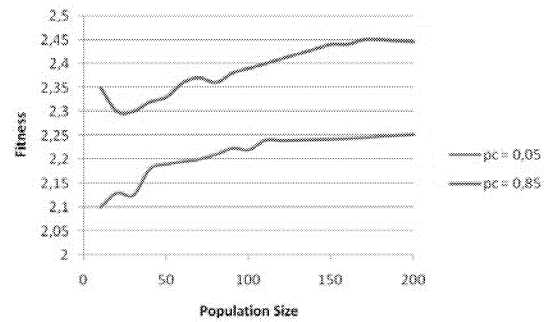


Figure 3 Average fitness value versus population size

It is demonstrated in Figure 4 that the fitness value increases as generation increases. A fast convergence rate is shown for the crossover due to moderate precedence relationships of tasks. It is observed from Figure 4 that the global optimum for the fitness value does not appear until several generations. Due to the multi criteria objectives the optimum of the project duration does not present the best solution for the project costs or the capacity utilization.

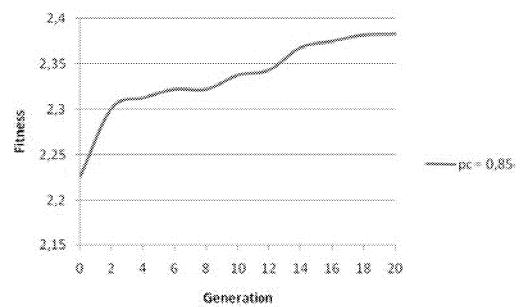


Figure 4 Fitness Value at population 50

It is shown in Figure 5 the project duration and cost of the fittest and worst individual of each generation.

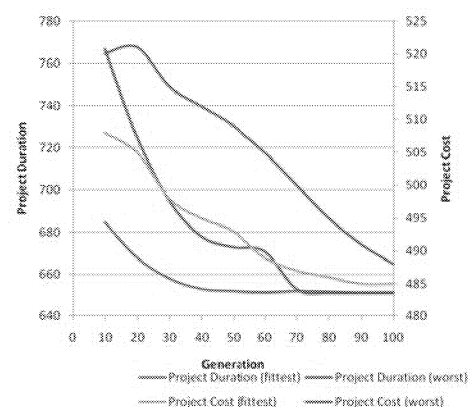


Figure 5 Project Duration and Cost for the fittest / worst individual of a generation

In this section we will also relate the performance of the proposed GA to a Petri net based simulation model for task scheduling (Kausch et al., 2007). A best solution for the RCMPSP is given by our

simulation model: 690 TU and 505 CU. It is observed that our GA is capable of significantly reducing the project duration, compared to the simulated project scenarios. The difference regarding the project costs was low due to the low variance of the wage of different actors.

The proposed GA-based scheduling method has demonstrated its advantage over the simulation approach in terms of simulation time, accuracy and efficiency in this particular test case.

## Stochastic Feedback

A project organization without any iterative execution of a task is often only a baseline schedule. In reality, some downstream tasks may be forced to be repeated due to changes in requirements of task's outcome. In order to accommodate this problem, this GA-based approach randomly generates a value based on the feedback probability and thus decides whether the feedback will be part of schedule or not. The model also considers that the probability of a feedback loop can be decrease or sometimes increase during multiple executions of a loop. Also it was implemented that the effort of a task during several iterations can vary. The value of the variance is based on the knowledge of the project manager and is calculated with the equation (4).

Random trials with possible different feedback tasks of the three projects are generated and an optimal schedule is obtained for each. A distribution function is found to best fit the resulting project distribution as shown in Figure 6. It is helpful to identify the most likely project duration range and provide a better understanding of how long the project may last. As such, we can conduct a sensitivity test on the project portfolio and evaluate how the three optimal project schedules are sensitive to changes in feedback structure.

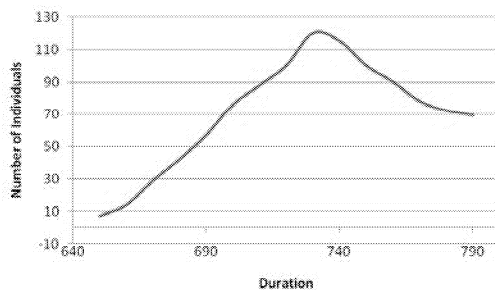


Figure 6 Project Duration for projects with stochastic feedback

## CONCLUDING REMARKS

This paper has proposed an implementation of Genetic Algorithms to solve a C3 model representation of the resource constrained multi-project scheduling problem for multi-criteria objectives. A population was initialized due to runs of a simulation model such that all individuals are precedence feasible. The novel characteristic of a genotype for the RCMPSP and GP mapping was introduced to maintain precedence and resource feasibility while obtaining the project duration, costs and degree of capacity utilization for fitness evaluation. The development of a novel GP mapping function was necessary due to the integration of uncertainties in the project model.

Good solutions were found however by using simple mutation and crossover operators. These genetic operators perform well for the continuous improvement of chromosomes over generations. It also accommodates feedback which is of paramount importance in a management of several concurrent development projects.

A great focus of future work should be on the integration of human behavior in the project model. This GA-based methodology can be easily extended to project models that include cooperation, coordination and communication processes between actors. In the latter case, preemption of tasks will be allowed. Therefore actors and resources will be available whenever tasks of a higher priority are ready to be performed. Finally, extensions to GA operators (crossover, mutation) to enlarge the performance can be made as well as the multi-objective fitness function can be revised to find better solutions for conflictive targets.

## REFERENCES

- Bäck, T., 1994. *Selective Pressure in Evolutionary Algorithms: A Characterization of Selection Mechanisms*. Proceedings of the First IEEE Conference on Evolutionary Computation, Vol. 1, 57–62.
- Brindle, A., 1981. *Genetic Algorithms for Function Optimization*. Doctoral dissertation, University of Alberta, Edmonton, Canada.
- Browning, T. R., and Eppinger, S. D., 2002. *Modeling Impacts of Process Architecture on Cost and Schedule Risk in Product Development*. In: IEEE Trans. Eng. Manage., 49(4), 428–442.
- Clark, K., Fujimoto, T., 1991. *Product Development Performance*. Harvard Business School Press, Boston, MA.
- Gärtner, T., Rohleder, N., Schlick, C.M. 2008 *Simulation of product change effects on the duration of development processes based on the DSM*, In: Proceedings of the 10th International DSM Conference, Stockholm, 11 and 12 November 2008, Hrsg.: Kreimeyer, M.; Lindemann, U.; Danilovic, M., Carl Hanser Verlag, Munich 2008,

- Garey, M. R., Johnson, D. S., 1979. *Computers and intractability: A guide to the theory of NP-completeness*. W. H. Freeman & Co., New York
- Goncalves, J.F., Mendes, J.J.M., Resende, M.G.C., 2008. *A genetic algorithm for the resource constrained multi-project scheduling problem*. European Journal of Operations Research, 189, 1171-1190.
- Grünert, T., Irnich, S. 2005. *Optimierung im Transport*, Band 1: Grundlagen, 182-244. Shaker Verlag GmbH, Aachen.
- Hartmann, S., 1998. *A competitive genetic algorithm for resource-constrained project scheduling*. Naval Research Logistics, 45, 733-750.
- Hölldt-Otto, K., Magee, C. L., 2006. *Estimating factors Affecting Project Task Size in Product Development – An Empirical Study*. IEEE Trans. Engineering Management, 53(1), 86-94.
- Kausch, B., Grandt, M., Schlick, C. 2007. *Activity-based Optimization of Cooperative Development Processes in Chemical Engineering*. In: SCSC 2007 "Summer Computer Simulation Conference", 15-18 July 2007, San Diego.
- Kelley, J.E. Jr., 1961. *Critical-Path Planning and Scheduling: Mathematical Basis*. In: Operations Research, 9(3), 296-320.
- KHosraviani, B., 2005. *An Evolutionary Approach for Project Organization Design: Producing Human Competitive Results Using Genetic Programming*. Doctoral Dissertation, Department of Civil and Environmental Engineering, Stanford
- Killich, S., Luczak, H., Schlick, C.M., Weissenbach, M., Wiedenmaier, S., Ziegler, J., 1999. *Behaviour & Information Technology*. 18, 5, 325-338.
- Kolisch, R., Padman, R., 2001. *An integrated survey of project deterministic scheduling*. In: International Journal of Management Science, 29(3), 249-272.
- Kolisch, R. (2000): *Integrated scheduling, assembly area- and part-assignment for large-scale, make-to-order assemblies*. International Journal of Production Economics, 64, 127-141.
- Kolisch, R., Hartmann, S. 1998. *Heuristic algorithms for solving the resource constrained project scheduling problem: classification and computational analysis*. In: Handbook on Recent Advances in Project Scheduling, Kluwer, Boston.
- Lenstra, J., Rinnooy, K., 1978. *Complexity of Scheduling under Precedence Constraints*. Operations Research, 26(1), 22-35.
- Liepins, G.E., Vose, M.D. 1990. *Representational issues in genetic optimization*. In: Journal of Experimental and Theoretical Artificial Intelligence, 2(2), 4-30.
- Linyi, D.; Yan, L., 2007. *A Particle Swarm Optimization for Resource-Constrained Multi-Project Scheduling Problem*. International Conference on Computational Intelligence and Security, doi:10.1109/CIS.2007.157.
- Malcolm, D.G., 1959. *Application of a Technique for Research and Development Program Evaluation*. In: Operations Research, 7(5), 646-669.
- Murata, T., Ishibuchi, H., 1994. *Performance Evaluation of Genetic Algorithms for Flow Shop Scheduling Problems*. In: Proceedings of the First IEEE Conference on Genetic Algorithms and their Applications Orlando, FL, June 27-29, 812-817.
- Nonobe, K., Ibaraki, T., 2001. *A Local Search Approach to the Resource Constrained Project Scheduling Problem to Minimize Convex Costs*. 4th Metaheuristics International Conference, Porto, Portugal.
- Schlick, C. M., Beutner, E., Duckwitz, S., Licht, T., 2007. *A Complexity Measure for New Product Development Projects*. In: Proceedings of the 19th International Engineering Management Conference, IEMC 2007, Managing Creativity: The Rise of the Creative Class. Austin, Texas, USA, 143-150.
- Shouman, M.A., Ibrahim, M.S., Khater, M., Forgani, A.A., 2006. *Genetic algorithm constraint project scheduling*. Alexandria Engineering Journal, Vol. 45, No. 3, 289-298.
- Tackenberg, S., Duckwitz, S., Karahancer, S., Kabuß, W., Kausch, B., Schlick, C.M., 2008. *An Activity- and Actor-Oriented Approach for the Project Engineering of Complex, Weakly Structured Projects*. In: ESM'2008 - The European Simulation and Modelling Conference October 27-29, Le Havre, France, Hrsg.: Bertelle, C.; Ayesh, A., EUROSIS-ETI, Ostend, Belgium
- Thierens, D., 1995. *Mixing in genetic algorithms*. Doctoral Dissertation, Katholieke Universiteit Leuven
- Whitfield, R. I., Duffy, A. H. B., Coates, G., Hills, W., 2003. *Efficient Process Optimization*, Concurr. Eng. Res. Appl., 11(12), 83-92.
- Yang, B., Geunes, J., O'Brien, W.J., 2001. *Resource-Constrained Project Scheduling: Past Work and New Directions*. Research Report 2001-6, Department of Industrial and Systems Engineering, University of Florida.
- Yassine, A. A., Meier, C., Browning, T. R. 2007a. *Design Process Sequencing With Competent Genetic Algorithms*. Transaction of the ASME, 129, 566-585.
- Yassine, A. A., Meier, C., Browning, T. R. 2007b. *Multi-Project Scheduling using Competent Genetic Algorithms*. Working Paper, University of Illinois
- Zhuang, M., Yassine, A. A., 2004. *Task Scheduling of Parallel Development Projects Using Genetic Algorithms*. In: Proceedings of DETC '04 ASME 2004, International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Salt Lake City, Utah, USA, 143-150.

## APPENDIX

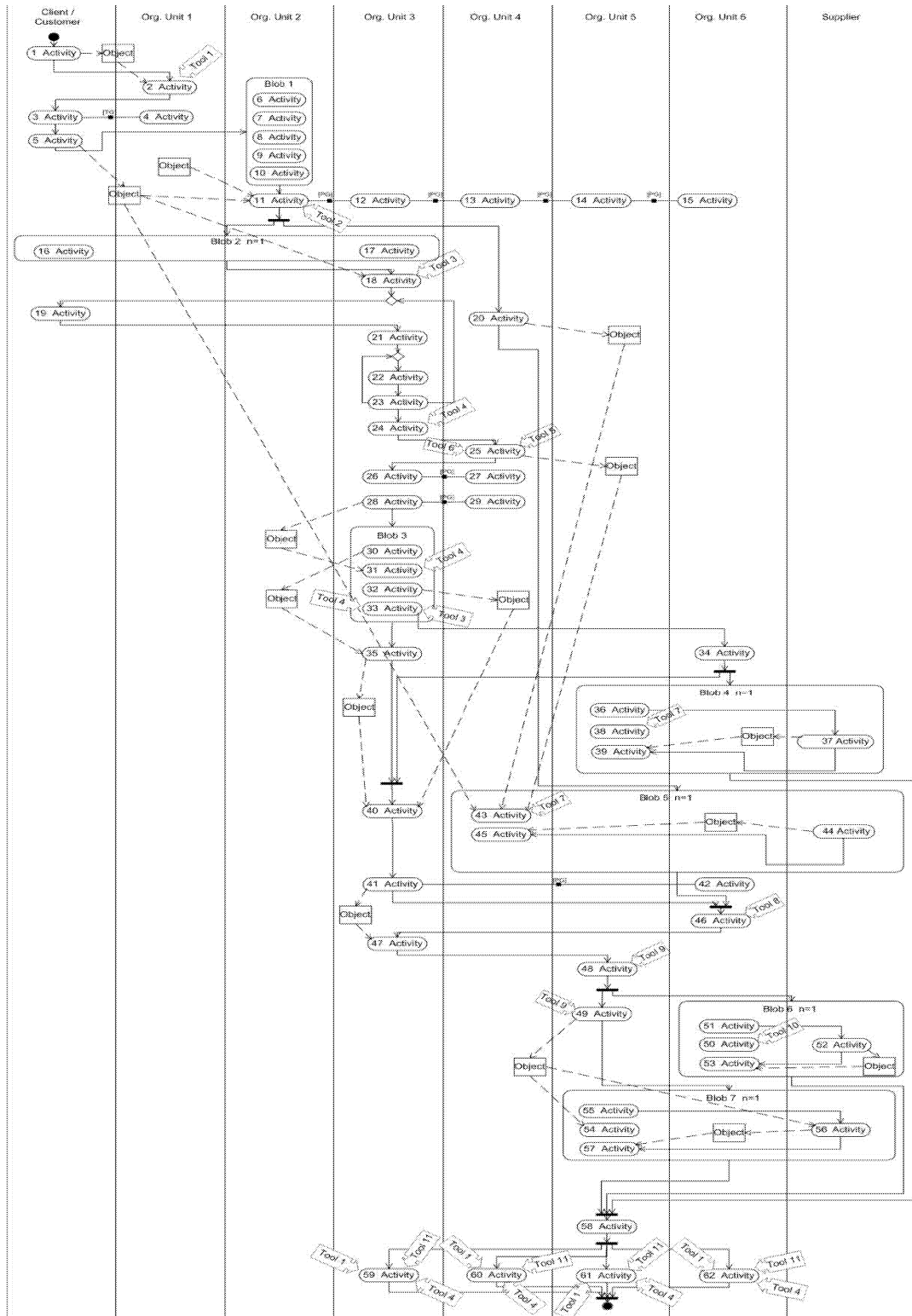


Figure A-1 C3 Model of development project





# **TEXTILE SIMULATION**



# STATISTICAL ANALYSIS OF THE RESULTS OF SOME TECHNOLOGIES FOR COTTON FIREPROOFING

Popescu Vasilica  
Textile Chemical & Finishing  
Department, Technical University  
"Gh. Asachi" Iasi, Romania, 53 Bd  
Mangeron, Iasi 700050  
vpopescu@tex.tuiasi.ro

Manea Liliana –Rozemarie  
Tech. and Textile Fabrics Design  
Department, Technical University  
"Gh. Asachi" Iasi, Romania, 53 Bd  
Mangeron, Iasi 700050  
lili191065@yahoo.com

Popescu Gabriel  
Machine Design & Mechatronics  
Department, Technical University  
"Gh. Asachi" Iasi, Romania, 61-63  
Bd Mangeron, Iasi 700050  
gpopescu65@yahoo.com

## KEYWORDS

Statistical analysis, Fischer function,  $\sigma^2$ , multiple comparison

## ABSTRACT

A fireproofing technology implies more phases and automatically there are more factors which can influence the efficiency. Cure temperature and concentration of Pekoflam DPN1 were studied. To study the qualitative and quantitative effects exerted by the above-mentioned possible influence factors, a mathematical method for the results analysis, namely, the bifactorial dispersional analysis with systematic effects, has been employed. The bifactorial dispersional analysis leads to the computation of three statistical tests F. The application of a C++ personal program for the bifactorial dispersional analysis with systematic effects, of the 4x6 type, concluded with a multiple comparison (based on the Scheffe test) has made us find out that both the concentration of the used fire-retardant (named Pekoflam DPN1) and the cure temperature have significant influences upon the fireproofing efficiency. Even the interaction of the two above-mentioned factors has a certain influence, although much smaller, than that of the enumerated individual factors. The significant positive effects are attributed, with 99% statistical confidence, to the concentrations of 100÷400 g/l fire-retardant and to the 160°C temperature while the negative effects are due to the concentrations of 500 and 600 g/l Pekoflam DPN 1 and to the 180°C cure temperature.

## INTRODUCTION

The treatment whose aim is the protection against ignition and burning is known as "fireproofing". The fireproofing of the cellulose materials is especially applied to the following items: protection clothing, children garments, bed linen, decorative items, furniture fabrics, etc. The efficiency of any fireproofing treatment depends on the flammability of the textile product (Alnishaenskin 1968), on the combustion and

post-incandescence. Flammability refers to the temperature and the heat necessary to ignite a textile material of a certain chemical composition and a given presentation. Combustion shows the development of the burning process after ignition, often accompanied by noxious gases and vapors releases.

The employed pad-dry-cure fireproofing technology implies many stages and factors which influence the fireproofing efficiency. The rendering evident of the factors with certain influences are done by means of the dispersion analysis.

The dispersion analysis, or the analysis of the variation, includes an ensemble of methods and techniques for grasping the causes which explain the variation of a phenomenon or process, for identifying the factors with significant influence, for measuring the main effects and the interaction of the factors. Since, as a principle, the dispersions of some statistical variables are not additive, one uses for decomposition the quadratic variations (deviations); in fact, the name of "variation analysis" comes from here (Baron 1979; Cojocaru et al.1986; Rancu and Tovissi 1963; Văduva 1970). The "dispersion" parameter characterizes as a rule the scattering of the values of an aleatory variable.

On the chemical processes from the textile industry there act simultaneously many parameters, each of them having a specific influence upon the desired result. Thus, by applying the dispersional analysis it is possible to separate and to test the effects caused by the variation of the technological parameters and to eliminate from the observation field the parameters whose variation is not technically significant. The mathematical principle of the dispersional analysis is based on the grouping together of all data noticed according to one or more criteria (factors) and on the rendering evident of the effects depending on the particular influence of the criteria the observations have been grouped upon.

Once the effects have been identified, the testing takes place by comparing the dispersions caused by the variable factors, with the dispersion produced by the random factors which act upon the studied process.

## EXPERIMENTAL PART

## Materials

Two identical series of 100% cotton material dyed with an Direct Blue dyestuff, having a 0.200 Kg/m specific weight, a 35 cm length and a 3 cm width, respectively, have been subjected to the fireproofing process.

The following substances have been used for fireproofing:

1. The Pekoflam DPN 1 which is an organic phosphorous compound with 455°C ignition temperature, 1.24 g/cm<sup>3</sup> density and 37 ml viscosity at 20°C. It is applied at room temperature and pH = 4;
2. The ortho-phosphoric acid is the catalyst and the accomplisher of the acid medium required by the active fireproofing substance, i.e. Pekoflam DPN 1;
3. The HML Cassurite is a 1,3,5 triazine-2,4,6-triamine, a methylol-melamimic compound with 50 % concentration and 115 g/cm<sup>3</sup> density;
4. Urea;
5. The JET Kerallon is a mixture of non-ionic surfactants.

## Experiments

The fireproof finishing is based on the pad-dry-cure technology which implies the following phases: padding – wringing – drying – cure – neutralization – washing (rinsing). The efficiency of the fireproofing treatment can depend on many factors among which we mention:

- The cure temperature (considered as A factor);
- The concentration of the fireproofing active agent (considered as B factor)
- The conditions of the neutralization (carbonate concentration, temperature and time period) and of the rinsing;
- The pad time period;
- The time period of the cure treatment.

We have experimented 24 fireproofing technological variants according to the following working formulae:

1. padding with  $X_1$  g/l Pekoflam DPN 1;  $X_1 \in [100 \div 600 \text{ g/l}]$ ; ( $X_1 = \text{B factor}$ ); 40 g/l HML cassiterite; 20 g/l ortho-phosphoric acid; 5 g/l urea; 2 g/l JET Kerallon
2. wringing at GS= 75% ;
3. drying at 120°C for 2 minutes;
4. cure at  $X_2^\circ\text{C}$ ;  $X_2 \in [150 \div 180^\circ\text{C}]$  for 2 minutes;  $X_2 = \text{A factor}$ .

5. neutralization with 2 g/l Na<sub>2</sub>CO<sub>3</sub> at 30 °C for 15 minutes.

After the fireproof finishing, the samples have been subjected to the fire-test according to STAS 8025-84.

There have been determined the following values: 1) the increase in weight of the samples after impregnation; 2) the increase of the fabric thickness after the fireproof finishing; 3) the length of the carbonized area and, implicitly, of the non-carbonized one; 4) the time period of the flame propagation; 5) the incandescence time period; 6) the loss in weight after the burning of the fireproofed fabric.

To study the qualitative and quantitative effects exerted by the above-mentioned possible influence factors, a mathematical method for the results analysis, namely, the bifactorial dispersional analysis with systematic effects, has been employed. The following steps have been carried out for the correct usage of this method: a rigorous planning of the experiments so that with any of the 6 above-mentioned

concentrations of Pekoflam DPN1, fireproof finishings at 150°, 160°, 170° and 180°C temperatures, respectively, should be accomplished, thus  $24 \times 2 = 48$  working variants being obtained (Table 1); in fact, each experiment has been done twice (a fact justified by the existence of two identical series subjected to the analysis) with a view to obtain two readings for each cell in the dispersional analysis.

Table 1: Coding of the experimented fireproofing variants  $V_{ijk}$  (where  $i$  = number of the A factor level;  $j$  = number of the B factor level;  $k \in [1, 2]$ ;  $k$  = number of the experiment from the respective cell

| Levels of the A factor<br>$i \in [1, 4]$ | Levels of the B factor<br>$j \in [1, 6]$ |                       |                       |                       |                       |                       |
|--|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
|  | Level 1<br>100<br>g/l                    | Level 2<br>200<br>g/l | Level 3<br>300<br>g/l | Level 4<br>400<br>g/l | Level 5<br>500<br>g/l | Level 6<br>600<br>g/l |
|  |  |                       |                       |                       |                       |                       |
| Level 1<br>150 °C                        | $V_{111}$                                | $V_{121}$             | $V_{131}$             | $V_{141}$             | $V_{151}$             | $V_{161}$             |
|  | $V_{112}$                                | $V_{122}$             | $V_{132}$             | $V_{142}$             | $V_{152}$             | $V_{162}$             |
| Level 2<br>160 °C                        | $V_{211}$                                | $V_{221}$             | $V_{231}$             | $V_{241}$             | $V_{251}$             | $V_{261}$             |
|  | $V_{212}$                                | $V_{222}$             | $V_{232}$             | $V_{242}$             | $V_{252}$             | $V_{262}$             |
| Level 3<br>170 °C                        | $V_{311}$                                | $V_{321}$             | $V_{331}$             | $V_{341}$             | $V_{351}$             | $V_{361}$             |
|  | $V_{312}$                                | $V_{322}$             | $V_{332}$             | $V_{342}$             | $V_{352}$             | $V_{362}$             |
| Level 4<br>180 °C                        | $V_{411}$                                | $V_{421}$             | $V_{431}$             | $V_{441}$             | $V_{451}$             | $V_{461}$             |
|  | $V_{412}$                                | $V_{422}$             | $V_{432}$             | $V_{442}$             | $V_{452}$             | $V_{462}$             |

## MATHEMATICAL METHOD

### The Bifactorial Dispersional Analysis

The dispersional analysis is a fundamental method of statistics which characterizes especially the variability and the factors determining it. The variability can be due to the existence of some factors with systematic influences, of some random factors with stronger fluctuation, and finally, of unavoidable local factors which determine a smaller fluctuation defined as "experimental fluctuation". The dispersional analysis is aimed for the separation of the "total variability" in the following: the variability due to the systematic factors, the variability of the factors with random effects, the "residual" variability (i.e. the difference up to the total variability) which in fact represents the experimental variability. The partial dispersions corresponding to the various factors are estimated from these variabilities and the significance of their ratios is computed by applying the F test. The main experimental data are, as a rule, grouped together according to various criteria and the effects upon the variability are being observed depending on these criteria, effects which are quantified as against the residual variability.

In a bifactorial dispersional analysis one can assume that the probing data admit the following configuration:

$$x_{ijk} = m + a_i + b_j + w_{ij} + e_{ijk} \quad (1 \leq i \leq h, 1 \leq j \leq q, 1 \leq k \leq n) \quad (1)$$

where:  $m$  = overall average value;  $a_i$  = main effect due to the "i" level of the A factor;  $b_j$  = main effect due to the "j" level of the B factor;  $w_{ij}$  = effect due to the interaction of the "i" level of the A factor with the "j" level of the B factor;  $e_{ijk}$  = errors

and the following hypotheses are satisfied:

a) the average values of all effects or interactions are null and this is why they revert to

$$\sum_{i=1}^h a_i = \sum_{j=1}^q b_j = \sum_{i=1}^h w_{ij} = \sum_{j=1}^q w_{ij} = 0 \quad (2)$$

$$1 \leq i \leq h, \quad 1 \leq j \leq q$$

b) the  $e_{ijk}$  errors are independent aleatory variables with the distribution  $N(0, \sigma^2)$ . No matter how exactly the unknown parameters are determined, there still exists the possibility of some errors which will affect the results; that is why the dispersal analysis is based on the determination of some confidence intervals (confidence limits) which should include, with a pre-established probability, the values of the estimated parameters.

Consequently, the dispersal analysis is the statistical method which, on the basis of the usage of the probing data, of the typical probing data as well as of the statistical tests, allows the separation and the testing of the effects caused by the parameters variation as well as the elimination from the observation field of the parameters whose variation is not technically significant (Gluck 1971). The statistical parameters of a population can be computed only approximately, the real values ranging between certain limits which vary according to the number of probing elements and the desired accuracy.

With the help of the statistical tests one can analyse if the difference which exists between the computed parameters and the real value of these parameters is significant for the conclusion to be drawn.

With this end in view, one uses the notion of null hypothesis, i. e. one first admits the hypothesis that there is no difference between the studied parameters, these parameters being equal one with the other, and, at the same time, equal with zero. One checks up if this hypothesis is valid.

If the computed probability is higher than a value chosen as critical (from the tables), this fact means that the difference between the parameters of the sample and the parameters of the probing population from which the sample has been taken, is at random and the null hypothesis is not accepted; in other words, this means that in reality there exists a significant statistical difference between the studied parameters.

If the computed probability is smaller than the chosen critical probability (from the tables), then the null hypothesis is accepted, therefore the studied parameters do not significantly influence the statistics.

The main tests employed in the dispersal analysis are: the F test (the Fisher-Snedecor test); the t test (the Student test); the  $\chi^2$  test.

There exist three models of bifactorial dispersal analysis, namely,

a) with random effect – when the cells are considered as a probing sample from a multitude of cells, so that the deviations related to these cells should be of the random type;

b) with systematic effects – the multitude of the cells is made up only from the analysed ones; there are no other cells and the data grouped according to those two factors determine the systematic-type deviations;

c) with mixed effects – when for the registered data, the deviations are at random according to a factor and systematical according to the other factor.

In this paper we have resorted to a C++ personal program for the bifactorial dispersal analysis with

systematic effects of the 4x6 type (i.e. 4 levels for the A factor and 6 levels for the B factor), a program whose logical diagram is shown in Figure 1. This method is concluded with a comparison of the dispersions according to the S (Scheffe) method based on contrasts.

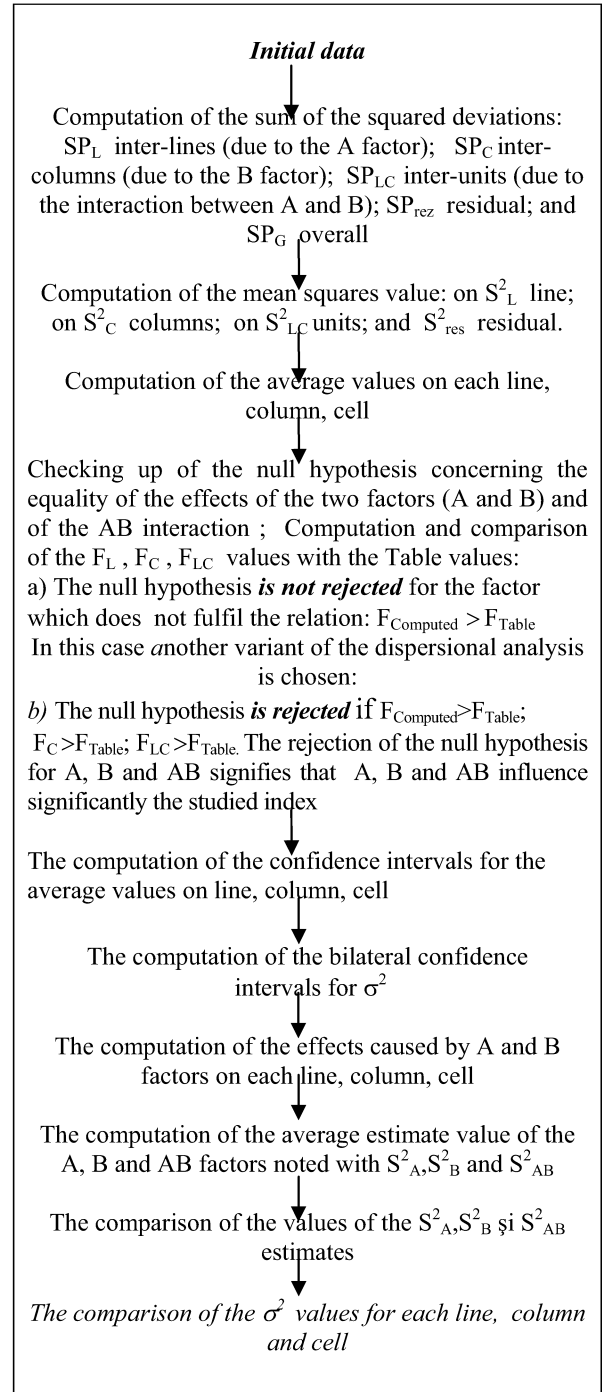


Figure 1: Logical diagram for the bifactorial dispersal analysis with systematic effects

## RESULTS AND DISCUSSIONS

The following values have been obtained by burning the control sample (non-fireproofed sample):

- a propagation time period of the flame of 13.75 seconds;
- an incandescence time period of 1 second;
- the length of the carbonized area  $L = 35$  cm (the sample burnt entirely);
- the mass after burning = 0.037g.

By burning two series of fireproofed samples under the same conditions we have obtained:

- 1) the propagation time period of the flame has varied between 0.2 and 3 seconds;
- 2) the incandescence time period = 0 – 0.5 seconds;
- 3) the values for the length of the carbonized area have ranged widely, from 1 ÷ 13 cm.

By applying the bifactorial dispersional analysis with systematic effects concerning the length of the carbonized area, the information shown in Table 2 have been obtained.

Table 2: Information referring to the bifactorial dispersional analysis applied for the study of the variation of the carbonized areas length

| Source of variation                  | Fischer function  | Estimation of dispersion                               |
|--------------------------------------|---|--|
| Line ( $A=Cure$ Temperature)         | $F_L=467.5719$  | $S^2_{Amin}=18.582444$<br>$S^2_{Amax}=18.617265$       |
| Column ( $B=Pekoflam$ Concentration) | $F_C=5115.2168$   | $S^2_{Bmin}=126.611051$<br>$S^2_{Bmax}=126.64587$      |
| Interaction $A$ with $B$             | $F_{LC}=197.689$  | $S^2_{Abmin}=7.867711$<br>$S^2_{Abmax}=7.902532$       |
| Errors inside the cells              | $F_{95,3,24}=3.01$<br>$F_{95,5,24}=2.62$<br>$F_{95,15,24}=2.11$ | $\sigma^2_{min}=0.019371$<br>$\sigma^2_{max}=0.054191$ |

According to the bifactorial dispersional analysis both the Pekoflam DPN1 concentration and the cure temperature influence the magnitude of the carbonized area, i. e. the fireproofing efficiency. One knows that a small length of the carbonized area is equivalent with a good fireproofing effect of the applied treatment.

According to the multiple comparison which certifies the accuracy of the dispersional analysis, all concentrations of Pekoflam DPN 1 lead, with 99% statistical confidence, to a certain degree of fireproofing; the greater the concentration is, the stronger this effect will be. In exchange, in what the "cure temperature" factor is concerned, the method of the multiple comparison shows the fact that only two temperatures influence the fireproofing efficiency, namely,  $T=160^\circ\text{C}$  and  $T=180^\circ\text{C}$ . From these two factors, i. e. the Pekoflam DPN 1 concentration and the cure temperature, the greatest effect is brought about by the concentration of the fireproofing agent, a fact also revealed by the values  $S^2_{Bmin}$  and  $S^2_{Bmax}$  which are much higher than  $S^2_{Amin}$  and  $S^2_{Amax}$ . The interaction of these two factors also has a certain contribution in accomplishing the fireproofing effect, but the exerted influence is much smaller as compared to those produced by each factor taken separately.

The minimum and maximum effects exerted by these two studied factors, computed by means of the dispersional analysis, are shown in Table 3.

Table 3: The effects minimum / maximum produced by the A and B factors

|     | The levels of the B factor |             |             |             |             |          |
|-----|----------------------------|-------------|-------------|-------------|-------------|----------|
|     | 100 g/l                    | 200 g/l     | 300 g/l     | 400 g/l     | 500g/l      | 600g/l   |
| min | +2.772<br>9                | +2.930<br>4 | +4.147<br>9 | +2.510<br>4 | -<br>6.5896 | -6..5896 |
| max | +3.043<br>8                | +3.201<br>3 | +4.418<br>8 | +2.781<br>3 | -<br>6.3187 | -6..3187 |
|     | The levels of the A factor |             |             |             |             |          |
|     | 150 °C                     | 160 °C      | 170 °C      | 180 °C      | -           | -        |
| min | -0.0140                    | +1.0693     | +0.3776     | +1.8390     | -           | -        |
| max | +0.1891                    | +1.2724     | +0.5807     | -1.6359     | -           | -        |

One can notice from table 3 that the width of the confidence interval for the effect produced by the B factor is 0.2709 while for the A factor is a little smaller, i.e. 0.2031. This fact means that the B factor has a greater influence upon the length of the carbonized area as one knows that a small width of the confidence interval is equivalent with a higher uniformity of the studied factor.

Figure 2 shows the dependence of the fireproofing efficiency upon the two studied factors (the concentration of the Pekoflam DPN1 fireproofing agent and the cure temperature), namely, the increase of the concentration of the fireproofing agent determines the appearance of a small carbonized area, irrespective of the temperature at which the cure phase has been carried out, a fact which is equivalent with an increase in the fireproofing efficiency.

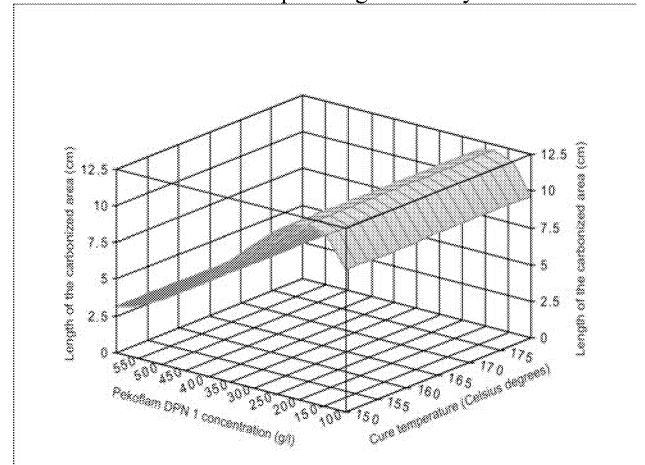


Figure 2: Length of the carbonized area

The efficiency of the fireproofing due to both the Pekoflam DPN 1 concentration and to the cure temperature factors, respectively, is also evident from the variation of the weight, of the thickness of the samples (after the pad-dry-cure treatment) (Figures 3 and 4). The irregularity of the burnt area from a fireproofed sample has imposed the finding of a solution for a fireproofing efficiency as high as possible; that is why the decrease of the mass of the sample fireproofed and subjected to burning has been studied (Figure 5); this decrease should be as small as possible to render evident an increased efficiency for the fireproofing treatment.

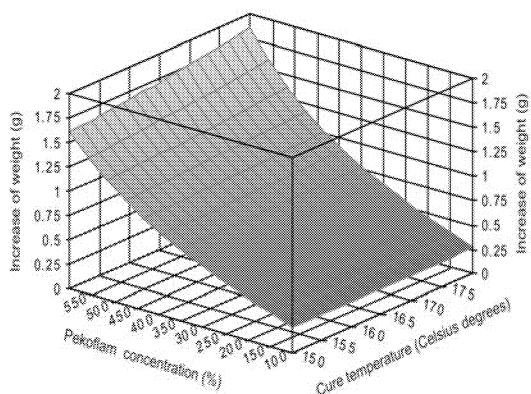


Figure 3: Increase of the samples weight after the fireproofing treatment

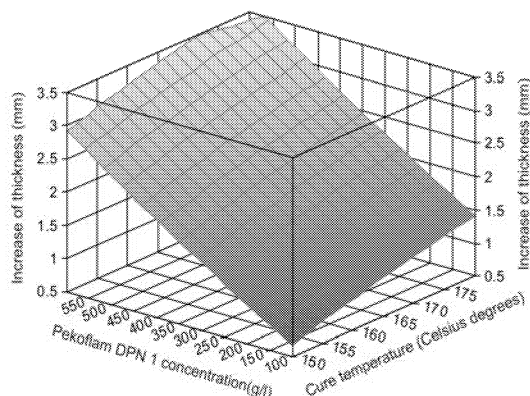


Figure 4: Increase of samples thickness after the fireproofing treatment

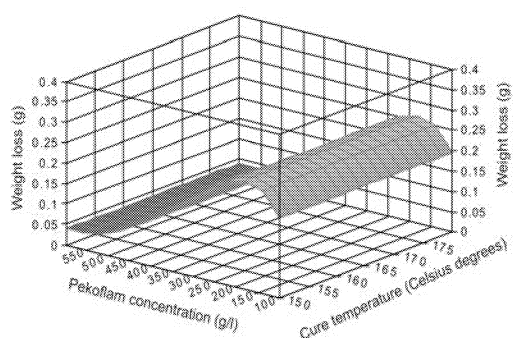


Figure 5: Weight loss of samples after burning

## CONCLUSIONS

As a result of the fireproofing treatment carried out with a phosphorus-based compound, i. e. the Pekoflam DPN 1, the following conclusions can be stated:

- one can assert, with 99% statistical confidence, that the fireproofing efficiency depends on the concentration of the active fireproofing agent and on the cure temperature;
- as a result of employing the dispersional analysis, we can affirm that the most significant effect is that of the Pekoflam DPN 1 concentration which determines the appearance of some positive influences for the 100÷400 g/l concentrations as well as negative effects for concentrations higher than 400 g/l Pekoflam DPN 1 (see Table 3);
- in what the cure temperature is concerned, only the 160°C temperature determines positive influences upon the fireproofing efficiency expressed in the length of the non-carbonized area and in the flame propagation time period, respectively

## REFERENCES

- Alnishaenskin, R., 1968. "Fireproofing possibilities on textiles" (in Francaise). *Les cahiers*, Ciba, no. 4, 35-46.
- Baron, T., 1979. *Statistical methods applied in production* (in Romanian), Pedagogical and Didactical Publishing House, Bucharest.
- Cojocaru, N., et al, 1986. *Statistical methods applied in textile industry* (in Romanian), Technical Publishing House, Bucharest.
- Gluck, A., 1971. *Mathematical methods in chemistry domain* (in Romanian), Technical Publishing House, Bucharest.
- Rancu, N., .Tovissi, L., 1963. *Statistical mathematic with application in production* (in Romanian), Academy Publishing House, Bucharest.
- Văduva, I., 1970. *Dispersional analysis* (in Romanian), Technical Publishing House, Bucharest.

## BIOGRAPHY

**POPESCU VASILICA** was born in Buzau, Romania and went to the Technical University "Gh. Asachi" from Iasi where she studied in textile chemical and finishing. After graduation, she worked for ten years in production and since 1998 is a lecturer Ph.D in Faculty of Textiles, Leather Engineering and Industrial Management, Department Textiles Chemical Finishing. Her interest area consists in the chemical finishing of textiles and statistical analysis.

**POPESCU GABRIEL** was born in Adjud, Romania and went to the Technical University "Gh. Asachi" from Iasi where he studied in. After graduation, he worked for 2 years in production and since 1998 is a lecturer Ph.D in Machine Design & Mechatronics Department, from Mechanical Engineering Faculty. His interest area consists in programming on PC, mathematical methods, contact elasto-plastic.

**If you have any questions regarding the preparation of manuscripts, it is best to have them answered now - contact the EUROSIS for any clarification.**

# OBJECTIVE HIERARCHIZATION OF SOME TECHNOLOGIES FOR COTTON FIREPROOFING

Popescu Vasilica  
Textile Chemical & Finishing  
Department, Technical University  
"Gh. Asachi" Iasi, Romania, 53 Bd  
Mangeron, Iasi 700050  
vpopescu@tex.tuiasi.ro

Manea Liliana –Rozemarie  
Tech. and Textile Fabrics Design  
Department, Technical University  
"Gh. Asachi" Iasi, Romania, 53 Bd  
Mangeron, Iasi 700050  
lili191065@yahoo.com

Popescu Gabriel  
Machine Design & Mechatronics  
Department, Technical University  
"Gh. Asachi" Iasi, Romania, 61-63  
Bd Mangeron, Iasi 700050  
gpopescu65@yahoo.com

## KEYWORDS

Chemical engineering, mathematical, hierarchization, matrix

## ABSTRACT

The pad-dry-cure fireproofing technology employed in this paper implies many stages and factors which influence the fireproofing efficiency. Twenty four technological fireproofing variants have been studied.

The development of mathematics offers the possibility to hierarchize the working variants according to certain principles by using the multi-criteria decision methods.

To hierarchize the technological variants there can be used one of the methods for the multi-attribute decisions with the cardinal preferences of the criteria. From all these methods we have employed the Onicescu one because this method implies the fact that all criteria have the same importance, thus eliminating the subjectivity which can appear when choosing and attributing the importance for each chosen criteria.

The hierarchization of the fireproofing treatments has been accomplished according to 3 criteria, all considered equally important, namely, 1) the increase of the samples thickness after the fireproofing process; 2) the length of the carbonized area; 3) the losses in the samples mass after burning.

The implementation of the Onicescu method in the field of the textile chemical finishing – and not only here – by means of the C++ program is possible, easy and fast while the results are relevant.

## INTRODUCTION

The treatment whose aim is the protection against ignition and burning is known as "fireproofing". The fireproofing of the textile materials is especially applied to the following items: protection clothing, children garments, bed linen, decorative items, furniture fabrics, etc. The efficiency of any fireproofing treatment depends on the flammability of the textile product (Alnishaenskin 1968; Grindea et al. 1981), on the combustion and post-incandescence.

Flammability refers to the temperature and to the amount of heat necessary to ignite a textile material of a certain chemical composition and a given presentation. Combustion shows the development of the burning process after ignition, and the post-incandescence is the phenomenon of burning without flame, being often accompanied by noxious gases and vapors releases.

The employed pad-dry-cure fireproofing technology implies many stages and factors which influence the fireproofing efficiency. Twenty four technological fireproofing variants have been studied.

The development of mathematics offers the possibility to hierarchize the working variants according to certain principles by using certain multi-criteria decision methods.

To hierarchize the technological variants there can be used one of the methods for the multi-attribute decisions with the cardinal preferences of the criteria (Andraşiu et al. 1986; Rancu et al. 1963), namely,:

a) *directs methods* which make up a function defined on the set of the variants with normal values and select the variants for which the function takes the highest value: the method of the linear attribution; the method of the simple additive weight; the method of the hierarchical additive weight; the diameters method; the Onicescu method.

b) *indirect methods* (e. g. the Electre method and its variants) which determine a hierarchy on the set of variants on the basis of a certain algorithm.

c) *methods which use the distance* and select a variant as close as possible to the ideal solution: the Topsis method; the deviation minimization method; the Saphier-Rusu method; the scores method.

From all these methods we have employed the Onicescu one because this method implies the fact that all criteria have the same importance, thus eliminating the subjectivity which can appear when choosing and attributing the importance for each chosen criteria.

The hierarchization of the fireproofing treatments has been accomplished according to 3 criteria, all considered equally important, namely, 1) the increase of the samples thickness after the fireproofing process; 2) the length of the carbonized area; 3) the losses in the samples mass after burning.



## EXPERIMENTAL PART

Twenty four samples of 100% cotton material dyed with an Direct Blue dyestuff, having a 0.200 Kg/m specific weight, a 35 cm length and a 3 cm width, respectively, have been subjected to the fireproofing process.

The following substances have been used for fireproofing:

1. *The Pekoflam DPN 1* which is an organic phosphorous compound with 455°C ignition temperature, 1.24 g/cm<sup>3</sup> density and 37 ml viscosity at 20°C. It is applied at room temperature and pH = 4; it is the active agent of fireproofing;
2. *The ortho-phosphoric acid* is the catalyst and the accomplisher of the acid medium required by the active fireproofing substance, i.e. Pekoflam DPN 1;
3. *The HML Cassurite* is a 1,3,5 triazine-2,4,6- triamine, an etherified methylol-melamimic compound with 50 % concentration and 1.15 g/cm<sup>3</sup> density;
4. *Urea*; and 5. *The JET Kerallon* is a mixture of non-ionic surfactants.

The fireproof finishing is based on the pad-dry-cure technology which implies the following phases: padding – wringing – drying – cure – neutralization – washing (rinsing). The efficiency of the fireproofing treatment can depend on many factors among which we mention: 1) the cure temperature; 2) the concentration of the fireproofing active agent; 3) the conditions of the neutralization (carbonate concentration, temperature and time period) and of the rinsing; 4) the pad time period; 5) the time period of the cure treatment.

We have experimented 24 fireproofing technological variants according to the following working formulae:

1. padding with 100, 200, 300, 400, 500 și 600 g/l Pekoflam DPN 1; 40 g/l HML Cassurite; 20 g/l ortho-phosphoric acid; 5 g/l urea; 2 g/l JET Kerallon
  2. wringing at GS = 75%;
  3. drying at 120°C for 2 minutes;
  4. cure at 150, 169, 170, 180°C for 2 minutes;
  5. neutralization with 2 g/l Na<sub>2</sub>CO<sub>3</sub> at 30 °C for 15 minutes.
- The samples have been firstly impregnated with each of the Pekoflam DPN 1 concentrations, then they have been cured in turn at the four mentioned cure temperatures. After the fireproof finishing, the samples have been subjected to the fire-test according to STAS 8025-84.

There have been determined the following values which represent the three criteria used: 1) the increase of the fabric thickness after the fireproof finishing; 3) the length of the carbonized area; 3) the weight loss after the burning of the fireproofed fabric.

### THE MATHEMATICAL FORMALISM EMPLOYED IN THE ONICESCU METHOD FOR MULTI-ATTRIBUTE DECISIONS

On assuming the notations  $V = \{V_1, V_2, V_3, \dots, V_m\}$  for a set of variants for the flow sheets and  $C = \{C_1, C_2, C_3, \dots, C_n\}$  for a set of criteria, one aims at the arranging of the elements of the V set according to all criteria (the elements of the C set) from the best to the worst ones (Andrașiu et al. 1986; Onicescu et al. 1979).

As against the  $C_j$ , ( $j = 1, 2, 3, \dots, n$ ) criteria, to each  $V_i$ , ( $i = 1, 2, 3, \dots, m$ ) variant is associated a vector which characterizes the result of the estimation of that variant as compared with the V set. The set of these vectors constitutes

the lines of an "A" matrix, called the consequences matrix. Thus, the A matrix completely characterizes the problem under study, a thing called "the problem of the multi-attribute criterial decision", a problem which can be either cardinal or ordinal. Any cardinal problem can be obviously reduced to an ordinal one by arranging the  $a_{1j}, a_{2j}, \dots, a_{mj}$  values, thus establishing a classification of the variants for each  $C_j$ , ( $j = 1, \dots, n$ ) criterion. The criteria are considered equi-important in the Onicescu method (version I). The A matrix has the following form (1):

$$A = \begin{matrix} & C_1 & C_2 & \dots & C_n \\ \begin{matrix} V_1 \\ V_2 \\ V_3 \\ \dots \\ V_m \end{matrix} & \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \end{matrix} \quad (1)$$

where:  $a_{ij}$  ( $i = 1, \dots, m$ ;  $j = 1, \dots, n$ ) represents the value obtained by each technological variant as against any important criterion.

One starts from the A consequences matrix and one arranges the variants as against the criteria; thus a new matrix (2), namely B, is being obtained:

$$B = \begin{matrix} & C_1 & C_2 & C_3 & \dots & C_n \\ \begin{matrix} V_1 \\ V_2 \\ V_3 \\ \dots \\ V_m \end{matrix} & \begin{pmatrix} L_{11} & L_{21} & L_{31} & \dots & L_{1n} \\ L_{21} & L_{22} & L_{23} & \dots & L_{2n} \\ L_{31} & L_{32} & L_{33} & \dots & L_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ L_{m1} & L_{m2} & L_{m3} & \dots & L_{mn} \end{pmatrix} \end{matrix} \quad (2)$$

where  $L_{ij}$  = the variant which holds the "i" position as against the "j" criterion.

Then one can write the C matrix (3) of the following type:

$$C = \begin{matrix} & 1 & 2 & 3 & \dots & m \\ \begin{matrix} V_1 \\ V_2 \\ V_3 \\ \dots \\ V_m \end{matrix} & \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \dots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \dots & \alpha_{2m} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \dots & \alpha_{3m} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{m1} & \alpha_{m2} & \alpha_{m3} & \dots & \alpha_{mm} \end{pmatrix} \end{matrix} \quad (3)$$

where  $\alpha_{ij}$  = how many times the "i" variant holds the "j" position while  $i$  and  $j \in \{0, 1, 2, \dots, m\}$ .

One makes up the function  $f: V \rightarrow R_+$  defined by eq. (4):

$$f(V_i) = \alpha_{11} \frac{1}{2^1} + \alpha_{12} \frac{1}{2^2} + \dots + \alpha_{1m} \frac{1}{2^m} \quad (4)$$

The hierarchization of the variants is given by the decreasing values of this function.

## RESULTS AND DISCUSSIONS

To be able to hierarchize the 24 technological variants one needs to cover all the stages imposed by the Onicescu method, version I.

The criterion 1 is a maximum one because an increase of the thickness after fireproofing means a greater quantity of absorbed active substance and, implicitly, a greater efficiency for the fireproofing effect. The other two criteria, i. e.  $C_2$  and  $C_3$ , are minimum ones because a decrease both of the length of the carbonized area and a loss of mass after burning are equivalent with good fireproofing effects.

The accomplishing of the A consequences matrix is based on the data shown in Table 1.

Table 1. The values of the equi-important criteria

| Cod of var.     | Technological Conditions |                | Echiimportant Criteria    |                                |                             |
|-----------------|--------------------------|----------------|---------------------------|--------------------------------|-----------------------------|
|                 | Conc. active agent [g/l] | Cure temp [°C] | Crit.1 Increase thickness | Crit. 2 Length carbonized area | Crit.3 Mass loss after burn |
| V <sub>1</sub>  | 100                      | 150            | 0.75                      | 7.50                           | 0.170                       |
| V <sub>2</sub>  | 100                      | 160            | 0.95                      | 9.00                           | 0.124                       |
| V <sub>3</sub>  | 100                      | 170            | 1.25                      | 12.0                           | 0.356                       |
| V <sub>4</sub>  | 100                      | 180            | 1.55                      | 10.5                           | 0.221                       |
| V <sub>5</sub>  | 200                      | 150            | 1.20                      | 10.6                           | 0.243                       |
| V <sub>6</sub>  | 200                      | 160            | 1.30                      | 11.9                           | 0.255                       |
| V <sub>7</sub>  | 200                      | 170            | 2.15                      | 10.8                           | 0.237                       |
| V <sub>8</sub>  | 200                      | 180            | 1.80                      | 12.1                           | 0.310                       |
| V <sub>9</sub>  | 300                      | 150            | 1.31                      | 9.20                           | 0.136                       |
| V <sub>10</sub> | 300                      | 160            | 0.96                      | 11.5                           | 0.235                       |
| V <sub>11</sub> | 300                      | 170            | 1.75                      | 10.9                           | 0.247                       |
| V <sub>12</sub> | 300                      | 180            | 1.35                      | 10.4                           | 0.210                       |
| V <sub>13</sub> | 400                      | 150            | 2.05                      | 10.3                           | 0.105                       |
| V <sub>14</sub> | 400                      | 160            | 3.10                      | 5.70                           | 0.036                       |
| V <sub>15</sub> | 400                      | 170            | 2.30                      | 9.80                           | 0.064                       |
| V <sub>16</sub> | 400                      | 180            | 2.35                      | 3.00                           | 0.014                       |
| V <sub>17</sub> | 500                      | 150            | 2.70                      | 3.60                           | 0.010                       |
| V <sub>18</sub> | 500                      | 160            | 2.55                      | 4.60                           | 0.062                       |
| V <sub>19</sub> | 500                      | 170            | 1.19                      | 4.50                           | 0.035                       |
| V <sub>20</sub> | 500                      | 180            | 3.50                      | 3.40                           | 0.015                       |
| V <sub>21</sub> | 600                      | 150            | 2.20                      | 4.50                           | 0.139                       |
| V <sub>22</sub> | 600                      | 160            | 2.15                      | 4.70                           | 0.070                       |
| V <sub>23</sub> | 600                      | 170            | 2.55                      | 1.30                           | 0.127                       |
| V <sub>24</sub> | 600                      | 180            | 2.75                      | 2.00                           | 0.035                       |

A hierarchization of the 24 variants (according to Table 1) without employing the Onicescu method is difficult because the values obtained for those 3 criteria are not based on strict increasing or decreasing rules. According to the Onicescu method those 24 technological variants are arranged as against those 3 criteria under study, thus the B matrix being obtained.

$C_1$                        $C_2$                        $C_3$

$$B = \begin{pmatrix} V_1 & V_{20} & V_{23} & V_{17} \\ V_2 & V_{14} & V_{24} & V_{16} \\ V_3 & V_{24} & V_{16} & V_{20} \\ V_4 & V_{17} & V_{20} & V_{24} \\ V_5 & V_{23} & V_{17} & V_{19} \\ V_6 & V_{18} & V_{21} & V_{14} \\ V_7 & V_{16} & V_{19} & V_{18} \\ V_8 & V_{15} & V_{18} & V_{15} \\ V_9 & V_{21} & V_{22} & V_{22} \\ V_{10} & V_{22} & V_{14} & V_{13} \\ V_{11} & V_7 & V_1 & V_2 \\ V_{12} & V_{13} & V_2 & V_{23} \\ V_{13} & V_8 & V_9 & V_{21} \\ V_{14} & V_{11} & V_{15} & V_1 \\ V_{15} & V_4 & V_{13} & V_9 \\ V_{16} & V_{12} & V_{12} & V_{12} \\ V_{17} & V_6 & V_4 & V_4 \\ V_{18} & V_3 & V_5 & V_{10} \\ V_{19} & V_5 & V_7 & V_7 \\ V_{20} & V_{19} & V_{11} & V_5 \\ V_{21} & V_9 & V_{10} & V_{11} \\ V_{22} & V_{10} & V_6 & V_6 \\ V_{23} & V_2 & V_3 & V_8 \\ V_{24} & V_1 & V_8 & V_3 \end{pmatrix}$$

The C matrix is of the following type:

$$C = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \end{matrix} \\ \begin{matrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_6 \\ V_7 \\ V_8 \\ V_9 \\ V_{10} \\ V_{11} \\ V_{12} \\ V_{13} \\ V_{14} \\ V_{15} \\ V_{16} \\ V_{17} \\ V_{18} \\ V_{19} \\ V_{20} \\ V_{21} \\ V_{22} \\ V_{23} \\ V_{24} \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

Using the equation (4), the following values were obtained:  
 $f(V_1) = 1/2^{11} + 1/2^{14} + 1/2^{24} = 0.0005492$ ;  $f(V_2) = 1/2^{11} + 1/2^{12} + 1/2^{24} = 0.0007324$ ;  $f(V_3) = 0.0000039$ ;  $f(V_4) = 0.0000457$ ;  $f(V_5)$

=0.0000066;  $f(V_6)$  =0.0000080;  $f(V_7)$  =0.000492;  $f(V_8)$  =0.0001221;  $f(V_9)$  =0.0001529;  $f(V_{10})$  =0.0000044;  $f(V_{11})$  =0.0000623;  $f(V_{12})$  =0.0000457;  $f(V_{13})$  = 0.0012511;  $f(V_{14})$  =0.2666015;  $f(V_{15})$  =0.0078734;  $f(V_{16})$  =0.2617187;  $f(V_{17})$  =0.59375;  $f(V_{18})$  =0.0273437;  $f(V_{19})$  =0.0390634;  $f(V_{20})$  =0.6875;  $f(V_{21})$  =0.0177001;  $f(V_{22})$  =0.004882;  $f(V_{23})$  =0.5315;  $f(V_{24})$  =0.4375

By arranging the values in a decreasing order, the hierarchization of the studied working variants is obtained:

$V_{20}; V_{17}; V_{23}; V_{24}; V_{14}; V_{16}; V_{19}; V_{18}; V_{21}; V_{15}; V_{22}; V_{13}; V_2; V_{11}; V_7; V_9; V_8; V_{11}; V_{12} = V_4; V_6; V_5; V_{10}; V_3$

According to this mathematical hierarchization method we have found out that the most recommended technological fireproofing variant is  $V_{20}$  which is based on a concentration of 500 g/l fireproofing agent and on a 180°C cure temperature. If the equipment used for the cure process does not allow the increase of temperature up to 180°C, then one can choose either the  $V_{17}$  variant which implies a smaller temperature, i. e. 150°C, for the same working concentration or the  $V_{23}$  variant, situated on the 3-rd position, based on a concentration of 600 g/l fireproofing agent and on a 170°C cure temperature, with good results in what the fireproofing efficiency is concerned.

## CONCLUSIONS

By employing a mathematical method for the hierarchization of the technological variants for the fireproofing process of the textile materials, the following results have been obtained:

- the determination of the decreasing order of the studied technological variants as against those three criteria which determine the fireproofing efficiency. By carrying out the steps imposed by the Onicescu method, the hierarchy of the studied variants is strictly mathematically determined;
- the most efficient fireproofing variant from the view points of the thickness acquired after this treatment, of the length of the carbonized area and of the weight losses after burning, respectively, is the  $V_{20}$  variant. This variant is based on a concentration of 500 g/l fireproofing agent and on a 180°C cure temperature, and is considered to be the best, but only from all the other studied 24 variants;
- the ensurance of the objectivity of the method by using some criteria considered to be equi-important. This is the way one can eliminate the subjectivity related to the attribution of the importance levels granted to the criteria as one does for the other methods with the multi-attribute decisions and with the cardinal preferences of the criteria;

- the implementation of the Onicescu method in the field of the textile chemical finishing is easy and the results are relevant.

## REFERENCES

- Andraşiu M.et all. 1986. *Multi- attribute decisions methods (in Romanian)*, Technical Publishing House, Bucharest.
- Alnishaenskin, R., 1968. "Fireproofing possibilities on textiles " (in Francaise) . *Les cahiers* , Ciba, no. 4, 35-46.
- Grindea, M., Grigoriu A., Hanganu. A., Puscas. E.L., 1981. *Textile Chemical Technology (in Romanian)*, Technical Publishing House, Bucharest.
- Onicescu O., Stefanescu V., 1979. *Elements of informational statistics with application (in Romanian)*, Technical Publishing House, Bucharest
- Rancu, N., .Tovissi, L.,1963. *Statistical mathematic with application in production (in Romanian)*, Academy Publishing House, Bucharest.

## BIOGRAPHY

**POPESCU VASILICA** was born in Buzau, Romania and went to the Technical University "Gh. Asachi" from Iasi where she studied in textile chemical and finishing. After graduation, she worked for ten years in production and since 1998 is a lecturer Ph.D in Faculty of Textiles, Leather Engineering and Industrial Management, Department Textiles Chemical Finishing. Her interest area consists in the chemical finishing of textiles and statistical analysis.

**MANEA LILIANA-ROZEMARIE** was born in Bacau, Romania and went to the Technical University "Gh. Asachi" from Iasi where she studied in department of Technology and Textile Fabrics Design. After graduation, she worked for tow years in production and since 1991 and she is a assoc. prof. Ph.D in Faculty of Textiles, Leather Engineering and Industrial Management, department of Technology and Textile Fabrics Design. Her interest area consists in the chemical finishing of textiles and statistical analysis.

**POPESCU GABRIEL** was born in Adjud, Romania and went to the Technical University "Gh. Asachi" from Iasi where he studied in. After graduation, he worked for 2 years in production and since 1998 is a lecturer Ph.D in Machine Design & Mechatronics Department, from Mechanical Engineering Faculty. His interest area consists in programming on PC, mathematical methods, contact elasto-plastic.



# **AUTHOR LISTING**



## AUTHOR LISTING

|                              |             |                              |         |
|------------------------------|-------------|------------------------------|---------|
| Abelha A. ....               | 203         | Farima D. ....               | 116     |
| Aguirre F. ....              | 97          | Fernandes A.V. ....          | 325     |
| Aleisa E.E. ....             | 165         | Ferrante A. ....             | 214/219 |
| Alfonseca M. ....            | 103         | Franko S. ....               | 243     |
| Aloui A. ....                | 13          |                              |         |
| Al-Qallaf B. ....            | 293         | Glazier J.A. ....            | 103     |
| Alsabei R. ....              | 251         | Gomez-Padilla A. ....        | 141     |
| Al-Saedy H. ....             | 157         | Gonçalves J.F.B. ....        | 183     |
| Ammann P. ....               | 21          | Grondin-Perez B. ....        | 257     |
| Angeli C. ....               | 52          |                              |         |
| Aspergh G. ....              | 214         | Haklıdır M. ....             | 243     |
| Assumma V. ....              | 196         | Hennequin S. ....            | 31/97   |
| Auriol G. ....               | 111         | Ho C.-S. ....                | 280     |
|                              |             | Hosseinzadeh A. ....         | 152     |
| Bahrami M. ....              | 152         | Hosseinzadeh H. ....         | 152     |
| Baron C. ....                | 111         | Hsiao F.-B. ....             | 280     |
| Belo D. ....                 | 203         | Huertas Quintero L.A.M. .... | 79      |
| Benne M. ....                | 257         | Hughes R.W.C. ....           | 209     |
| Boström H. ....              | 65          | Hvala N. ....                | 318     |
| Brodsky A. ....              | 21          |                              |         |
|                              |             | Ion I. ....                  | 237     |
| Cabanellas Becerra J.M. .... | 227         | Iza F. ....                  | 315     |
| Camilleri L. ....            | 42          | Jayanta Singh Y. ....        | 157     |
| Cano-Moreno J.D. ....        | 227         | Jeffery D. ....              | 191     |
| Castañeda-Marroquín C. ....  | 103         | Jia Y. ....                  | 122     |
| Chabriat J.-P. ....          | 257         |                              |         |
| Chatzinikolaou A. ....       | 52          | Kasper R. ....               | 232     |
| Chirita C. ....              | 237         | Kaveh-Baghbaderani B. ....   | 309     |
| Conley W. ....               | 39          | Kißner H. ....               | 26      |
| Conway P.P. ....             | 79          | Köksal S. ....               | 243     |
| Culita J. ....               | 5           | Kong M. ....                 | 315     |
| Curteza A. ....              | 116         | Kukanja D. ....              | 318     |
|                              |             | Kulkarni A. ....             | 309     |
|                              |             | Kung C.-C. ....              | 280     |
|                              |             |                              |         |
| Da Frè M. ....               | 222         | Labajo Tirado C. ....        | 227     |
| Damour C. ....               | 257         | Lai M. ....                  | 277     |
| Das D.B. ....                | 293         | Lan C.E. ....                | 280     |
| de la Puente A.O. ....       | 103         | Lenk P. ....                 | 133     |
| Del Prete A. ....            | 86          | Lo Y.-P. ....                | 315     |
| Dias L. ....                 | 127/144     | Lupu L. ....                 | 116     |
| Doicin C. ....               | 237         |                              |         |
| Drikakis D. ....             | 277         | Machado J. ....              | 203     |
| Dudas C. ....                | 65          | Mahmood B.S. ....            | 262     |
| Dutu I. ....                 | 269         | Majed R.E. ....              | 262     |
|                              |             | Manea L.-R. ....             | 347/352 |
| Elia A. ....                 | 86          | Marin A. ....                | 237     |
| Esteves E.F. ....            | 173/183     | Masoumi A. ....              | 57      |
|                              |             | McDonald M. ....             | 191     |
| Facchin P. ....              | 214/219/222 | McLean C. ....               | 21      |

## AUTHOR LISTING

|                          |             |                           |         |
|--------------------------|-------------|---------------------------|---------|
| Mehrotra S.C. ....       | 157         | Swat M. ....              | 103     |
| Meng Q. ....             | 91          | Syberfeldt A. ....        | 178     |
| Mindán J.F. ....         | 227         |                           |         |
| Miranda M. ....          | 203         | Tackenberg S. ....        | 332     |
| Miteva T. ....           | 318         | Turki S. ....             | 31      |
| Mitroi M. ....           | 269         |                           |         |
| Monfared R. ....         | 79          | Vasiliu C. ....           | 269     |
|                          |             | Versteeg H. ....          | 287/301 |
| Nagy Z.K. ....           | 251         | Vicente H. ....           | 325     |
| Najibi A. ....           | 57          | Vik P. ....               | 127     |
| Nassehi V. ....          | 251/309     | Visonà Dalla Pozza L. ... | 214/222 |
| Neves J. ....            | 203/325     | Vitetta A. ....           | 196     |
| Ng A. ....               | 65          | Vlasenko D. ....          | 232     |
|                          |             |                           |         |
| Olatunji O. ....         | 293         | Wang Z. ....              | 26      |
| Oliveira E.C. ....       | 173/183     | Ward J. ....              | 315     |
| Ourbih ex Tari M. ....   | 13          | West A.A. ....            | 79      |
|                          |             | Wischnewski R. ....       | 133     |
| Pereira G. ....          | 127/144     |                           |         |
| Perera T. ....           | 122/209     | Yang M.-H. ....           | 280     |
| Persson L. ....          | 178         | Yang S.-H. ....           | 91      |
| Popescu G. ....          | 347/352     | Yuqing F. ....            | 47      |
| Popescu T.C. ....        | 269         |                           |         |
| Popescu V. ....          | 347/352     | Zheng P. ....             | 191     |
|                          |             |                           |         |
| Qaddoum K. ....          | 70          |                           |         |
|                          |             |                           |         |
| Raposo R. ....           | 144         |                           |         |
| Razali S. ....           | 91          |                           |         |
| Rezg N. ....             | 97          |                           |         |
| Rizzato E. ....          | 219/222     |                           |         |
| Romanin Jacur G. ....    | 214/219/222 |                           |         |
| Romano G. ....           | 86          |                           |         |
| Rossetti R.J.F. ....     | 173/183     |                           |         |
| Roßmann J. ....          | 133         |                           |         |
|                          |             |                           |         |
| Saleem W. ....           | 47          |                           |         |
| Salmaso L. ....          | 219/222     |                           |         |
| Sauer N. ....            | 31          |                           |         |
| Savsar M. ....           | 165         |                           |         |
| Schlick C.M. ....        | 332         |                           |         |
| Schneider S. ....        | 332         |                           |         |
| Seager R. ....           | 315         |                           |         |
| Segura Velandia D.M. ... | 79          |                           |         |
| Shaik A.Q. ....          | 287/301     |                           |         |
| Shao G. ....             | 21          |                           |         |
| Siems M. ....            | 26          |                           |         |
| Stefanoiu D. ....        | 5           |                           |         |
| Stoica O. ....           | 116         |                           |         |