

**12<sup>th</sup> INTERNATIONAL INDUSTRIAL  
SIMULATION CONFERENCE  
2014**

**ISC'2014**

**EDITED BY**

**Amos Ng**

**and**

**Anna Syberfeldt**

**JUNE 11-13, 2014**

**SKÖVDE, SWEDEN**

**A Publication of EUROSIS-ETI**





12<sup>th</sup> Industrial Simulation Conference 2014

SKÖVDE, SWEDEN

JUNE 11-13, 2014

Organised by

ETI- The European Technology Institute

Sponsored by

EUROSIS - The European Simulation Society

In Cooperation  
With

University of Skövde

Co-Sponsored by

Ghent University

Hosted by

University of Skövde

Skövde, Sweden

## EXECUTIVE EDITOR

**PHILIPPE GERIL  
(BELGIUM)**

## EDITORS

### **General Conference Chairs**

Amos Ng, University of Skovde, Skovde, Sweden  
Leif Pehrsson, University of Skovde, Skovde, Sweden  
Anna Syberfeldt, University of Skovde, Skovde, Sweden

### **Past Conference Chairs**

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

### **Publications Chairs**

Peter Lawrence, Swinburne University, Australia  
Yan Luo, NIST, Gaithersburg, USA

## INTERNATIONAL PROGRAMME COMMITTEE

### **Discrete Simulation Methodology, Languages and Tools**

Jonathan Berglund, Chalmers University of Technology, Gothenburg, Sweden  
Helge Hagenauer, Universitaet Salzburg, Salzburg, Austria  
Sophie Hennequin, ENIM, Metz Cedex, France  
Bjorn Johansson, Chalmers University of Technology, Gothenburg, Sweden  
Sarka Kvetonova, Brno University of Technology, Brno, Czech Republic  
Erik Lindskog, Chalmers University of Technology, Gothenburg, Sweden  
Goreti Marreiros, Polytechnic of Porto, Porto, Portugal  
Stefano Marrone, Seconda Universita di Napoli, Naples, Italy  
Jiri Safarik, University of West Bohemia, Plzen, Czech Republic

### **Ambient Intelligence and Simulation**

Selwyn Piramuthu, University of Florida, Gainesville, USA

### **Simulation in Manufacturing**

Pascal Berruet, Universite Bretagne Sud, Lorient, France  
Peter Byrne, Dublin City University, Dublin, Ireland  
Ana Camacho, UNED, Madrid, Spain  
Luis Camarinha-Matos, Universidade Nova de Lisboa, Lisbon, Portugal  
Eduardo Castellano, IKERLAN Technol. Res. Centre, Mondragon-Arrasate, Spain  
Remy Dupas, Universite Bordeaux 1, Bordeaux, France  
Frank Heinze, (RIF) e.V., Dortmund, Germany  
Pouria Homayonifar, Acona Flow Technology, Skien, Norway  
Carlo Meloni, Politecnico di Bari - DEE, Bari, Italy  
Pascal Meyer, Forschungszentrum Karlsruhe, Karlsruhe, Germany  
José A V Oliveira, University of Minho, Braga, Portugal  
Marina Valles, UPV, Valencia, Spain  
Roland Wischnewski, (RIF) e.V., Dortmund, Germany  
Farouk Yalaoui, Universite de Technologie de Troyes, Troyes Cedex, France

### **Simulation in Steel Manufacturing**

Brian Hollocks, Bournemouth University, Bournemouth, United Kingdom

### **Simulation in Automotive Systems**

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

## INTERNATIONAL PROGRAMME COMMITTEE

### **Simulation in Robotics**

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  
Martin Mellado, UPV, Valencia, Spain  
Bogdan Raducanu, Computer Vision Centre, UAB, Barcelona, Spain  
Krzysztof Skrzypczyk, Silesian University of Technology, Gliwice, Poland  
Sergiu-Dan Stan, Technical University of Cluj-Napoca, Cluj-Napoca, Romania

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

Teresa Alvarez, University of Valladolid, Valladolid Spain  
Christos Bouras, University of Patras, Patras, Greece  
Silvia Mirri, University of Bologna, Bologna, Italy  
Maurizio Palesi, Università di Catania, Catania, Italy  
Fernando Boronat Seguí, UPV, Valencia, Spain  
Renate Sitte, Griffith University, Gold Coast, Australia

### **Simulation in Electronics Manufacturing**

Theresa Roeder, San Francisco State University, USA  
Gerald Weigert, Dresden University of Technology, Germany

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

Orhan Gemikonakli, Middlesex University, London, United Kingdom

### **Simulation in Computer Science**

Lipika Deka, Tezpur University and IIT Guwahati, India  
Ernst Kessler, NLR, Amsterdam, The Netherlands  
Wolfgang Kreutzer, University of Canterbury, Christchurch, New Zealand

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

El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium  
Christian Almeder, European University Viadrina, Frankfurt (Oder), Germany  
Maria Sameiro Carvalho, University of Minho, Guimaraes, Portugal  
Isabel Garcia Gutierrez, Univ. Carlos III de Madrid, Madrid, Spain  
Martin Grunow, TU Munich, Munich, Germany  
Thomas Hanne, Univ. of Applied Sciences Northwestern Switzerland, Olten, Switzerland  
Dmitry Ivanov, Chemnitz University of Technology, Chemnitz, Germany  
Peter Lawrence, Swinburne University, Lilydale, Australia  
Marie-Ange Manier, UTBM, Belfort, France  
Herve Manier, UTBM, Belfort, France  
Roberto Montemanni, IDSIA, Manno-Lugano, Switzerland  
Jaap Ottjes, TU Delft, Delft, The Netherlands  
Guilherme A B Pereira, University of Minho, Braga, Portugal  
Cyril Ray, Ecole Navale, Brest, France  
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

### **Simulation in the aviation sector**

Gabriel Lodewijks, Delft University of Technology, Delft, The Netherlands

### **Hospital Logistics Simulation**

**Track Chair:** Giorgio Romanin-Jacur, University of Padova, Vicenza, Italy  
Antonio Abelha, Universidade do Minho, Braga, Portugal  
Joel Colloc, Université du Havre, Le Havre, France  
Jose Machado, University of Minho, Braga, Portugal  
Peter Summons, University of Newcastle, Australia

## INTERNATIONAL PROGRAMME COMMITTEE

### **Complex Systems Modelling**

**Track Chair:** Igor N Litvine, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa  
Frantisek Capkovic, Slovak Academy of Sciences, Bratislava, Slovak Republic  
Alexandre Nketsa, LAAS, Toulouse, France  
Miguel Rocha, University do Minho, Braga, Portugal  
Alfonso Urquia, UNED, Madrid, Spain

### **Simulation in Aerospace**

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

### **Marine Simulation**

Sergeij Kalashnikow, DANFOSS, Austria

### **Simulation in Industrial Design and Product Design**

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

### **Simulation in Engineering Processes**

Chrissanti Angeli, Technological Institute of Piraeus, Athens, Greece  
Alejandra Gomez Padilla, University of Guadalajara, Mexico  
Jan Studzinski, Polish Academy of Sciences, Warsaw, Poland  
Joao Tavares, University of Porto, Porto, Portugal  
Henk Versteeg, Loughborough University, Loughborough, United Kingdom

### **Civil and Building Engineering**

Alistair Borthwick, Oxford University, Oxford, United Kingdom  
Graham Saunders, Loughborough University, Loughborough, United Kingdom

### **Simulation in Energy and Power Systems**

Sergeij Kalashnikow, DANFOSS, Austria  
Janos-Sebestyen Janosy, KFKI Atomic Energy Research Institute, Budapest, Hungary

### **Simulation in Multibody Systems**

**Track Chair:** Bernd Schaefer, DLR, Wessling, Germany  
Ignacio Garcia-Fernandez, University of Valencia, Valencia, Spain  
Jose Diaz Lopez, Anoto AB, Lund, Sweden

### **Simulation in Chemical, Petroleum and Mining Engineering**

Diganta Das, Loughborough University, Loughborough, United Kingdom  
Mohamad R. Riazi, Kuwait University, Kuwait

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

Roberto de Beauclair Seixas, IMPA, Rio de Janeiro, Brazil  
Carlos Palau, UPV, Valencia, Spain  
Matthias Reuter, CUTEC GmbH, TU-Clausthal, Clausthal, Germany

### **Virtual Reality and Graphical Simulation in Industrial Simulation**

**Track Chair:** Guodong Shao, NIST, Gaithersburg, USA  
Emilio Camahort, Universidad Politecnica de Valencia, Valencia, Spain  
Anders Hast, University of Gavle, Gavle, Sweden  
Fabrizio Lamberti, Politecnico di Torino, Turin, Italy  
Christoph Laroque, University of Paderborn, Paderborn, Germany  
Sudhir Mudur, Concordia University, Montreal, Canada  
Marta Pla-Castells, Universidad de Valencia, Valencia, Spain  
Marcos A. Rodrigues, Sheffield Hallam University, Sheffield, United Kingdom

## INTERNATIONAL PROGRAMME COMMITTEE

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

Peter Kvasnica, Alexander Dubcek University of Trencin, Trencin, Slovakia

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

Roberto Revetria, University of Genoa, Genoa, Italy

Agustin Yague, Technical University of Madrid, Madrid, Spain

### **Simulation and Training**

Manuel Alfonso, Universidad Autonoma de Madrid, Madrid, Spain

Wenji Mao, Chinese Academy of Sciences, Beijing, P.R. China

Eshan Rajabally, Loughborough University, Loughborough, United Kingdom

Gerhard Schreck, Fraunhofer IPK, Berlin, Germany

### **Workshops**

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

**Track Chair:** Anna Syberfeldt, University of Skovde, Skovde, Sweden

Paul Davidsson, Blekinge Institute of Technology, Ronneby, Sweden

Petr Hanacek, Brno University of Technology, Brno, Czech Republic

Jairo Montoya Torres, Universidad de la Sabana, Chia, Columbia

### **NANOSIM**

Clemens Heitzinger, Cambridge University, Cambridge, United Kingdom

Yong K. Kim, University of Massachusetts Dartmouth, Dartmouth, USA

Javier Marin, ETSI, University of Malaga, Malaga, Spain

### **Workshop Augmented Reality and Pervasive Systems in Simulation**

Alessandro Genco, University of Palermo, Palermo, Italy

### **Workshop on Simulation in Lean Manufacturing**

**Track Chair:** Leif Pehrsson, Volvo Car Corporation, Skovde, Sweden

El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium

Hendrik Van Landeghem, Ghent University, Ghent, Belgium

### **Simulation-based Optimization in Industry**

Amos H.C. Ng, University of Skovde, Skovde, Sweden

Anna Syberfeldt, University of Skovde, Skovde, Sweden

### **OPENSIM**

Philippe Geril, ETI Bvba, Ostend, Belgium

© 2014 EUROSIS-ETI

Responsibility for the accuracy of all statements in each peer-referenced paper rests solely with the author(s). Statements are not necessarily representative of nor endorsed by the European Simulation Society. Permission is granted to photocopy portions of the publication for personal use and for the use of students providing credit is given to the conference and publication. Permission does not extend to other types of reproduction, nor to copying for incorporation into commercial advertising nor for any other profit-making purpose. Other publications are encouraged to include 300- to 500-word abstracts or excerpts from any paper contained in this book, provided credits are given to the author and the conference.

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

All articles published in these Proceedings have been peer reviewed

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

Selected papers of this conference are published in scientific journals.

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

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

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

EUROSIS-ETI Publication  
**ISBN: 978-90-77381-83-0**  
**EAN: 978-90-77381-83-0**

## PREFACE

Dear participants,

Welcome to Sweden and the 12th annual Industrial Simulation Conference (ISC), hosted by the School of Engineering Science at the University of Skövde, 11-13 June 2014!

Over the years, ISC has become an annual forum for gathering international researchers and practitioners together to exchange research outcomes, results and experiences as well as to discuss new challenges and ideas within a very wide spectrum of simulation-related research and industrial applications. ISC'2014 is no exception. Apart from many applications in manufacturing as usual, this year we will have presentations related to petroleum refineries, energy systems, water supply networks, customer services and personnel scheduling, underwater robotics and even "outer space croquet"!

Skövde is a small Swedish city in terms of population, but is known as one of the most important industrial cities in the country. Both Volvo Trucks and Volvo Cars operate their engine manufacturing and assembly facilities inside the city. The companies together, produce over half a million engines per year. Inside the plant of Volvo Trucks in Skövde, is where the largest foundry in Nordic Europe located. On the other side, within the engine plant of Volvo Cars, you find one of the world's most advanced automated guided vehicle systems. We greatly acknowledge the sponsorship of Volvo Cars to ISC'2014 by offering a factory visit on the second day of the conference.

With the integration of artificial intelligence and other computational intelligence techniques, simulation has become a highly effective and practical optimization and decision-support tool in industry. Therefore, it is our great pleasure to have one of the world-renowned experts in the area of evolutionary simulation optimization, Professor Juergen Branke, to provide a keynote speech related to this exciting research area. On the other hand, Sweden has some distinguished research communities in the areas of modelling, simulation, virtual manufacturing and their industrial applications. Within this opportunity we have the pleasure to invite two of the representative research leaders in Sweden, namely, Professor Rikard Söderberg and Professor Peter Fritzson, to share their rich expertise and application experiences in two different domains. We gratefully appreciate the participations of these three keynote speakers!

Last but not least, we would like to express our special thanks to Philippe Geril, Secretary General of EUROSIS, who has taken the responsibility for most of the organizational matters. Many thanks for his hard work and never-ending enthusiasm so that the ISC'2014 is possible.

Once again, we wish you warmly welcome and hope that you will enjoy the conference and your stay in Skövde/Sweden!

Conference Chairs  
Amos Ng and Anna Syberfeldt





<b>Preface .....</b>	<b>IX</b>
<b>Scientific Programme .....</b>	<b>1</b>
<b>Author Listing .....</b>	<b>111</b>

## **SIMULATION TOOLS**

### **Comparison of SIMIO and ARENA Simulation Tools**

António Vieira, Luís Dias, Guilherme Pereira and José Oliveira .....5

### **Initial Bubble and Dew Point Temperature Estimations for Flash Calculations of Hydrocarbon mixtures**

Serdar Çelebi, Gizem Kuşoğlu, Ümmühan Canan, Mehmet Ş. Kaya, Şerife Baş Hekimoğlu and Mustafa Bakır.....14

### **Buffer Overflow Simulation in Self-Similar Queuing Networks with Finite Buffer Capacity**

Izabella Lokshina and Evan Schiele .....17

### **A Virtual Testbed for Underwater Robotics – Application to Control Design for AUVS**

Linus Atorf, Martin Krehel, Juergen Rossmann and Gaurav S. Sukhatme .....24

## **GENETIC ALGORITHMS**

### **Some Algorithms supporting the Water Network Management by use of Simulation of Network Hydraulic Model**

Jan Studzinski and Marek Kurowski .....33

### **Optimization of Water Supply Network Rehabilitation using Genetic Algorithms**

Bartłomiej Fajdek, Marcin Stachura and Jan Studziński .....38

## **BUSINESS SIMULATION**

### **The Business of Outer Space Croquet**

William Conley .....45

### **Interactive Simulation for Discovering Investment Scenarios in Energy Systems**

N.B. Szirbik, G.B. Huitema, R.H. van der Burg and J.C. Wortmann .....53

## CONTENTS

<b>Modelling customer preference heterogeneity to iPad attributes using a Finite Mixture procedure</b>	
Liberato Camilleri and Keith Lia .....	56

<b>Evolutionary Simulation Optimization of Personnel Scheduling</b>	
Martin Andersson, Anna Syberfeldt, Amos Ng and Victor Bengtsson .....	61

## DECISION SUPPORT SYSTEMS

<b>Interactive and Intelligent Decision Support in Manufacturing using Simulation Based Innovization and Cloud Computing</b>	
Ingemar Karlsson, Amos Ng and Anna Syberfeldt .....	69

<b>Framework for an Adaptive decision Support System for Industrial Shop Floor Operators</b>	
Magnus Holm, Göran Adamson, Lihui Wang and Philip Moore .....	75

## MANUFACTURING OPTIMIZATION AND EVALUATION

<b>Aggregated Discrete Event Modelling for Simulation and Optimisation of Manufacturing Systems</b>	
Leif Pehrsson, Simon Lidberg, Marcus Frantzén, Tehseen Aslam and Amos Ng .....	83

<b>A Case Study of Evolutionary Simulation Based Optimization in Aircraft Engine Manufacturing</b>	
Anna Syberfeldt, Patrik Gustavsson, Joakim Svantesson and Torgny Almgren .....	91

<b>Practical Production Systems Optimization using Multiple-Choice Sets and Manhattan Distance Based Constraints Handling</b>	
Jacob Bernedixen and Amos H.C. Ng .....	97

<b>Study of the Impact of Production Quantity Loaded to a Production Line on Apparel Production System Performance using Simulation</b>	
Lottie L.C. Mak, W.K. Wong and S.Y.S. Leung .....	104

# **SCIENTIFIC PROGRAMME**



# **SIMULATION TOOLS**



# COMPARISON OF SIMIO AND ARENA SIMULATION TOOLS

António Vieira<sup>(a)</sup>, Luís Dias<sup>(b)</sup>, Guilherme Pereira<sup>(b)</sup>, José Oliveira<sup>(b)</sup>

<sup>(a) (b)</sup> University of Minho, Campus Gualtar, 4710-057, Braga, Portugal.

<sup>(a)</sup> antonio6vieira@gmail.com, <sup>(b)</sup> luis.dias / guilherme.pereira / jose.oliveira @algoritmi.uminho.pt

**KEYWORDS:** Simulation, Arena, Simio

## ABSTRACT

The purpose of this paper is to establish a comparison between Simio and Arena, helping a vast community of simulation practitioners to gain access to advanced modelling capabilities to address complex problems. Several aspects were compared, such as: concept of simulation models, animation development, modelling philosophies, Simio libraries and Arena templates, concept of entities, interface of the tools and Simio objects versus Arena blocks. The comparison was consolidated through the analysis of two case studies where the authors aimed to emphasize the way each simulation tool addresses some important issues related to model construction. The several compared aspects indicate the many advantages of using the more recent tool. Thus, this object-oriented tool appears to have all the conditions to trigger a widespread paradigm shift in the way practitioners build models.

## 1. INTRODUCTION

Simulation modelling is being widely used for performance improvement of many systems (Hlupic, 2000, Hlupic and Paul, 1999). Consequently, the number of simulation tools available is also increasing and tools comparison becomes a required task. Yet, “most of scientific works related to tools comparison/reviews analyse only a small set of tools and usually evaluating several parameters separately avoiding to make a final judgement due to the subjective nature of such task” (Dias et al., 2007). One of such works was developed by Hlupic and Paul (1999). The authors compared a set of simulation tools, distinguishing between users of software for educational purpose and users in industry. In his turn, Hlupic (2000) developed “a survey of academic and industrial users on the use of simulation software, which was carried out in order to discover how the users are satisfied with the simulation software they use and how this software could be further improved”. Dias and Pereira et al. (2007, 2011) compared a set of tools based on popularity on the internet, scientific publications, WSC (Winter Simulation Conference), social networks and other sources. “Popularity should never be used alone otherwise new tools, better than existing ones would never get market place” (Dias et al., 2007). However, a positive correlation may exist between popularity and quality, since the best tools have a greater chance of being more popular. The author’s final classification indicated that the most popular tool

at the date was Arena. The only new tool on the ranking was Simio which obtained a good classification, meaning that this tool has good odds of becoming more popular and widely used in the future.

The development of Simio and Arena simulation tools was leaded by the same author: Dennis Pegden. Thus, it is normal that there are some resemblances between them. “However care is required to ensure that your knowledge of Arena does not become a handicap that prevents exploiting the full power of Simio” (Pegden, 2013a). Dennis Pegden (2007) exposed a comparison between the two tools with the purpose to “help experienced Arena users transition from Arena to the new Simio” (Pegden, 2013a). His white paper focused mainly on exposing some differences in the concepts of the two tools regarding: models, entities and resources concepts, animation development and modelling philosophies. Yet, some practical differences from the two tools were not addressed. To that end, this paper intends to compare both tools, taking into account several relevant factors but also enlightening the learning of Simio.

Chapter 2 presents a review over the analysed literature. In chapter 3, a comparison over the two simulation tools is made, regarding: general concepts, interface of the tools and Simio objects and Arena blocks. Lastly, two case studies are presented and analysed in chapter 4. In the final chapter, the main conclusions of the conducted work are drawn.

## 2. LITERATURE REVIEW

### 2.1. Arena

In 1993 SIMAN and CINEMA (simulation languages) were combined into a single tool: Arena (<http://www.erlang.com.br/arena.asp>). This tool is a simulation environment consisting of module templates, built around SIMAN language constructs, as well as other facilities and the CINEMA animation package (Altiok and Melamed, 2010). Thus, when an Arena model is created it is implemented in SIMAN code which is then compiled and run without any need to write programming code. SIMAN consists of blocks and elements. Blocks are basic logic constructs that represent operations (e.g. seize block). Elements are objects that represent facilities such as resources, queues and tallies (Altiok and Melamed, 2010). In 1995 the first version of Arena for Windows 95 was released. It was the first to run in 32-bit systems. From 2000 on, after being acquired by Rockwell, the software received a huge development boost and new versions, in

increasingly shorter time periods, were launched. Nowadays, the software is considered the most popular simulation tool in the world (Dias et al., 2007, Pereira et al., 2011). Since Arena already has many years of existence, a lot of published documents is available (Altioik and Melamed, 2010, Kelton et al., 2002). However, the same does not apply to Simio.

## 2.2. Simio

Simio, developed in 2007 (Vik et al., 2010), is based on intelligent objects (Sturrock and Pegden, 2010, Pegden, 2007, Pegden and Sturrock, 2011). These “are built by modellers and then may be used in multiple modelling projects. Objects can be stored in libraries and easily shared” (Pegden, 2013b). Unlike other object-oriented systems, in Simio there is no need to write any programming code, since the process of creating a new object is completely graphic (Pegden and Sturrock, 2011, Pegden, 2007, Sturrock and Pegden, 2010). The activity of building an object in Simio is identical to the activity of building a model. In fact there is no difference between an object and a model (Pegden, 2007, Pegden and Sturrock, 2011). A vehicle, a customer or any other agent of a system are examples of possible objects and, combining several of these, one can represent the components of the system in analysis. Thus, a Simio model “looks” like the real system (Pegden and Sturrock, 2011, Pegden, 2007). This fact can be very useful, particularly while presenting the results to someone non-familiar to the concepts of simulation.

In Simio, the model logic and animation are built in a single step (Pegden and Sturrock, 2011, Pegden, 2007). This feature is very important, because it makes the modelling process very intuitive (Pegden and Sturrock, 2011). Moreover, the animation can also be useful to reflect the changing state of the object (Pegden, 2007). In addition to the usual 2D animation, Simio also supports 3D animation as a natural part of the modelling process (Sturrock and Pegden, 2010). To switch between 2D and 3D views the user only needs to press the 2 and 3 keys of the keyboard (Sturrock and Pegden, 2010). Moreover, Simio provides a direct link to Google Warehouse, a library of graphic symbols for animating 3D objects (Sturrock and Pegden, 2010, Pegden and Sturrock, 2011).

Simio offers two basic modes for executing models: the interactive and the experimental modes. In the first it is possible to watch the animated model execute, which is useful for building and validating the model. In the second, it is possible to define one or more properties of the model that can be changed, in order to see the impact on the system performance (Sturrock and Pegden, 2010).

According to Pegden (2007): although Simio incorporates a number of innovative features in pursuit of this goal, “only time will tell if this tool has bridged the many practical issues that must be addressed to trigger a widespread paradigm shift in the way practitioners build models” (Pegden, 2007).

Currently there are not many studies that use Simio for modelling systems. Even so, it is possible to find some studies that used this tool for other types of problems. Akhtar et al. (2011) studied the role of consanguineous marriages in causing congenital defects. Li and Wang (2011) developed a micro simulation model to evaluate the performance and service level of a ticket office. Vik et al. (2010) used Simio to model a logistic system design of a cement plant. Brown and Sturrock (2009) also used this tool to improve a set of production processes. Lastly, Kai et al. (2011) used Simio to explore simulation of casualty treatment in wartime.

## 3. COMPARISON OF THE TOOLS

This chapter first concentrates on different views of the conceptual philosophy of both tools when developing a simulation model. Thereafter, it explains different approaches of both tools as far as the interface with the user is concerned. Lastly, this chapter shows how the behaviour of Simio objects could be addressed with Arena blocks - and this corresponds to the implementation issue, where the authors aim to analyse the practical aspects of building simulation models.

### 3.1. General Concepts Comparison

- **Simulation model concept:** In Arena when a user refers to his “model” he is referring to the Arena simulation model. Yet, in Simio, a model is simply an object that can be instantiated in other models.

- **Animation development:** In Arena a user animates his model as a 2-step process: first he draws process flows for the model, and then, in a separate area of the same drawing space, he adds levels, animated routes and others, that are linked back to the process flow (Pegden, 2013a). On the contrary, in Simio, the user drags objects to a drawing space. Since they represent the physical components of the system (Pegden and Sturrock, 2011, Pegden, 2007) the objects define both the logic and the animation of the model. Thus, modelling and animation are done as a single step (Pegden, 2013a). Moreover, Simio provides a direct link to Google Warehouse, a library of graphic symbols for animating 3D objects (Sturrock and Pegden, 2010, Pegden and Sturrock, 2011).

- **Modelling philosophies:** In Arena, a model is built by using the process orientation paradigm. In this philosophy, the user defines elements that hold the state of the system, and build process flows using blocks that perform actions on the elements. These blocks are passive and are only activated by the arrival of an entity (Pegden, 2013a). On the other hand, Simio is a multi-paradigm modelling tool, in the sense that it supports both object orientation and process orientation. In fact, the ability to mix object-based and process modelling within the same model is one of the unique and very powerful features of Simio (Pegden, 2013a). This way, users can use the faster capacities of the object paradigm and the more flexible capacities of the process orientation (Pegden, 2013a). Thus, a user builds object-



based models by thinking in terms of the physical objects in the system (machines, conveyors, etc.). These objects are placed in the Facility and interact with each other based on their internal logic. The process orientation is used in the Process panel in which the user constructs Arena-like process flows (Pegden, 2013a). These processes are used to either customize the behaviour of an existent object, or to create new object definitions (Pegden, 2013a). However, there are some differences between the terminology of the processes in Simio and in Arena, in the sense that in the first a process is comprised of steps, elements, and tokens that flow through a process executing steps that alter the state of one or more elements. Hence, steps are like Arena blocks (Pegden, 2013a).

- **Libraries versus templates:** An Arena template is a set of hierarchical blocks (modules) that can be placed in process logic. In contrast, a Simio library is a collection of object definitions for placing objects in the facility. This defines a new library that can be used by other models, and so on. Thus, Simio libraries and Arena template panels share the basic notion of user-defined library, but they differ in the modelling orientation that they are designed to support (Pegden, 2013a).

- **Entities concept:** Entities in Arena are part of a model of the system and their only purpose is to carry information (attributes) and to execute a process. In fact every entity in an Arena model must have exactly the same attributes (Pegden, 2013a). In Simio, entities have object definitions, thus have their own intelligent behaviour and can make decisions, such as reject requests, decide to take a rest, etc. Moreover, each entity has a token that corresponds to it and executes a process. Thus, an entity in Arena corresponds to a token of Simio (Pegden, 2013a).

## 3.2. Interface

### 3.2.1. Arena

Arena possesses 2 other tools incorporated: Input analyser and Output analyser. While the first fits a distribution to a sample data, the latter is a tool for analysing the data resulted from a simulation process.

There are 3 main regions that can be identified in the main Arena window:

#### 1) Project bar

Located on the left side, it contains several templates that can be attached or detached to the Project bar. Templates are a set of blocks pre-defined or user-defined, i.e., a collection of modelling tools. From those, basic process, advanced transfer and advanced process stand out. More information on Arena templates can be found at (Vieira, 2013).

To build a simulation model with Arena a user needs to use modules from the above mentioned templates. There are 2 types of modules: Flowchart blocks and data modules (Garrido, 2009). The user can place blocks on the model window and connect them to form a flowchart that describes the system he is

modelling. Data modules are data in spreadsheet-like format that enables the user to edit some information.

#### 2) Model window flowchart view

This region is located on the right side of Arena. It is the workspace for the simulation model and will contain all the model graphics, flowcharts, animation, and other drawings.

#### 3) Model window spread sheet view

Located on the right-hand side and below the flowchart view, it shows the model data and some details of the blocks being used/selected.

### 3.2.2. Simio

In Simio there are 3 areas that are always visible and can be seen in Figure 1: the ribbons, the browse panel and the tabbed panel views. These areas are described in more detail in (Vieira, 2013).

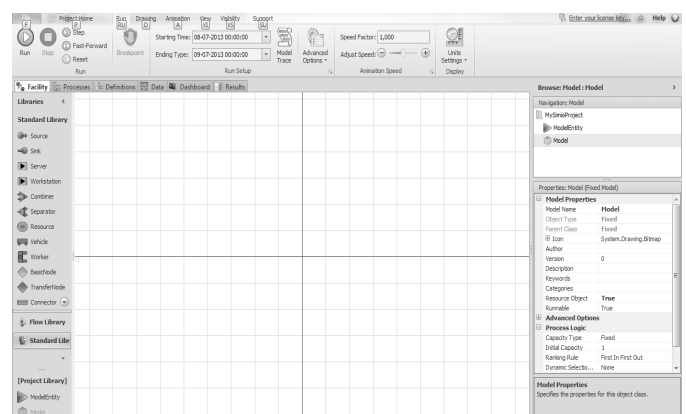


Figure 1: General view of Facility

The tabbed areas are divided in: Facility, Process, Definitions, Data and Results. More information on the panels can be found at (Vieira, 2013).

### 3.3. Objects versus Blocks

This comparison will be made by trying to model the same behaviour of Simio objects of the Standard library, resorting to some Arena blocks. The objects belonging to the Standard library are:

**Source:** This object is responsible for creating entities. Figure 2 illustrates a Source object.

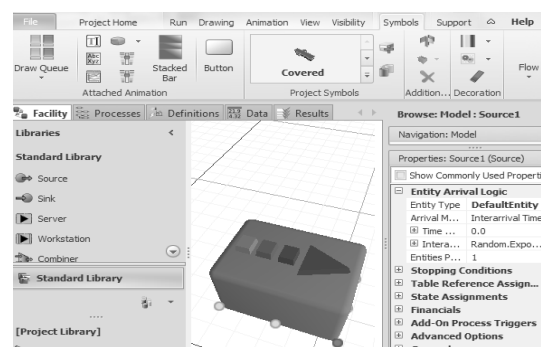


Figure 2: Source object

This object is comprised by an output node, an output buffer queue and the object itself. It is possible to establish a comparison between this object and the Create block of Arena, since both can define the entity type, Interarrival times, number of entities per arrival,

maximum number of arrivals and the time offset until the first arrival. However, in Simio it is possible to edit many more properties like assign values to states, assign add-on processes, make table references and state assignments, change animation of created entities and others. Figure 3 represents a Create block of Arena.

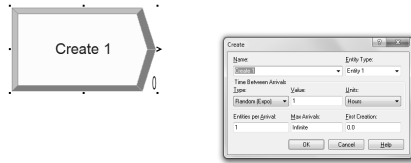


Figure 3: Create block

**Sink:** Figure 4 illustrates a Sink object. This object is responsible for eliminating entities from the system.

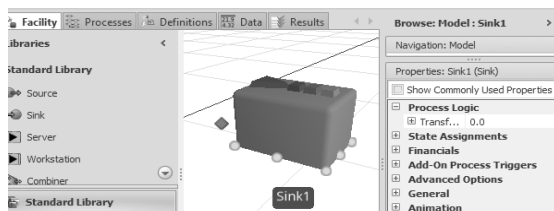


Figure 4: Sink object

This object is comprised by an input node, an input buffer queue and the object itself. The Dispose block of Arena has the same goal of this object. The Simio object also allows the definition of add-on processes, state assignments and others. Figure 5 displays a Dispose block.



Figure 5: Dispose block

**Server:** “The Server object can be used to model a single server or a single processing center with multiple identical servers, depending upon the capacity specified for the processing station” (www.simio.com). Figure 6 shows a Server object which is comprised by an input node, an input buffer queue, a processing station, an output buffer queue, an output node and the object itself.

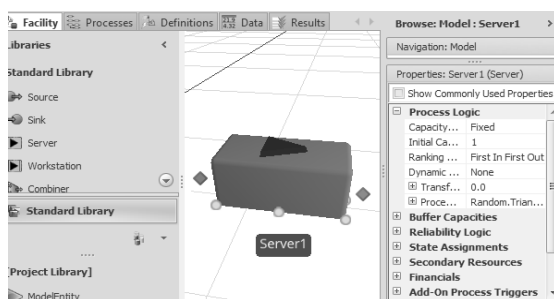


Figure 6: Server object

The Process block of Arena has the same objective of this object, since both model processes with a determined processing time. In a Process block, the user has to specify the type of process (e.g. seize, seize delay release), the allocated resource, the processing time and needs to assign the resource type. In a Server, this is

done in a more natural way, since the user does not need to assign a specific type of resource to be seized. Despite having the same goal, the Simio object allows a user to edit a superior set of properties like secondary resources, failures, state assignments, add-on processes, and others. Figure 7 illustrates a Process block.

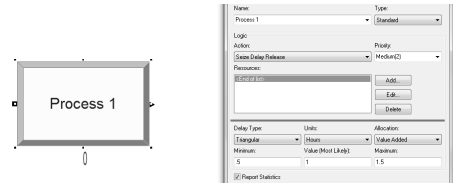


Figure 7: Process block

**Workstation:** The most complex object of the Standard Library. It is similar to a Server except that it models the processing station in far more detail, since the latter is represented by an operation divided into 3 activities: setup, processing, and teardown. All the entities moving through the Workstation will perform each of these activities. Figure 8 illustrates a Workstation object. It is not possible to establish a comparison between this object and a single block of Arena. In order to model this object, a great number of blocks would have to be used.

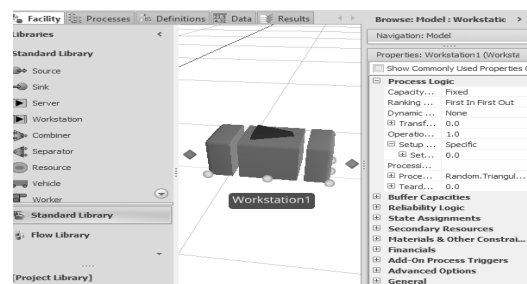


Figure 8: Workstation object

**Combiner:** Matches multiple entities, groups them into a batch, and then attaches the batched members to a parent entity. Figure 9 illustrates a Combiner.

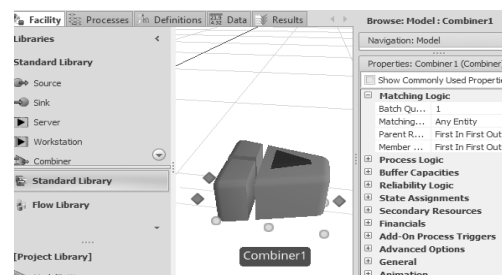


Figure 9: Combiner object

This object is comprised by a parent input node, a member input node, a parent input buffer queue, a member input buffer queue, a processing queue, an output buffer queue, an output node and the object itself. In Arena it is possible to model this behaviour by using the blocks Match to synchronize the entities and then, the Batch block to attach those entities together. Nevertheless, in Simio there is the possibility to visualize the member entities of a parent entity as they travel through the model. Figure 10 illustrates the

addressed situation. In this image a tray (parent entity) waits for 2 cakes (member entities) to be combined. The cakes can be visualized through the creation of a batch member queue for the tray entity.

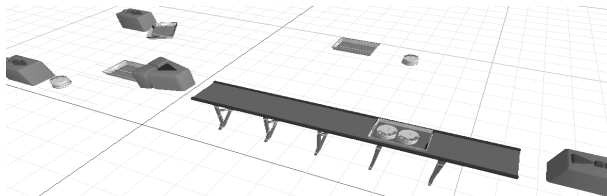


Figure 10: Combiner example in Simio

The Combiner also offers the possibility to edit properties like failures definition, add-on processes, state assignments, secondary resources, capacity types and others. Figure 11 displays the two blocks: Batch and Match of Arena.

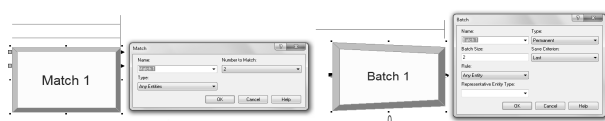


Figure 11: Match and batch blocks

**Separator:** Either separates batched members from a parent entity, or makes copies of an entity. Figure 12 illustrates a Separator object, which is comprised by a main object, an input node, an input buffer queue, a Processing queue, a parent output buffer queue and a member output buffer queue.

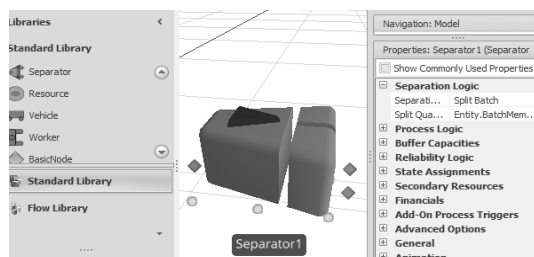


Figure 12: Separator object

A comparison between this object and the Separate block of Arena can be established. Apart from the common goal of the object and the block, the first also offers the possibility of editing further properties like define failures, add-on processes, state assignments, secondary resources, capacity types and others. Figure 13 illustrates a Separate Arena block.

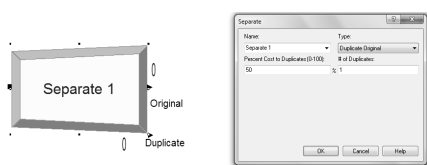


Figure 13: Separate block

**Resource:** Generic object with capacity that can be seized and released by other objects. Entities do not pass through this object, unlike the previous ones. In fact, the placement of this object on the Facility, only intends to declare the existence of a resource type that can be seized and released. In Arena, a user defines

resource types through the Resource data module, in order to achieve the same behaviour. Despite this, in Simio the concept of resources is quite different and much more robust than in Arena, due to the fact that any object can seize and release any other object. In this object, a user can also define failures, add-on processes among others. Figure 14 shows a Resource object.

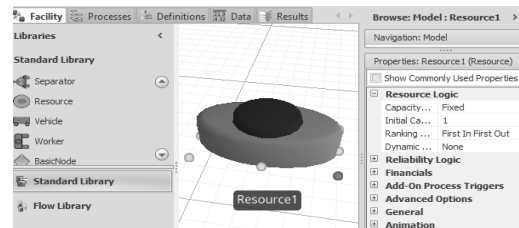


Figure 14: Resource object

**Vehicle:** Used to model devices that follow a fixed route (e.g. bus, train, etc.), or respond to dynamic requests for pickups (e.g. taxi, AGV, etc.). Similarly to the Resource object, the placement of this object on the Facility only intends to declare the existence of a vehicle type. Therefore this object is not connected to any object and thus entities do not pass through it. Figure 15 illustrates a Vehicle object.

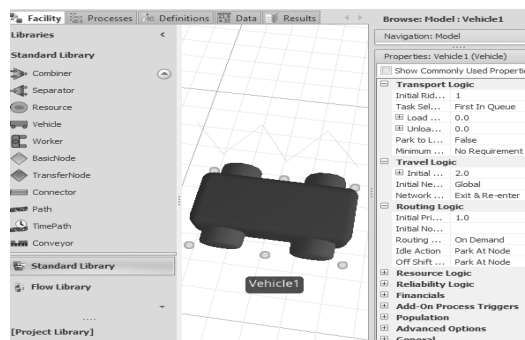


Figure 15: Vehicle object

This object is comprised by a ride station queue and the object itself. Arena also possesses the concept of vehicles to transport entities though these cannot follow fixed routes. Additionally, in Arena a transporter can only transport an entity at a time, whilst in Simio this can be done several entities at a time. To model transports in Arena it is necessary to use the template AdvancedTransfer, more specifically the blocks Request, Transport and Free. This object also allows a user to define failures, population number and others. Figure 16 illustrates these blocks combined, in order to model a transport.

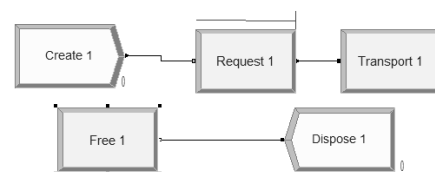


Figure 16: Using Arena blocks to model a transport

**Worker:** Defines a moveable resource that may be seized, released or used to transport entities between nodes. In contrast to the Vehicle that supports a Routing

Type which can be on demand or fixed, the Worker always operates on demand, i.e., the Worker always waits for either a visit request or a transport request. Additionally, unlike the Vehicle, the Worker has the ability to follow a work schedule and the Worker always assigns priority to seize visit requests over transport requests. This object is comprised by a ride station queue and the object itself. In Arena, the notion of moveable resources does not exist. Figure 17 shows a Worker object.

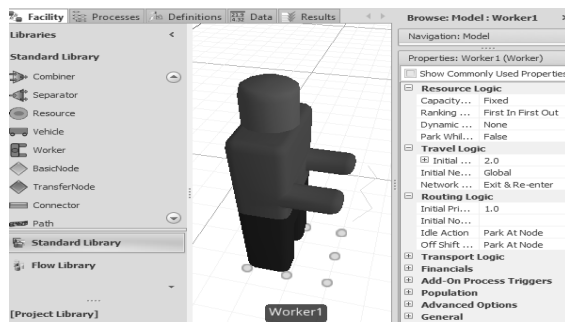


Figure 17: Worker object

**BasicNode:** Models a simple intersection between multiple links and can also be used as input nodes of objects. This object cannot model changes of destination. In Arena, a block with a similar concept to the nodes of Simio is the Station block. Figure 18 shows a BasicNode object and Figure 19 illustrates a Station block.

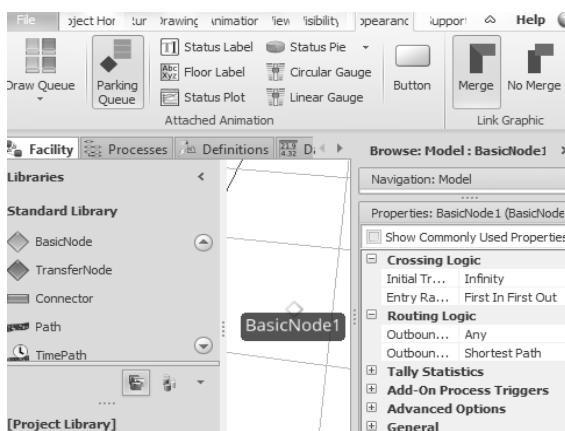


Figure 18: BasicNode object

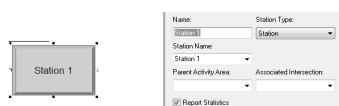


Figure 19: Station block

**TransferNode:** Models a complex intersection for travel mode. Unlike the previous object, this can model changes of destination. Additionally, the TransferNode can be used as output nodes of objects. Figure 20 illustrates a TransferNode object.

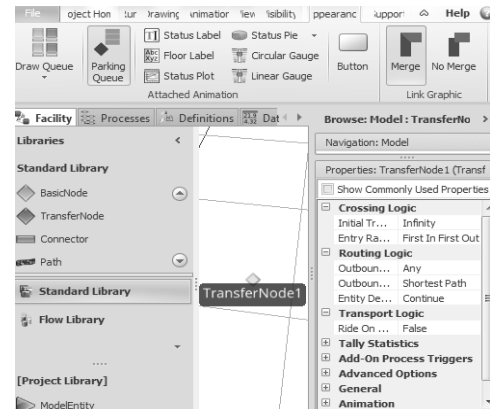


Figure 20: TransferNode object

**Connector:** Represents a simple zero-time travel link between 2 nodes. In Arena, the same goal is achieved by connecting 2 blocks, using the connect option. Figure 21 displays this object connecting a Source to a Sink.

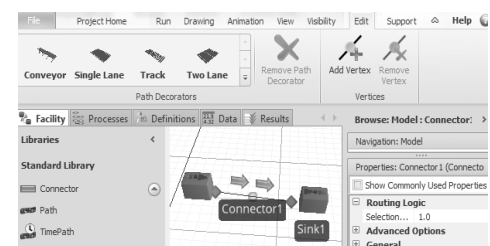


Figure 21: Connector object

**TimePath:** Used to transfer entities between 2 nodes with a specified travel time. Figure 22 shows a TimePath connecting a Source to a Sink.

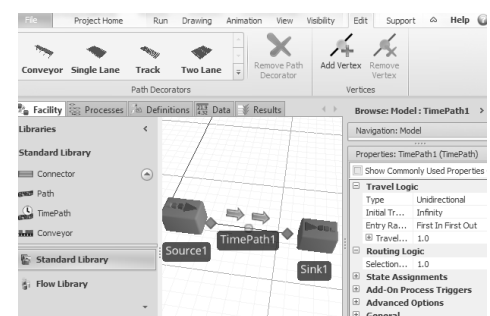


Figure 22: TimePath object

In Arena, to model the same behaviour, it is necessary to use the template AdvancedTransfer, more specifically, the blocks Station and Route. Figure 23 illustrates an example of these blocks being used.

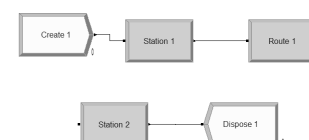


Figure 23: Using Arena blocks to model routes

**Path:** Represents links over which entities may move independently, at their own speed rates. Figure 24 shows a Path object, connecting a Source to a Sink. In Arena it is not possible to achieve the same behaviour of this object.

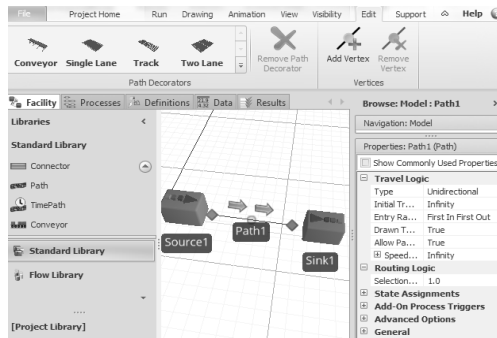


Figure 24: Path object

**Conveyor:** Entities traveling on this kind of connection do not “move“. Their movement is done by a conveyor that can be accumulating or non-accumulating. Figure 25 illustrates a Conveyor between a Source and a Sink.

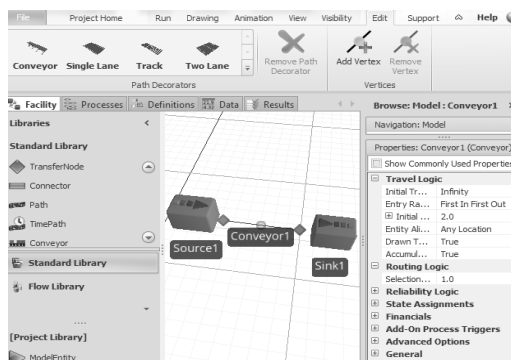


Figure 25: Conveyor object

In Arena, to model the same behaviour it is necessary to use the AdvancedTransfer template, more specifically, the blocks Station, Access and Convey. Figure 26 illustrates a conveyor being modelled in Arena.

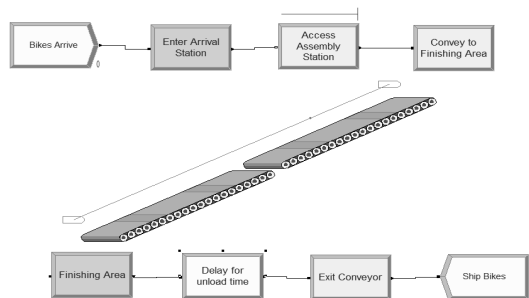


Figure 26: Using Arena blocks to model conveyors

## 4. CASE STUDIES

This chapter intends to introduce two case studies and analyse the way both simulation tools address the same problem. First, using a basic example - then, adapting it to the use of transports, which is a very important aspect of the development of any simulation model regarding the representation of a real operating system.

### 4.1.1. Basic Problem

• **Problem description:** This problem consists on a situation where trucks arrive at a factory and need to

unload its merchandise. Each truck is loaded with malfunctioning TVs that have to be repaired by repairmen. After a TV is repaired, it has to go to the inspection, where the inspectors will evaluate its condition. If the TV has no more defects, it goes to a truck parked outside. The trucks wait for repaired TVs and then leave the system.

• **Basic problem in Arena:** This system was modelled using Arena. Figure 27 shows the developed simulation model.

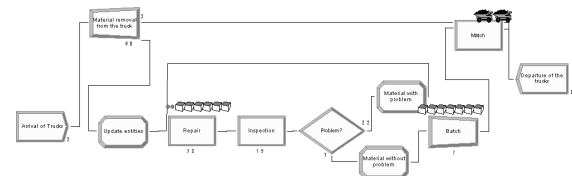


Figure 27: Basic problem in Arena

For modelling this system, two types of entities were used: one to represent the trucks and another to represent its merchandise. To create entity types, a user needs to use the Data module Entity. The Create block “Arrival of trucks” only creates entities of the type truck. After the creation of a truck, the Separate block “Material removal from the truck” is responsible for separating the truck and its merchandise, though, the merchandise will be animated with the same image as the truck. Since this does not correspond to what is intended, the Assign block “Update entities” updates the image of the entity of the merchandise. After being separated from its merchandise, the truck waits for the repaired TVs on the Match block. Thus, the TVs need to be repaired by the Process “Repair”. In this process, an entity seizes a repairman, delays the allocation for a specified process time correspondent to the time needed to repair a TV and then releases the repairman, allowing his allocation to another entity. When the reparation is concluded, the TV follows to the “Inspection” process. In it, an inspector is seized by an entity for a specified process time and then is released, in order for the inspector to be allocated to another entity. This process evaluates whether or not the TV still has any defects. Nearly 25% of the entities that pass through this inspection fail on the test and thus, need to be repaired again. This situation is modelled by the block Decide “Problem?”. Before repeating the “Repair” process, the image of a red ball is assigned to these entities. In the Batch block, the fixed TVs that will go to a truck wait until a specified value of TVs is reached. After the number is reached, the truck leaves the system with the merchandise, through the Dispose block.

• **Basic problem in Simio:** Figure 28 shows the developed model in Simio.

By looking at the model developed in Simio, the most notable difference to the same model developed in Arena is the animation. Simio’s model is much more realistic, in fact, it “looks” like the real system. However, there are several more differences that will now be addressed. Firstly, in Simio a user may associate an array of symbols to an object. Thus, in this model

two symbols were associated to the TV: a regular TV for TVs that haven't been repaired or TVs with no problems and a red TV for TVs with defects. Secondly, the creation of an entity type is done by simply dragging a ModelEntity object to the Facility. In this case, two of these objects were dragged: the truck and the TV. Another difference is the fact that in Arena, in order to model the change of destination of TVs that need to be repaired anew, the block Decide needs to be used. Yet, in Simio, the same goal can be achieved by adding a Path between the output node of the "Inspection" and the input node of the "Repair" and editing the respective probabilities of each destination. Additionally, any object can perform state assignments when entities enter or before leaving the object. In this sense, when entities fail the inspection and need to be repaired again, the Path connecting the two objects assigns the symbol red TV to the image of the entity. Lastly, in Simio only the Combiner object is necessary to model the Match and Batch blocks. In fact, this makes much sense, since the two blocks are used together almost every time in Arena. The parent input buffer queue displays the trucks waiting for the merchandise and the member input buffer queue displays the TVs waiting to be combined. When the later reaches a specified number, the first truck of the parent input buffer queue leaves the system through the Sink.

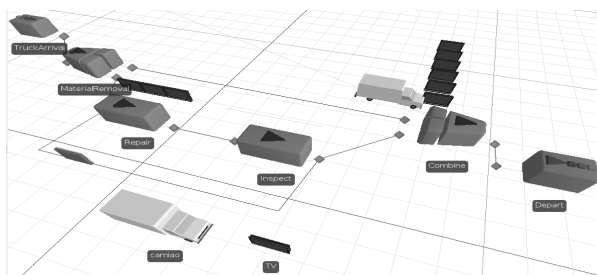


Figure 28: Basic problem in Simio

#### 4.1.2. Problem with Transports

- **Problem description:** This problem corresponds to the previous one with the introduction of transports. More specifically, in this case study, the merchandise removal from the trucks that arrive at the system and the loading of the trucks with repaired TVs is done with the help of forklifts. Also, when a TV needs to be inspected or to repeat the reparation process, it is placed on a conveyor.

- **Problem with transports in Arena:** Figure 29 illustrates the developed model in Arena and Figure 30 shows the correspondent model animation.

As can be seen, the physical model is created separately from the logical model.

- **Problem with transports in Simio:** Figure 31 shows the developed model in Simio.

As can be seen, the same model can be built using a lower number of components. This is due to the fact that, for instance, to model the Vehicles (forklifts) and the Conveyors it is only necessary to drag the correspondent objects to the Facility and edit their

properties. On the other hand, in Arena a great number of blocks need to be used to achieve the same goal.

To develop the considered case studies in both tools, the authors concluded the modelling in Simio was done in a simpler, faster and more intuitive way. Regarding the animation of the models, by examining the figures related to the case studies, it is clear that Simio models "look" like the real system. Lastly, there are some systems that can be modelled in Simio and are impracticable to achieve in Arena. Such a case can be consulted in (Vieira, 2013).

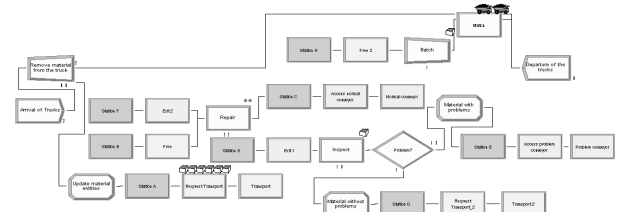


Figure 29: Problem with transports in Arena



Figure 30: Animation of the problem with transports

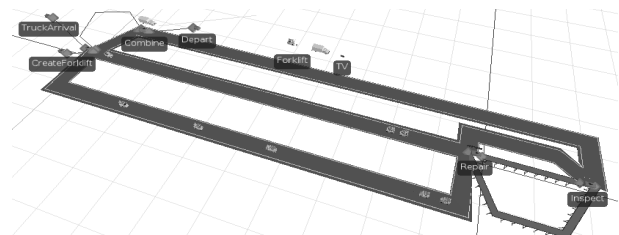


Figure 31: Problem with transports in Simio

## 5. CONCLUSIONS

Simulation modelling is being widely used for performance improvement of many systems (Hlupic, 2000, Hlupic and Paul, 1999). Hence, the comparison of simulation tools becomes a required task. In this context, Dias et al. (2007, 2011) compared a set of tools based on the popularity of most commercial tools. The author's final classification indicated that the most popular tool at the date was Arena and the only new tool on the "top 20" was Simio, both developed by the same authors: Dennis Pegden and David Sturrock. Thus, it should be normal that there are some resemblances between them. In this context, this paper intended to compare both tools taking into account several factors, such as: concepts of simulation model, animation development, modelling philosophies, Simio libraries versus Arena templates, entities concept, interface of the tools and Simio objects versus Arena blocks. Lastly, two case studies were addressed, in order to analyse the way both problems should be modelled on each tool.

Once the comparison of the tools was concluded, several aspects can be highlighted. Firstly, the Arena interface is simpler than Simio's. Regarding published documentation, Arena is highly more referenced than Simio. However, the latter is a much more recent tool.

The most visible difference between the models of Arena and Simio is the animation. Whilst in Arena the animation is developed in a separated step of the modelling, in Simio the modelling and the animation are done as a single step and the direct link with Google Warehouse makes the models very similar to the real systems. Even so, there are more differences in the systems modelling approach of each tool. Namely, Arena uses the process orientation while Simio is a multi-paradigm tool and its main feature is the ability to model intelligent objects and everything in Simio is an object. Consequently, Simio's entities are objects with their own intelligent behaviour and can make decisions, reject requests, etc. Moreover, entities have tokens that correspond to them and execute processes, while in Arena the only purposes of entities is to carry information (in their attributes) and to execute processes (Pegden, 2013a). Thus, an entity in Arena is similar to a token in Simio (Pegden, 2013a).

The comparison of objects in the Standard library of Simio with the blocks in the Basic Process template of Arena showed some resemblances between them, in the sense that most Arena blocks can be modelled by a correspondent Simio object, making it a very intuitive tool to use for Arena users. However, there are cases, in which the user needs to use a great number of blocks to achieve the same goal of a single Simio object. However, a great number of advantages of the Simio objects are noted, like the possibility of assigning values to states, add-on processes, arrays of animation, failures and schedules without the need to create new objects. Furthermore, some cases can be easily and quickly modelled in Simio than in Arena, for instance the utilization of conveyors or transports in Simio is achieved by simply dragging the correspondent objects to the Facility while in Arena a greater number of blocks need to be used.

The development of the chosen case studies showed that it is possible to model the same model of Arena, in Simio. Nevertheless, it required less effort and time to model the considered examples, in Simio. Moreover, Simio model's shape mimics the real system's layout, having required a lower abstraction level to develop.

An example of a case study developed in Simio, where it would be impracticable to model the same system in Arena was also given.

The compared aspects indicate the many advantages of using Simio. However, there are some down-sides typical of a recent tool. For instance, in Simio a user cannot create a clock to have a better perspective of the simulation time passing. In Arena this is very simple to add to a model. Another feature that Arena possesses and Simio does not is the incorporated tools: Input and Output Analyser. Also, at the date this paper was prepared there were some features that were not completely implemented (e.g. acceleration of entities). Nevertheless, this object oriented tool appears to have all the conditions to "trigger a widespread

paradigm shift in the way practitioners build models" (Pegden, 2007).

## ACKNOWLEDGEMENTS

This work has been supported by FCT – *Fundação para a Ciência e Tecnologia* in the scope of the project: PEst-OE/EEI/UI0319/2014.

## REFERENCES

- AKHTAR, N., NIAZI, M., MUSTAFA, F. & HUSSAIN, A. 2011. A discrete event system specification (DEVS)-based model of consanguinity. *Journal of Theoretical Biology*, 285, 103-112.
- ALTIOK, T. & MELAMED, B. 2010. *Simulation Modeling and Analysis with ARENA*, Elsevier Science.
- BROWN, J. E. & STURROCK, D. 2009. Identifying Cost Reduction and Performance Improvement Opportunities Through Simulation. Proceedings of the 2009 Winter Simulation Conference: M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, eds.
- DIAS, L., PEREIRA, G. & RODRIGUES, G. 2007. A Shortlist of the Most Popular Discrete Simulation Tools. *Simulation News Europe*, 17, 33-36.
- GARRIDO, J. M. 2009. *Object Oriented Simulation: A Modeling and Programming Perspective*, Springer-Verlag.
- HLUPIC, V. Simulation software: an Operational Research Society survey of academic and industrial users. *Simulation Conference*, 2000. Proceedings. Winter, 2000 2000. 1676-1683 vol.2.
- HLUPIC, V. & PAUL, R. 1999. Guidelines for selection of manufacturing simulation software. *IIE Transactions*, 31, 21-29.
- KAI, Z., RUICHANG, W., JIE, N., XIAOFENG, Z. & HAIJIAN, D. Using Simio for wartime casualty treatment simulation. IT in Medicine and Education (ITME), 2011 International Symposium on, 9-11 Dec. 2011 2011. 322-325.
- KELTON, W. D., SADOWSKI, R. P. & SADOWSKI, D. A. 2002. *Simulation with Arena*, McGraw-Hill School Education Group.
- LI, J. & WANG, L. Microscopic simulation on ticket office of large scale railway passenger station. *Advanced Forum on Transportation of China (AFTC 2011)*, 7th, 22-22 Oct. 2011 2011. 41-47.
- PEGDEN, C. D. Simio: A new simulation system based on intelligent objects. *Simulation Conference*, 2007 Winter, 9-12 Dec. 2007 2007. 2293-2300.
- PEGDEN, C. D. 2013a. An Introduction to Simio for Arena Users. Simio. White paper. Available online at: <http://www.simio.com/resources/white-papers/For-Arena-Users/An-Introduction-to-Simio-For-Arena-Users.htm>.
- PEGDEN, C. D. 2013b. Intelligent objects: the future of simulation. Simio. White paper. Available online at: <http://www.simio.com/resources/white-papers/Intelligent-objects/Intelligent-Objects-The-Future-of-Simulation-Page-1.htm>.
- PEGDEN, C. D. & STURROCK, D. T. Introduction to Simio. Proceedings - Winter Simulation Conference, 2011 Phoenix, AZ. 29-38.
- PEREIRA, G., DIAS, L., VIK, P. & OLIVEIRA, J. A. 2011. Discrete simulation tools ranking: a commercial software packages comparison based on popularity.
- STURROCK, D. T. & PEGDEN, C. D. Recent innovations in Simio. Proceedings - Winter Simulation Conference, 2010 Baltimore, MD. 21-31.
- VIEIRA, A. 2013. *Master Thesis. Micro simulation to evaluate the impact of introducing pre-signals in traffic intersections*. Department of Production and Systems - University of Minho.
- VIK, P., DIAS, L., PEREIRA, G., JOS, #233, OLIVEIRA & ABREU, R. 2010. Using simio for the specification of an integrated automated weighing solution in a cement plant. *Proceedings of the Winter Simulation Conference*. Baltimore, Maryland: Winter Simulation Conference.

# INITIAL BUBBLE AND DEW POINT TEMPERATURE ESTIMATIONS FOR FLASH CALCULATIONS OF HYDROCARBON MIXTURES

Serdar Çelebi, Gizem Kuşoğlu, Ümmühan Canan, Mehmet Ş. Kaya, Şerife Baş Hekimoğlu, Mustafa Bakır  
R&D Department  
Turkish Petroleum Refineries Corporation  
Körfez, Kocaeli  
Turkey  
E-mail: serdar.celebi@tupras.com.tr

## KEYWORDS

Correlation, Bubble-Dew Point, Flash, Hydrocarbons

## ABSTRACT

Bubble and dew point temperatures and process pressure correlations were developed to eliminate the large deviations in the bubble and dew point temperature calculations of hydrocarbon mixtures. The correlation gives good initial equilibrium constant (K) estimates for the thermodynamic calculations. Developed correlation is based on actual bubble and dew point temperatures covering a wide range of petroleum assays such as light straight run naphta (LSRN), heavy straight run naphta (HSRN), kerosene, light atmospheric diesel (LAD) and heavy atmospheric diesel (HAD). The correlation uses a simple power law depending upon the process pressure and a hydrocarbon physical property, mean average boiling point (MeABP). Correlation constants were obtained for different types of thermodynamic methods and gave good estimates for the initial K values. The TÜPRAŞ Simulation Platform uses the initial estimates for any flash procedure of hydrocarbons and evaluates the stream conditions to determine its bubble and dew point temperatures.

## INTRODUCTION

The TÜPRAŞ Simulation Platform is developed by R&D Department for simulating crude oil unit and includes thermodynamic models library and necessary flash calculations. Flash calculation of a hydrocarbon mixture gives the process information at a specified temperature and pressure or at a specified temperature and vapor mole fraction. Flash calculations are crucial mathematical operations to determine the stream properties before evaluating them in a refinery unit. In the simulation tool, refinery unit models use this information for further calculations.

The flash calculation of a dew or bubble temperature of a stream is performed by using the given information such as process pressure and vapor mole fraction of the defined hydrocarbon assay. The flash calculation mentioned in

literature is given as scheme in Figure 1 (Elliott and Carl 2012; Walas 1985). In order to calculate the bubble or dew temperature of a stream, a good initial estimate of temperature should be assumed in order to begin with a better estimate for K values. Some of the previously developed correlations for K values take into account the pressure and composition effects (Ahmed 2007). In our work, initial K value equation (Wilson correlation) is chosen as;

$$K_i = \frac{P_{ci}}{P} \exp[5.37(1 + \omega_i) \left(1 - \frac{T_{ci}}{T_{bb} \text{ or } T_{dew}}\right)] \quad (1)$$

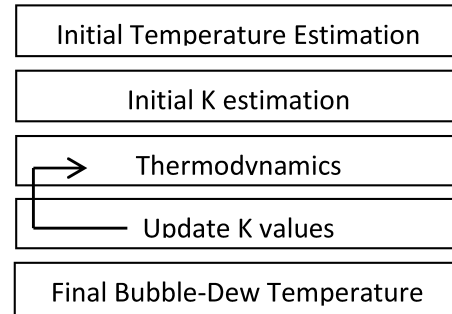


Figure 1: Bubble and Dew Point Calculation Procedure

Where  $K_i$  is the equilibrium constant of pseudo component  $i$ ,  $\omega_i$  is the acentric factor of pseudo component  $i$ ,  $P_{ci}$  and  $T_{ci}$  are the critical pressure (bar) and temperature ( $^{\circ}\text{R}$ ) for each pseudo component respectively (Wilson 1968). Riazi-Daubert correlation is used to obtain critical pressure and critical temperature of pseudo components (Riazi and Daubert 1987; API 1997). Also, correlation suggested by Pitzer is used for acentric factor (Pitzer et al. 1955). In Equation (1),  $T_{bb}$  stands for boiling point temperature and can be replaced by  $T_{dew}$  if dew temperature calculation is interested in.

In Equation (1), a bubble or dew point temperature estimate is required in order to initialize K values. Generally, for  $T_{bb}$ , it is advised to use a temperature slightly below the boiling point temperature of the most volatile component or slightly above the boiling point temperature of the least volatile component of the hydrocarbon assay (Vidal 2003).



However, this rule of thumb is not applicable at pressures far from the atmospheric pressure since calculated K values deviates too much at this pressures. At the end of the first iteration, the calculation of the bubble or dew point temperature is far away from the actual one. Therefore, in each iteration K values tend to increase or decrease too much and flash calculations mostly diverge.

In this work, a simple correlation is developed to start with good initial temperature estimates for the flash calculations. In the literature, initial temperature is determined according to boiling point temperature of volatile pseudo component in petroleum fraction. It remains constant for the same petroleum fraction while pressure is changing. However, dew and bubble point temperatures depend on the system pressure. The resulting correlation includes the effect of the pressure and petroleum fraction properties.

## CORRELATION DEVELOPMENT

A simple correlation was developed to make better initial estimates for K values of hydrocarbon mixtures. To do this, five different petroleum assays (LSRN, HSRN, Kerosen, LAD and HAD) were selected. The ASTM D86 experimental data of the assays is given in Table 1 with their mean average boiling points (MeABP) and American Petroleum Institute (API) gravities.

Table 1: ASTM D86 Data of Hydrocarbon Mixtures

Volume %	LSRN R <sup>0</sup>	HSRN R <sup>0</sup>	KERO R <sup>0</sup>	LAD R <sup>0</sup>	HAD R <sup>0</sup>
1	545.7±20	630.8±26	791.4±31	861.6±39	1014.8±42
10	560.1±6	661.2±9	815.7±17	917.4±13	1066.1±16
30	569.1±13	680.8±7	825.2±10	959.0±15	1101.5±14
50	579.9±11	703.2±8	840.7±8	991.0±12	1136.8±12
70	592.5±12	729.6±10	864.1±13	1027.0±9	1172.4±17
90	606.9±8	762.4±12	910.5±9	1096.7±12	1218.5±20
95	610.5±10	780.6±9	933.6±12	1140.2±10	1240.1±13
MeABP	578.0±10	700.0±12	845.0±15	987.0±19	1131.0±20
API	85.6±5.2	60.4±4.5	46.6±3.7	38.1±2.4	28.8±1.3

Table 2: Pressure & Dew Points of Hydrocarbon Mixtures

	LSRN	HSRN	KERO	LAD	HAD
Pressure Bar	$T_{dew}$ R <sup>0</sup>	$T_{dew}$ R <sup>0</sup>	$T_{dew}$ R <sup>0</sup>	$T_{dew}$ R <sup>0</sup>	$T_{dew}$ R <sup>0</sup>
0.5	542.2	682.7	821.3	987.2	1116.9
2.0	617.3	767.6	920.8	1091.4	1236.7
4.0	664.8	820.8	983.1	1156.3	1310.6
6.0	696.6	856.2	1024.5	1199.0	1359.0
8.0	721.1	883.5	1056.4	1231.5	1395.7
10.0	741.5	905.9	1082.7	1258.1	1425.3
15.0	781.8	949.8	1133.9	1308.1	N.A.
20.0	813.1	983.3	1172.7	N.A.	N.A.
25.0	839.0	1010.1	N.A.	N.A.	N.A.

From the ASTM D86 data of the assays, bubble and dew point temperatures were determined at the specified pressure by using good initial estimates of bubble and dew temperatures. In bubble point calculations, total vapor mole fraction ( $n_v$ ) is known to be zero and in the dew point calculations it is one since dew point is the point where first liquid droplet begins to condense. In Table 2, dew point temperatures (in Rankine) of hydrocarbon assays are listed against its process pressure and the data is plotted in Figure 2.

The correlation developed here is based on the data obtained from the flash calculations of hydrocarbon mixtures at the specified pressure and total vapor mole fraction. If a power law fitting procedure is applied to the data in Figure 2, one can obtain a relation between  $T_{dew}$  and P as the following;

$$T_{dew} = m \times P^n \quad (2)$$

Where  $T_{dew}$  is in Rankine and P is in Bar.

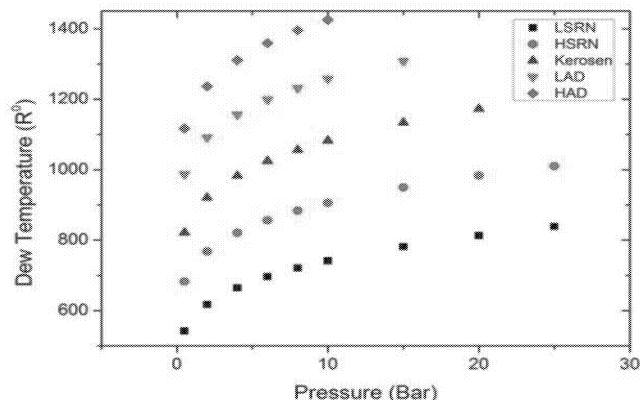


Figure 2: Dew Point Temperature & Pressure for Hydrocarbons.

Other than pressure, there should be at least one important physical characterization factor for petroleum assays in equation (2). Mean average boiling point is one of the important physical parameter which characterizes the petroleum assay and it is described by;

$$MeABP = (SG \times K_w)^3 \quad (3)$$

Where, SG stands for specific gravity and  $K_w$  stands for Watson characterization factor.

In order to embed the MeABP data into equation 2, the constants m and n were plotted against MeABP (Figure 3 and Figure 4).

Linear fitting was applied to the data in Figure 3 and Figure 4 and the following equations were obtained.

$$m = a \times MeABP + b \quad (4)$$

$$n = c \times MeABP + d \quad (5)$$

Where  $MeABP$  is in Rankine in equations (4) and (5).

Surprisingly, the m value is linearly correlated with MeABP whereas n value has some small deviations from

linearity. From linear fitting, dew temperature constants can be determined easily and the equation (2) becomes;

$$T_{\text{dew or bubble}} = (a \times \text{MeABP} + b) \times P^{(c \times \text{MeABP} + d)} \quad (6)$$

Where, T and MeABP are in Rankine and P is in Bar.

Correlation parameters a, b, c and d are reported in Table 3 for bubble-dew temperatures calculated by different kind of thermodynamic methods such as Soave Redlich Kwong (SRK) and Peng Robinson (PR) equation of states and Chao Seader (CS), Grayson Streed (GS).

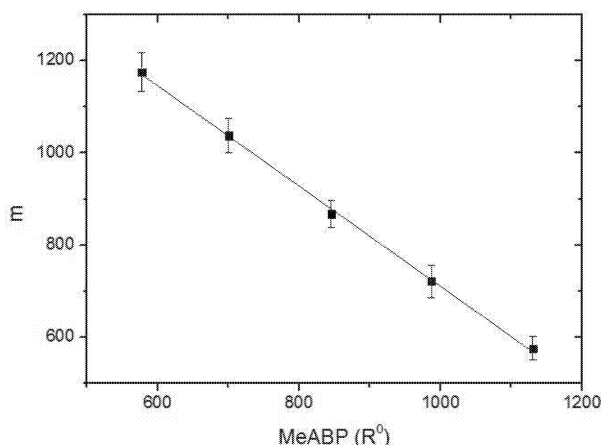


Figure 3: m & MeABP relation

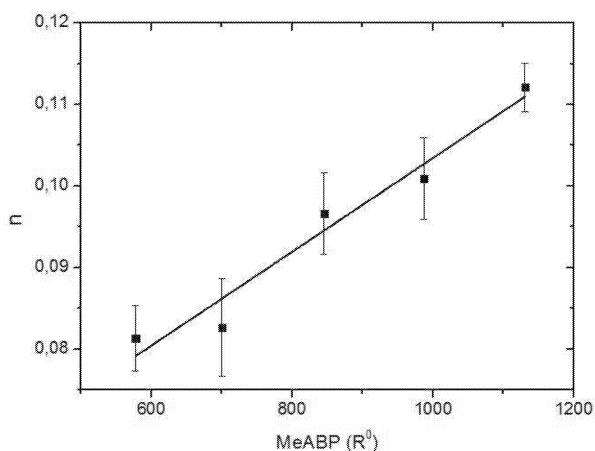


Figure 4: n & MeABP relation

Table 3: Constants of equation 4 for different thermodynamic methods.

T.M.	Temp.	a	b	c	d
SRK	Bubble	-0.983	1654.6	0.0000369	0.07856
	Dew	-1.089	1799.3	0.0000576	0.04583
PR	Bubble	-0.985	1657.1	0.0000382	0.07789
	Dew	-1.092	1802.2	0.0000588	0.04513
CS	Bubble	-0.971	1646.1	0.0000313	0.08330
	Dew	-1.319	2029.9	0.0001103	0.00681
GS	Bubble	-0.992	1668.2	0.0000595	0.05100
	Dew	-1.105	1820.0	0.0000830	0.01560

Although different constants were obtained for equation (6), estimation of initial temperatures for bubble and dew point calculations can be applied for any thermodynamic method. The initial estimates obtained from the correlation given in this work make the convergence of flash calculations faster and more reliable.

## CONCLUSIONS

A simple correlation, that uses system pressure and MeABP of hydrocarbon mixtures, has been developed for the initial estimate of the bubble and dew point temperatures of petroleum assays. The correlation gives values that are very close to bubble and dew point temperatures of mixtures. Hence, the flash calculations converge faster.

Although very promising results have been obtained, as future work, it is intended to improve the correlation both by including new properties such as API, Watson K etc. and by performing the statistical analysis for a longer period of time.

## REFERENCES

- Ahmed, T. 2007. *Equations of State and PVT Analysis*. Gulf Publishing, Houston, Texas, 339-341.
- American Petroleum Institute. 1997. *Technical Data Book-Petroleum Refining*. API Publishing Services, Washington, D.C.
- Elliott, R. and T.L. Carl. 2012. *Introductory Chemical Engineering Thermodynamics*. Prentice Hall, New York, 830-841.
- Pitzer, K.S.; D.Z. Lippmann; R.F. Jr. Curl; C.M. Huggings and D.E. Petersen. 1955. "The Volumetric and Thermodynamics Properties of Fluids, II. Compressibility Factors, Vapor Pressure and Entropy of Vaporisation." *J. of American Chemical Society*, Vol.77, 3433-3440.
- Riazi, M.R. and T.E. Daubert. 1987. "Characterization Parameters for Petroleum Fractions." *Ind. Eng. Chem. Res.*, Vol.26 No.24, 755-759.
- Vidal, J. 2003. *Thermodynamics: Applications in Chemical Engineering and the Petroleum Industry*. Editions Technip, Paris, 194.
- Walas, S.M. 1985. *Phase Equilibria in Chemical Engineering*. Butterworth, USA, 318-324.
- Wilson, G. 1968. "A modified Redlich-Kwong EOS, application physical data calculation". *Paper No15C, 65<sup>th</sup> AIChE National Meeting* (Cleveland, Ohio, May 4-7).

# BUFFER OVERFLOW SIMULATION IN SELF-SIMILAR QUEUING NETWORKS WITH FINITE BUFFER CAPACITY

Izabella Lokshina  
SUNY Oneonta  
324B Netzer Administration Bldg.  
Oneonta, NY 13820, USA  
E-mail: Izabella.Lokshina@oneonta.edu

Evan Schiele  
SAP America, Inc  
89 Headquarters Plz,  
Morristown, NJ 07960, USA  
E-mail: evan\_schiele@hotmail.com

## KEYWORDS

Telecommunications, Long-Range Dependent Processes, Self-Similar Queuing Network, Finite Buffer Capacity, Buffer Overflow Simulation, RESTART/LRE.

## ABSTRACT

The paper recommends an approach to estimate effectively the probability of buffer overflow in self-similar queuing networks with finite buffer capacity. Simulations with stochastic (or long-range dependent) traffic source models are conducted. A new efficient algorithm, based on the RESTART method, is developed and applied to accelerate the buffer overflow simulation in a finite buffer single server model under long-range dependent self-similar traffic load with different buffer sizes. Numerical examples and simulation results are provided.

## INTRODUCTION

The growth of broadband networks and the Internet has been exponential in terms of users and telecommunication systems as well as in traffic in recent years. The currently used high-speed communication networks are able to support a wide range of multimedia applications, such as audio, video and computer data that differ significantly in their traffic characteristics and performance requirements. One promising goal for telecommunications developers is to build a unified high-speed communication network platform capable of carrying diverse traffic, including long-range dependent self-similar teletraffic, and supporting diverse levels of quality of service (QoS). (Radev and Lokshina 2010; Lokshina 2012).

Recent studies of the high-speed communication network traffic have clearly shown that teletraffic (technical term, identifying all phenomena of transport and control of information within the high-speed telecommunication networks) exhibits long-range dependent self-similar properties over a wide range of time scales (Lokshina 2012).

Therefore, self-similar queuing systems are appropriate reference models for the high-speed communication network traffic as they provide capacity to estimate the network performance, allocate resources, and ensure the QoS. The impact of self-similar models on queuing and network performance is very significant, and the main trends in such findings are associated with efficient

simulation of real teletraffic data in the high-speed communication networks and, also, with analyzing self-similar queuing models and protocols in practical traffic scenarios.

Furthermore, recent studies of rare events in the high-speed communication networks confirmed that rare events take place in telecommunication networks also when evaluating reliability and performance. The frequency of rare events is very small, e.g.  $10^{-6}$  or less; however, the rare event probability can be used to simulate, estimate and analyze many queuing and network characteristics. Self-similar queuing systems are suitable reference models being used in different methodologies and techniques to accelerate rare event simulation in high-speed communication networks. Estimation of buffer overflow probability, which in recent telecommunications studies is acknowledged as a rare event, using Monte Carlo simulation requires a very long computing time and cannot easily be implemented (Rubino and Tuffin 2009). Instead, estimation of buffer overflow probability with Importance Sampling (IS) and Cross-Entropy algorithms that accelerate rare event simulation with Markov chains (Radev and Lokshina 2006-b; Lokshina and Bartolacci 2012) can be applied to a wide range of optimization tasks. This paper will take a different, new approach with estimation of buffer overflow in SSM/M/1/B self-similar queuing systems with finite buffer capacity applying steady-state simulation.

The remaining of this paper is organized as follows:

- Self-similar processes in diverse high-speed communication network traffic are discussed in section 2.
- Section 3 explains the unique properties of long-range dependent self-similar processes that do not degenerate with increase of the non-overlapping batch size to infinity in difference with traditional processes, which are normally used in modeling IP queuing network traffic that degenerate.
- Section 4 describes the buffer overflow simulation in SSM/M/1/B self-similar queuing systems with long-range dependent processes.
- Section 5 describes the limited relative error algorithm for buffer overflow simulation in SSM/M/1/B self-similar queuing systems with finite buffer capacity.
- Simulation results and performance evaluation are given in section 6.

- Finally, conclusions and future insights are provided in section 7, followed by acknowledgements and references.

## SELF-SIMILAR PROCESSES IN DIVERSE HIGH-SPEED COMMUNICATION NETWORK TRAFFIC

Self-similar teletraffic exhibits long-range dependent self-similar properties over a wide range of time scales and is usually observed in LAN and WAN, where superposition of strictly independent alternating ON/OFF traffic models, whose ON- or OFF-periods have heavy-tailed distributions with infinite variance, can be used to model aggregate queuing network traffic that exhibits long-range dependent self-similar behavior, typical for measured LAN traffic over a wide range of time scales (Radev and Lokshina 2010; Lokshina 2012).

Long-range dependent self-similar teletraffic is also observed in ATM networks: when arriving at an ATM buffer, it results in a heavy-tailed buffer occupancy distribution, and a buffer cell loss probability decreases with the buffer size not exponentially, like in traditional Markovian models, but hyperbolically (Radev and Lokshina 2006-a; Radev and Lokshina 2010).

Furthermore, long-range dependent self-similar or long-range dependent teletraffic is observed in the Internet as many characteristics can be modeled using heavy-tailed distributions, including the distributions of traffic times, user requests for documents, and document sizes. In IP with TCP self-similar queuing networks the transfer of files or messages shows that the reliable transmission and flow control mechanisms serve to maintain long range dependent structure included by heavy-tailed file size distributions (Bobbio et al. 2003).

Long-range dependent self-similar video traffic provides possibility for developing models for Variable Bit Rate (VBR) video traffic using heavy-tailed distributions (Radev and Lokshina 2006-a; Radev and Lokshina 2010). Therefore; we can clearly see that impact of self-similar models on the queuing and network performance is very significant.

The properties of long-range dependent self-similar teletraffic are very different from properties of traditional models based on Poisson, Markov-modulated Poisson, and related processes. More specifically, while tails of the queue length distributions in traditional teletraffic models decrease exponentially, those of self-similar teletraffic models decrease much slower.

Therefore, the use of traditional models in high-speed communication networks characterized by long-range dependent self-similar processes can lead to incorrect conclusions about the queuing and network performance. Traditional models can lead to over-estimation of the queuing and network performance, insufficient allocation of communication and data processing resources, and consequently difficulties in ensuring the QoS.

Self-similarity can be classified into two types: deterministic and stochastic (Radev and Lokshina 2010; Lokshina 2012).

In the first type, deterministic self-similarity, a mathematical object is assumed to be self-similar (or

fractal) if it can be decomposed into smaller copies of itself. That is, deterministic self-similarity is a property, in which the structure of the whole is contained in its parts.

This work is focused on stochastic self-similarity. In that case, probabilistic properties of self-similar processes remain unchanged or invariant when the process is viewed at different time scales. This is in contrast to Poisson processes that lose their burstiness and flatten out when time scales are changed (Radev and Lokshina 2010).

However, the time series of self-similar processes exhibit burstiness over a wide range of time scales. Self-similarity can statistically describe teletraffic that is bursty on many time scales (Lokshina 2012).

One can distinguish two types of stochastic self-similarity. A continuous-time stochastic process  $Y_t$  is strictly self-similar with a self-similarity parameter  $H$  ( $1/2 < H < 1$ ), if  $Y_{ct}$  and  $c^H Y_t$  (the rescaled process with time scale  $ct$ ) have identical finite-dimensional probability for any positive time stretching factor  $c$ .

This definition, in a sense of probability distribution, is quite different from that of the second-order self-similar process, observed at the mean, variance and autocorrelation levels (Lokshina 2012).

The process  $X$  is asymptotically second-order self-similar with  $0.5 < H < 1$ , if for each  $k$  large enough  $\rho_k^{(m)} \rightarrow \rho_k$ , as  $m \rightarrow \infty$ , where  $\rho_k = E[(X_t - \mu)(X_{t+k} - \mu)]/\sigma^2$ .

In this work the exact or asymptotic self-similar processes are used in an interchangeable manner, which refers to the tail behavior of the autocorrelations (Radev and Lokshina 2010; Lokshina 2012).

## LONG-RANGE DEPENDENT SELF-SIMILAR PROCESSES

Recent studies of high-speed communication networks have evidently shown that the most striking feature of some second-order self-similar processes is that the accumulative functions of the aggregated processes do not degenerate as the non-overlapping batch size  $m$  increasing to infinity. Such processes are known as Long-Range Dependent (LRD) processes (Bobbio et al 2003; Radev and Lokshina 2010; Lokshina 2012).

This is in contrast to traditional processes used in modeling IP queuing network traffic, all of which include the property that the accumulative functions of their aggregated processes degenerate with the non-overlapping batch size  $m$  increasing to infinity, i.e.,  $\rho_k^{(m)} \rightarrow 0$  or  $\rho_k^{(m)} = 0$  ( $|k| > 0$ ), for  $m > 1$ .

The equivalent definition of long-range dependence is given as (1).

$$\sum_{k=-\infty}^{\infty} \rho_k = \infty \quad (1)$$

Another definition of LRD is presented as (2),

$$\rho_k \sim L(t)k^{-(2-2H)}, \text{ as } k \rightarrow \infty \quad (2)$$

where  $1/2 < H < 1$  and  $L(\cdot)$  slowly varies at infinity, i.e. for all  $x > 0$  it could be determined as (3).

$$\lim_{t \rightarrow \infty} \frac{L(xt)}{L(t)} = 1 \quad (3)$$

The Hurst parameter  $H$  characterizes the relation in (2), which specifies the form of the tail of the accumulative function. One can show that is true for  $1/2 < H < 1$ , as given in (4).

$$\rho_k = \frac{1}{2} \left[ (k+1)^{2H} - 2k^{2H} + (k-1)^{2H} \right] \quad (4)$$

For  $0 < H < 1/2$  the process is Short-Range Dependent (SRD) and could be presented as (5).

$$\sum_{k=-\infty}^{\infty} \rho_k = 0 \quad (5)$$

For  $H=1$  all autocorrelation coefficients are equal to one, no matter how far apart in time the sequences are. This case has no practical importance in real telecommunications network traffic modeling. If  $H > 1$ , then (6) is true.

$$\rho_k = \begin{cases} 1 & \text{for } k = 0 \\ \frac{1}{2} k^{2H} g(k^{-1}) & \text{for } k > 0 \end{cases} \quad (6)$$

where

$$g(x) = (1+x)^{2H} - 2 + (1-x)^{2H} \quad (7)$$

One can see that  $g(x) \rightarrow \infty$  as  $H > 1$ . If  $0 < H < 1$  and  $H \neq 1/2$ , then the first non-zero term in the Taylor expansion of  $g(x)$  is equal to  $2H(2H-1)x^2$ . Therefore, (8) is true.

$$\rho_k / (H(2H-1)k^{2H-2}) \rightarrow 1, \text{ as } k \rightarrow \infty \quad (8)$$

In the frequency domain, an essentially equivalent definition of LRD for a process  $X$  with given spectral density (9),

$$f(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} \rho_k e^{ik\lambda} \quad (9)$$

is that in the case of LRD processes, this function is required to satisfy the following property (10),

$$f(\lambda) \sim c_{f_1} \lambda^{-\gamma}, \text{ as } \lambda \rightarrow 0 \quad (10)$$

where  $c_{f_1}$  is a positive constant and  $0 < \gamma < 1$ ,  $\gamma = 2H-1 < 1$ .

As a result, LRD manifests itself in the spectral density that obeys a power-law in the vicinity of the origin. This implies that  $f(0) = \sum_k \rho_k = \infty$ . Consequently, it requires a spectral density, which tends to  $+\infty$  as the frequency  $\lambda$  approaches 0.

For a Fractional Gaussian Noise (FGN) process, the spectral density  $f(\lambda, H)$  is given by (11),

$$f(\lambda, H) = 2c_f (1 - \cos(\lambda)) B(\lambda, H) \quad (11)$$

with  $0 < H < 1$  and  $-\pi \leq \lambda \leq \pi$ , where (12) is true,

$$c_f = \sigma^2 (2\pi)^{-1} \sin(\pi H) \Gamma(2H+1) \quad (12)$$

$$B(\lambda, H) = \sum_{k=-\infty}^{\infty} |2\pi k + \lambda|^{-2H-1}$$

and  $\sigma^2 = \text{Var}[X_k]$  and  $\Gamma(\cdot)$  is the gamma function.

The spectral density  $f(\lambda, H)$  in (11) complies with a power-law at the origin, as shown in (13),

$$f(\lambda, H) \rightarrow c_f \lambda^{1-2H}, \text{ as } \lambda \rightarrow 0 \quad (13)$$

where  $1/2 < H < 1$ .

## BUFFER OVERFLOW SIMULATION IN SSM/M/1/B SELF-SIMILAR QUEUEING SYSTEMS WITH LONG-RANGE DEPENDENT PROCESSES

Let us first think that the first buffer has a finite capacity  $N$ . In this case the state space of the driving process  $(X_t)$  is finite in  $\{0, \dots, N\}$ . Let us consider Markov additive process  $(X_t, Z_t)$ .

As we have previously described in Section 3, there is a significant difference in the queuing and network performance between traditional models of teletraffic, such as Poisson processes and Markovian processes, and those exhibiting long-range dependent self-similar behavior. More specifically, while tails of the queue length distributions in traditional models of IP network traffic decrease exponentially, those of self-similar traffic models decrease much slower (Radev and Lokshina 2010; Lokshina 2012).

Let us consider the potential impact of traffic characteristics, including the effects of self-similar behavior on queuing and network performance, protocol analysis, and network congestion controls.

Steady-state simulation of self-similar queuing system includes:

- Generation of self-similar traffic (Radev and Lokshina 2010; Lokshina 2012);
- Simulation of self-similar queuing process (Lokshina 2012); and
- Simulation of the overflow probability (Lokshina 2012; Lokshina and Bartolacci 2012).

This can be demonstrated with buffer overflow simulation in SSM/M/1/B queuing systems ( $B < \infty$ , i.e. queuing systems with the finite buffer capacity) with long-range dependent self-similar queuing processes.

In this case, the difference with M/M/1/B queuing system is that the arrival rate  $\lambda_j$  into SSM/M/1/B queuing system is not a constant value. It depends on the sequential number of time-series  $i$ , the total number of observations  $n$  and the Hurst parameter  $H$ , which determine the rate of self-similarity. The analyzed SSM/M/1/B queuing system has exponential service times with constant rates  $1/\mu$  as is shown in Figure 1.

The flow balance equations are given below (Giambene 2005; Lokshina 2012):

$$\begin{aligned} \lambda_j &= \lambda(i, n, H); & j &= 1, 2, \dots, B \\ \lambda_j &= 0; & j &\geq B+1 \\ \mu_j &= \mu; & j &= 1, 2, \dots, B+1 \end{aligned} \quad (14)$$

This system is stable with a throughput  $\rho = \frac{\lambda(i, n, H)}{\mu} < 1$ .

Let us consider two separated cases:  $\rho=1$ , and  $\rho \neq 1$ . For  $j = 0, 1, 2, \dots, B$  the distribution of the number of flows in the system is  $P_j = \rho^j P_0$ , which is determined according to

$$P_j = \frac{\rho^j(1-\rho)}{1-\rho^{B+1}}; \quad \rho \neq 1 \quad (15)$$

$$P_j = \frac{1}{B+1}; \quad \rho = 1$$

Therefore, the rate at which the flows are blocked and lost is  $\lambda P_B$ . The self-similar queuing process is described with the steady-state simulation scheme (Radev and Lokshina 2006), presented in Figure 2.

The self-similar traffic can be generated and the sequence of arrivals is obtained. The fixed length of self-similar traffic is extracted by fixing the number of observations.

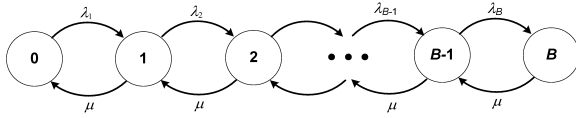


Figure 1: State Transition Diagram for a SSM/M/1/B Self-Similar Queuing System

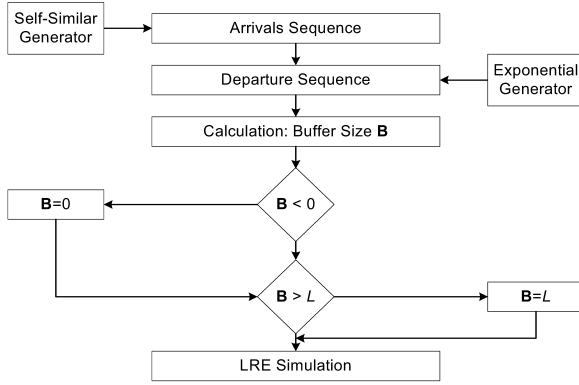


Figure 2: Steady-state Simulation of Self-similar Queuing Process

As the service process is Markovian, the sequence of departures has exponential distribution, generated with an inverse transform generator (Lokshina 2012).

The next step is the calculation of the buffer size. If the service size is greater than the size of arrivals, then the buffer size  $B=0$ , as it is impossible to have a negative buffer size. In cases when the buffer size is greater than the overflow  $L$ , i.e.  $B>L$ , the traffic is lost, therefore we have made an assumption that  $B=L$ .

The simulation is performed with splitting (Lokshina and Bartolacci 2012) using a variant based on the RESTART method (Georg and Schreiber 1996), where any chain is split by a fixed factor when it hits a level upward, and one of the copies is tagged as the original for that simulation level. When any of those copies hits that same level downward, if the copy is the original it just continues its path, otherwise it is killed immediately. This rule applies recursively, and the method is implemented in a depth-first fashion, as follows: whenever there is a split, all the non-original copies are simulated completely, one after the other; then the simulation continues for the original chain (Villen-Altamirano and Villen-Altamirano 2006).

The reason for eliminating most of the paths that go downward is to reduce the work. The buffer size calculations being made for all sequences provide the opportunity to estimate the overflow probability using the RESTART method with the Limited Relative Error (LRE) algorithm.

### LIMITED RELATIVE ERROR ALGORITHM FOR BUFFER OVERFLOW SIMULATION IN SSM/M/1/B SELF-SIMILAR QUEUING SYSTEMS WITH FINITE BUFFER CAPACITY

The limited relative error algorithm helps to determine the complementary cumulative function of arrivals at single server buffer queues with Markov processes. In order to describe the LRE principles for steady-state simulation in Discrete-Time Markov Chains (DTMC), let us consider a homogeneous two-node Markov chain, which is extended to regular DTMC, consisting of  $(k+1)$  nodes with states, respectively  $S_0, S_1, \dots, S_k$ , as shown in Figure 3.

We obtain the random generated sequence  $x_1, x_2, \dots, x_t, x_{t+1}, \dots$  for  $x=0, 1, \dots, k$ , for which a transition for state  $S_j$  at the time  $t$  exists, e.g.  $x_t=j$  and there are no constraints to the parameters of the transition probabilities:

$$p_{ij} = P(j | i); \quad (i, j) = 0, 1, \dots, k; \quad \sum_{j=1}^k p_{ij} = 1 \quad (16)$$

There are no absorbing states  $S_i$  at  $p_{ii}=1$  for all stationary probabilities  $P_j$   $j=0, 1, \dots, k$ , which satisfy the constraint condition:

$$0 \leq P_j < 1; \quad \sum_{j=0}^k P_j = 1 \quad (17)$$

The cumulative distribution  $F(x)$  can be presented as:

$$\left. \begin{aligned} F(x) &= F_i; \quad (i-1) \leq x < i; \quad i = 1, 2, \dots, k+1; \\ F_i &= \sum_{j=0}^{i-1} P_j; \quad F_0 = 0; \quad F_{k+1} = 1; \end{aligned} \right\} \quad (18)$$

In order to simulate the  $(k+1)$  nodes of Markov chain, the complementary cumulative distribution  $G(x)=1-F(x)$  that is more significant, can be determined along with the local correlation coefficient  $\rho(x)$  through the limited relative error approach. After having the homogeneous two-node Markov chain defined as shown in Figure 1, with changing the states  $n$  times, an estimation of the local correlation coefficient  $\hat{\rho}(x)$  can be obtained, which connects the number of transitions through a dividing line  $a_i \approx c_i$ , with the total number of observed events  $l_i = n - d_i$  ( $\beta = 0, 1, \dots, i-1$ ) at left side, and  $d_i$  at right side ( $\beta = i, i+2, \dots, k$ ).

The value of simulated complementary cumulative distribution  $\hat{G}_i$  can be defined directly by using relative frequency  $d_i/n$ , if there is enough number of samples:

$$n \geq 10^3; \quad (l_i, d_i \geq 10^2); (a_i, c_i, l_i - a_i, d_i - c_i) \geq 10 \quad (19)$$

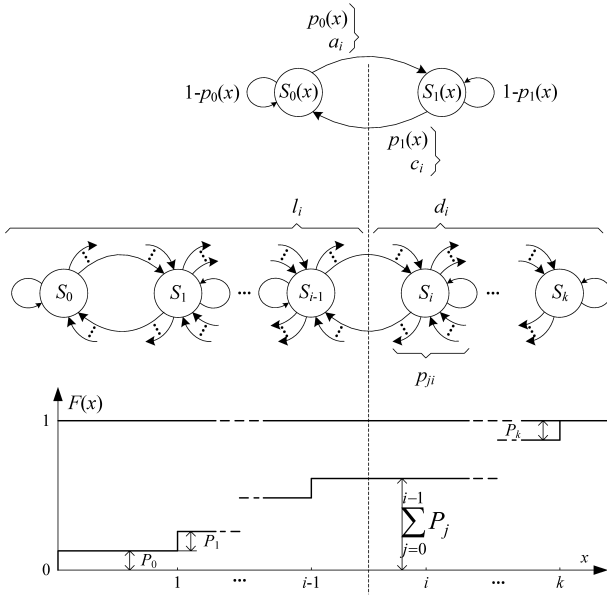


Figure 3: Cumulative Function  $F(x)$  for  $(k+1)$ -node Markov Chain

The posterior equations can be used for the complementary function  $\hat{G}(x)$ , the average number of generated values of  $\hat{\beta}$ , the local correlation coefficient  $\hat{\rho}(x)$ , the correlation coefficient  $\text{Cor}[x]$  and the relative error  $\text{RE}[x]$ :

$$\hat{G}(x) = \hat{G}_i = d_i/n \quad \hat{\beta} = \frac{1}{n} \sum_{i=1}^k d_i \quad (20)$$

$$\hat{\rho}(x) = \hat{\rho}_i = 1 - \frac{c_i/d_i}{1 - d_i/n} \quad \begin{matrix} i-1 \leq x < i \\ i = 1, \dots, k \end{matrix}$$

$$\text{Cor}[x] = \text{Cor}_i = (1 + \hat{\rho}_i)(1 - \hat{\rho}_i) \quad \text{RE}[x]^2 = \text{RE}_i = \frac{1 - d_i/n}{d_i} \cdot \text{Cor}_i$$

The main advantage of this approach is that the relationships between transitions  $c_i$  are obtained with routine statistical calculations.

The necessary total number of simulation trails  $n$  is determined with the maximal relative error  $\text{RE}_{\max}[x]^2$  and with the less value of the function  $G(x)$ , presented as  $G_{\min} = \hat{G}_k$  in approximation equation:

$$n = \frac{(1 - G_{\min})}{G_{\min} \cdot \text{RE}_{\max}[x]^2} \approx \frac{\text{Cor}_k}{\hat{G}_k \cdot \text{RE}_{\max}[x]^2}; \quad (21)$$

$$\text{Cor}_k = \frac{1 + \hat{\rho}_k}{1 - \hat{\rho}_k}$$

This procedure can be described with a standard version of limited relative error algorithm for random discrete sequences of buffer arrivals:

**Step 1:** Initialization of minimal and maximal values of the simulation parameter.

**Step 2:** Estimation and management of the simulation time.

**Cycle  $L_1$ :** Determine the current variable for calculating the Markov chain, e.g.  $\omega := \beta$ ; generate a new value for  $\beta$  with given distribution.

Increase the number of state  $h_\beta$ .

If the condition  $\beta < \omega$  is true, then increase the number of transitions  $c_{\beta+1}$  while it reaches the value of  $c_\omega$ .

**Cycle  $L_2$ :** Determine the total number of events at the left part  $l_s$  and at the right part  $d_s$  of the Markov chain and number of transitions  $a_s := c_s$ ; check on the constraint condition (19) for the index  $i=s$ .

If the constraint condition (19) is true, then calculate the posterior values of the local correlation coefficient  $\hat{\rho}_s$  and relative error  $\text{RE}[x]$  with use of (20). Calculate whether the relative error  $\text{RE}[x] \leq \text{RE}_{\max}[x]$ .

If  $s < k$ , then leave the cycle  $L_2$ .

If the index  $s=k$  is reached, then leave the cycle  $L_1$  and increase the index of the simulation time  $s := s+1$ ;

**Step 3:** Printing out the experimental results for  $i=1, 2, \dots, k$ . The results for the total frequency  $d_i$  are determined according to (22):

$$d_i = \sum_{j=1}^k h_j \quad \text{for } i=0, 1, \dots, k \quad \text{where } d_0 = n \quad (22)$$

The values of the complementary function  $\hat{G}_i$ , the local correlation coefficient  $\hat{\rho}_s$  and the relative error  $\text{RE}[x]$  are calculated as given in (20).

## SIMULATION RESULTS AND PERFORMANCE ANALYSIS

As an example, the overflow probability of an SSM/M/1/B self-similar queuing system has been simulated with different characteristics of long-range dependent self-similar arrival processes. In order to demonstrate the effects of self-similarity on the buffer overflow probability, the obtained experimental results were compared with the complementary cumulative distribution in the traditional single server finite buffer queue M/M/1/B. The obtained results in a logarithmic scale are given in Figure 4.

In order to get representative and steady results the sequences of 10 000 observations were used. With the suggested LRE algorithm the values of complementary cumulative function  $G(x)$  for different buffer sizes were calculated. The calculations were provided with the step  $m=4$ . One can see in Figure 4 that the increasing Hurst parameter has led to an insignificant decrease in the overflow probability. For example, for the value of Hurst parameter  $H=0.6$  the overflow probability was  $G(L)=1.045 \cdot 10^{-1}$ , and for  $H=0.9$  it was  $G(L)=5.6 \cdot 10^{-2}$ .

On the other hand, the overflow probability of self-similar queuing system has increased significantly in comparison with the theoretical M/M/1/B self-similar queuing system, for which  $G(L)=4.79 \cdot 10^{-5}$ .

After that, the simulation was repeated for SSM/M/1/B self-similar queuing system by using long-range dependent self-similar arrival process with  $H=0.6$  and different buffer sizes. The obtained results for buffer size  $B=40$ ,  $B=60$  and  $B=80$  are shown in Figure 5. One can see that since the buffer size was increased twice, the overflow probability

has been changed simply by about two orders of magnitude – from  $1.045 \cdot 10^{-1}$  to  $6.4 \cdot 10^{-3}$ .

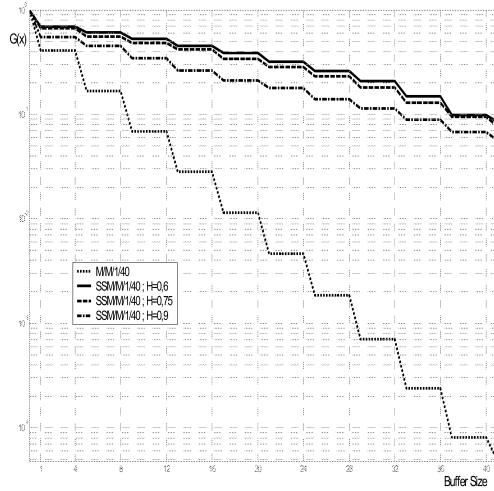


Figure 4: Buffer Overflow Probability ( $L=41$ ) in SSM/M/1/40 Self-similar Queuing System

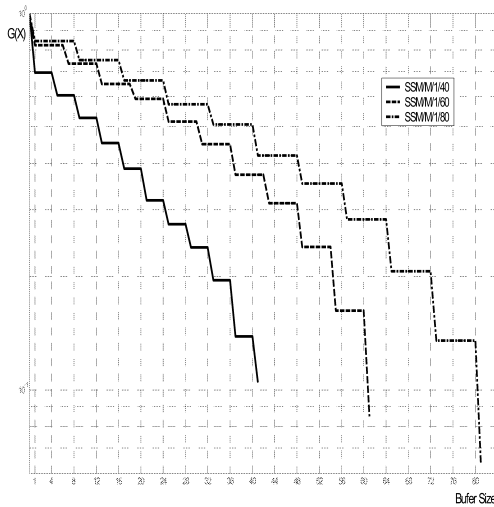


Figure 5: Buffer Overflow Probability in SSM/M/1/B Self-similar Queuing System with Different Buffer Sizes

Finally, it was confirmed that in order to design a single server finite buffer model with long-range dependent self-similar arrival processes, the buffer size has to be increased many times in order to decrease the overflow probability.

## CONCLUSIONS

High-speed communication network traffic exhibits long-range dependent self-similar properties over a wide range of time scales that are very different from the properties of traditional models. Self-similar models are appropriate for queuing network traffic as they provide capacity to

estimate the network performance and quality, allocate resources, and ensure the QoS.

Steady-state simulation of self-similar queuing processes was conducted using the RESTART method. Efficient algorithm to simulate the buffer overflow probability was developed and applied in order to generate self-similar queuing network traffic with a limited relative error.

The models of SSM/M/1/40 self-similar queuing system were used with different parameters of long-range dependent self-similar arrival processes and different buffer sizes. Simulations and numerical results were shown and analyzed.

The resulting recommended approach is appropriate and particularly efficient being used to evaluate reliability and performance in high-speed communication networks, while higher performance networks must be described by lesser buffer overflow probabilities.

## ACKNOWLEDGEMENTS

The authors would like to thank our colleague Cees J. M. Lanting from CSEM, Switzerland for his time, thoughtful insights and review during the preparation of this paper.

## REFERENCES

- Bobbio, A.; Horváth, A.; Scarpa, M.; and M. Telek. 2003. "Acyclic discrete phase type distributions: Properties and a parameter estimation algorithm." *Performance Evaluation*, Vol. 54, No. 1, 1-32.
- Georg, C. and F. Schreiber. 1996. "The RESTART/LRE method for rare event simulation." In *Proceedings of the 1996 Winter Simulation Conference*. Coronado, CA, 390-397.
- Giambene, G. 2005. *Queuing Theory and Telecommunications: Networks and Applications*. Springer, NY, 585 p.
- Lokshina, I. 2012. "Study about Effects of Self-similar IP Network Traffic on Queuing and Network Performance." *Int. J. Mobile Network Design and Innovation*, Vol. 4, No. 2, 76-90.
- Lokshina, I. and M. Bartolacci. 2012. "Accelerated Rare Event Simulation with Markov Chain Modelling in Wireless Communication Networks." *Int. J. Mobile Network Design and Innovation*, Vol. 4, No. 4, 185-191.
- Radev, D. and I. Lokshina. 2006-a. "Performance Analysis of Mobile Communication Networks with Clustering and Neural Modelling." *Int. J. Mobile Network Design and Innovation*, Vol. 1, Nos. 3/4, 188-196.
- Radev, D. and I. Lokshina. 2006-b. "Rare Event Simulation with Tandem Jackson Networks." In *Proceedings of the Fourteen International Conference on Telecommunication Systems: Modeling and Analysis (ICTSM 2006)*, Penn State Berks, Reading, PA, 262-270.
- Radev, D. and I. Lokshina. 2010. "Advanced Models and Algorithms for Self-Similar Network Traffic Simulation and Performance Analysis." *Journal of Electrical Engineering*, Vol. 61, No. 6, 341-349.
- Rubino G. and B. Tuffin. 2009. *Rare Event Simulation using Monte Carlo Methods*, UK, John Wiley & Sons, 278p.
- Villen-Altamirano, M. and J. Villen-Altamirano. 2006. "On the efficiency of RESTART for multidimensional systems." *ACM Transactions on Modeling and Computer Simulation*, Vol. 16, No. 3, 251-279.



## **AUTHOR BIOGRAPHIES**

**IZABELLA LOKSHINA**, PhD is Professor of MIS and Chair of Management, Marketing and Information Systems Department at SUNY Oneonta. Her main research interests include complex system modeling (telecommunications and queuing networks) and artificial intelligence; Izabella.Lokshina@oneonta.edu

**EVAN SCHIELE**, PMP, MBA is a Program Delivery Manager with SAP America. His main research interests are complex system modeling (business applications of ICT) and SAP; evan\_schiele@hotmail.com

# A VIRTUAL TESTBED FOR UNDERWATER ROBOTICS – APPLICATION TO CONTROL DESIGN FOR AUVS

Linus Atorf  
Martin Krehel  
Juergen Rossmann  
Institute for Man-Machine Interaction  
RWTH Aachen University  
Ahornstr. 55, 52074 Aachen, Germany  
email: {atorf, rossmann}@mmi.rwth-aachen.de

Gaurav S. Sukhatme  
Robotic Embedded Systems Laboratory  
University of Southern California  
Ronald Tutor Hall, RTH426  
3710 McClintock Ave  
Los Angeles, CA 90089-9121  
email: gaurav@usc.edu

## KEYWORDS

3D Simulation, Multibody Simulation, Fluid Dynamics, Control Design, AUVs, Virtual Testbeds

## ABSTRACT

Developing and operating autonomous underwater vehicles (AUVs) is a complex, hazardous, and expensive task. In order to help engineers reduce the involved costs and risks, we establish a Virtual Testbed for underwater robotics. We provide a concept for holistic and realistic 3D simulation systems including dynamic rigid body simulation for underwater scenarios as well as actuator and sensor simulation, which can be used for development, testing, and evaluation of AUV models. The real-world problem of control design for AUV reference tracking is utilized as example and tackled by developing three different linear controllers for the REMUS AUV. Finally, we perform simulated experiments to evaluate and compare the different controllers.

## INTRODUCTION

Vast parts of the oceans have not yet been explored, while operating underwater remains a big challenge with many unsolved problems. Advances in autonomous underwater vehicles (AUVs) and the associated reduction of reliance upon human operators enables more efficient work underwater. In general, robots are highly complex systems with high operating costs, which is especially true for underwater applications. Developers of robotic systems switch more and more to simulation-based approaches to reduce operating costs and to minimize development time (Rossmann et al. (2013a)).

This contribution presents how established dynamic multibody simulations can be upgraded to provide a 3D environment for developing and evaluating virtual underwater robotic applications. After providing a theoretical summary of the steps necessary to construct the dynamic equations of motion for rigid bodies underwater, an application from control theory is chosen as example to illustrate how the Virtual Testbed (VTB)

for underwater robotics can be used to solve real-world problems in practice. On that account, three different controller designs for AUV reference tracking are developed and evaluated.

## A Virtual Testbed for Underwater Robotics

VTBs cover the modeling and simulation of the interdependencies between the prototype's internal components as well as the interaction between the digital prototype and its environment (Rossmann and Schluse (2011)). The environmental model should be able to resemble typical scenarios from underwater robotics, such as exploration missions or maintenance tasks.

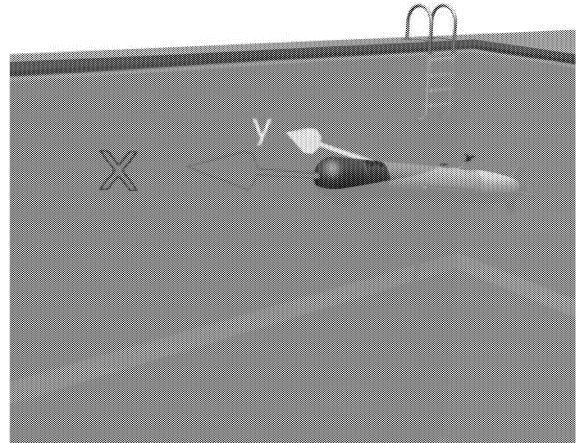


Figure 1: 3D simulation of the REMUS AUV surfacing in a virtual test basin. The body coordinate system is highlighted.

The development of complex systems such as robots usually can not be carried out individually for each subcomponent, but requires a holistic development approach, regarding also their interplay of different modules under varying conditions (Rossmann et al. (2013b)). This is especially important if real testbeds for the target scenario are not feasible, expensive, or involve the risk of human injuries—all those conditions are true for under-

water scenarios.

Two of the main aspects are the early test of interfaces between subsystems and the comprehensive and simultaneous verification of soft- and hardware components under different operating conditions including the case of component failures. Existing interfaces for advanced sensor simulation can already be used (see Emde et al. (2011)) or ground truth data from the simulation system can directly be accessed as stand-in while more complex sensor and error models are developed.

VTBs allow for a cost-efficient engineering during the overall life cycle of robotic systems in different levels of detail. Target scenarios with a first simple prototype can be created rapidly to answer fundamental design questions. Such a system can be evaluated very early under nearly all important general conditions, while alternative designs can be analyzed at short notice. In the ongoing design process, components and subprocesses that have been identified as important can be modeled down to the desired level of detail (Rossmann et al. (2013b)). Figure 1 shows the VTB with a simulated 3D model of the REMUS AUV.

## Multibody Simulation

The following section describes how to extend an existing rigid body dynamic simulation for underwater applications, i.e. rigid bodies submerged in ambient fluid. In specific, resulting forces and torques resulting from fluid dynamics are calculated and applied to the affected bodies. In our case, the multibody simulation is based on a maximum coordinates formulation using Lagrange Multipliers (Stewart and Trinkle (2000)). Velocity-based constraints are used for features such as e.g. joint mechanism and non-penetration requirement. More details are given in Jung et al. (2011).

## UNDERWATER SIMULATION OF RIGID BODIES

This section deals with the dynamic equations of motion of AUVs including effects caused by the actuation system as well as by moving through ambient fluid. The structure of the equations is very similar to the well-known dynamic equation of motion for rigid bodies

$$\boldsymbol{\tau}_{RB} = \mathbf{M}_{RB}\dot{\mathbf{v}} + \mathbf{C}_{RB}(\mathbf{v})\mathbf{v} . \quad (1)$$

The following sections are a summary of a well-known approach, presented in e.g. Fjellstad (1994), Fossen (1994; 2002), Hoang (2006).

Generally speaking, the effects caused by moving through ambient fluid can be divided into the these terms :

- Added mass, Coriolis, and centripetal force and torque, collected in  $\boldsymbol{\tau}_A$ ,
- Damping force and torque  $\boldsymbol{\tau}_D$ ,

- Force and torque caused by gravity and buoyancy, combined into  $\boldsymbol{\tau}_{GB}(\mathbf{r})$ ,
- Environmental forces and torques  $\boldsymbol{\tau}_E$  (such as ocean current, perturbations, etc.),
- Output  $\boldsymbol{\tau}$  generated by the actuation system.

Hence the total force and torque acting on a rigid body in a fluid is given by:

$$\boldsymbol{\tau}_{RB} = -\boldsymbol{\tau}_A - \boldsymbol{\tau}_D - \boldsymbol{\tau}_{GB}(\mathbf{r}) - \boldsymbol{\tau}_E + \boldsymbol{\tau}$$

In the following sections each term is discussed separately. In favour of simplicity, we assume a current-free environment with neglectable perturbations, so  $\boldsymbol{\tau}_E \approx \mathbf{0}$ .

## Added Mass and Coriolis Terms

When moving a rigid body through water, fluid particles around the vehicle have to be accelerated which causes their kinetic energy to increase. Due to conservation of energy, the work needs to be accomplished by the vehicle which leads to a decrease of the vehicle's kinetic energy. This effect can be mathematically described by the *added mass matrix*  $\mathbf{M}_A \in \mathbb{R}^{6 \times 6}$  and the according *Coriolis-matrix*  $\mathbf{C}_A(\mathbf{v}) \in \mathbb{R}^{6 \times 6}$ , yielding the same form as (1) (note that  $\mathbf{v} \in \mathbb{R}^6$  contains translational and rotational velocity):

$$\boldsymbol{\tau}_A = \mathbf{M}_A\dot{\mathbf{v}} + \mathbf{C}_A(\mathbf{v})\mathbf{v}. \quad (2)$$

Generally speaking, a submerged body behaves as if its mass and inertia tensor were increased.  $\mathbf{C}_A(\mathbf{v})$  can be calculated from  $\mathbf{M}_A$ . The coefficients of  $\mathbf{M}_A$  usually have to be determined experimentally (see e.g. Triantafyllou and Hover (2003)). For rigid bodies with three planes of symmetry,  $\mathbf{M}_A$  becomes diagonal.

## Hydrodynamic Damping

In addition to the effects of added mass, dissipative forces and torques occur which are mainly a function of the vehicle's velocity and geometry. The main reasons for hydrodynamic damping are laminar and turbulent skin friction due to boundary layer flow as well as viscous damping due to vortex shedding (Fjellstad (1994)). Vortex shedding is an oscillating flow that occurs if a body passes through a fluid at certain velocity (Batchelor (2000)). In general, it is complicated to separate these effects, but it is possible to collect coefficients in the form  $\mathbf{D}(\mathbf{v}) = \mathbf{D} + \mathbf{D}_n(\mathbf{v})$  for most applications. The matrix  $\mathbf{D} \in \mathbb{R}^{6 \times 6}$  represents *linear damping*, whereas  $\mathbf{D}_n \in \mathbb{R}^{6 \times 6}$  applies to *nonlinear damping*. The hydrodynamic damping matrix  $\mathbf{D}(\mathbf{v})$  becomes diagonal for an AUV which has three planes of symmetry. The resulting damping force and torque are

$$\boldsymbol{\tau}_D = \mathbf{D}(\mathbf{v})\mathbf{v} . \quad (3)$$

## Weight and Buoyancy

The weight force  ${}^I\mathbf{f}_G$  acts in the center of gravity  $\mathbf{r}_G = (x_G, y_G, z_G)^T$ , the buoyancy force  ${}^I\mathbf{f}_B$  in the center of buoyancy  $\mathbf{r}_B = (x_B, y_B, z_B)^T$ . Weight force and buoyancy force are given with respect to world coordinates

$${}^I\mathbf{f}_G = (0, 0, -mg)^T, \quad {}^I\mathbf{f}_B = (0, 0, \rho g V)^T,$$

where  $m$  is the vehicle's mass,  $g$  the standard gravitational acceleration,  $V$  the amount of water displaced by the vehicle and  $\rho$  the fluid density. To account for surface effects, the volume of displaced water is calculated according only to the submerged parts of the rigid body. Using the definition of torque  $\mathbf{m} = \mathbf{r} \times \mathbf{f}$  the restoring force and torque can be expressed as

$$\boldsymbol{\tau}_{GB}(\mathbf{r}) = \begin{pmatrix} \mathbf{f}_G + \mathbf{f}_B \\ \mathbf{r}_G \times \mathbf{f}_G + \mathbf{r}_B \times \mathbf{f}_B \end{pmatrix}. \quad (4)$$

## Actuation System

Usually, the actuation system of AUVs consists of propellers and fins. In the following we present the different parts of the actuation system.

### Propellers

The force and torque of single-screw propellers are a non-linear function of the vehicle's velocity  $\mathbf{v}$  and the control vector  $\mathbf{n} := (n_1|n_1|, \dots, n_m|n_m|)^T$ , where  $n_i$  denotes the speed of rotation of the  $i$ -th propeller. A *bilinear model* can be obtained which is determined by two matrices  $\mathbf{B}_1 \in \mathbb{R}^{6 \times m}$  ( $m$  propellers) and  $\mathbf{B}_2(\mathbf{n}) \in \mathbb{R}^{6 \times 6}$

$$\boldsymbol{\tau} = \mathbf{B}_1 \mathbf{n} - \mathbf{B}_2(\mathbf{n}) \mathbf{v},$$

which have to be identified experimentally.

### Fins

The general, scalar formula for the fin lift is given by e.g. Prestero (1994):

$$f_{\text{fin}} = \frac{1}{2} \rho c_L A \delta v^2, \\ m_{\text{fin}} = r_{\text{fin}} f_{\text{fin}}.$$

where  $c_L$  is the fin lift coefficient,  $A$  the fin area,  $\rho$  the fluid density,  $\delta$  the relative fin angle and  $v$  the relative fin velocity. In general, the forces and torques are highly non-linear functions of parameters like the *Reynolds number*. As reasonable simplifications (assuming the location of the fin, with respect to body coordinates, to be  $\mathbf{r}_{\text{fin}} = (x_{\text{fin}}, 0, 0)^T$  and the lift coefficient to

be constant) we obtain for small angles

$$f_y = \frac{1}{2} \rho_W c_L A_{\text{fin}} [v_x^2 \delta_r - v_x v_y - x_{\text{fin}} v_x v_\psi], \\ f_z = -\frac{1}{2} \rho_W c_L A_{\text{fin}} [v_x^2 \delta_s - v_x v_z + x_{\text{fin}} v_x v_\theta], \\ m_\theta = \frac{1}{2} \rho_W c_L A_{\text{fin}} x_{\text{fin}} [v_x^2 \delta_s + v_x v_z - x_{\text{fin}} v_x v_\theta], \\ m_\psi = \frac{1}{2} \rho_W c_L A_{\text{fin}} x_{\text{fin}} [v_x^2 \delta_r - v_x v_y - x_{\text{fin}} v_x v_\psi],$$

where the indices  $r$  and  $s$  denote rudder and stern, respectively.

## Dynamic Equation of Motion

By combining (1) and (2) to

$$\mathbf{M} := \mathbf{M}_{RB} + \mathbf{M}_A, \quad \mathbf{C}(\mathbf{v}) := \mathbf{C}_{RB}(\mathbf{v}) + \mathbf{C}_A(\mathbf{v}),$$

the equations of motion for rigid bodies can be extended with the discussed terms (3) and (4). With respect to the body coordinate system, we obtain

$$\mathbf{M}\dot{\mathbf{v}} + \mathbf{C}(\mathbf{v})\mathbf{v} + \mathbf{D}(\mathbf{v})\mathbf{v} + \boldsymbol{\tau}_{GB} = \boldsymbol{\tau}. \quad (5)$$

For dynamic multibody simulation in maximal coordinates, (5) has to be transformed from body to world coordinates. The transformation matrix  $\boldsymbol{\Omega} \in \mathbb{R}^{6 \times 6}$  is given by

$$\dot{\mathbf{r}} = \boldsymbol{\Omega} \mathbf{v} \Leftrightarrow \mathbf{v} = \boldsymbol{\Omega}^{-1} \dot{\mathbf{r}},$$

and its time derivative is

$$\ddot{\mathbf{r}} = \dot{\boldsymbol{\Omega}} \mathbf{v} + \boldsymbol{\Omega} \dot{\mathbf{v}} \Leftrightarrow \dot{\mathbf{v}} = \boldsymbol{\Omega}^{-1} (\ddot{\mathbf{r}} - \dot{\boldsymbol{\Omega}} \boldsymbol{\Omega}^{-1} \dot{\mathbf{r}}).$$

Hence, the equation of motion (5) in world coordinates becomes:

$${}^I\mathbf{M}(\mathbf{r})\ddot{\mathbf{r}} + {}^I\mathbf{C}(\mathbf{r}, \mathbf{v})\dot{\mathbf{r}} + {}^I\mathbf{D}(\mathbf{r}, \mathbf{v})\dot{\mathbf{r}} + {}^I\boldsymbol{\tau}_{GB}(\mathbf{r}) = {}^I\boldsymbol{\tau}(\mathbf{r})$$

Equivalents in world coordinates are denoted with  ${}^I(\cdot)$ , which are defined as

$${}^I\mathbf{M}(\mathbf{r}) = \boldsymbol{\Omega}^{-T} \mathbf{M} \boldsymbol{\Omega}^{-1}, \\ {}^I\mathbf{C}(\mathbf{r}, \mathbf{v}) = \boldsymbol{\Omega}^{-T} (\mathbf{C}(\mathbf{v}) - \mathbf{M} \dot{\boldsymbol{\Omega}} \boldsymbol{\Omega}^{-1}) \boldsymbol{\Omega}^{-1}, \\ {}^I\mathbf{D}(\mathbf{r}, \mathbf{v}) = \boldsymbol{\Omega}^{-T} \mathbf{D}(\mathbf{v}) \boldsymbol{\Omega}^{-1}, \\ {}^I\boldsymbol{\tau}_{GB}(\mathbf{r}) = \boldsymbol{\Omega}^{-T} \boldsymbol{\tau}_{GB}, \\ {}^I\boldsymbol{\tau}(\mathbf{r}) = \boldsymbol{\Omega}^{-T} \boldsymbol{\tau},$$

with the notation  $(\boldsymbol{\Omega}^{-1})^T \equiv \boldsymbol{\Omega}^{-T}$ . Note that in case of Euler angles,  $\boldsymbol{\Omega}_a$  does not exist for  $\theta \in \{\frac{\pi}{2}, \frac{-\pi}{2}\}$ .

## APPLICATION

### Control problem of AUVs

The control problem of underwater robots consists of two parts: guidance or path planning on the one hand,

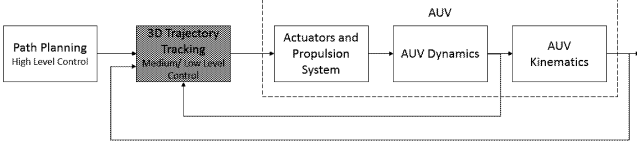


Figure 2: Control problem specification.

and autopilot design or reference tracking (see Figure 2) on the other hand. We want to show how the VTB for underwater robotics can be used to design controllers and test them in 3D simulation in addition to commonly used rapid control prototyping environments. Apart from testing the control system, we obtain the opportunity to evaluate its performance and compare different control strategies. Most torpedo-shaped AUVs are under-actuated systems, which are highly dynamic, nonlinear and coupled. The control problem can be formulated extending the definition of the *characterization vector* (see e.g. Fossen (2002))

$$\mathbf{r}_{cv} = (\mathbf{r}_d, \dot{\mathbf{r}}_d, \ddot{\mathbf{r}}_d)^T.$$

This vector is the output of a path planning algorithm and the input for the reference tracking controller. The feedback terms may contain translational and angular velocity, position and orientation.

### Design of control systems for reference tracking

For the controller design we use data of the torpedo-shaped REMUS AUV, obtained from Prestero (1994). We develop and investigate three different linear control approaches; two controllers based on decoupling the system and a third controller using a fully coupled system representation.

#### Decoupled, basic PI control

When decoupling the system (using Euler angles), we can identify three subsystems; see Table 1.

Table 1: Decoupled subsystems.

Subsystem	State variables
Speed system	$v_x(t)$
Steering system	$v_y(t), v_\psi(t), \psi(t)$
Diving system	$v_z(t), v_\theta(t), \theta(t), z(t)$

First, we design a control system only based on PI controllers. The main control objective is to minimize each component of the control error  $\mathbf{e}(t) = (v_{x,d} - v_x, \psi_d - \psi, z_d - z)^T$ . The speed subsystem is represented by a first order differential equation after decoupling and linearization

$$(m + A_1)\dot{v}_x + D_1 v_x + D_{1n}|v_{x,d}|v_x = b_1|n|n,$$

which can be controlled sufficiently using a PI controller. To use the procedure for the steering and diving system we have to introduce more tightening assumptions. The basic PI controller can be improved by compensating for the resulting restoring force and torque as described in the following section. First parameter approximations are obtained using the *Ziegler-Nichols method*.

#### Decoupled, hybrid control

The control performance of the speed system using PI control is sufficient, hence we remain with the PI control structure. For the steering- and the diving subsystems we use state feedback control, therefore the control system is referred to as *hybrid* control. The nonlinear equation of motion for the steering motion can be decoupled and linearized and we obtain a matrix equation of motion

$$\underbrace{\begin{pmatrix} m + A_2 & 0 & 0 \\ 0 & I_z + A_6 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{=: \mathbf{M}_s} \begin{pmatrix} \dot{v}_y \\ \dot{v}_\psi \\ \dot{\psi} \end{pmatrix} + \underbrace{\begin{pmatrix} D_2 + D_{2n}|v_{y,d}| & (m + A_1)v_{x,d} & 0 \\ (A_2 - A_1)v_{x,d} & D_6 + D_{6n}|v_{\psi,d}| & 0 \\ 0 & -1 & 0 \end{pmatrix}}_{=: \mathbf{D}_s} \begin{pmatrix} v_y \\ v_\psi \\ \psi \end{pmatrix} = \begin{pmatrix} f_y \\ m_\psi \\ 0 \end{pmatrix},$$

which can be rearranged to the state space representation  $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$  using

$$\mathbf{A} := -\mathbf{M}_s^{-1}\mathbf{D}_s, \quad \mathbf{B} := \mathbf{M}_s^{-1}, \\ \mathbf{x} := (v_y, v_\psi, \psi)^T, \quad \mathbf{u} := (f_y, m_\psi, 0)^T.$$

The LQR algorithm can be used to obtain an optimal feedback matrix. The weighting matrices  $\mathbf{Q}, \mathbf{R}$  of the cost function  $J = \int_0^\infty \frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \frac{1}{2} \mathbf{u}^T \mathbf{R} \mathbf{u} dt$  should satisfy the following relations

$$\mathbf{Q} = \text{diag}\{q_1, q_2, q_3\}, \quad q_1 \gg q_2 \gg q_3 \\ \mathbf{R} = \text{diag}\{r_1, r_2, r_3\}, \quad q_2 \gg r_1, r_2, r_3.$$

These relations are reasonable, considering the side-slip velocity  $v_y$  is not directly controllable. For the steering subsystem the controller is set up in a cascaded structure as shown in Figure 3, which allows to induce the *cross track error* as an additional control variable. It is defined as the euclidean distance from the center of gravity to the desired trajectory in the  $(x, y)$ -plane. The outer cross track error control loop compensates for errors caused by linearization and the simplification assumptions. The linearized equations of motion for the diving subsystem can be generated in the same way using the small angle approximation for  $\theta$ . We get the state space representation  $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$ , with

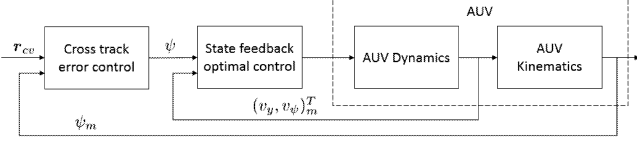


Figure 3: Block diagram of steering control system.

$\mathbf{x} := (v_z, v_\theta, \theta, z)^T$  and  $\mathbf{u} := (f_z, m_\theta, 0, 0)^T$ . We use state feedback optimal control (LQR) for the diving subsystem again. Additionally, we have to consider the constant force and torque produced by the difference between weight and buoyancy, which is regarded by applying a feed forward control command, so it follows the control law

$$\boldsymbol{\tau} = \boldsymbol{\tau}_{\text{feedforw.}} + \boldsymbol{\tau}_{\text{feedback}},$$

where  $\boldsymbol{\tau}_{\text{feedforw.}} = (0, 0, c_{\text{feedforw.}}, 0, 0, 0)^T$  and  $c_{\text{feedforw.}}$  can be found during simulation.

#### Coupled, state feedback

It is possible to obtain a fully coupled time invariant state space representation of the equation of motion (Fossen (1994))

$$\begin{pmatrix} \dot{\mathbf{x}}_1 \\ \dot{\mathbf{x}}_2 \end{pmatrix} = \begin{pmatrix} -\mathbf{M}^{-1}[\mathbf{C} + \mathbf{D}] & -\mathbf{M}^{-1}\mathbf{G} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{M}^{-1} \\ \mathbf{0} \end{pmatrix} \mathbf{u},$$

where the matrices  $\mathbf{M}, \mathbf{C}, \mathbf{D}, \mathbf{G}$  are the Jacobian matrices of the linearization process. To obtain a time invariant form we have to use several assumptions. Under general operating conditions, we want  $v_{y,d} = v_{z,d} = v_{\phi,d} = 0$  which reflects zero side slip velocity and roll rate. We can assume that the desired velocity in forward direction is  $v_{x,d} = \text{const.}$  Usually, the maneuvers do not contain rapid diving parts, that means  $v_{\theta,d} \approx 0$ . For steering maneuvers a valid approximation is given by  $v_{\psi,d} \approx \text{const.}$  Especially for steady state it follows  $v_{\theta,d} = v_{\psi,d} = 0$ ,  $\theta_d = \phi_d = 0$  and  $\psi_d = \text{const.}$  We use the LQR algorithm to find an optimal feedback matrix.

## EVALUATION

Beside numerical analysis, some experiments have been performed to verify the implemented simulation for underwater applications. Figure 4 depicts the dynamic behavior of the REMUS AUV, resulting from an initial orientation of  $\theta_0 = -20^\circ$ . We do not apply any actuator forces or torques. Considering the AUV to be positively buoyant the magnitude of the velocity in  $z$ -direction increases until an equilibrium between the resulting buoyancy force and the damping forces is reached (*terminal velocity*). Due to the high coefficients of added mass and hydrodynamic damping we see a damped oscillation in

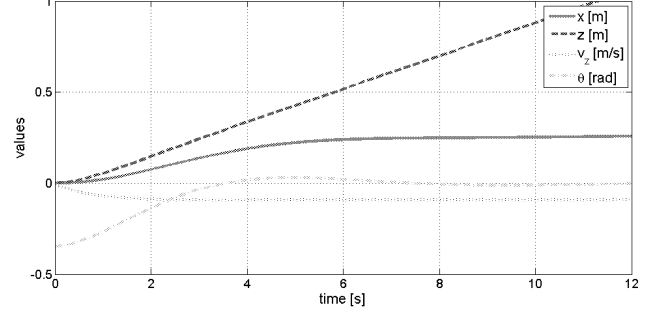


Figure 4: Dynamic behavior of REMUS AUV, relative to initial position with orientation  $\theta = -20^\circ$ .

$\theta$  caused by restoring forces, due to the displacement of location of center of gravity and center of buoyancy. Additionally a motion in  $x$ -direction is observed, caused by the small coefficients of added mass and hydrodynamic damping in that direction.

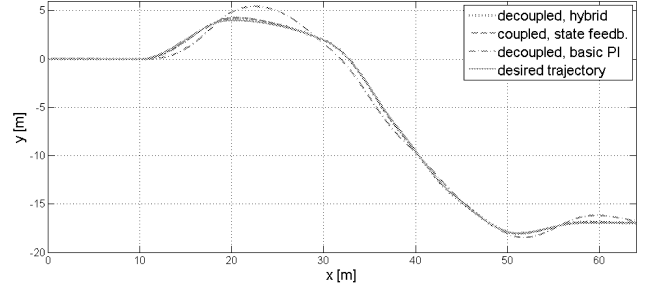


Figure 5: Example trajectory in horizontal plane.

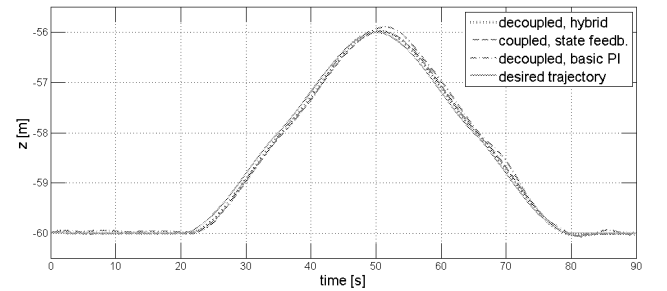


Figure 6: Example trajectory in vertical plane.

In order to compare the implemented control systems, we use the root-mean-square error (difference in euclidean distance between reference and achieved trajectories) defined as

$$\text{RMSE} := \sqrt{\frac{\sum_{i=1}^n \|\mathbf{r}_i - \mathbf{r}_{d,i}\|_2^2}{n}}, \quad \mathbf{r}, \mathbf{r}_d \in \mathbb{R}^3, \quad (6)$$

where  $\mathbf{r}_{d,i}$  are points of the desired trajectory. Considering the assumptions used for obtaining linear and time invariant equations of motion, the cross track

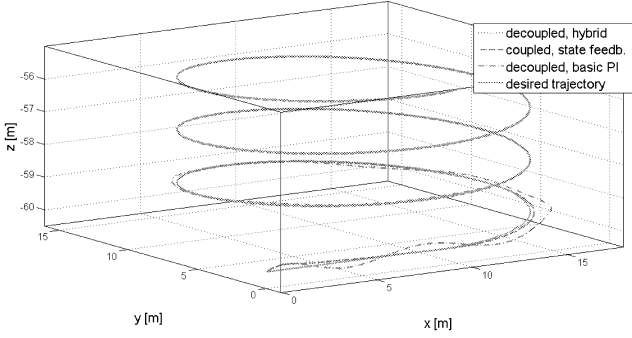


Figure 7: Helix-shaped example trajectory in 3D.

error controller has been set up in a cascaded structure for the basic PI control system and the fully coupled state feedback controller as well. Figures 5, 6, and 7 show trajectories in the  $(x, y)$ -plane, in the  $(x, z)$ -plane, as well as a helix-shaped motion in 3D. The resulting errors are presented in Table 2.

Table 2: Errors (RMSE in meters) for horizontal, vertical, and helix-shaped example trajectories.

Control system	Horiz.	Vert.	Helix
Decoupled, basic PI	0.0398	0.0039	0.0265
Decoupled, hybrid	0.0130	0.0021	0.0131
Coupled, state feedb.	0.0190	0.0031	0.0180

## CONCLUSION

In this contribution we have presented how the concept of Virtual Testbeds can be applied to scenarios from underwater robotics. We have shown how established dynamic multi-body simulations can be upgraded to account for effects of ambient fluids acting on submerged rigid bodies. Three possible approaches to the real-world problem of designing control algorithms for AUV reference tracking have been developed and evaluated, underlining the usefulness of VTBs for underwater robotics.

The evaluation shows that the basic PI control system is sufficient for steady state operation but fails in very transient regions. As one may not expect, the decoupled hybrid control system achieves best performance even compared to the fully coupled system. This is mainly induced by the very tight assumptions used to obtain a linear time invariant representation. The cascaded structure for steering motion and the feed forward part for diving motion are satisfying ways to compensate for non-linearities and decoupling assumptions. It would be interesting to compare these structures to non-linear control approaches. Nevertheless, the linear, decoupled, hybrid, and cascaded control structures can be recommended for AUV reference tracking.

## REFERENCES

- Batchelor G.K., 2000. *An Introduction to Fluid Dynamics*. Cambridge University Press.
- Emde M.; Rossmann J.; Sondermann B.; and Hempe N., 2011. *Advanced Sensor Simulation In Virtual Testbeds: A Cost-Efficient Way To Develop And Verify Space Applications*. In *AIAA SPACE, California*.
- Fjellstad O.E., 1994. *Control of unmanned underwater vehicles in six degrees of freedom: a quaternion feedback approach*. Ph.D. thesis, The Norwegian Institute of Technology, University of Trondheim.
- Fossen T.I., 1994. *Guidance and Control of Ocean Vehicles*. John Wiley & Sons.
- Fossen T.I., 2002. *Marine Control Systems: Guidance, Navigation, and Control of Ships, Rigs and Underwater Vehicles*. Norwegian University of Science and Technology and Marine Cybernetics As.
- Hoang N.Q., 2006. *Zur Modellierung und Regelung ferngesteuerter Unterwasserfahrzeuge, Verkehrstechnik/Fahrzeugtechnik*, vol. 12, no. 624. Fortschritt Berichte VDI. ISBN 3183624125.
- Jung T.; Rast M.; Kaigom E.; and Romann J., 2011. *Fast VR Application Development Based on Versatile Rigid Multi-Body Dynamics Simulation*. In *ASME 2011 IDETC/CIE*. Washington, DC.
- Presterio T., 1994. *Verification of a Six-Degree of Freedom Simulation Model for the REMUS Autonomous Underwater Vehicle*. Ph.D. thesis, University of California at Davis, USA.
- Rossmann J.; Guiffo Kaigom E.; Atorf L.; and Schlette C., 2013a. *From space to manufacturing industry: New approaches of eRobotics to serve today's and future manufacturing needs*. In *15. ASIM Fachtagung, Paderborn. AM 147*, vol. 316, 417–426.
- Rossmann J.; Kaigom E.; Atorf L.; Rast M.; and Schlette C., 2013b. *A Virtual Testbed for Human-Robot Interaction*. In *UKSim 2013, UK*. 277–282.
- Rossmann J. and Schluse M., 2011. *Virtual Robotic Testbeds: A Foundation for e-Robotics in Space, in Industry - And in the Woods*. In *DeSE*. 496–501.
- Stewart D. and Trinkle J.C., 2000. *An Implicit Time-Stepping Scheme for Rigid Body Dynamics with Coulomb Friction*. In *IEEE ICRA 2000*.
- Triantafyllou M.S. and Hover F.S., 2003. *Maneuvering and Control of Marine Vehicles*. Ph.D. thesis, Department of Ocean Engineering, MIT, USA.





# **GENETIC ALGORITHMS**



# SOME ALGORITHMS SUPPORTING THE WATER NETWORK MANAGEMENT BY USE OF SIMULATION OF NETWORK HYDRAULIC MODEL

Jan Studzinski, Marek Kurowski  
Polish Academy of Sciences  
Systems Research Institute  
Newelska 6, 01-447 Warsaw, Poland  
[studzins@ibspan.waw.pl](mailto:studzins@ibspan.waw.pl), [makur@xl.wp.pl](mailto:makur@xl.wp.pl)

## KEYWORDS

IT systems for waterworks, water net revitalization, water leaks detection, water age and chlorine concentration computing.

## ABSTRACT

The short description of an ICT system for computer aided management of communal water networks that has been developed at the Systems Research Institute of Polish Academy of Sciences (IBS PAN) is presented (Studzinski 2013A) in this paper. Several waterworks management tasks requiring mathematical modelling, optimization and approximation algorithms can be solved using this system. Static optimization and multi-criteria algorithms are used for solving more complicated tasks like calibration of the water net hydraulic model, water net optimization and planning, control of pumps in the water net pump stations etc. (Stachura et al. 2012). But some of the management tasks are simpler and can be performed by means of repetitive simulation runs of the water net hydraulic model. The water net simulation, planning of the SCADA system, calculation of water age and chlorine concentration in the water net, localization of hidden water leaks occurring in the network and planning of water net revitalization works are the examples of such tasks executed by the ICT system. They are described in this paper.

## INTRODUCTION

Three essential goals that can be reached by computer aided management of municipal water networks are reduction of costs and simplification of waterworks operation as well as improving the quality of drink water supplied to the city. Main problems connected with the water network management are water losses caused by the network damages, unsuitable water pressures on the end user nodes caused by inappropriate work of pump stations installed on the network or by wrong planning of the water net, and a bad quality of produced water caused by incorrect control of the network or by inaccurate planning of water net revitalization. All these problems can be solved in relatively simple way by using new informatics technologies and this idea led to the concept of an integrated ICT system for complex management of communal water networks. The system developed at IBS PAN is now tested in some Polish waterworks.

## ICT SYSTEM DESCRIPTION

The ICT system for water networks consist of four main modules, i.e. of GIS (*Geographical Information System*), SCADA (*Supervisory Control And Data Acquisition*) and CIS (*Customer Information System*) systems and of Computational Module (CM) with several algorithms of applied mathematics (Fig. 1). GIS, SCADA and CIS systems are adapted from

outside and treated as sources of all data describing the water net investigated and its functioning. All calculations solving the tasks of water net management are executed by the Computational Module applications. The communication between the modules and the applications inside the CM module is done via data files.

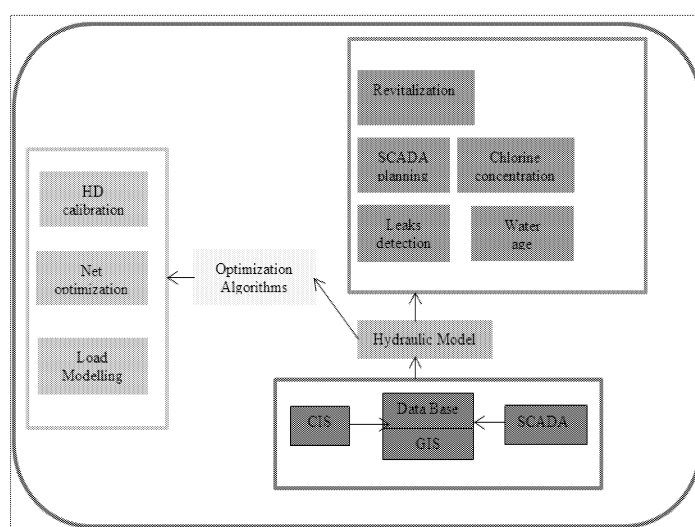


Figure 1. Structure and functions of the ICT system.

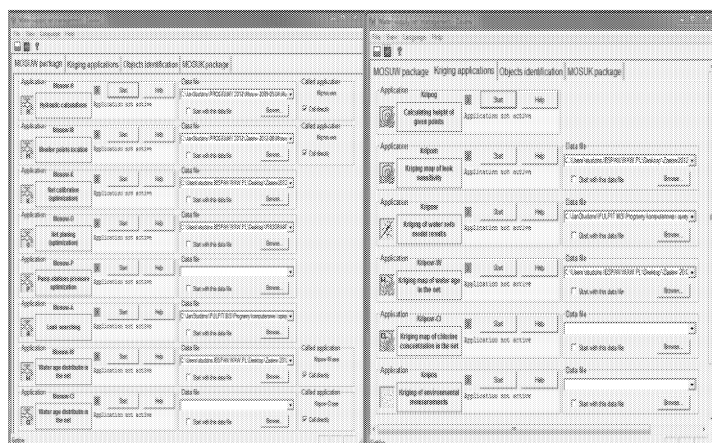


Figure 2. Optimization and kriging applications of the Computational Module of the ICT system.

The CM module includes presently 20 applications with the algorithms of mathematical modelling (Studzinski et al. 2013), kriging approximation (Studzinski and Bogdan 2007) and multi-criteria optimization (Straubel and Holznagel 1999). Concerning the structure of the ICT system the programs are divided into three groups for solving the tasks of optimization, kriging approximation and mathematical modelling (Fig. 2).

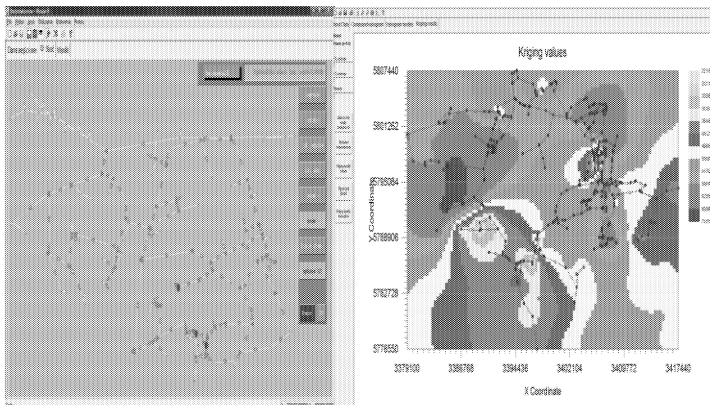
But basing on functionality the programs can be separated in three other groups responsible for hydraulic calculation of the water net, for solving the management tasks for which the optimization is needed and for solving the tasks when only repetitive simulation runs of the hydraulic model must be done (Fig. 1). In the paper only the programs of the last group will be discussed.

## MANAGEMENT TASKS SOLVED BY MEANS OF SIMULATION

One can see that the ICT system presented consists of the programs operating on three levels of the system functionality. The first level contains programs collecting the data concerning the investigated water net and these are GIS, CIS and SCADA systems. On the second level there are programs solving the management tasks with repetitive simulation runs of the water net where the main operating routine is the water net hydraulic model. The third level consists of programs using optimization algorithms for solving more complicated tasks of the water net management. In the paper six programs of the second level of the system functionality are described.

### 1. Hydraulic calculation of the water net

Hydraulic model of the water net plays the main role in the ICT system. It uses in its calculations the data collected in GIS, CIS and SCADA systems and is used as data source by all simulation and optimization programs of the ICT system. The model itself may contain as many as several thousand linear and nonlinear algebraic equations depending on the size of the water network modelled. These equations are solved using the Newton-Raphson algorithm. The network graph is exported to the model from the Branch Data Base of the GIS system and the data concerning the water consumption at the end user nodes of the water net are supplied from CIS. With the model the main parameters characterizing the water net, i.e. the flows and pressures of the water are calculated and with them the state of the network functioning can be quantitatively assessed. Additionally the state of the water network can be assessed qualitatively by using the kriging approximation for designing the colour maps of flow and pressure distributions in the network (Fig. 3).



**Figure 3.** Hydraulic calculation of the water net; results of the hydraulic model (left) and kriging approximation designing the pressure distribution in the network.

### 2. SCADA system planning

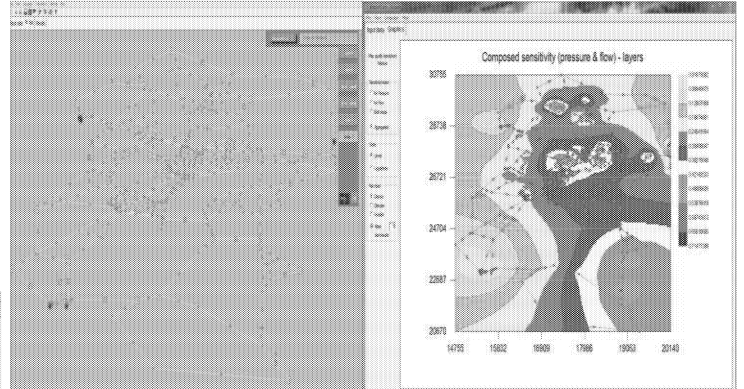
The exactness of hydraulic calculation of the water network depends strictly on the quality of the calibration of the water net model. On the other hand the proper calibration depends on the rightness of the SCADA system located on the network. The

best solution would be the installation of devices for flow and pressure measurements on each pipe and each node of the water net, but this is impossible from the economical point of view. The SCADA system for proper monitoring the water network and the exact calibration of the water net model must be specially planned. It means that with minimal number of measure devices a possibly maximal amount of information concerning the flow and pressure changes in the water net can be obtained. To reach this goal the measure devices of SCADA shall be installed in characteristic points of the network in which the parameter changes occurring even far away from these points can be recorded. To find these characteristic points of the water net successive simulation runs of water leaks in the nodes of the network are performed and the sensitivity against pressure and flow changes is calculated with the following formulas (Straubel and Holznagel 1999):

$$SP_m = \frac{\sum_{k \neq m} (\text{abs}(P_m - p_m) / \max(P_m, p_m)) L_{km}}{\sum_{k \neq m} L_{km}}, \quad SQ_m = \frac{\sum_{k \neq m} (\text{abs}(Q_m - q_m) / \max(\text{abs}(Q_m), \text{abs}(q_m))) L_{km}}{\sum_{k \neq m} L_{km}}$$

where:  $k$  – node with the water leak simulated,  $m$  – potential measurement point,  $P_m$  or  $Q_m$  – pressure or flow value in the node  $m$  without leak simulation,  $p_m$  or  $q_m$  – pressure or flow value in the node  $m$  with the leak simulation,  $L_{km}$  – distance between the nodes  $k$  and  $m$ ,  $SP_m$  and  $SQ_m$  – sensitivity values against pressure and flow changes in the water net nodes.

After the water net sensitivity is calculated the measurement points with highest sensitivity values are chosen for the SCADA system in number depending on the amount of the finances in disposition (Fig. 4).



**Figure 4.** Calculation of the water net sensitivity and the sensitivity map designed with the kriging approximation.

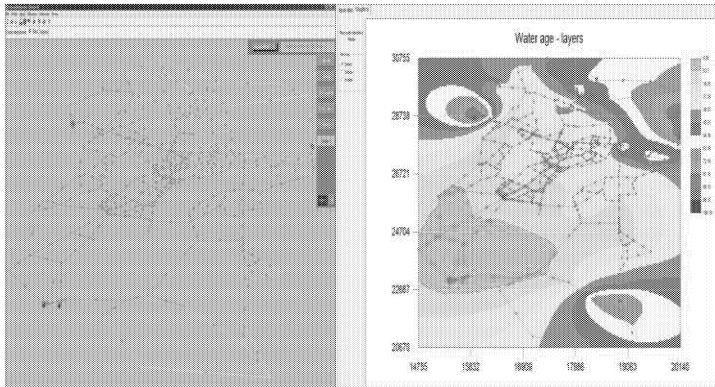
### 3. Water age calculation

Water age in the net decides of the water quality which is worsening with the extending time of water staying in the network, especially in older water nets. While calculating the water age the hourly curves of water consumption in the network nodes and the formulas for mixing the water in nodes and pipes of the net must be given. The water age values are usually established after a couple of days of the simulated work of the water net. For the water mixing the weighted average formula with (1) equal or (2) linear or (3) exponential growing weight coefficient  $w_i$  is used:

- (1)  $w_i = 1,0$
- (2)  $w_i = 1,0 + 0,050 \cdot t$
- (3)  $w_i = \exp(+0,030 \cdot t)$

with  $t$  – time in days.

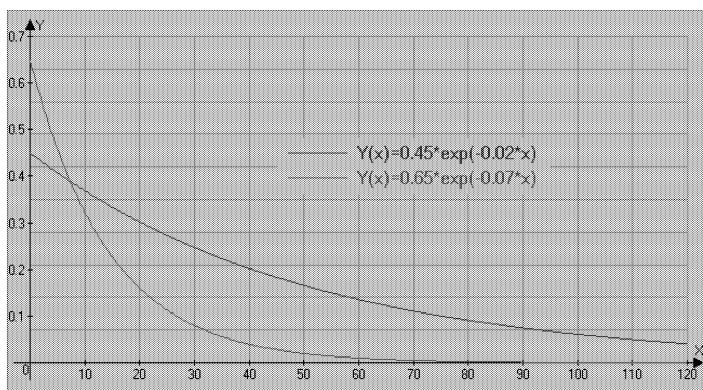
At the start of the program the water age in all nodes and pipes of the network are equal to zero. For the successive time steps the hydraulic calculations are made and the water age for successive water packets moving in the pipes is calculated depending on the water consumption amounts in the end nodes. The mixing of the water occurs in the nodes and there the water changes its age while in the pipes the water packets are only pushed. As the flow velocities in each pipe and the pipe lengths are known one can calculate exactly the movements of the water packets in all pipes for each time step. As the result many water packets with water of different age can be located in the pipes. The results of water age calculation for exemplary water net are shown in Fig. 5.



**Figure 5.** Calculation results of water age and its distribution in the water net designed by kriging approximation.

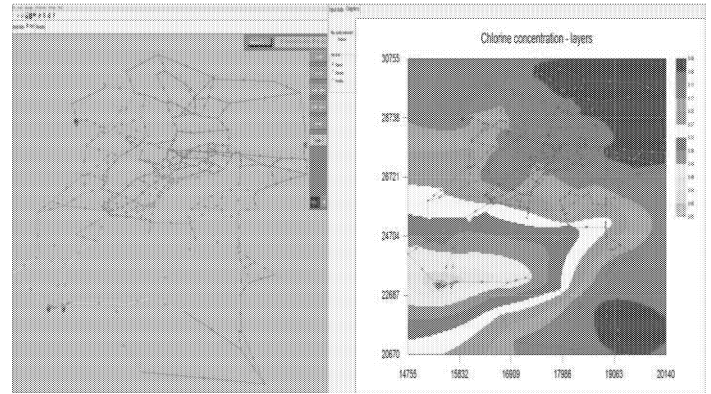
#### 4. Calculation of chlorine concentration

With the water age calculated the natural decrease of chlorine concentration occurring with the time in the network can be computed. The functions for this decrease depends on the chemical composition of the produced water and on the material structure of the water net pipes and the suitable coefficients have to be adapted individually to the specific conditions of the waterworks investigated.



**Figure 6.** Exemplary functions of dropping of chlorine concentration Y in [mg/l] with time X in [h] in network pipes.

In the ICT system presented the exemplary exponential functions used for calculating the dropping of chlorine concentration in the network are shown in Fig. 6. The function coefficients can be easily changed while using the system in other waterworks. Some exemplary results of calculating the distribution of chlorine concentration for a Polish waterworks are shown in Fig. 7. Different concentration values are marked at the water net graph with different colours. For the fast assessment of chlorine contents in the scale of the whole network the map of chlorine distribution is designed by means of kriging approximation.



**Figure 7.** Calculation of chlorine distribution in the water net and its distribution in the network designed by kriging approximation.

#### 5. Water leaks detection and localization

Detection and localization of water leaks in the water net and especially of these hidden ones is one of the most important problems for the waterworks management because of financial costs caused by the water losses. The algorithms that could solve this problem are very desirable by the waterworks management although their implementation into operational practice is connected with costs which are fairly high. There are in general three approaches dealing with this problem.

The first and simplest approach consists in installation of a dense SCADA system on the water net and this SCADA density enables to detect all changes in flow or pressure values in the network which differs essentially from the standard ones. With this approach the water leaks detection is possible but their more exact localization is unfortunately, very difficult.

The second and most complicated approach consists in adding the water net hydraulic model to the SCADA system and in development of a neuronal classifier with which the potential places of water leaks can be fast localized (Rojek and Studzinski 2014). In this approach the hydraulic model describes the real water net and the neuronal classifier models the hydraulic model. It means that identification of two algebraic models of different arts must be performed.

The third approach is an indirect solution between two earlier approaches and it consists of the following steps:

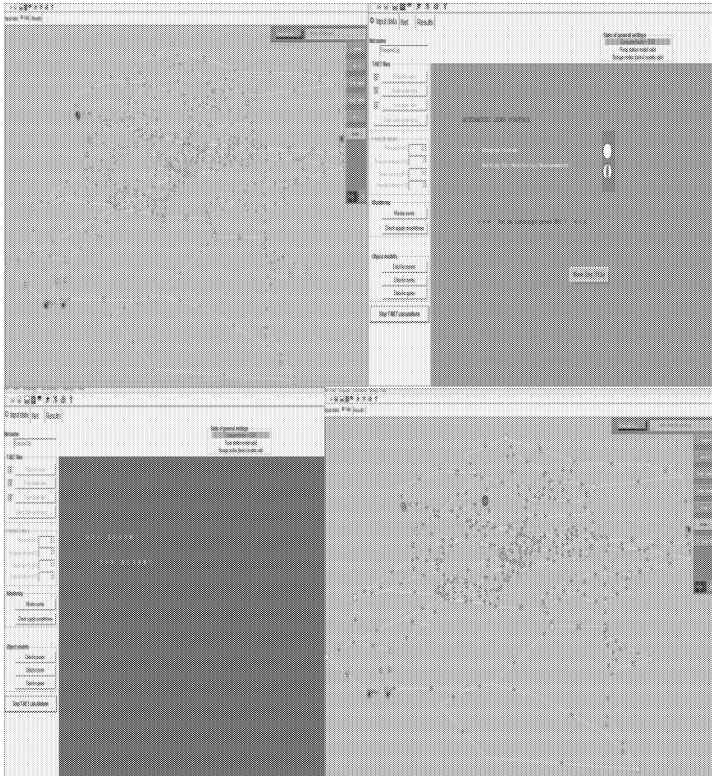
- Development of SCADA system for the water net investigated using the program for SCADA system planning with which the characteristic, most sensitive points can be defined for the network.
- Development of the water net hydraulic model and its calibration with the data gathered by SCADA.
- Development of standard curves of water flows and pressures for all monitoring points of the SCADA system; to do it the moving averages from the last several days can be used.
- Waiting for alarm signal coming from SCADA and informing about deviation between the standard and current value of water flow or pressure observed and recorded at any monitoring point.
- Recording the current distributions of flows and pressures measured on the monitoring points.
- Performing simulation runs of the hydraulic model with water leaks simulated successively in all nodes and pipes of the water net.

- Recording the flow and pressure distributions calculated for the monitoring points for all water leak simulations.
- In the set of the calculated flow and pressure distributions finding out these ones which are most similar to the just measured values; as the criterion for comparing the different value distributions (1) the sum of errors for all monitoring points or (2) minimal value of the maximal error noted on the monitoring points can be used:

$$(1) Cr = \min_i \sum_{k=1}^K (Q_k - Q_{ki})^2 + \min_i \sum_{k=1}^K (P_k - P_{ki})^2$$

$$(2) Cr = \min_i (\max_k |Q_k - Q_{ki}| + \max_k |P_k - P_{ki}|)$$

This third approach of water leaks detection and localization proved to be very effective by testing it on the real data from some Polish waterworks and the illustration of some of its steps is shown in Fig. 8.



**Figure 8.** Water leaks localization; clock wise from the left up: water network with monitoring points; ICT system in stand up state; alarm signal arriving; water leak localized.

## 6. Water net revitalization

Planning the revitalization of the water net can use different approaches for changing the network pipes. In the first case the network revitalization means the exchange of some pipes against the ones with bigger diameters what shall improve, i.e. heighten the water pressures at the end user nodes of the network. To do it the pipes for the exchange have to be chosen and their diameters must be changed by means of an optimization algorithm. The solution of such task with the use of a genetic optimization algorithm is presented for example in (Fajdek et al. 2014).

In another approach to the revitalization task the exchange of some pipes in the water net is needed because of their wrong technical state and usually against the pipes with the same diameters. The goal of this approach is to reduce the liability of the network to break down and, in result, to reduce the potential water losses in the water net. The susceptibility of water nets to

accidents can cause in older municipal waterworks the water losses reaching even up to 30% of the water production and that means essential financial losses for the enterprise (Saegrov 2000). Such revitalization tasks are important for both the water and wastewater nets and in the latter case the reconstruction of the damaged sewage canals protects the environment that would be polluted by the sewage escaping from them. To solve revitalization tasks for wastewater networks the fuzzy sets invented by Zadeh (Zadeh 1965) can be used. Such the solution has been applied in (Sluzalec and Ziolkowski 2013) where for modelling the wastewater network the SWMM model (Rossman 2012) developed by EPA (*US Environmental Protection Agency*) is adopted.

In our case of the ICT system the water net revitalization is realized as the task with the second approach. While planning the revitalization one must decide which pipes are to be exchanged to minimize the water net susceptibility to accidents and at the same time to secure proper functioning of the whole network. The following factors are taken into consideration when choosing the set of pipes to be replaced:

- Technical state of the pipes characterized by their roughness.
- Current durability of the pipes calculated as the difference between the year of pipe construction and the normative pipe durability.
- Pipe liability to break down in percent defined on the base of historical data concerning the pipe damages.
- Risk of the water losses calculated as the pressure in the pipe modified by the pipe diameter:  $p * (1 + d / 500)$ .
- Costs of the pipe revitalization which consists of two components: the costs of the pipe installation and the costs of buying the new pipes.

In order to select the pipes for revitalization from the whole set of the water net pipes the revitalization indicator is calculated from the following formula:

$$IR = w_c * Cn + w_t * (1.0 - Tn) + w_a * An + w_s * Sn$$

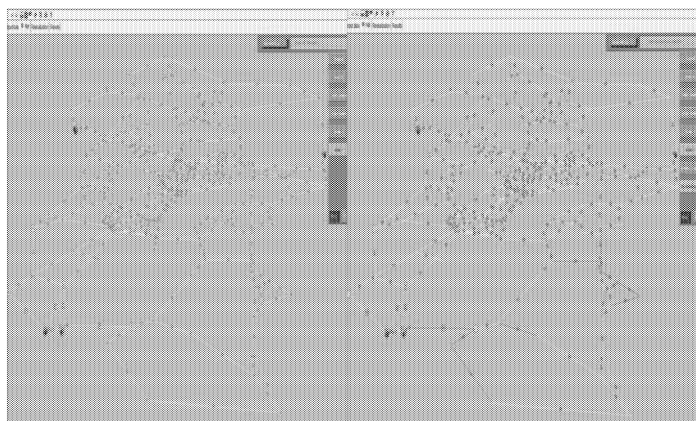
where  $w_c$ ,  $w_t$ ,  $w_a$  and  $w_s$  are weights coefficients,  $Cn$  means pipe roughness,  $Tn$  means current pipe durability,  $An$  is pipe liability to break down and  $Sn$  is risk of the water losses defined for the pipe concerned. The weights coefficients can be chosen arbitrary by the program user and all factors in the formula are normalized in the standardized range of values from 0,0 to 1,0.

After the indicators are calculated for all pipes a ranking list for them can be prepared according to the diminishing indicator values. For the sorted pipes also the costs of their revitalization can be defined considering two costs components. Depending on the financial funds which are at the management disposal one can make choice of the set of pipes for the exchange taking the pipes from the top part of the ranking list and summarizing the costs of their revitalization up to the funds limit.

When the pipes to be exchanged are already selected then the effects of the planned revitalization can be verified by performing the hydraulic calculation for the whole water net with roughness values equal to null for the selected pipes. When the revitalization action is done then the vulnerability of the water net to the accidents will be reduced and the water pressures in some end user nodes will be enlarged.



Some exemplary results of the revitalization planning are shown in Fig. 9 where the pipes selected for the exchange are marked with the green colour.



**Figure 9.** Water net revitalization; water net graph before (left) and after revitalization planning with the pipes indicated for changing.

## SUMMARY

In the paper some algorithms supporting the management of municipal water networks have been presented. The algorithms use in their calculations only hydraulic model of the water net and with the simulation runs of this model several useful management tasks can be realized. These tasks are connected indeed only with planning the water net, like SCADA planning and revitalization algorithms, and with informing about the water net functioning, like calculations of network hydraulics, water age and chlorine concentration, but nevertheless they are important for correct water net operation. More complicated tasks like water net optimization or pumps or tank control need for their solution more sophisticated methods like multi criteria optimization algorithms. An important condition of effective operation of the algorithms described is however their using in strict cooperation with GIS and SCADA systems in frame of a united ICT system. Such the solution is more expensive than individual use of only water net hydraulic models but it makes sure that the management tasks will be done fast, easy, suitably and faultless. Such systems for waterworks are under development at the Systems Research Institute of the Polish Academy of Sciences.

## ACKNOWLEDGEMENT

The paper presented has been realized in the frame of the research project co-financed by the European Union from the European Regional Development Fund, Sub-measure 1.3.1. "Development Projects"; project title: "IT system supporting the optimization and planning of production and distribution of water intended for human consumption in the sub-region of the central and western province of Silesia"; project ref no POIG.01.03.01-14-034/12.

## REFERENCES

- Fajdek, B., Stachura, M., Studzinski, J. 2014. "Optimization of water supply network rehabilitation using genetic algorithms", ISC'2014 Conference, Skovde.
- Rojek, I., Studziński, J. 2014. "Comparison of different types of neuronal nets for failures location within water supply networks", Maintenance and Reliability, Vol. 16, No 1, 2014, 42-47.
- Rossman, L. 2012. „Storm Water Management Model (SWMM) – User's manual", Version 5.0.022", 2012. Online <http://www.epa.gov/nrmrl/wswrd/wq/models/swmm>.
- Saegrov, S. 2005. „Care-W – Computer Aided Rehabilitation for Water Networks", IWA Publishing, Alliance House, London.
- Sluzalec A., Ziolkowski A. 2013. "Planning revitalization of sewage systems using the rules of decision making fuzzy logic", ISC 2014 Conference, Ghent.
- Stachura M., Fajdek B. Studzinski J. 2012. „Model based decision support system for communal water networks", ISC'2012 Conference, Brno.
- Straubel R., Holznagel B. 1999. „Mehrkriteriale Optimierung fuer Planung und Steuerung von Trink- und Abwasser-Verbundsystemen", Wasser•Abwasser, 140, No. 3, pp. 191-196.
- Studzinski J. 2013A. "Entwicklung eines IT-Systems zum komplexen Management von kommunalen Wassernetzen". In: Simulation in Umwelt- und Geowissenschaften (J. Wittmann, M. Mueller, Hrsg.), ASIM-Mitteilung AM 146, 2013, 187-202.
- Studzinski J. 2013B. "IT system for computer aided management of communal water networks by means of GIS, SCADA, mathematical models and optimization algorithms", ICT4S Conference, Zurich.
- Studzinski J., Bartkiewicz L., M. Stachura 2013. "Development of mathematical models for forecasting hydraulic loads of water and wastewater networks". EnviroInfo 2013, Hamburg (B.Page, A.G.Fleischer, J.Goebel, V.Wohlgemuth, eds.), Shaker Verlag Aachen, 736-748.
- Studzinski J., Bogdan L. 2007. "Application of kriging algorithms for environmental and engineering parameters approximation". In: O. Hryniewicz, J. Studzinski, M. Romaniuk (eds.) Environmental Informatics and Systems Research, EnviroInfo'2007 Conference, Warsaw, Vol. 1, Shaker Verlag Aachen, 185-192.
- Zadeh, L.A. 1965. „Fuzzy sets", Information and Control No.8 338-353.

## BIOGRAPHIES

**Marek KUROWSKI** was born in Warsaw, Poland. He studied Electronic Engineering at Warsaw University of Technology, in 1969 he received his M. Sc. Degree in the field of Automatic Control and Robotics. After studies he started to work at the Systems Research Institute of the Polish Academy of Sciences and simultaneously he established his microenterprise WINDSOFT dealing with programming of complex ICT systems.

**Jan STUDZIŃSKI** was born in Warsaw, Poland. He studied Electrical Engineering at Warsaw University of Technology and Applied Mathematics at Warsaw University, in 1989 he received his Ph. D. degree in the field of Automatic Control and Robotics, in 2005 a D. Sc. degree. Currently he works at the Systems Research Institute of the Polish Academy of Sciences.

# OPTIMIZATION OF WATER SUPPLY NETWORK REHABILITATION USING GENETIC ALGORITHMS

Bartłomiej Fajdek

Warsaw University of Technology  
Institute of Automatic Control and  
Robotics

Ul. Sw. A. Boboli 8, 02-525 Warsaw  
E-mail: m.stachura@mchtr.pw.edu.pl

Marcin Stachura

Warsaw University of Technology  
Institute of Automatic Control and  
Robotics

Ul. Sw. A. Boboli 8, 02-525 Warsaw  
E-mail: b.fajdek@mchtr.pw.edu.pl

Jan Studziński

Polish Academy of Science  
Institute of System Research

Ul. Newelska 6, 01-447 Warsaw  
E-mail: jan.studzinski@ibspan.waw.pl

## KEYWORDS

water supply system, optimal rehabilitation, genetic algorithms

## ABSTRACT

Nowadays, genetic algorithms (GA) have become one of the preferred water system designed optimization techniques for many researchers (Cheung 2003; Prasad et al. 2004; Zheng 2013; Bureerat et al. 2013; Jin Xi et al. 2008; Kurek and Ostfeld 2013; Siew et al. 2014; Ostfeld 2013; Wang et al. 2014). The main reason is that GA have the ability to deal with complex nonlinear optimization problems. This paper presents the description of the IT solutions which have been applied in a computer simulation system for optimal rehabilitation of water distribution system. The purpose of this paper is to illustrate the development an optimization model for determining the rehabilitation and replacement of water distribution system components in particular pipes. The Genetic Algorithm (GA) was used to solve the altered multi-objective optimization model. The GA is integrated with the hydraulic simulation program called WDS (Water Distribution System). The short description of the implementation of the WDS software is provided. The WDS software can be used to design new networks, analyse existing networks, perform hydraulic simulation of the networks and etc. The analysis were conducted on a simple hypothetical network of correct planning and implementing network rehabilitation.

## INTRODUCTION

Optimal rehabilitation of water supply system has been a one of the main research subjects for a long time. Most of the water distribution systems were developed to operate during a specific time period. However, during the operation of a water distribution system a number of failures occurs, which are mainly caused by deterioration of pipes an hydraulic components. Thus, it is very important to improve the operation of the water distribution system. Improvement in a system performance can be achieved through replacing, rehabilitating or repairing selected pipes or other components of the system. A method that counter the risk, that is, to minimize the probability of failure of the water supply system is implemented in the form of complex integrated information systems. Nowadays there is a growing number of systems for that purpose. Mainly these are the systems

dedicated to a particular, specified solutions. Moreover the systems that are distributed by renowned international companies are too expensive especially for Polish companies and their functionality is usually limited. Implementation of such systems usually entail the need to hire highly qualified specialists from abroad. Moreover, managers of water utilities face numerous difficulties on account of the fact that expert scientific knowledge is often not readily accessible and is not available in an user friendly manner. Present study attempts to developed a decision support system that will perform the task of correct planning and implementing network rehabilitation. Water network rehabilitation is a complex problem. It is a non-linear, multiobjective problem.

## WATER DISTRIBUTION SYSTEM MODELING SOFTWARE

The WDS software was used for modeling and simulation of the water distribution system. The developed water network modeling environment is based on one of the most famous and appreciated modules called EPANET, which was created by the US EPA (United States Environmental Protection Agency). The EPANET application comes with a package of libraries written in C/C++ which allows to customize the computational engine to suit needs. It is possible to use the supplied function in the own 32-bit applications that can be created in C/C++, Delphi, Pascal, Visual Basic. A specially designed C++ library was created, which integrates WDS application with the computational engine of EPANET library. The developed WDS software in relation to Epanet has been expanded with a number of additional options. The WDS software is an attempt to develop a fully integrated decision support system that will perform both organizational-financial tasks (stock management, accounting, customer service companies, etc.) as well as technical aspects (computer simulations, process optimization, process control, process monitoring etc.). Modern automation systems are integrated with decision support systems and company management systems. In general, company management can be divided into the following layers (Korbicz and Kościelny 2010): measurement and controlling action layer, direct digital control level, supervisory control level, the supervisory, co-ordination and process optimization level, production management layer, business management layer. The proposed decision support system is located mainly in the group of functions for production management. During the design process of the WDS software the following



assumptions were made: the primary task of the system is to perform several functions for managing production of water and the management of water supply company, the system design should be modular to allow its easy extension, selected functions of the system will be available remotely via the Internet (without the need to install dedicated software on the user's machine). The decision support system consists of several interacting software modules. The general structure of presented system is shown in Fig. 1.

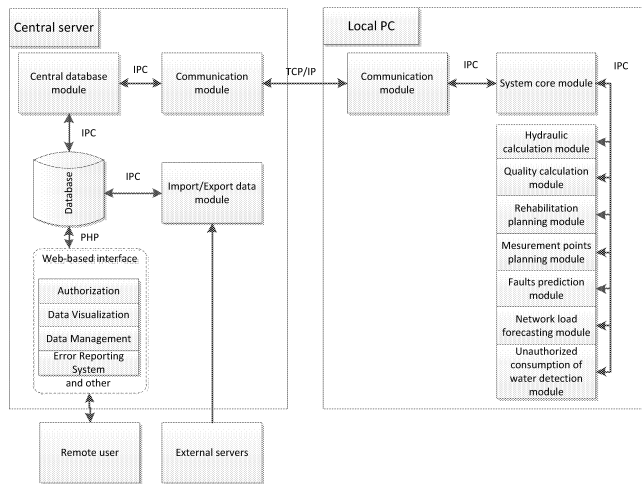


Figure 1: The Simplified Structure of System.

One can distinguish the following elements of the system (Stachura, Fajdek and Studziński 2012):

- a) **system core module** – provides the essential system components that allow operation and integration of all other modules of the system; it contains the links to all the optional modules and includes graphical user interface implementation (the mechanism by which the user communicate with the system), a core module is built around the central database module,
- b) **communication module** – provides the interface to allow the communication between system core module (located on a local PC) and central database module (located on a central server of the system),
- c) **import/export data module** – provides the interface to import/export data to/from external servers; eg import data from GIS (Geographic Information System), import data from water company database etc.,
- d) **central database module** – in the system MySQL database is running on the central server; it is used to store all process related information such as: configuration and internal data of system modules, detailed information on water supply networks, archival information about network demands, archival information about faults in the network etc.; all necessary tables of relational database have been created in SQL (Structured Query Language) and are linked to each other using a primary key,
- e) **hydraulic calculation module** – it allows to perform calculations of water flow in each pipe, the pressure at each node, the height of water in each tank in a network; the hydraulic model is described by linear and nonlinear algebraic equations; to solve equations the hybrid node-loop (Gradient Method)(Todini and Pilati 1987) approach has been applied; the hydraulic calculation module was implemented using EPANET programming toolkit,

- f) **quality calculation module** – it allows to perform calculations of water quality at each node and pipe in a network; water quality solver are based on the principles of conservation of mass coupled with reaction kinetics (Rossman and Boulos 1996); water quality analysis allows to compute a concentration of a chemical species (eg chlorine concentration) throughout the network during the simulation period; in addition water age and source tracking can also be simulated; the water quality calculation module was implemented using EPANET programming toolkit,
- g) **measurement points planning module** – to select correct measurement points for monitoring the water supply network an algorithm given in (Straubel and Holznagel 1999) was implemented; It consist of the following steps: multi criteria optimization for determining the number and location of measurement points (with the following criteria: minimum number of points, maximum sensitivity of leak location, minimum installation cost of measuring points), execution of numerical simulations of hydraulic model for a standard state of the network to determine nodal pressures at all nodes and flows in all section of the network for a given average network load, sorting potential locations in order of their sensitivity to pressure changes in a network,
- h) **faults and unauthorized consumption of water detection module** – the system implements two methods for detecting faults: by comparing the actual demands with archival data form database (obtained from monitoring system), and by comparing actual demands with values obtained from hydraulic simulations of a network,
- i) **faults prediction module** – the are three methods of predicting the possible faults: by estimating a risk of a fault and proposing the activities for the case of fault, by predicting faults and elimination of possible causes of their occurrence and by locating the faults that just occurred so as to minimize losses caused by them; the module to work properly requires a solid monitoring system and an accurate hydraulic model of the network,
- j) **network load forecasting module** –the module implements an algorithm to minimize the energy consumption of waterworks pumps by their time distributed switching on and off and under additional consideration of their wear time and the necessary exchange against the new machines;
- k) **rehabilitation planning module** - it allows to perform the following tasks: identification of technical condition of water distribution system and its individual parts, identifying the most vulnerable for faults sections, generating revitalization scenarios, identifying investment possibilities of a company; module requires a large amounts of archival data that contain information on faults that occurred on a water distribution network and an accurate hydraulic model of the network,
- l) **web-based interface** – the set of modules that run on the www server and web browser side responsible for realizing the graphical user interface to all the data stored in the central database; the module will perform several basic tasks such as: visualization and management of data stored in the central database which includes information from the monitoring system (process variables) and management related data (eg faults

history, revitalization history etc), user management (including the division into the different roles and groups) and advanced permission management system etc.; the module will be implemented using a Web-desktop technique, which is based mainly on JavaScript and Ajax technology; the server part of the module will be implemented in PHP language; using the module we can remotely access the system through the Internet network without installing any additional software.

The GA algorithm that is used to perform the task of optimal network rehabilitations, has been incorporated in a WDS software.

## REHABILITATION ALGORITHM

A well designed water supply system should ensure that water delivered for each customer in the required quantity with the required pressure. This means that under extreme operating conditions occurring in rush hours and hours when there is maximum demand for water, the pressure at each point of the network should be in range between required economic pressure and the maximum permissible pressure. During the maintenance of each water supply system, hydraulic parameters of water pipes change, causing an increase in the hydraulic resistance of the pipes. In order to meet the requirements, that is, to supply customers water with the required pressure, it is necessary to increase pumps head in order to overcome the additional pressure loss caused by the increase of pipes resistance. Another solution to this problem is renovation or replacement of selected network pipes.

Designed algorithm for planning and implementing network rehabilitations is implemented in an iterative form and can be divided into the following parts (Fig. 2):

- Select network nodes in which too small pressure was observed. The parameters of these nodes ( pressure waveform) will be one of the quality indicators of the selected solution.
- Select network links that may be exchanged. This means that those sections that have been identified by an expert as possible to exchange and may affect the pressure in the considered (point a) nodes.
- For each of the selected links, the cost of the technical work related to the replacement and to the replacement of the road surface, if necessary, should be specified.
- Determine catalog of possible for installation network links that contains: diameter, roughness.
- Select network links marked for exchange. Assign the new values of diameters, roughness, etc. (from the catalog of links). Verify the effect of changes using computer simulation.

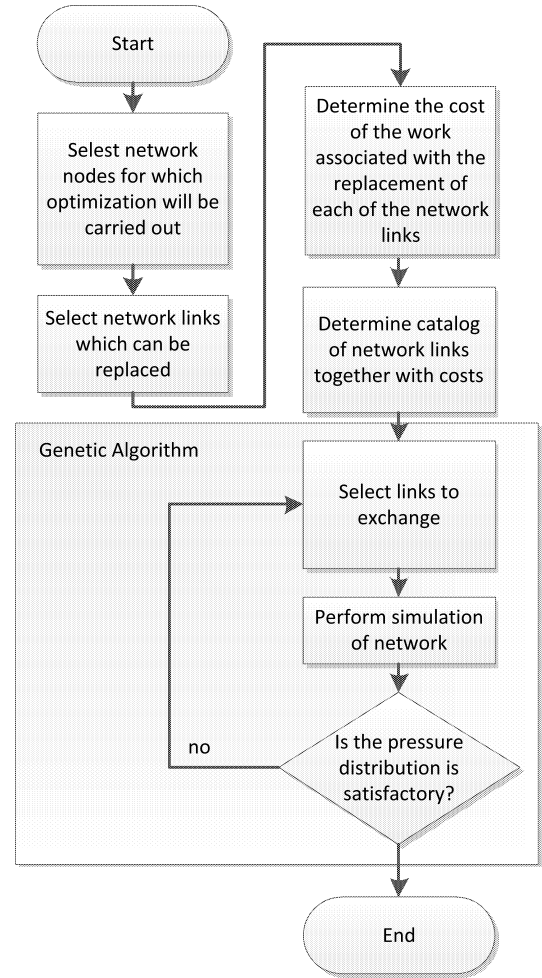


Figure 2: Flowchart of the Algorithm to Optimize the Network Structure.

The presented algorithm requires expert knowledge in the initial phase of the selection of network links for which a replacement is possible. Point (e) can be performed manually or using the optimization algorithm. In the present work a genetic algorithm was used. The following quality indicator was used:

$$Q = \sum_{i=1}^n \Delta p_i k_p + \sum_{j=1}^m k_{wj} + l_j k_{kj} \quad (1)$$

$$\begin{cases} \Delta p_i = 0 & \text{jezeli } p_{min} \leq p_i \leq p_{max} \\ \Delta p_i = p_i - p_{max} & \text{jezeli } p_i > p_{max} \\ \Delta p_i = p_{min} - p_i & \text{jezeli } p_i < p_{min} \end{cases} \quad (2)$$

where:

$Q$  – value of the optimized indicator,

$n$  – number of observed nodes,

$\Delta p_i$  – difference between the set and observed pressure at  $i$  node,

$p_i$  – pressure in  $i$  node,

$p_{min}, p_{max}$  – respectively lower and upper pressure from the assumed range, to which we should bring pressure in  $i$  node,

$k_p$  – weight factor determining the effect of pressure on the objective function,

$l_j$  – length of the  $j$  link,

$k_{kj}$  – unit cost of the  $k$  type link selected from the catalog of links that could be inserted in place of link  $k$ .

The simplest way of determining the best configuration for pipes exchange is to use the exhaustive search method. This method is simple to implement, but it requires checking all existing combinations of subsets of changes for new pipes, selected from the catalog. Using this method a global solution can be obtained, but it is only effective for a set of data with a small number of network sections. Even a small increase in the number of possible pipes exchange causes the exhaustive search to become very inefficient. Therefore the genetic algorithm was used for the selection process. The practical application of evolutionary algorithms requires the adoption of an appropriate presentation method for the links that are selected for exchange. The most natural is to save them in a vector of integers. The essence of this form is to create a one-dimensional array with a number of columns equal to the number of selected links. Each of the array element can take values from 1 to K (number of links in the catalog). During each calculation of the objective function simulation is executed, in which each link from the table is replaced by the link from the catalog.

### APPLICATION EXAMPLE

The rehabilitation study of the hypothetical network in figure 3 was performed. The network was taken from Epanet2 software examples.

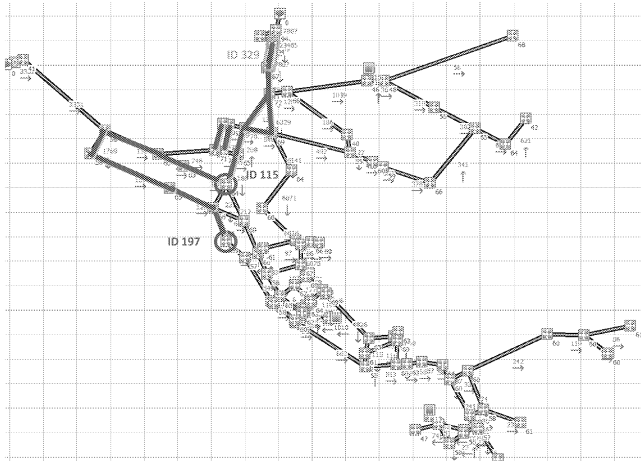


Figure 3: Hypothetical Network.

On the network map (Fig. 3) two junctions (ID: 197 and ID: 115) are marked in which too little pressure was observed. The parameters of these nodes (pressure waveform) will be one of the quality indicators of the selected solution. The waveforms of pressure in the selected nodes before the rehabilitation process are shown in Figure 4 and 5.

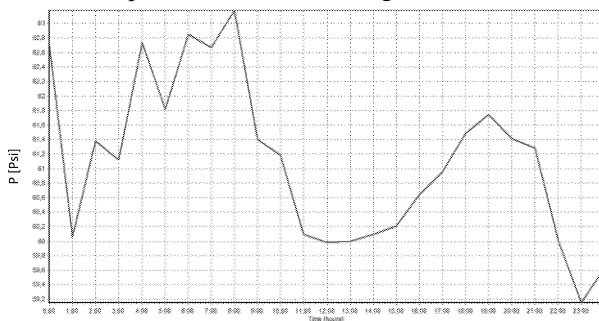


Figure 4: Pressure waveform in node with id 197.

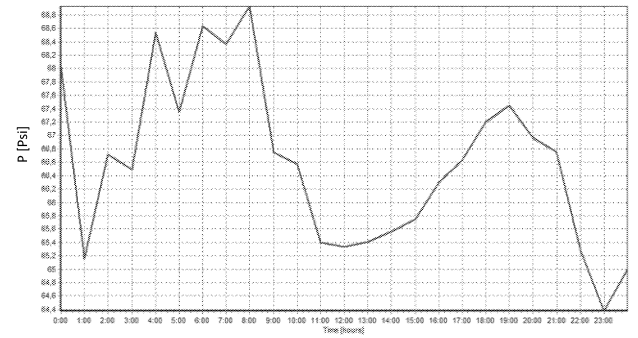


Figure 5: Pressure waveform in node with id 115.

The network links that may be exchanged are marked on the network map. The links with the following IDs were selected: 225, 111, 109, 105, 103, 107, 115, 119, 121, 122, 125, 123, 120, 329, 303, 301.

The links that are in the catalog are presented in Table 1.

Table 1: Catalog of links

ID	Diameter [inch]	Roughness	Lay cost
1	8	130	2
2	10	130	2,2
3	16	130	3
4	10	100	2,5
5	20	130	3,5
6	40	130	4

For this application example, following parameters were chosen for GA algorithm:

- population size: 100,
- crossover probability: 0.7,
- mutation probability: 0.01,
- generation size: 2000.

The following results were obtained from the rehabilitation process of the hypothetical network: only one pipe with ID 329 was selected for exchange with the pipe with ID 6 from the catalog. The pipe is marked on the network map (Figure 3). The rehabilitation process was repeated 10 times. For each of the tests the same pipe was selected for exchange. The new pressure waveform, for selected network junctions obtained from the rehabilitation process, are presented in Figure 6 and 7.

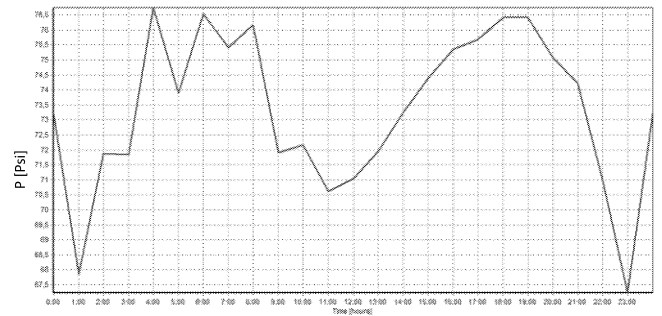


Figure 6: Pressure waveform in node with id 197.

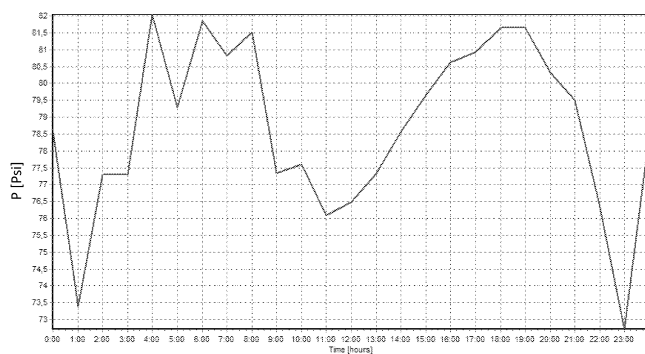


Figure 7: Pressure waveform in node with id 115.

From the results it can be observed that after the rehabilitation process the average pressure in two nodes was increased. In the node with ID: 197 from approx. 60 [psi] to 73 [psi] and in the node with ID: 115 from approx. 66 [psi] to 78 [psi].

## CONCLUSION

The paper presents the description of the IT solutions which have been applied in a computer simulation system for optimal rehabilitation of a water distribution system. The Genetic Algorithm (GA) was used to solve the altered multi-objective optimization model. Moreover, the short description of the implementation of the WDS simulation software was provided. The implemented system performance was presented using a simple example of rehabilitation process applied to the hypothetical network. The results confirm the effectiveness of the algorithm for automatic rehabilitation of the water distribution system. In the near future an implementation of the proposed system is planned in Głubczyce (town in southern Poland, app. 13 000 citizens).

## REFERENCES

- Bureerat, Sujin, and Krit Sriworamas. "Simultaneous topology and sizing optimization of a water distribution network using a hybrid multiobjective evolutionary algorithm." *Applied Soft Computing* 13.8 (2013): 3693-3702.
- Cheung, Peter B., et al. "Multiobjective evolutionary algorithms applied to the rehabilitation of a water distribution system: a comparative study." *Evolutionary Multi-Criterion Optimization*. Springer Berlin Heidelberg, 2003.
- Jin, Xi, et al. "Multi-objective optimization of water supply network rehabilitation with non-dominated sorting Genetic Algorithm-II." *Journal of Zhejiang University SCIENCE A* 9.3 (2008): 391-400.
- Korbicz, J., Kościelny, J.M. 2010. Modeling, Diagnostics and Process Control: Implementation in the DiaSter System. Springer, Berlin, 384 p.
- Kurek, Wojciech, and Avi Ostfeld. "Multi-objective optimization of water quality, pumps operation, and storage sizing of water distribution systems." *Journal of environmental management* 115 (2013): 189-197.
- Montero, C., Villanueva, A., Raclavsky, J., Sulcova, V., Leif Sigurd, H., Freni, G. 2002. Wastewater rehabilitation technology survey. *Computer Aided Rehabilitation of Sewer Networks* (CARE-S).
- Ostfeld, Avi, Nurit Olikier, and Elad Salomons. "Multi-Objective Optimization for Least Cost Design and Resiliency of Water

Distribution Systems." *Journal of Water Resources Planning and Management* (2013).

- Prasad, T. Devi, and Nam-Sik Park. "Multiobjective genetic algorithms for design of water distribution networks." *Journal of Water Resources Planning and Management* 130.1 (2004): 73-82.
- Siew, Calvin, Tiku T. Tanyimboh, and Alemtsehay G. Seyoum. "Assessment of Penalty-Free Multi-Objective Evolutionary Optimization Approach for the Design and Rehabilitation of Water Distribution Systems." *Water Resources Management* 28.2 (2014): 373-389.
- Stachura, M., Fajdek, B., Studziński, J. 2012. Model Based Decision Support System for Communal Water Works. *Industrial Simulation Conference 2012* (Brno, June 4-6).
- Todini, E., Pilati, S. 1987. A gradient method for the analysis of pipe networks. *International Conference on Computer Applications for Water Supply and Distribution* (Leicester Polytechnic, September 8-10).
- Rossman, L.A., Boulos, P.F. 1996. Numerical methods for modeling water quality in distribution systems: A comparison. *J. Water Resour. Plng. and Mgmt.*, Vol. 122, No. 2, 137-146.
- Straubel, R., Holznagel, B. 1999. Mehrkriteriale Optimierung fuer Palnung und Steuerung von Trink- und Abwasser-Verbundsystemen. *Wasser Abwasser*, Vol. 140, No. 3, 191-196.
- Wang, Qi, Dragan A. Savić, and Zoran Kapelan. "Hybrid metaheuristics for multi-objective design of water distribution systems." *Journal of Hydroinformatics* 16.1 (2014).
- Zheng, Feifei, Aaron C. Zecchin, and Angus R. Simpson. "Self-adaptive differential evolution algorithm applied to water distribution system optimization." *Journal of Computing in Civil Engineering* 27.2 (2012): 148-158.

## BIOGRAPHIES

**BARTŁOMIEJ FAJDEK** was born in Warsaw, Poland. He studied at Warsaw University of Technology, in 2007 he received a M. Sc. Degree in the field of Automatic Control and Robotics. After studies he started to work at the Warsaw University of Technology.

**MARCIN STACHURA** was born in Szczecin, Poland. He studied at Maritime Academy of Szczecin, in 2004 he received a B. Sc. degree in the field of studies in the field of Mechanics and Machine Construction. Next he moved to Warsaw, and studied at Warsaw University of Technology. In 2007 he received M. Sc. degree in the field of Automatic Control and Robotics. After studies he started to work at the University. He received Ph. D. degree in the field of Automatic Control and Robotics in 2011.

**JAN STUDZIŃSKI** was born in Warsaw, Poland. He studied at Warsaw University of Technology, in 1989 he received a Ph. D. degree in the field of Automatic Control and Robotics, in 2005 a D. Sc. degree. Currently he works at the Institute of System Research of the Polish Academy of Sciences.

## ACKNOWLEDGEMENTS

The paper presented has been realized in frame of the research project financed by the Polish National Centre for Science (NCN), project ref. no. N N519 6521 40.

# **BUSINESS SIMULATION**



# THE BUSINESS OF OUTER SPACE CROQUET

William Conley  
Austin E. Cofrin School of Business  
480B Wood Hall  
University of Wisconsin at Green Bay  
Green Bay, Wisconsin 54311-7001  
U.S.A.  
[Conleyw@uwgb.edu](mailto:Conleyw@uwgb.edu)

## KEYWORDS

Complex systems, nonlinear optimization, environmental defense, business, organization

## ABSTRACT

Asteroids hitting the earth and exploding causing considerable devastation are rare events. However, they are possibly becoming more frequent because of the increasing number of asteroids in the asteroid belt near the planet Jupiter (every random crash of two asteroids breaks them up into more, smaller fragments that can still be dangerous to the earth). Presented here will be four nonlinear optimization problems, all directed to at least partially solving this dangerous asteroid problem for the earth. The multi stage Monte Carlo optimization solution approach will be used to solve the four examples. Additionally, the point will be made that even though asteroid events on the earth are infrequent they are potentially very devastating and can be environmental catastrophes. One only needs to look at the February 2013 asteroid explosion near Chelyabinsk, Russia to realize that this problem will be an ever present danger until the space powers of the world take more aggressive action to prevent these disasters from ever occurring. The idea with “outer space croquet” is to ram into the asteroid to knock it off line without using bombs. The business side of this endeavor will also be presented.

## INTRODUCTION

The four examples presented will be a 12 asteroid attack on a big bad asteroid that is headed toward the earth. The 12 asteroids will smash into the bad asteroid to knock it off its collision course line with the earth.

Additional examples will include an engine tuning optimization problem where a new rocket engine used for asteroid defense is adjusted for optimal performance. Also, an optimization problem, including the purchase of dozens of new telescopes at minimum cost to be used to search for the dangerous asteroids will be presented.

These will be followed by a mass attack of hundreds of rockets (launched from the earth) to intercept a dangerous asteroid and knock it off of its line to the earth. The idea of knocking a dangerous asteroid off line (rather than trying to bomb it) is like the lawn game of croquet where opponents knock each other’s croquet ball off its intended line. This “outer space croquet” is already being “played” in the asteroid belt near Jupiter, with random crashes of asteroids into each other. The world’s space powers should start playing this game to better protect the earth from a future asteroid disaster like the ones that hit in Russia in 1908 and 2013 (where many people were injured). Let us look at some examples after a few preliminaries about “croquet” and multi stage Monte Carlo optimization.

## THE GAME OF OUTER SPACE CROQUET

The lawn game of croquet involves each player hitting a round wooden ball around a fixed course through a series of wickets. The first player to hit his or her ball through all of the wickets in order wins. However, a very effective defensive strategy is for the players to hit their balls into the ball of the player who is leading, to knock it off its intended line. Another good strategy is to go ahead to a future wicket and block it with one’s ball to stop the leader from ever getting through there.

Outer space croquet is using these ideas, to ram into the big bad dangerous asteroid (or rock) with other asteroids or rocks (that are already in outer space) to knock the bad asteroid off its dangerous line. Additionally, with enough telescopes to map the intended paths of the dangerous asteroids, the earth can send asteroids and rockets down the bad asteroid’s intended course to block its path to the earth. Let us look at four examples after a brief discussion of multi stage Monte Carlo optimization.

## MULTI STAGE MONTE CARLO OPTIMIZATION

Multi stage Monte Carlo optimization (MSMCO) is like regular Monte Carlo optimization in that a random sample of thousands of feasible solutions is looked at and the best one is stored. However, with MSMCO that is just stage one. Then centered about this “best answer so far,”

another set of thousands of feasible solutions is looked at in a reduced search region centered about this best answer so far. Then a stage three search in a further reduced region (centered about the best answer produced in stage two) is performed. This process is continued for as many stages as necessary until the multi stage Monte Carlo process converges to the exact optimal or a useful approximation. The two hundred variable famous test problem number 33 (in the mathematics literature) was solved for its exact optimal (Conley, 1991a) using this process.

### **IS THE ASTEROID THREAT INCREASING?**

The conventional wisdom is that so many, many milleniums ago there were only a few hundred asteroids being held in check by Jupiter's enormous gravity. Now it appears that there are millions of asteroids up there (of course of a smaller size on the average). However, because they are smaller in size, that seems somewhat comforting at first glance. It means though, that because there are so many of them, random collisions between them are much more likely now than ages ago. These random collisions are probably the chief cause of asteroids heading in the general direction of the earth from time to time.

The seriousness of this is the fact that even a relatively small asteroid exploding over the earth could cause a terrible catastrophe. One author (Vergano, 2013) writing about the asteroid explosion over Chelyabinsk, Russia also noted that another asteroid of about 50 meters or less missed the earth by 17,000 miles at about the same time. The prediction was that if that had hit the earth's atmosphere and exploded it would have been equivalent to a 4 megaton blast that could possibly cause global panic and tremendous damage.

Even though smaller asteroids are less dangerous than bigger ones, it is still possible that a "small" one could be devastating. Note also that the Hubble telescope was only functioning in space for a little while and it observed a collision between two asteroids near Jupiter (Maugh, 2010) and (DeVorkin and Smith, 2008).

This should be a wake-up call to the human race that these infrequent "asteroid events" are dangerous and may become more common. Building and deploying 100 new telescopes on the earth and orbiting the earth, the moon, Mars, and Venus would be a start toward being able to find and chart the path of all of these potentially dangerous asteroids, comets and meteors.

The world (and all the countries' space programs) have much of the technology (and we can develop the rest) needed to deflect these dangers away from the earth.

However, it is absolutely vital that we know where each and every one of the dangerous objects are (and what their paths of motion are). Then we will stand a much better chance in defending our planet from a disaster from outer space.

### **ANOTHER WARNING FROM THE MOON**

Over the years and centuries so many people have looked at the earth's moon and wondered what caused the rough surface that faces the earth. Was it now extinct volcanoes or asteroids, comets and meteors constantly smashing into it (or both)?

Once the space age began the astronauts who orbited the moon noticed that the so-called "far side" of the moon was very much rougher than the side that always faces the earth. (Note one side of the moon does constantly face the earth because the moon takes the same amount of time to rotate once as it does to revolve once around the earth). This observation by the astronauts tends to point toward the far side of the moon being bombarded more frequently by comets, meteors, and asteroids than the near side which faces the earth. This must be because an asteroid headed toward the near side of the moon is probably more likely to get pulled into a collision with the earth than one that is headed toward the far side of the moon.

Also, with no vegetation on the moon to cover the results of these impacts, the record is there for all to see and compare now that we can see all of the extra damage on the far side of the moon. Given that the earth's gravity is so much more powerful than the moon's, the earth must have been hit by asteroids, etc., so many more times than the moon. However, our oceans and vegetation cover up most of the evidence from these collisions.

This should serve as an additional warning to humanity to deal with the "asteroid" threat now and in the future. We must be prepared, by having more telescopes to find and map the paths of all of these harmful objects. Also, ways of diverting them from colliding with the earth must be developed and deployed. Some new technology may need to be developed and deployed also.

Doing nothing about the danger is unacceptable now and it is unacceptable for all future generations. We must be prepared internationally to prevent this threat from ever harming the earth again.

### **THE SAFE KINETIC ENERGY APPROACH**

The Apollo asteroids are the ones that cross the earth's orbital path. The smaller ones would of course totally or partially burn up in the earth's atmosphere and do little or



no harm. However, a large one, possibly 100 feet or more in diameter, could do much harm to the earth. The one that wiped out the dinosaurs (as the theory goes) almost destroyed the earth. The 1908 asteroid explosion in Russian Siberia leveled a forest of a bit less than 1000 square miles (Stone, 2008) and it was only about 150 feet in width.

The conventional wisdom is that if a large asteroid showed up in the future, the earth could suffer a major catastrophe (Stone, 2008). Presented here is a way to potentially prevent this type of disaster from ever occurring. It is introduced and discussed with a hypothetical example.

## THE FUTURE TWELVE ASTEROID ATTACK

About 35 or 40 years from now a big bad asteroid (about 2000 feet across) is headed dangerously toward the earth. However, the earth, fortunately, has spent considerable time, effort, expertise and money on preparing to deal with an asteroid event of this potentially disastrous nature. Various chemical and bombing schemes are going to be the second line of defense (which hopefully will not have to be used). The asteroid attack center that the world established on Mars (with much foresight) in 2030 will coordinate the attack with “safe” kinetic energy weapons that do not have to be launched from the earth because they are already in outer space.

After the first successful landing of a spacecraft on an asteroid (NASA mission) (Bell and Mitton, 2002) in February 2000, the world has become quite adept at using asteroids for mining and landings, etc. Therefore, twelve smaller asteroids were selected and space probes landed on them and fitted them with the latest high tech engines and guidance systems controlled by the Mars asteroid Central Outpost with some help from Florida, Houston, Europe’s French Guiana launching site, Russia, Japan, China and India’s two bases at Tumba in Kerela and Vishakhapatnam on the Bay of Bengal.

The plan is for the twelve smaller asteroids (all 50 to 100 feet across) to have their engines turned on at precisely the correct time after New Year’s so all 12 of the little asteroids smash into the big bad asteroid at one minute after midnight on January 31 of next year at the point (10000000, 10000000, 10000000) above Mars (the units are in miles). This mass kinetic energy bombing attack will be perpendicular to the line of motion of the large dangerous asteroid to maximize the likelihood that it will be knocked a half of a degree or a degree off course from its path toward the earth. That should be enough to drive it far away from the earth. Notice the goal is not to destroy it, but to knock it off course. If the attack should break the big asteroid up into smaller pieces that are still

deemed dangerous, an armada of rockets will leave the earth to bomb the pieces into harmless pebbles before they get too captured by earth’s gravity.

Specifically, the third 10,000,000 miles coordinate is on the Z-axis that goes from the north pole of Mars to the North Star. The other two axes are on a plane tangential to Mar’s North Pole, with all three axes perpendicular to each other. Therefore, the 12 small asteroids will hit the big asteroid about  $(300)^5$  million miles from Mars (or 17,320,508 miles above Mars) at one minute after midnight (Greenwich Mean Time) on January 31 of the new year at the previously described (10000000, 10000000, 10000000) point.

Therefore, the equations of motion of the twelve smaller asteroids (for the first thirty-one days of the new year) are:

$$\begin{aligned} D_{1,i} &= 15,500,000 - 100,000i - (3.9 - .1i)X_i \\ D_{2,i} &= 14,240,000 - 240,000i - (3.5 - (12-i).11)X_i \\ D_{3,i} &= 13,170,000 - 170,000i - (3.1 - .2i)X_i \end{aligned}$$

for  $i=1,2,3 \dots 12$  and the  $D_1, D_2, D_3$  motion components are the X, Y, Z coordinates for each asteroid respectively and the  $X_i$ ’s for  $0 \leq x_i \leq 2,592,060$  ( $i=1,2,3, \dots 12$ ) are the times in seconds into the new year that the twelve little asteroids have their engines turned on and head straight toward the (10000000, 10000000, 10000000) crash point. Therefore, what should those  $X_i$ ’s times be if the asteroid speeds toward their target are speed  $(s)=(20,200-500i)$  miles per hour for the first 1,550,000-50000i miles traveled with the engine on and speed  $(s)=18000+340i$  miles per hour after that (to the target) for  $i=1, 2, 3, \dots 12$  little asteroids.

A 35-stage run of the general purpose multi stage Monte Carlo optimization simulation solution technique produced the following solution in Table 1.

Table 1: Solution Printout

	$X_i$	error <sub>i</sub>
1	1953655.88013	.00021
2	1969554.23776	.000031
3	1977575.30642	.000054
4	1977480.71786	.000039
5	1969340.76262	.000021
6	1953588.77505	.000022
7	1931015.28859	.000023
8	1902695.82568	.000022
9	1869870.24379	.000030
10	1833808.13345	.000037
11	1795694.36927	.000014
12	1756554.61553	.000004
		.0003 total error in seconds

## SOLUTION EXPLANATION

Little asteroid one will have its engine turned on and head for the rendezvous point at 1953655.88013 seconds into the new year G.M.T. There are  $60 \times 60 \times 24 = 86,400$  seconds in a day.

Therefore  $1953655.88013 / 86400 = 22.61176$  days into the new year or January 22<sup>nd</sup> at 40 minutes and 56 seconds after 2 pm in the afternoon G.M.T.

Therefore, in a little more than a week later little asteroid one will smash into the big asteroid at the previously mentioned point of (10000000, 10000000, 10000000) above Mars at the same time as the other 11 little asteroids will with very little error in time (.000021 of a second). Similarly, the 11 days and times for the other little asteroids could also be determined.

A hypothetical 12 asteroid attack on a big asteroid that threatened the earth in the future was presented and discussed, with the goal of making this kind of asteroid defense (and other means of defense) available in the future before we need them. Let's use the space agencies of the world, in partnership with business and industry, in an international effort to see that the earth is forever safe from asteroids in the future. The solution technique used here (Multi Stage Monte Carlo Optimization) MSMCO is general purpose and is also featured in (Conley, 2012) on a million variable nonlinear optimization problem.

## TELESCOPES

Three different suppliers of telescopes (for use in searching for dangerous comets, meteors and asteroids) have been identified. The three different suppliers are located in the country or regions 1, 2, and 3 which have the largest and best developed space programs. The plan is to buy a total of no more than 1000 telescopes of each kind with at least 50 coming from each of the three suppliers (to keep the three big space powers involved long term in telescope development). The specifications are such that telescope 1 costs 5 million dollars each for one or  $5x_1^{.89}$  where  $x_1$  is the total number purchased. It has a 95% chance of working with no problems. Telescope 2 costs 6 million dollars each and has a 97% chance of working with no problems. However, telescope 3 only costs 1 million dollars but has an 85% chance of working with no problems.

Each telescope will be deployed to search for dangerous comets, meteors and asteroids and map their orbit. Given a budget of 3 billion dollars, how many of each telescope should be purchased so that the average success rate of

working is 90%? We solve the telescope system of equations

$$\begin{aligned} 5x_1^{.89} + 6x_2 + x_3 &= 3000 \text{ millions of dollars} \\ (.95x_1 + .97x_2 + .85x_3) / (x_1 + x_2 + x_3) &= .90 \\ \text{subject to} \\ 50 \leq x_i &\leq 1000 \text{ for } i = 1, 2, 3 \end{aligned}$$

With a multi stage Monte Carlo optimization (MSMCO) computer program to

Minimize

$$C = \left| 5x_1^{.89} + 6x_2 + x_3 - 3000 \right| + \left| .95x_1 + .97x_2 + .85x_3 - .90(x_1 + x_2 + x_3) \right|$$

Subject to

$$50 \leq x_i \leq 1000 \text{ for } i = 1, 2, \text{ and } 3$$

With a 21 stage MSMCO simulation optimization run drawing 25,000 feasible solutions at each stage.

Two runs of the program yielded the solutions

$x_1 = 439$ ,  $x_2 = 194$ ,  $x_3 = 172$  with equation errors of .01 for equation 1 and .07 for equation 2

and  $x_1 = 308$ ,  $x_2 = 253$ ,  $x_3 = 662$  with equation errors of .06 for equation 1 and .01 for equation 2.

Buying, building and deploying more land based and outer space based telescopes like the Hubble one (DeVorkin and Smith, 2008) is one of the most important aspects of asteroid comet and meteor defense. The article by (Morrison, Harris and Boslough, 2013) mentions that one of the problems in detecting the Chelyabinsk meteor (or asteroid rock) was that it was hidden by the sun. Therefore, it is essential that we have orbiting telescopes around our moon and Mars and its moons, etc. to give us more views for early warning.

## ASTEROID ROCKET ENGINE ADJUSTMENT

A new powerful rocket engine that is inexpensive and relatively small and can be mass produced and deployed for asteroid defense has been developed. The plan is to land on the big bad asteroid (as the USA's NASA landed on an asteroid in February of 2000) (Bell and Mitton, 2002) and attach one or more of the engines to aim them so that they alter its orbit to drive the asteroid away from the earth. Additionally, the engine may be used in quantity for attack rockets that will be sent out to deflect the dangerous asteroid away from the earth.

Hundreds of test runs of the rocket engine were done by setting the 14 controls at the values 1, 2, 3 ...150 and running the rocket and measuring its thrust. The resulting fitted equation to be maximized (for maximum thrust using a 16 stage MSMCO simulation optimization drawing 1,000,000 sample answers at each stage) is

$$\begin{aligned} \text{Thrust} = & 2,003,376 - x_1^2 + 228x_1 \\ & -x_2^2 + 116x_2 - x_3^2 + 75x_3 \\ & -x_4^2 + 162x_4 - x_5^2 + 194x_5 \\ & -x_6^2 + 246x_6 - x_7^2 + 76x_7 \\ & -x_8^2 + 290x_8 - x_9^2 + 126x_9 \\ & -x_{10}^2 + 84x_{10} - x_{11}^2 + 236x_{11} \\ & -x_{12}^2 + 104x_{12} - x_{13}^2 + 72x_{13} \\ & -x_{14}^2 + 204x_{14} \end{aligned}$$

subject to  $x_{11} + x_{12} + x_{13} + x_{14} \leq 300$  for safety because the engineers feel exceeding 300 could cause the engine to burn out or explode.

A short run on a desk top P.C. produced the solution for the control settings of

$$\begin{aligned} x_1 &= 107 & x_2 &= 51 \\ x_3 &= 68 & x_4 &= 74 \\ x_5 &= 97 & x_6 &= 123 \\ x_7 &= 38 & x_8 &= 145 \\ x_9 &= 63 & x_{10} &= 42 \\ x_{11} &= 116 & x_{12} &= 50 \\ x_{13} &= 34 & x_{14} &= 100 \end{aligned}$$

Maximum thrust = 1,893,550

Therefore, setting controls 1, 2, 3, ... 14 at 107, 51, 68 ... 100 should produce an engine thrust of about 1,893,550 pounds and still have  $x_{11} + x_{12} + x_{13} + x_{14} \leq 300$  to ease the engineers concerns about the overheating of the engine.

## LARGE ATTACK ON A BIG BAD ASTEROID

A big bad asteroid is a threat to explode on the earth. The ideal situation would be to hit it hundreds of millions of miles away up in the near Jupiter asteroid belt. However, it was not noticed until it got away from there.

Therefore, the International Asteroid Defense Group is going to launch up to 1500 of its attack rockets (which are currently in a holding pattern) orbiting the moon. It is estimated that (because the attack rocket's cargo bays are loaded with small asteroids to add mass), if 900 of them strike the big bad asteroid that should be enough force to be 99+% sure of diverting the bad asteroid from its earth-threatening path. Table 2 gives a summary of the 12 rocket types available orbiting our moon.

Table 2: Summary of 12 Rocket Types Available

Rocket Type	Replacement Cost Per Rocket (in Millions of Dollars)	Chance of Hitting The Target
1	$(100 - .30)x_1$	.98
2	$(100 - .32)x_2$	.97
3	$(100 - .34)x_3$	.96

4	$(100 - .36)x_4$	.95
5	$(100 - .38)x_5$	.94
6	$(100 - .40)x_6$	.93
7	$(100 - .42)x_7$	.92
8	$(100 - .44)x_8$	.91
9	$(100 - .46)x_9$	.90
10	$(100 - .48)x_{10}$	.89
11	$(100 - .50)x_{11}$	.88
12	$(100 - .52)x_{12}$	.87

How many of each of the 12 different rocket types should be launched to assure success (.999 probability) and minimize cost given that they will fire the rocket escape engines and launch at least 35 of each type of rocket but no more than 350 of each kind?

## MULTI STAGE MONTE CARLO OPTIMIZATION

An 18 stage multi stage Monte Carlo optimization (MSMCO) drawing 1,000,000 answers at each stage produced a solution of 94.262 billion dollars for the following

$$\begin{aligned} x_1 &= 350 & x_2 &= 179 & x_3 &= 43 & x_4 &= 60 \\ x_5 &= 43 & x_6 &= 43 & x_7 &= 39 & x_8 &= 38 \\ x_9 &= 36 & x_{10} &= 38 & x_{11} &= 35 & x_{12} &= 36 \end{aligned}$$

where the  $x_i$ 's for  $i = 1, 2, 3 \dots 12$  are the number of rockets of each type to launch in our mass rocket attack on the bad asteroid to be pretty sure that at least 900 hit the target.

The 900 figure has been the estimated amount of rocket crashes into the asteroid necessary to guarantee a successful deflection of it away from the earth.

## ASTEROID CORRAL AT THE MOON AND ELSEWHERE

The idea of an asteroid corral has been mentioned earlier by (Conley, 2011). However, here the idea is that as the asteroid threat may be increasing (and Chelyabinsk shows the danger that even small asteroids present), perhaps several asteroids corrals at the moon, on one of the Mars moons and perhaps even on a moon of Jupiter might be good future projects to protect the earth. Even though these could be costly long term international projects, the money would be well spent by a management team with considerable expertise. (A top notch international business management team is essential.)

These small asteroids orbiting the moon (with engines attached to them) could then go on attack to knock any

earth-threatening asteroid off of its intended path to the earth.

Also, rockets could be in these holding pattern orbits, with big empty cargo bays to collect some of these asteroids to add rocket mass. The rockets themselves could be sent to smash into the dangerous asteroid to knock it away from the earth.

Additionally, “Hubble like” telescopes could be put in these orbits to search the skies for dangerous objects. This would give the astronomers on earth more views of our solar system and perhaps better prepare the world to protect our earth.

## **FUNDING AND EXPERTISE**

The international effort to make the earth safe from dangerous exploding comets, meteors and asteroids should look to the business and industrial worlds and the scientific communities and governments to organize the business of “outer space croquet” (using the rocks that are already in outer space) to block the path of the big bad asteroid headed toward the earth. Alternatively the rocks up there could just ram into the big bad rock to knock it off of its line, if it is on a collision course with the earth. Both of these moves are common in the enjoyable lawn game of croquet.

The outer space “croquet balls” are already up there in our solar system (so heavy launches from the earth aren’t necessary to carry the rocks). The rockets that go into space just collect and organize the rocks that are up there to make sure the earth wins the game of OUTER SPACE CROQUET. Another great source of help in defending and protecting the earth could be wealthy industrialists, business people and scientists who are interested in space exploration (Vergano, 2012).

## **CHELYABINSK, RUSSIA**

The asteroid or meteor that exploded in February, 2013 about 50 miles from downtown Chelyabinsk, Russia (a city of over a million people) sent a shock wave (Kramer, 2013) and (Stone, 2013) into Chelyabinsk that caused wide spread damage and hundreds of injuries to people. The devastating explosion of a 150 foot wide asteroid in Siberia in 1908 (Stone, 2008) that leveled a forest of about one thousand square miles could not have been prevented, given the telescopes and “rockets” available in 1908.

However, the one over Chelyabinsk 105 years later possibly could have been prevented (Conley, 2010) (Conley, 2011), and (Vergano, 2013) if more telescopes were in place on the earth and the moon and orbiting the

earth and the moon. This could be in conjunction with rockets and asteroids ready to attack the bad asteroid to prevent it from ever reaching the earth.

One possibility is to land on the dangerous asteroid and mount an engine or engines on it that can be fired to change its orbit so it misses the earth. Also (Wichowski, 2013) recommends placing the engines so that they fire in the opposite direction to the asteroid’s path to the earth to slow it down. This would also change its orbit and additionally, if it were going slow enough, it would never explode, completely minimizing the danger to the earth.

One of the world’s more well-known newspapers had its lead front page article (complete with color photographs of the explosion and subsequent devastation) on the asteroid disaster in Chelyabinsk (Barry and Kramer, 2013). More planning and cooperation internationally between business, industry and governments with the space programs could perhaps prevent these dangerous events from ever occurring. Then the reporting on them could be moved to the back pages of newspapers reflecting the mitigation and/or elimination of this problem.

## **CONCLUSION**

Thousands of researchers, engineers, and business people are either currently involved in asteroid defense or are considering getting into this business. The international effort, guided by the world’s space powers, is the approach being advocated, even though many individuals are also working on the problem. All four problems were solved with the multi stage Monte Carlo optimization (MSMCO) solution technique. Additional examples of MSMCO’s power and versatility are in (Conley, 1989), (Conley, 1991b), (Conley, 2007) and (Conley, 2008).

Presented here were four hypothetical problems. A twelve rocket attack on a dangerous asteroid was launched from Mars to deflect it from the earth. A new rocket engine tuning optimization problem was presented. A minimum cost purchase of a large number of telescopes to be used in asteroid mapping was discussed. Additionally, a mass rocket attack from the earth was launched toward an earth-threatening asteroid.

The business and industrial world will be heavily involved in asteroid defense. Businesses will make profits while helping to remove an extremely dangerous long term threat to our planet once and for all.

The author has former students from Chelyabinsk, Russia. This asteroid threat is real. Business, industry and academia can solve this problem, or prevent it from ever occurring again. One hopeful sign is that the world in

general is starting to take this asteroid problem more seriously (Micklethwait, 2013) and survey ways for prevention and/or mitigation of these potentially devastating disasters coming from outer space.

There are hundreds of large enough asteroids (or comets or meteors) cruising through our solar system that could devastate our civilizations if they exploded in the earth's atmosphere. However (Stone, 2013), says that none of them are headed close enough to the earth to be a problem.

However, what about the ones that our telescopes have not detected yet? Additionally, the new threats from the "outer space croquet" being played at random in the asteroid belt (Mars and Jupiter areas) could emerge at any time. We need hundreds more telescopes to watch our sky. The former director of NASA's manned space flight center, Eugene Kranz's book title Failure is Not an Option in dealing with the famous Apollo 13 flight could and should apply to asteroid defense (Kranz, 2000).

## REFERENCES

- Barry, E. and Kramer, A. 2013. "Meteor Explodes, Raining Fragments on Siberia." *New York Times*, Saturday, Feb 16<sup>th</sup> Edition, New York, p. 1.
- Bell, J. and Mitton, J. 2002. ASTEROID RENDEZVOUS: NEAR Shoemaker's Adventures at Eros. Cambridge University Press, Cambridge, England.
- Conley, W. 1989. "A Shipping Route Minimization Technique with Thirty and One Hundred Point Examples." *Journal of the Japanese Society of Simulation Technology*, Vol. 8, pp. 40-44.
- Conley, W. 1991a. "Multi Stage Monte Carlo Optimization Applied to a Two Hundred Point Traveling Salesman Problem." In Proceedings of 1991 Summer Computer Simulation Conference, Baltimore, 1991a. SCS, San Diego, pp. 145-151.
- Conley, W. 1991b. "Programming an Automated Punch or Drill." *International Journal of Systems Science*, Vol. 22, No. 11, pp 2039-2056.
- Conley, W. 2007. "Simulation Optimization and Correlations with Multi Stage Monte Carlo Optimization." *The International Journal of Systems Science*, Vol. 38, No. 12, pp. 1013-1019.
- Conley, W. 2008. "Ecological Optimization of Pollution Control Equipment and Planning from a Simulation Perspective." *International Journal of Systems Science*, Vol. 1, pp. 1-7.
- Conley, W. 2010. "The Large Asteroid Problem from an Industrial Management Perspective." *Industrial Simulation Conference, ISC 2010*, Budapest, June 9-12. EUROSIS-ETI, Ghent, Belgium, pp. 50-54.
- Conley, W. 2011. "Momentum, Probability and Staffing Calculations for Comet, Meteor and Asteroid Defense." In Proceedings of 2011 Simulation Conference, Gemarich, Portugal. EUROSIS, Ostend, Belgium, pp. 29-36.
- Conley, W. 2012. "A Million Variable Nonlinear Transportation Problem." In Proceedings of the 2012 Industrial Simulation Conference. ISC2012, June 2012, Brno Czech Republic, EUROSIS, Ostend, Belgium, pp. 113-121.
- DeVorkin, D. and Smith, R. 2008. Hubble Imaging Space and Time. National Geographic Society, Washington, D.C.
- Kramer, A. 2013. "After Assault from Heavens, *New York Times Sunday*, Feb. 17, 2013 Edition, p. 6.
- Kranz, E.F. 2000. Failure is Not an Option. Simon & Schuster, New York and London.
- Maugh, T. H. 2010. "Aftermath of Asteroid Collision Examined." *Los Angeles Times*, Feb. 5, 2010.
- Micklethwaite, J. 2013. "Defenders of the Earth: Stopping Asteroid Strikes." *The Economist*, June 29, 2013, p. 72.
- Morrison, D., Harris, A. and Boslough, M.. "The Chelyabinsk Event of February 15, 2013." *Skeptical Inquirer*, Vol. 37, No. 4, July/August 2013, Amherst, New York, pp. 27-31.
- Stone, R. "Target Earth." *National Geographic*, Washington, D.C. August 2008, pp. 134-149.
- Stone, R. 2013. "Siberian Meteor Spurs Dash for Data, Calls for Safeguards." *Science*, Vol. 339, March 8, 2013, p. 1135.
- Vergano, D. 2012. "Ultra Rich Soar to New Heights." *U.S.A. Today*. Thursday, May 10<sup>th</sup>, 2012, Arlington, Virginia, p. 3A.
- Vergano, D. 2013. "Experts: With Enough Warning, Earth Can Handle Asteroid Strikes." *U.S.A. Today*,

Friday, Feb. 16 Edition, Arlington, Virginia, p. D3.

Wichowski, M. 2013. Personal communication.

#### **AUTHOR BIOGRAPHY**

**WILLIAM CONLEY** received a BA in mathematics (with honors) in 1970, an MA in mathematics in 1971 from Western Michigan University and an MS in statistics in 1973 and a PhD in mathematics (computer statistics) in 1976 from the University of Windsor. He has taught mathematics, statistics and computer programming in universities for 35 years. He is currently professor emeritus of Business Administration and Mathematics at the University of Wisconsin at Green Bay. The developer of multi stage Monte Carlo optimization and the CTSP correlation statistics, he is the author of five books and 203 publications worldwide. He is a member of the American Chemical Society, a fellow in the Institution of Electronics and Telecommunication Engineers and a senior member of the Society for Computer Simulation. Career highlights include presenting a paper at a National Aeronautics and Space Administration (NASA) conference in Houston and a speech at the NASA Goddard Space Flight Center outside of Washington, D.C. on the occasion of the tenth anniversary of the data center.

# INTERACTIVE SIMULATION FOR DISCOVERING INVESTMENT SCENARIOS IN ENERGY SYSTEMS

N.B. Szirbik

G.B. Huitema, R.H. van der Burg, J.C. Wortmann

Department of Operations

Faculty of Economics and Business

University of Groningen

The Netherlands

{n.b.szirbik, g.b.huitema, r.j.van.der.burg, j.c.wortmann}@rug.nl

## KEYWORDS

Energy, Management games, Model design, Gaming,  
Interactive simulation

## ABSTRACT

This paper shortly presents the advantages of using an interactive simulation platform for the study of multi-stakeholder investment scenarios in the energy industry, where real or emulated stakeholders participate in simulation sessions as players. The aspect of flexibility in energy system is discussed and the link to investments in energy infrastructure is explained. The platform elements are presented, and some research outcomes are concluding the paper.

## INTRODUCTION

Given the current state of the art of the various technologies used in the energy industry and the related research, there are multiple opportunities and scenarios for investment. Companies in this industry and also governments have all a high stake in the exploration of how to invest in future energy infrastructures. One view is that the investments should lead to the transition from a traditional, hierarchical energy system or grid to a decentralized, *smart energy system* (National Energy Technology Laboratory, 2007). The main long-term objectives for this evolutionary change are meeting global and local sustainability goals, saving energy and resources, and in the end, a reduction of high-value and risky investments. From a very general point of view, a smart energy system today has become an approach in which the physical grid goes in parallel with an overlaying information network enabling *flexibility*. This is achieved (partially) by accommodating the integration of renewables, the usage of locally distributed energy resources, and smart and distributed (local) matching of demand and supply.

One of our current research directions is oriented towards the long term investment patterns that affect the operational flexibility of an energy system that expands its biogas facilities (production, storage, conversion, transport, and trade). To study these patterns, we are using complex simulation models that are capturing the structure and behavior of the extant gas infrastructure, and also attempting to emulate the behavior of the players and stakeholders in the gas (and gas to electricity) trade. We have realized early that one of the most difficult features to verbalize, operationalize, and formally model for simulation, but also for practical

operational purposes, is the (quasi non) functional feature of flexibility. This is a feature that is extremely challenging to operationalize and quantify. One of the reasons is that there is a too diverse universe of discourse if flexibility is discussed in strategic terms, medium time horizon terms, and operational terms, all in the same time. At the operational level in an energy system, in very simple words, flexibility means the “degree of choice” that the operational decision makers in the system have when faced with common problems like surges in demand, intermittence in supply, technical failures, predictions of highly negative outcomes in the near future, etc. However, one can immediately link the aspect of flexibility from operational to higher levels, because the very nature and composition of the infrastructure in the energy system (which is a strategic, long term investment-driven issue) may give an intrinsic potential for better flexibility. Amongst many, there are three crucial elements that contribute to the increase in flexibility in any energy system: the capacity for energy storage, physical network connectivity (including conversion capabilities between various energy carriers), high speed (automated even) trading at multiple levels of granularity and locality, from big volumes to micro-volumes, respectively from regional level to household level. There is a strong interconnection between these three elements (Pelletier and Wortmann, 2009), for example stored energy can be traded in advance, and in the absence of connectivity, energy cannot be transferred from storage to the point of demand.

Another fact that was early clear to us, is that it is extremely difficult -basically impossible- to model the strategic, investment-driven behavior of the human components in the system, the ones who are responsible for investment. The single solution we found to achieve a degree of realism in the behavior of the investment related stakeholders is to let real humans play their roles themselves, which means to transform a batch type of simulation into an *interactive simulation* (like a sort of serious game, but one that has a very realistic physical model of the infrastructure, and also an economics-based representation of the markets involved). The purpose of such a simulation becomes twofold. On the one hand, we study the investment patterns and the changes in the simulated business network of stakeholders and players that bring the highest levels of flexibility in the future – due to infrastructural and business relationships changes. On the other hand, we give the real stakeholders and players an opportunity to participate in the interactive simulation, and learn how to better collaborate in a way that make them understand better each other and discover goals

(like better future flexibility) that exist outside their weltanschauung, but which in the end contribute in the long term to their own goals.

## REAL-LIFE EXAMPLE OF INVESTMENT SCENARIO

### Example: Flexibility in Biogas Industry

An interesting and relevant real-life scenario to be enacted and analyzed by an interactive simulation is the investment process in a form of flexibility in a biogas energy system. The starting point of this scenario assumes a structural imbalance in the energy system of various types of gas (fossil, synthetic, and biogas), due to the lack of connectivity, conversion, trading, and storage. An investment in a specific flexibility related asset can be a way to deal with this imbalance. Flexibility options are regarded differently by the stakeholders involved, for example traders need the capacity for energy storage, grid operators are interested in various forms of physical network connectivity, and brokers need forms of high speed trading, or interruptible contracts. It is expected that such investments in flexibility are not straight forwardly involving only one stakeholder, but these decisions are affected by a form of *network externalities* (Katz and Shapiro, 1986), making it a complex multi-stakeholder investment problem.

This crucial issue of multiple stakeholders with different and potentially conflicting interests has a decisive influence on the investment process and outcomes, making it interesting for analysis notwithstanding. All kind of questions may rise when it comes to such multi-stakeholders investments in flexibility, like for example: Which are the stakeholders who face the most significant problems due to imbalances? Who takes (or should take) the initiative to invest in a form of flexibility? Which form of flexibility suits best each stakeholder? Is it for stakeholders interesting and relevant to cooperate to make the long term investments fruitful and

have a win-win outcome? Which stakeholders have incentives to cooperate and which do not? Which stakeholder has more influence on investment decisions and the outcome of the scenario? What types of decisions and negotiations foster or discourage investments? What problems occur and what type of constructs are negotiated and agreed? Enacting this scenario in multiple interactive simulations sessions can give interesting insights in the potentially best investment patterns, and also yield some answers to the above mentioned questions.

## INTERACTION SIMULATION PLATFORM

### Development of the simulation platform

To study these questions, an interactive simulation platform (ISP) has been developed in a partnership consisting of academic and industrial partners. It consists of three main parts: the physical representation, the underlying physical and economical models, and the knowledge and behavior capture system. To realize an interactive simulation session, an investment scenario skeleton is used as the starting point for the development of the physical representation. The skeleton scenario formally establishes who the stakeholders are, what physical elements exist a priori, and what commodities are traded. For example, in the biogas industry the main stakeholders incorporated in the skeleton scenario are a number of biogas producers, consumers, a regulator, a network operator, and a fossil gas producer/energy retailer. The commodities involved are biogas, natural gas, electricity (that produced out of biogas) and the economic output produced by factories of the consumers. The physical elements incorporated in the ISP include gas infrastructures and other related assets like pipes (several quality and capacity variants), gas conversion and upgrading plants, digesters and storage for biogas, gas using factories, and natural gas sources. A map where all these assets can be placed visualizes these physical elements, as in the figure

You are test3@gmail.com, playing grid operator. Round 1: Jan 2022 Leave Game

New Chat  
Public Chat  
GGP1 - Grid op. X

Hi grip operator  
Hello  
I would like to have a connection with your grid to inject my biogas. How much do you charge for the connection fee and capacity tariff?  
You can connect if your biogas meets our standards. Therefore you need a substation. The connection fee is 200 and the capacity tariff is 10 per unit per round.  
thanks

Send

Bank Account  
Cash 2650  
Loan 0  
Balance 2650

+ New Loan  
Invoices  
Grid  
Substations

Your turn will end in 9:34. The next turn is for the balancing shipper. Finish Turn >  
Press F1 for Help



above. The situation on the map is a real time visualization of the current state of the interactive simulation and it changes when stakeholders interact and realize new investments. In addition to these physical elements the communication and control features exist which enable stakeholders to communicate with each other (chat function), and provide the opportunity to deploy assets, and make investments in order to influence the outcome of the simulation session. Besides these, there is some general information provided, like for example the number of the playing round (a simulation session is a series of rounds), the playing time left of a round, the expected amount of natural gas present, etc. A screenshot of the 'simulation board', as it appears to the playing stakeholders is depicted in the figure above. The simulation is running on a server and each player can access its board via any internet browser. The researchers play roles as observers, experimenters who introduce new elements in the simulation on the fly, and there is one role for the simulation master (named Deus ex Machina, or DEM), who controls the whole session and can intervene if necessary.

### Implementation aspects

Behind the user interface, there are economic and gas flow models that are implementing the dynamics of the simulation. These models link into an interactive, coherent simulation platform all the physical elements, the commodities and the roles of the stakeholders/agents - all with their specific characteristics and some regulated behavior. At the end of every round, the models compute all the gas flow that are realizable with the given infrastructure and also provide financial data in the form of invoices and settlements of accounts and contracts of players. The economic and gas flow models, and the technical characteristics of the physical elements have been defined beforehand by studying a series of potential investment scenarios, and each novel scenario skeleton needs such a study and setup. The scenario skeleton is modeled by using an interaction model, defined in the agent interaction modelling language TALL (Stuit and Szirbik, 2009). One of the main novelties in our approach is that it is augmented with Natural Language Processing (NLP) techniques for analyzing the interactions (chats) that take place between the stakeholders during the session (Ittoo et al, 2013). These chats can be "private" (stakeholder-to-stakeholder chat "rooms") or "public" in a general chat "room".

In order to execute an interactive simulation session, the available players are assigned by the DEM to the specifically defined roles for stakeholders. The session is played in multiple rounds of a pre-determined length of time, mimicking a preset number of years in real life. The strategy of using the ISP can vary from having sessions in a distributed and asynchronous setting, or in a central and synchronous setting with all players together in a room, like a LAN party. Synchronously played, a session may take several hours, with rounds during 15 or 20 minutes for example. Asynchronously played, a session can take several weeks or even months to be played, with separate rounds for each day (or longer). When the session has ended, the NLP analysis and various post-session statistics (e.g. percentage of green gas, transaction flows, financial performance, intention and interest of stakeholder, etc.) provide

information and specific characteristics of the session. This can help to gain in-depth insights in the intentions, decisions, and performed activities of players, together with their related consequences. Since the process of capturing and analyzing the stakeholders' interactions via NLP, and also the post-session statistics generation are completely automated, this makes immediate analysis and discussions between players after a synchronous session possible.

### CONCLUSION

The development of the ISP has gone through a series of successive prototypes (from 1 to 4), each realized in periods of a half a year. Starting from prototype 2, it was possible to execute relevant simulation sessions with the participation of players that can emulate realistically the behavior of real stakeholders. Multiple investment patterns have been studied, and insights in the intentions and decision making of the players have been analyzed.

### FUTURE WORK

In the immediate future, we need to operationalize the aspect of collaborative investment behavior between stakeholders, and see what the relation of this aspect is to the before-mentioned aspect of flexibility. Another aspect to expand in more detail in the future is the effect of network externalities on investments in energy flexibility.

### ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of Wilrik Mook and Vincent Velthuis in developing the ISP and in providing valuable feedback that have improved the quality of this article.

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

### REFERENCES

- Ittoo, A, N.B. Szirbik, G.B. Huitema, and J.C. Wortmann. 2013. "Simulation gaming and natural language processing for modelling stakeholder behavior in energy investments." *In Proceedings of the 2013 Industrial Simulation Conference*, Ghent University, Ghent, Belgium, 23-27.
- Katz, M.L. and C. Shapiro. 1986. "Technology Adoption in the Presence of Network Externalities". *Journal of Political Economy* 94, No. 4, 822-841.
- National Energy Technology Laboratory. 2007. "A system view of the modern grid". Technical report. U.S. Department of Energy - Office of Electricity Delivery and Energy Reliability.
- Pelletier, C. and J.C. Wortmann. 2009. "A risk analysis for gas transport network planning expansion under regulatory uncertainty in Western Europe". *Energy Policy* 37, No. 2, 721-732.
- Stuit, M., and N.B. Szirbik. 2009. "Towards agent-based modelling and verification of collaborative business processes: An approach centred on interactions and behaviours". *International Journal of Cooperative Information Systems* 18, No. 3-4, 423- 479.

# Modelling customer preference heterogeneity to iPad attributes using a Finite Mixture procedure

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

## KEYWORDS

Market Segmentation, Latent Class Mixture Model, EM algorithm, Maximum Likelihood Estimation

## ABSTRACT

The identification of segments in strategic market planning has long been recognized as a powerful tool to understand consumer behaviour. An approach that has managerial appeal in addressing market heterogeneity is by assuming that customers can be grouped in a number of unobserved homogeneous segments where customers in each cluster have similar purchasing behaviours. This paper describes the different procedures in affecting market segmentation focusing more on the Finite Mixture approach, while the application addresses heterogeneity issues in customer preferences when purchasing iPads given demographic and product-related predictors.

## 1. INTRODUCTION

Traditionally, market segmentation have been conducted either by using priori segmentation in which the number of segments are determined before the data collection or post hoc segmentation in which the segments are identified by forming groups of consumers that are homogeneous along a set of measured characteristics. One of the most used post hoc methods is the two-stage approach in which a conjoint regression model is fitted for each respondent and utilities (regression coefficients) are estimated for each level of each attribute for every person. Segments are then generated by conducting cluster analysis of the individual-level utilities. The main problem with the two-stage approach is that different clustering techniques produce different segments in which the initial utility estimation method and the subsequent cluster analysis optimize different and unrelated objective functions.

In response to the limitations of these traditional clustering methods, several integrated conjoint segmentation methods were proposed where the estimation and the segmentation stage are conducted concurrently. Hagerty (1985) proposed

a Q-type factor analysis to partition the respondents and showed that the method reduces the variance of individual parameter estimates without unduly increasing the bias of the estimates. Kamakura (1988) uses the same general approach by pooling respondents who are similar in terms of their conjoint full-profile responses, but employs an agglomerative clustering algorithm. He showed that his approach improves predictive accuracy at the individual respondent level. (Ogawa 1987) presented a stochastic logit framework to model rank order responses. The model uses a hierarchical, non-overlapping clustering method and estimation and segmentation are conducted concurrently. (DeSarbo et al., 1989) proposed a clusterwise regression procedure that uses a simulated annealing algorithm for optimization. (Spath 1982) proposed a non-hierarchical, clusterwise regression procedure to identify homogeneous groups in terms of the relationship between dependent and independent variables. (Wedel and Kistemaker 1989) proposed a generalization of the clusterwise regression by extending Spath's method to handle more than one observation per individual. Their procedure uses an exchange algorithm, developed by Banfield and Bassil to maximize the likelihood and yields non-overlapping, non-hierarchical segments. (Wedel and Steenkamp 1991) used a fuzzy clusterwise regression algorithm to partition the data by minimizing the residual sum of squares criterion, which represents the sum of the distances of subjects from the regression equations in all clusters.

The development of new techniques for segmentation in the area of finite mixture (latent class) models stands out to be the most far-reaching developments in the early 90's. The work of (Kamakura and Russell 1989), (DeSarbo et al., 1992) and (Wedel and DeSarbo 1995) brought major changes in market segmentation applications in theory and practice. Finite mixture models address heterogeneity through a discrete distribution where estimation is carried out by maximizing the likelihood function. The main advantage of these models is that they address market heterogeneity by assuming a number of unobserved clusters.

Managers seem to be comfortable with the idea of market segments, and the models tend to do well in identifying useful groups. Another advantage of latent class models is that they enable statistical inference where estimation and segmentation are carried out simultaneously. A study conducted by Vriens, Wedel, and Wilms (1996) found that finite mixture models had the best overall performance of nine conjoint clustering methods (which included both post hoc and integrated conjoint segmentation methods) in terms of parameter recovery, segment membership recovery and predictive accuracy.

Recent changes in the market environment presented new challenges and opportunities for market segmentation. The introduction of micro marketing, direct marketing and mass customisation enabled marketers to customize their products or services to very small groups of customers. This implied that estimation and predicted responses to marketing variables had to be conducted at the individual level rather than the segment level. Bayesian estimation methods in marketing have gained popularity in the last ten years and are used extensively in various marketing problems. Besides providing a set of techniques that allow for the development and analysis of complex models they can estimate models at the individual level in which heterogeneity is addressed through a continuous rather than a discrete distribution. While the conceptual appeal of Bayesian methods have long been recognized, the recent popularity arises from computational and modeling breakthroughs. Hierarchical Bayesian estimation was rarely used in the past due to the fact that it could only be applied to simple models since the class of models for which the posterior inference could be computed was no larger than the class of models for which exact sampling results were available. The technical problems in applying the method to complex models seemed insurmountable.

During the last ten years, simulation methods, particularly Markov chain Monte Carlo (MCMC) methods have overcome these computational constraints for a wide range of marketing models. The classic work of (Roberts and Casella 2004), (Gelman et al., 2004) and (Rossi et al., 2006) contributed considerably towards this shift in interest in Bayesian estimation. A study conducted by (Andrews, Ansari and Currim 2002) compares the relative efficiency of Finite Mixture and Hierarchical Bayes conjoint analysis models in terms of fit, prediction, and parameter recovery. The authors show that both modelling techniques are equally effective in recovering individual-level parameters and predicting rating evaluations. They found that the two modelling techniques produce good parameter estimates both at the individual and segment levels. Moreover, the authors show that the two models are robust to violations of underlying assumptions and that traditional individual-level models tend to overfit the data.

## 2. FINITE MIXTURE MODEL FRAMEWORK

Let the random variables  $\mathbf{y}_j = (y_{jk})$  for  $j = 1, \dots, n$  and  $k = 1, \dots, K$ , belong to a super-population which constitutes a mixture of a finite number ( $I$ ) of sub-populations in proportions  $\pi_1, \dots, \pi_I$ , where it is not known in advance from which class a particular vector of observation arises. The probabilities  $\pi_i$  follow the constraint:

$$\sum_{i=1}^I \pi_i = 1, \pi_i \geq 0, i = 1, \dots, I \quad (1)$$

Assume that the conditional probability density function of  $y_{jk}$  given that  $y_{jk}$  comes from class  $i$ , takes the form:

$$f_{j|k|i}(y_{jk} | \theta_{ijk}, \lambda_i) = \exp \left\{ \frac{y_{jk} \theta_{ijk} - b(\theta_{ijk})}{a(\lambda_i)} + c(y_{jk}, \lambda_i) \right\} \quad (2)$$

for specific functions  $a(\cdot)$ ,  $b(\cdot)$  and  $c(\cdot)$  conditional upon class  $i$  and  $y_{jk}$  are independently distributed with canonical parameters  $\theta_{ijk}$  and means  $\mu_{ijk}$ . The dispersion parameter  $\lambda_i$  is assumed to be a known constant over observations in class  $i$ , while  $a(\lambda_i) > 0$ . The predicted value  $\mu_{ijk}$  is linked to the linear predictor  $\eta_{ijk}$  through the link function  $g(\cdot)$  such that in class  $i$ :

$$\eta_{ijk} = g(\mu_{ijk}) \quad (3)$$

where the linear predictor comprises  $P$  covariates  $\mathbf{X}_1, \dots, \mathbf{X}_P$  where  $\mathbf{X}_p = (\mathbf{X}_{jkp})$ ,  $p = 1, \dots, P$  and the parameter vectors  $\beta_i = (\beta_{ip})$  in class  $i$ .

$$\eta_{ij} = \sum_{p=1}^P X_{jp} \beta_{ip} \quad (4)$$

Conditional upon class  $i$ , a generalized linear model consists of a specification of the distribution of the response variable  $y_{jk}$ , a linear predictor,  $\eta_{ijk}$  and a function  $g(\cdot)$  which links the random component to the systematic component. The unconditional probability density function of an observation vector  $\mathbf{y}_j$  can then be expressed in the finite mixture model form:

$$f_j(y_j | \Phi) = \sum_{i=1}^I \pi_i \prod_{k=1}^K f_{j|k|i}(y_{jk} | \beta_i, \lambda_i) \quad (5)$$

where  $\Phi' = (\pi', \beta', \lambda')$ ,  $\pi = (\pi_1, \dots, \pi_I)'$ ,  $\lambda = (\lambda_1, \dots, \lambda_I)'$  and  $\beta = (\beta'_1, \dots, \beta'_I)'$ . To estimate the parameter vector  $\Phi$  we formulate the likelihood for  $\Phi$ :

$$L(\Phi; y) = \prod_{j=1}^n f_j(y_j | \Phi) \quad (6)$$

An estimate of the parameter vector  $\Phi$  is obtained by maximising the above likelihood equation with respect  $\Phi$  subject to the constraint (1), using the EM algorithm (Dempster, Laird and Rubin 1977). Once an estimate of  $\Phi$  is obtained, estimates of the posterior probability  $\alpha_{ij}$ , that observation  $j$  comes from the latent class  $i$  can be calculated for each observation vector  $\mathbf{y}_j$  by using Bayes' theorem given by:

$$\alpha_{ij}(\mathbf{y}_j|\Phi) = \frac{\pi_i \prod_{k=1}^K f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i)}{\sum_{i=1}^I \pi_i \prod_{k=1}^K f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i)} \quad (7)$$

The EM Algorithm iterates between an expectation E-step and a maximization M-step. To derive the EM Algorithm, we introduce unobserved data  $z_{ij}$  indicating if observation  $j$  belongs to latent class  $i$ , such that  $z_{ij} = 1$  if  $j$  comes from class  $i$  and  $z_{ij} = 0$  otherwise. It is assumed that these  $z_{ij}$  are independent and identically distributed and have a multinomial distribution.

$$f_j(\mathbf{z}_j|\pi) = \prod_{i=1}^I \pi_i^{z_{ij}} \quad (8)$$

where the vector  $\mathbf{z}_j = (z_{1j}, \dots, z_{Ij})'$ . We denote the matrix  $(\mathbf{z}_1, \dots, \mathbf{z}_n)'$  by  $\mathbf{Z}$  and the matrix  $(\mathbf{X}_1, \dots, \mathbf{X}_p)$  by  $\mathbf{X}$ . It is assumed that the observed data  $\mathbf{y}_{jk}$  given unobserved data  $\mathbf{z}_j$  are conditionally independent and that  $\mathbf{y}_{jk}$  given  $\mathbf{z}_j$  has the density function:

$$f(\mathbf{y}_{jk}|\mathbf{z}_j) = \prod_{i=1}^I f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i)^{z_{ij}} \quad (9)$$

So the observations  $\mathbf{y}_{jk}$  comprise the incomplete data set and the unknown observations  $z_{ij}$  are treated as missing data. Hence the complete data set combines  $\mathbf{X}$  and  $\mathbf{Z}$  and the complete-data log-likelihood can be formed by using the equations (8) and (9).

$$\ln L_c(\Phi; \mathbf{y}, \mathbf{Z}) = \sum_{j=1}^n \sum_{k=1}^K \sum_{i=1}^I z_{ij} \ln f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i) + \sum_{j=1}^n \sum_{k=1}^K \sum_{i=1}^I z_{ij} \ln \pi_i$$

The complete log-likelihood  $\ln L_c(\Phi; \mathbf{y}, \mathbf{Z})$  is maximized using an iterative EM algorithm. In the E-step the complete log-likelihood is replaced by its expectation calculated on the basis of the provisional estimates of  $\Phi$  from the previous iteration. In the M-step the expectation of the complete log-likelihood is maximized with respect to the parameter vector  $\Phi$  to obtain new updated parameter estimates. The E-step and M-step are then alternated repeatedly until the iterative procedure converges and no further improvement in the likelihood function is possible. Dempster, Laird and Rubin (1977) proved that the EM algorithm provides monotone increasing values of the complete log likelihood.

In the E-Step the expectation of the complete log-likelihood is calculated with respect to the conditional distribution of the unobserved data  $\mathbf{Z}$  given the observed data  $\mathbf{y}$  and provisional estimates of  $\Phi$ .  $E[\ln L_c(\Phi; \mathbf{y}, \mathbf{Z})]$  can be obtained by replacing  $z_{ij}$  in  $\ln L_c(\Phi; \mathbf{y}, \mathbf{Z})$  by their expected values,  $E(z_{ij}|\mathbf{y}, \Phi)$ . To obtain this expectation, we first calculate the conditional distribution of  $\mathbf{y}_j$ , given  $\mathbf{Z}$ , which is:

$$f(\mathbf{y}_j|\mathbf{Z}, \Phi) = \prod_{i=1}^I \left( \prod_{k=1}^K f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i) \right)^{z_{ij}} \quad (10)$$

By using Bayes' theorem, we can derive the conditional distribution of  $z_{ij}$  given  $\mathbf{y}_j$  by using equations (10) and (8), which in turn can be used to calculate the required conditional expectation given by:

$$E(z_{ij}|\mathbf{y}_j, \Phi) = \frac{\pi_i \prod_{k=1}^K f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i)}{\sum_{i=1}^I \pi_i \prod_{k=1}^K f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i)} \quad (11)$$

This is identical to the posterior probability  $\alpha_{ij}(\mathbf{y}_j|\Phi)$  in equation (7). Estimates of the posterior probabilities  $\hat{\alpha}_{ij}$  are obtained by evaluating equation (11) using the current estimates of  $\beta$  and  $\lambda$ .

The M-step maximizes the expectation of the complete log-likelihood with respect to the parameter vector  $\Phi$  after replacing the unobserved data  $\mathbf{Z}$  in  $\ln L_c(\Phi; \mathbf{y}, \mathbf{Z})$  by their current expected values  $\hat{\alpha}_{ij}$ :

$$E[\ln L_c(\Phi; \mathbf{y}, \mathbf{Z})] = \sum_{i=1}^I \sum_{k=1}^K \sum_{j=1}^n \hat{\alpha}_{ij} \ln f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i) + \sum_{i=1}^I \sum_{k=1}^K \sum_{j=1}^n \hat{\alpha}_{ij} \ln \pi_i$$

The maximization of  $E[\ln L_c(\Phi; \mathbf{y}, \mathbf{Z})]$  with respect to  $\pi$  subject to the constrain (1), is solved by maximizing the augmented function:

$$\sum_{i=1}^I \sum_{k=1}^K \sum_{j=1}^n \hat{\alpha}_{ij} \ln \pi_i - \mu \left( \sum_{i=1}^I \pi_i - 1 \right) \quad (12)$$

where  $\mu$  is a Langrangian multiplier. Setting the derivative of (12) with respect to  $\pi_i$  equal to zero and solving for  $\pi_i$

$$\hat{\pi}_i = \sum_{j=1}^n \hat{\alpha}_{ij} / n \quad (13)$$

Maximization of  $E[\ln L_c(\Phi; \mathbf{y}, \mathbf{Z})]$  with respect to  $\beta$  and  $\lambda$  is equivalent to independently maximizing each of the following  $I$  expressions:

$$L_i^* = \sum_{j=1}^n \sum_{k=1}^K \hat{\alpha}_{ij} \ln f_{jk|i}(\mathbf{y}_{jk}|\beta_i, \lambda_i) \quad (14)$$

### 3. APPLICATION

The finite mixture model was used to identify factors that influence the customer choices when buying iPads and identify the product attributes that most influence the consumers in buying the product. In this application, the three selected iPad attributes included the price, capacity and connectivity. This survey was designed and devised on Kwik Survey (an online survey questionnaire) where a number of iPads profiles having distinct attributes were generated and these profiles had to be assessed on a 7-point Likert scale where 1 corresponds to 'Not worthy' and 7 corresponds to 'Extremely worthy'. A rating scale was selected since it expresses the intensity of a preference better than a ranking scale. The target population for this survey were university students. The respondents were asked to provide demographic information, including their gender, age and whether they owned an iPad.

The three selected iPad attributes included the capacity of the iPad (16GB, 32GB and 64GB), connectivity (Wi-Fi and Wi-Fi plus 3G) and price (€500, €600, €700 and €800). These three attributes were chosen on the merit that they are found in literature to be the most pertinent when compared to the other attributes, such as colour and size. A full-profile method and full factorial design were chosen for the data collection method yielding a total of 24 distinct profiles. The sample of 364 participants who completed the online questionnaire included a larger proportion of females (55.5%) than males. Around 70% of the university students had less than 24 years and only a third owned an iPad.

To identify the optimal number of segments, the finite mixture model was fitted several times each time changing the number of segments from 1 to 4. For each solution the BIC criterion was computed. Table 1 shows that the three-segment solution is the one which minimizes the criterion.

Number of segments $K$	Deviance $(-2 \log L)$	Number of parameters $d$	BIC
1	21192	7	21233
2	19203	14	19286
3	19133	21	19257
4	19097	28	19262

**Table1:** BIC value for each segment solution

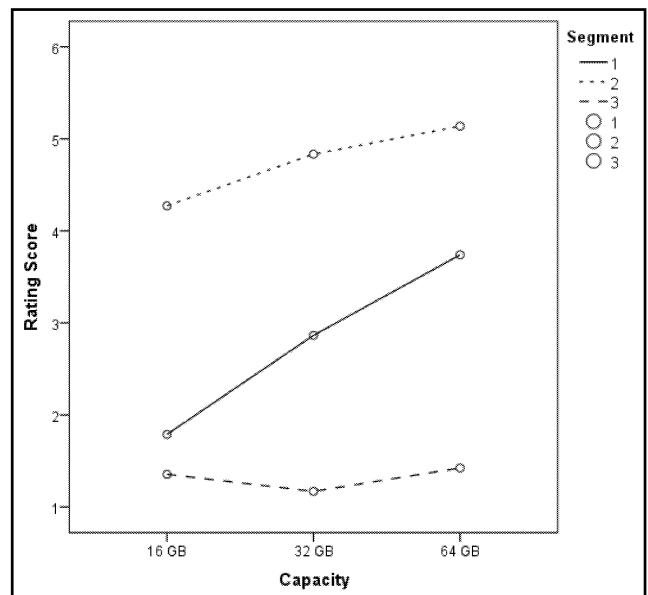
### 4. RESULTS OF FINITE MIXTURE ANALYSIS

Posterior probabilities were computed for each respondent and each person was allocated to the segment with highest posterior probability. 212 respondents were allocated to segment 1, 111 students to segment 2 and the remaining 41 participants to segment 3. Segment 1 included a larger proportion of females, aged between 17 and 19 years and owned an iPad. Segment 2 comprised a larger proportion of males, aged between 20 and 23 years and owned an iPad. Segment 3 included a larger proportion of males, aged at least 24 years and did not own an iPad. Table 2 displays the parameter estimates and standard errors for each segment solution.

Parameter estimates	Standard Error	Term
3.739	0.033	Segment(1)
5.137	0.045	Segment(2)
1.424	0.075	Segment(3)
2.478	0.036	Price(1).Segment(1)
0.740	0.048	Price(1).Segment(2)
0.824	0.080	Price(1).Segment(3)
1.538	0.036	Price(2).Segment(1)
0.504	0.048	Price(2).Segment(2)
0.387	0.080	Price(2).Segment(3)
0.625	0.036	Price(3).Segment(1)
0.209	0.048	Price(3).Segment(2)
0.080	0.080	Price(3).Segment(3)
-1.950	0.031	Capacity(1).Segment(1)
-0.865	0.041	Capacity(1).Segment(2)
-0.069	0.069	Capacity(1).Segment(3)
-0.874	0.031	Capacity(2).Segment(1)
-0.304	0.041	Capacity(2).Segment(2)
-0.254	0.069	Capacity(2).Segment(3)
-0.709	0.025	Connectivity(1).Segment(1)
-0.395	0.034	Connectivity(1).Segment(2)
-0.130	0.057	Connectivity(1).Segment(3)

**Table 2:** Parameter estimates and standard errors

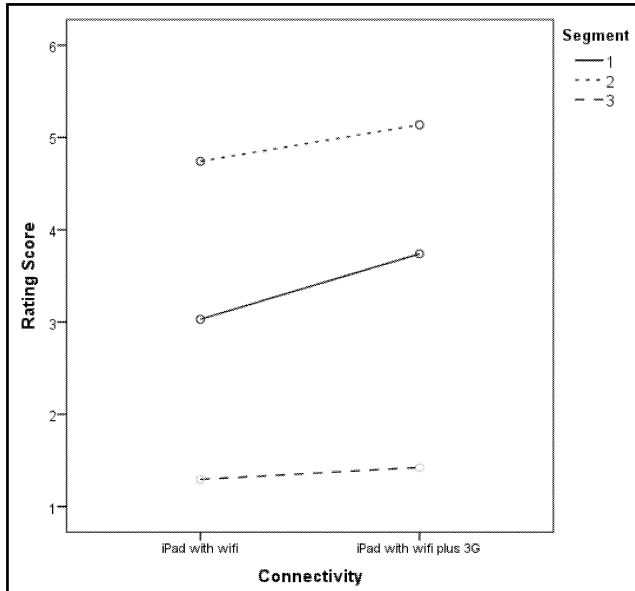
Segmentation is effective if it is identifiable and accessible. These segments are meaningless if they are not described and defined. Figures 1, 2 and 3, show the mean rating scores provided by respondents in different segments for different profile manifestations categorized by the levels of capacity, connectivity and price.



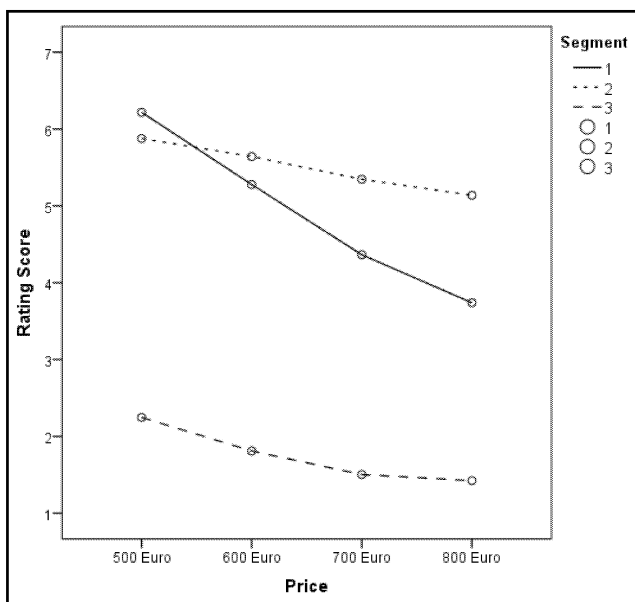
**Figure 1:** Mean rating score by cluster membership and iPad capacity

Respondents in Segment 1 worth iPads more if they have higher capacity, faster connectivity and are less expensive. Respondents in Segment 2 behave similarly to those in Segment 1 because they value iPads more if they have higher capacity, faster connectivity and are cheaper in price. Though, on average, they are providing higher rating scores

and are discriminating less between the iPad attributes categories since changes in their mean rating scores are less conspicuous compared to those in Segment 1. Respondents in Segment 3 are providing very low rating scores. They are not price sensitive and hardly discriminate between the iPad features since their mean rating scores vary marginally for different profile manifestations.



**Figure 2:** Mean rating score by cluster membership and iPad connectivity



**Figure 2:** Mean rating score by cluster membership and iPad price

## REFERENCES

- Andrews, R.L., Ansari, A. and Currim, I.S. (2002), Finite Mixture Conjoint versus Hierarchical Bayes, *Journal of Marketing Research*, 19, 1, 87-98.
- Dempster, A., Laird, N. and Rubin, D. (1977), Maximum likelihood from incomplete data via the EM algorithm *Journal of the Royal Statistical Society*, B, 39, 1-38.
- DeSarbo, W.S., Oliver, R.L. and Rangaswamy, A. (1989), A Simulated Annealing Methodology for Clusterwise Linear Regression, *Psychometrika*, 54, 707-736.
- DeSarbo, W., Wedel, M., Vriens, M. and Ramaswamy, V. (1992), Latent class metric conjoint analysis, *Marketing Letters*, 3, 3, 273-288.
- Gelman, A., Carlin, J. and Rubin, D. (2004), *Bayesian Data Analysis* (2<sup>nd</sup> edn). Chapman & Hall CRC
- Hagerty, M.R. (1985), Improving the predictive Power of Conjoint Analysis: Use of Factor and Cluster Analysis *Journal of Marketing Research*, 22, 168-184.
- Kamakura, W. (1988), A Least Squares Procedure for Benefit Segmentation with Conjoint Experiments, *Journal of Marketing Research*, 25, 157-167.
- Kamakura, W. and Russell, G. (1989), Probabilistic choice model for market segmentation and elasticity structure, *Journal of Marketing Research*, 26, 379-390.
- Ogawa, K. (1987), Approach to Simultaneous Estimation and Segmentation in Conjoint Analysis, *Marketing Science*, 6, 66-81.
- Roberts, C. and Casella G. (2004), *Monte Carlo Statistical Methods* (2<sup>nd</sup> edn). New Your Springer-Verlag.
- Rossi, P., Allenby, G. and McCulloch, R. (2006), *Bayesian Statistics and Marketing* (2<sup>nd</sup> edn). John Wiley & Sons
- Spath, H. (1982), A Fast Algorithm for Clusterwise Linear Regression, *Computing*, 29, 175-181.
- 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), Mixture likelihood Approach for Generalized Linear Models, *Journal of Classification*, 12, 1-35.
- Wedel, M. and Kistemaker, C. (1989), Consumer Benefit Segmentation using Clusterwise Linear Regression, *Journal of Research Marketing*, 6, 45-49.
- Wedel, M. and Steenkamp, J.B. (1991), Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Segmentation, *Journal of Market Research*, 28, 385-396.

## 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.

# EVOLUTIONARY SIMULATION-OPTIMIZATION OF PERSONNEL SCHEDULING

Martin Andersson  
Anna Syberfeldt  
Amos Ng  
University of Skövde  
P.O. 408 SE-54148 Skövde,  
Sweden  
E-mail: martin.andersson@his.se

Victor Bengtsson  
Posten AB  
Terminalvägen 24  
Solna, Sweden

## KEYWORDS

Simulation, optimization, personnel scheduling, hill climbing, NSGA-II, case study.

## ABSTRACT

This paper presents a simulation-optimization system for personnel scheduling. The system is developed for the Swedish postal services and aims at finding personnel schedules that minimizes both total man hours and the administrative burden of the person responsible for handling schedules. For the optimization, the multi-objective evolutionary algorithm NSGA-II is implemented. The simulation-optimization system is evaluated on a real-world test case and results from the evaluation shows that the algorithm is successful in optimizing the problem.

## INTRODUCTION

Posten is the Swedish postal services, owned 60% by the Swedish state and 40% by the Danish state. Core business comprises distribution of post/mail and logistics, and Posten is one of the largest actors in these areas in the Nordic region. With more than 30 000 employees and revenue of approximately 25 billion SEK, Posten is also one of the largest corporations in Sweden. Posten is mandated to make postal services available all over the nation and handles approximately 20 million mails per work day. The mandate states that mail must be collected and delivered on every workday and at least five days a week nationwide. Furthermore, Posten must guarantee that at least 85 percent of priority mail posted before a specific time must be delivered during the following workday – wherever it is addressed to in Sweden.

As the Nordic postal market is fully deregulated, mail business is a highly competitive market. New operators are becoming established, especially in larger cities, and can compete with lower fees as they are not covered by the same regulations as Posten. Facing national and international actors operating in the same business areas puts high demands on efficient mail operations, and additional pressure arises from the legal directives that Posten is obligated to follow, specifying that mail operations must be fast, reliable, and cost-efficient. These challenges are not unique for the Swedish postal market, but postal administrations all over the world face the same requirement

to continuously analyze and improve their services (Larsen 2003).

This paper describes a project that attempts to improve Posten's processes by developing a simulation-optimization system for personnel scheduling. The current process for creating schedules is manual, and creating efficient schedules with low personnel overhead is difficult. The aim of this project is to replace major parts of this manual process with an automatic simulation-optimization system. The system is not meant to totally replace the human expert, but rather to act as a decision support tool. The reason for not replacing the human is that there are many small, but complicated, details that will be hard to implement into the system.

The project described in the paper is a collaboration between the University of Skövde and Posten. While the university is responsible for the development of the system and the design of the optimization algorithm, Posten contributes with process knowledge and knowledge of soft parameters that are crucial for the acceptance of the system. Posten also contributes with a real-world test case that allows the system to be tested on realistic data sets.

## PROBLEM

The purpose of this system is to create personnel schedules and assign people to those schedules, so that a given personnel demand is met with the lowest total scheduled time. This is an instance of a generalized set covering problem and it can be formulated as an integer problem. Usually the schedule composition is incorporated into the linear model, either fully (Bard 2003) or partially (Easton, 1999). A linear model with arbitrary schedules would be prohibitively difficult to solve. That's why the linear models are usually formulated as to only allow a smaller, sometimes predefined, subset of possible schedules.

## SOLUTION APPROACH

The difference in this system from other solution approaches involving linear models is that fewer restrictions are placed on the schedules. Instead of using the linear model for the schedule composition it is done by a metaheuristic, simplifying the linear model and allowing for more complex schedules. This comes at the cost of a greatly increased

search space. To handle this large search space a parallelized approach is taken, allowing the system to efficiently utilize large computing resources. The personnel schedules created by the metaheuristic must follow government laws, union regulations and company policies.

## OPTIMIZATION INPUTS

There are three inputs to the optimization, they are: the personnel demand, schedule types and shifts. The personnel demand is represented as a list of integers, where each integer is how many people that are needed at a specific time. Schedule types specify which types of employment that are acceptable. It's possible to limit how many percent of the total amount of personnel that are assigned to each schedule type. For example, 70% of all personnel should be full time employees. Work shifts are constraints placed on the schedules themselves, like earliest start time and the maximum amount of continuous work. For example, a day and night shift can be specified in order to prevent a single schedule to have work scheduled sometimes during the day and sometimes during the night.

## OPTIMIZATION

The optimization procedure consists of two main parts. The first part is a metaheuristic that creates a set of schedules, which are then evaluated by the second part; a simulation implemented using a linear programming solver.

Two objectives should be considered in the optimization: a) to minimize the total amount of scheduled man hours, and b) to minimize the total amount of schedules (in order to reduce the administrative burden of handling a large amount of schedules). The two objectives are usually in conflict, because the more schedules that can be used the easier it is to fit the schedules to the personnel demand.

The difficulty with conflicting optimization objectives is that there is usually no single optimal solution with respect to both objectives. Instead of a single optimum, there is a set of optimal trade-offs between the conflicting objectives, called Pareto-optimal solutions (Deb, 2004). In order to manage multiple objectives, specific optimization algorithms have been suggested. Instead of only seeking a single optimum, these algorithms maintain a set of Pareto-optimal solutions. One of the most efficient algorithms for Pareto optimization is the elitist non-dominated sorting genetic algorithm (NSGA-II) (Deb et. al, 2000). Due to the algorithm's proven efficiency, this is the optimization algorithm selected for the study. Further details are given in the next subsection.

### NSGA-II Algorithm

NSGA-II is basically a genetic algorithm extended with features for handling multiple trade-off solutions (Deb et al., 2002). Like an ordinary genetic algorithm, NSGA-II maintains a population of solutions (in this case schedules) and refines these solutions through generations. In the refinement process, crossover and mutation operators are used to create offspring solutions that become part of the next generation. The basic steps involved in evolution

algorithms are presented below (a complete description of genetic algorithms can be found in Bäck et al., 1997).

Initialize population

Evaluate the fitness of solutions in the population

**Repeat**

Select solutions to reproduce

Form a new generation of population through crossover and mutation

Evaluate the new solutions

**Until** terminating condition

Two vital parts in NSGA-II, as in all evolutionary algorithms, are the crossover and mutation operators. The implementation of these operators is further described in the next two sections.

### Crossover Operator

The crossover operator is used to exchange schedules between solutions. It takes two solutions as input (the parents) and produces two solutions as output (the children). To select the schedules to pick from the first parent the crossover operator adds all schedule indices into a list, shuffles it and then it randomly selects a subset of that list to insert into the first child, the remaining schedules are added to the second child. Then the same procedure is performed for the second parent. The end result is two children that are a mix of the schedules from the two parents.

### Mutation Operators

There are seventeen different mutation operators; they differ from the crossover in that they make smaller changes, usually only affecting one schedule. These operators were chosen based on small-scale experiments, which demonstrated that each operator could produce beneficial mutations. A more thorough analysis on the effects of the different operators is out of scope of this paper.

Abbreviation; schedule (S), work period (WP).

- Add one S, based on WPs from other Ss.
- Add one S, from randomly generated WPs.
- Remove one S.
- Add one WP to a S.
- Remove one WP from a S.
- Shorten one WP.
- Extend one WP.
- Pick a S and create a new S for each WP.
- Pick a S and create a new S from a subset of all WPs.
- Pick a S and merge a subset of all other Ss into that one. Remove the Ss that are completely merged.
- Move one WP backward, cannot move beyond the preceding WP.
- Move one WP forward, cannot move beyond the succeeding WP.
- Move one WP, not limited by other WPs.
- Move one WP from one S to another S.
- Move the day rest.
- Move the week rest.
- Rearrange the breaks on one WP.



## SIMULATION

For evaluating schedules generated by the optimization algorithm, a simulation is used. The simulation is implemented using C++ based on the software library Cbc (<https://projects.coin-or.org/Cbc>) which is an open-source mixed integer programming solver. The task of the linear programming solver is to assign people to schedules. This is an integer problem, since it's not possible to assign, for example, 4.6 people to a schedule. The issue is that solving linear integer problems is computationally expensive. One method of mitigating this is to relax the integer problem (and round solution values up), thereby sacrificing solution quality for solution speed. This is the approach taken in this paper, since the relaxed problem is at least an order of magnitude faster to solve, easily overcoming the loss of solution quality.

From an overall perspective, the simulation takes a set of schedules as input, finds the optimal assignment of people to these schedules given a specific demand (using linear programming) and eventually returns the total amount of man hours and feasibility (if the schedule is valid, i.e. fulfills basic constraints). Based on the results returned, the optimization algorithm generates a new set of schedules and sends these to the simulation model. This process of generating and evaluating schedules (shown in Figure 1) then continues until the user-defined stopping criterion is met.

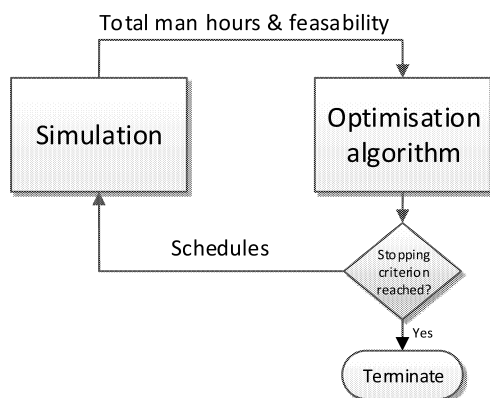


Figure 1. Simulation-Optimization Loop

For the optimization to find sufficiently good solutions, a large amount of simulations is also needed. Although the simulation is fast (one simulation run takes approximately 30 milliseconds), the time consumption of the optimization process becomes quite large due to the high complexity of the problem. To reduce the time consumption, a parallel approach with multiple simulation agents is adopted. This approach is further described in the next section.

## PARALLELIZATION

For increased efficiency, a number of parallel computing resources are used in the simulation-optimization. Contrary to many other systems, not only the simulation but also the optimization algorithm is distributed and run in parallel. This is done because when the simulation is fast and/or the amount of available computing resources is large, there comes a point where the optimization algorithm can't

generate solutions fast enough to keep the computing resources completely busy. To scale the system beyond this point the algorithm must be able to run in parallel instances. Parallel instances of an algorithm that don't share solutions are equivalent to running multiple serial replications of the same algorithm. This can be useful to speed up a series of replications, but it won't increase the speed of single run. In order to do that solutions must be shared between the algorithms so they can benefit from each other's work.

There are three major parallelization strategies for metaheuristic algorithms, which fall into three distinct hierarchical levels, see Figure 2 (Talibi, 2009). On the algorithmic level multiple independent or cooperating algorithms are used to achieve parallelization. This is problem independent and it also has the potential of affecting the behavior of the algorithm. The iteration level concerns the parallelization of a single algorithm. An example of this would be in a population based algorithm the evaluation of solutions could be performed in parallel. This is also problem independent but it does not affect the behavior. The last level is the solution level; this concerns the parallelization of one single solution evaluation. Either the objective function is divided into several parts that can be executed in parallel or the input data is partitioned. This is problem dependent since both these approaches are tightly coupled to the problem at hand.

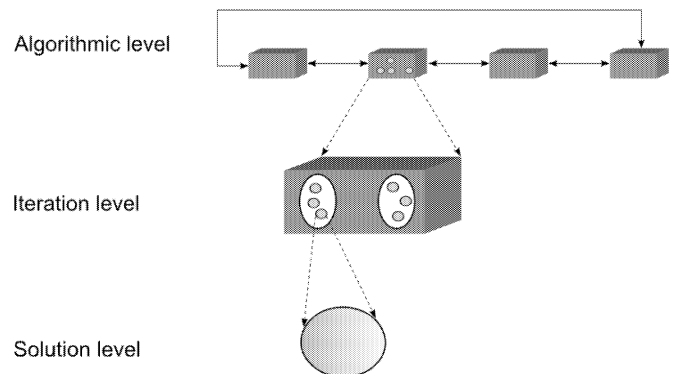


Figure 2. Parallelization Levels

In this system the parallelization is both on the algorithmic level (creating solutions in parallel) and on the iteration level (evaluating solutions in parallel). There is no parallelization on the solution level. By having the parallelization on both the algorithmic and iteration levels it's possible to achieve a system that can adapt to most problem instances (problems differ in the time they take to create and evaluate solutions), which allows for high utilization of computing resources by avoiding bottlenecks in either the schedule generation or assigning people to schedules tasks.

When an optimization instance finds a better solution it is always directly sent to all other currently running optimization instances. This means that there's only a small time window where the instances differ in what solutions they have. The advantage is that each optimization instances is always operating on the global best solutions (or something very close to it), with the disadvantage that there are a lot of traffic between the optimization instances in the beginning when it is easy to find better solutions. That

disadvantage is mitigated by the fact that it is possible to run multiple optimization instances on the same computer, which communicate by message passing via system memory in order to avoid excessive network traffic.

## EVALUATION

This section describes the evaluation of the simulation-optimization system.

### Test case

A real-world test case is being used for the evaluation. The goals of the test case are defined as:

- Make sure that the system generates schedules that obey laws and regulations.
- Find the optimal parameter settings to improve the effectiveness of the system.
- Evaluate the quality of the schedules generated by the system.
- Find the requirements of the graphical user interface.

For the test case, one of the departments at Postens mail hub in Gothenburg was chosen. The hub has about 150 employees working in three shifts. As input to the test case, the personnel demand for one month was collected with a time resolution of half an hour, together with actual schedules that was used at the time.

### Baseline Comparison

For assessment of the performance of the NSGA-II optimization algorithm, an additional optimization algorithm is implemented for comparison. The additional algorithm implemented is hill climbing, which is selected since it is well-known, simple to implement and often used for baseline comparisons. Hill Climbing is an iterative algorithm that belongs to the family of local search algorithms (Russell and Norvig, 2003). The algorithm starts with a random solution to the problem and attempts to find a better solution by mutating (changing) the solution. If the mutation produces a better solution, it is kept and the procedure is repeated until no further improvements can be made. A drawback of hill climbing is that it can get stuck in local optima's. This is somewhat mitigated by the fact that the version used in this paper is modified to work with two objectives. The modification is that instead of only improving one solution the entire non-dominated set is simultaneously improved. This helps to preserve some diversity among the solutions.

The hill climbing is implemented with the same mutation operators as the NSGA-II algorithm and also allocated the same number of total simulation evaluations. For increased efficiency, the hill climbing algorithm is implemented to constrain the amount of schedules each solution can have to a predetermined amount. This is done because even though increasing the amount schedules will generally allow for better solutions, this is only true up to a point. Further, this point is usually well beyond what is considered a reasonable amount of schedules. This fact is utilized by the hill climbing, which allows it to store every solution with a

unique amount of schedules and then individually improve these solutions, using the second objective which is the total amount of scheduled time.

## Results

The optimization results achieved by the NSGA-II algorithm and the hill climbing are presented in Table 1. As shown in the table, the results indicate that NSGA-II outperforms the hill climbing algorithm.

Table 1. Results (best solution found)

Algorithm	Total man hours (minimize)	No. of schedules (minimize)
NSGA-II	20 676	31
Hill climbing	24 251	41

The results from the optimization are also shown in Figure 3. In the figure, the y-axis represents the total amount of man hour (as this is considered the most interesting of the two objectives) and the x-axis represents number of evaluations. As can be seen in the figure, the progress of the two algorithms differs mostly in the beginning of the search and is quite similar towards the end.

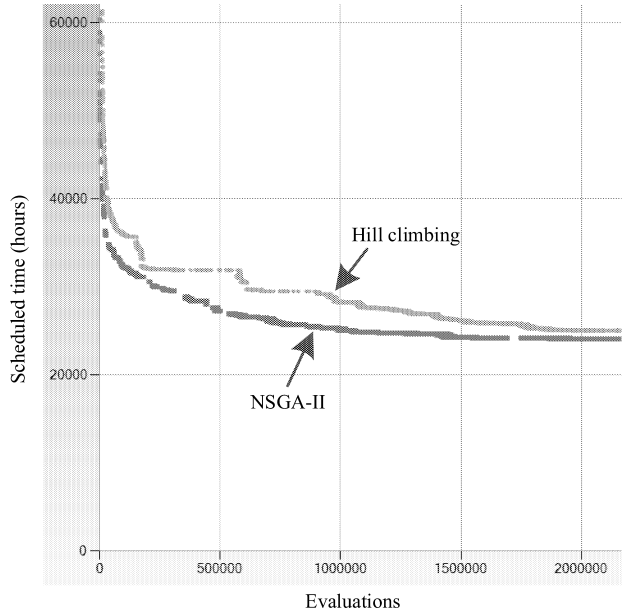


Figure 3. Results over time

It is worth to notice that from an algorithmic perspective the hill climbing algorithm is considerable less time consuming than the NSGA-II. While the NSGA-II algorithm took 79 minutes to run in the real-world test case, the hill climbing algorithm took only 32 minutes (with the exact same number of simulation evaluations allocated for both algorithms). The reason that hill climbing is more efficient from an algorithmic perspective is because it avoids computationally expensive operations like non-dominated sorting and crowding distance calculations that is used in the NSGA-II algorithm.

Posten has evaluated the schedules obtained by the simulation-optimization system and they see an advantage in

that the system heavily reduces the human effort associated with creating schedules. Compared to the manual approach of creating schedules used at Posten so far, a lot of time and effort can be saved for the person responsible for creating schedules. The optimization also makes it easy to obtain schedules with certain focuses, for example, a minimum number of full-time schedules, as the decision makers' preferences are considered in the optimization process.

## SUMMARY AND FUTURE WORK

This paper presented a simulation-optimization system for improving personnel schedules in the Swedish postal services. The optimization aims at finding schedules that minimizes both the total man hours and the administrative burden of the person responsible for handling schedules by reducing the number of schedules. The two objectives of minimizing man hours and number of schedules are conflicting; because the more schedules the easier it is to fit the schedules to the personnel demand. The difficulty with conflicting optimization objectives is that there is usually no single optimal solution with respect to both objectives, and to handle this difficulty a specialized multi-objective algorithm called NSGA-II is implemented. For evaluating solutions, the NSGA-II algorithm uses a simulation based on the software library Cbc. The simulation-optimization system is evaluated on a real-world test case obtained from one of Posten's mail hubs in Sweden having about 150 employees. The schedules being the outcome of the test case has been evaluated by the company and the results look very promising. These results have so far only been verified theoretically, but in the future optimized schedules can hopefully be implemented and evaluated also in reality. Even though the system is not yet in use by Posten, the most promising application is to provide information for strategic decisions. For example, what it would cost to increase the number of full time employees or what is the cost tradeoff between using different number of shifts.

An improvement of the system would be to programmatically integrate it with the personnel handling systems used at Posten. Such integration would eliminate the need of manually providing input regarding personal demand, available personnel, etc before an optimization is started. With integration, the optimization can be fully automatic and save even more human efforts.

## REFERENCES

- Bard, J. F., Binici, C., & desilva, A. H. (2003). Staff scheduling at the United States postal service. *Computers & Operations Research*, 30(5), 745-771.
- Bäck, T., Fogel, D. and Michalewicz, Z. (eds) 1997. *Handbook of Evolutionary Computation*. Oxford University Press.
- Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2000). A fast and elitist multi-objective genetic algorithm NSGA-II, KanGAL Report 2000001, Indian Institute of Technology Kanpur, India.
- Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*. VI(2), 182-197.

- Deb, K. 2004. *Multi-objective optimization using evolutionary algorithms*. Chichester: John Wiley & Sons, Ltd.
- Easton, F. F., & Mansour, N. (1999). A distributed genetic algorithm for deterministic and stochastic labor scheduling problems. *European Journal of Operational Research*, 118(3), 505-523.
- S.J. Russell, P. Norvig. 2003. *Artificial Intelligence: A Modern Approach*. 2<sup>nd</sup> ed, Upper Saddle River, New Jersey: Prentice Hall, 111-114.
- Larsen, N.E. 2003. Simulation – a key tool to accelerate and add confidence to postal network configuration. In *Proceedings of the 2003 Winter Simulation Conference*, 1585-1592. Orlando, FL.
- Talbi, E.-G. 2009. *Metaheuristics: From Design to Implementation*. John Wiley & Sons, 478-504.

## BIOGRAPHIES

**MARTIN ANDERSSON** is a Ph.D candidate at the University of Skövde, Sweden. He has a B.Sc. in Computer Science and a M.Sc. in Automation Engineering from the University of Skövde. His research interests include simulation-based optimization and parallel metaheuristics. His email address is martin.andersson@his.se.

**ANNA SYBERFELDT** is a senior researcher at the University of Skövde, Sweden. She holds a Ph.D in Computer Science from the De Montfort University, UK and a Master's degree in Computer Science from the University of Skövde, Sweden. Her research interests include simulation-based optimization, artificial intelligence, metaheuristics, and advanced information technology with applications in logistics and manufacturing. Her email address is anna.syberfeldt@his.se.

**AMOS NG** is a Professor at the University of Skövde, Sweden. He holds a Ph.D. degree in Computing Sciences and Engineering. His research interests lies in applying multiobjective optimization for production systems design & analysis. His e-mail address is amos.ng@his.se.

**VICTOR BENGTTSSON** is an Optimization- & Simulation Specialist at Posten AB. He holds a Master degree in Mathematics from the University of Linköping, Sweden. His research interests includes production simulation, simulation-based optimization, multi-objective optimization, set covering optimization.

## ACKNOWLEDGEMENTS

This work has been carried out within the SOL project which is partially financed by the Knowledge Foundation (KK Stiftelsen), Sweden. The author gratefully acknowledge the Knowledge Foundation for the provision of research funding.



# **DECISION SUPPORT SYSTEMS**



# INTERACTIVE AND INTELLIGENT DECISION SUPPORT IN MANUFACTURING USING SIMULATION-BASED INNOVIZATION AND CLOUD COMPUTING

Ingemar Karlsson  
Amos Ng  
Anna Syberfeldt  
Virtual Systems Research Center  
University of Skövde  
P.O. 408, SE-54148 Skövde, Sweden  
E-mail: [ingemar.karlsson@his.se](mailto:ingemar.karlsson@his.se)

## KEYWORDS

Simulation, Innovization, Decision support, Cloud computing.

## ABSTRACT

Simulation-based innovization is a method for extracting knowledge from a simulation model and optimization. This method can help decision makers to make high-quality decisions for their manufacturing systems so as to enhance the competitiveness of companies. Nevertheless, the simulation-based innovization process can be computationally costly and having these resources in-house can be expensive. By running the process in a cloud environment instead, the company only pays for the resources they are using. This paper proposes the concept of a cloud-based computing platform that can run the simulation-based innovization process and discuss its possibilities and challenges.

## INTRODUCTION

When designing and improving a manufacturing problem, a series of decisions have to be made. These decisions are about, for example, layout, manning, logistics and level of automation, can pose a big challenge for decision makers involved. The complex combinations and interactions between all components within the manufacturing system make it hard for a decision maker to understand the behavior of the system. To achieve the best overall performance of the system is a very complex task because it involves the selection of the optimal design based on the design variables on hand, for example buffer sizes, machine properties and so on.

This paper proposes a computer platform that helps the decision maker to define the problem and perform different analysis on the problem with the help of simulation models. Because of the complexity of manufacturing systems, the simulation models can also be complex and can therefore take time to run. The platform concept

described in this paper solves this by using distributed computing in a cloud environment.

The complexities and problems with decision making on manufacturing problems are discussed in the next section.

## Innovization

One of the hurdles for decision making in manufacturing problems is to organize all available data and convert them into knowledge that the decision maker can use. Even though the optimal solutions are known, there can be a lot of knowledge embedded in these data. By extracting knowledge from these data, the company can learn more about their manufacturing system than just aiming at choosing an optimal solution to a problem.

One method to perform this, is by doing a post-optimality analysis of pareto-optimal solutions from multi-objective optimization (Deb and Srinivasan, 2006). They call the task of identifying the main principles in a set of pareto-optimal solutions for innovization (innovation via optimization). This task has previously been done manually by studying the pareto-set but there are recent studies that have tried to do this automatically by applying data mining techniques, which have shown promise in, for example engineering problems (Bandaru and Deb, 2011).

Data mining is a collection of algorithms and techniques for extracting knowledge from data sets. By integrating data mining techniques with innovization, the innovization task can become even more efficient for decision support for manufacturing problems. This has been shown in previous work, e.g. (Ng et al., 2009, 2011). By integrating interactive methods for supplying knowledge back into the optimization loop, the optimization process can be even more efficient, see Ng et al. (2013).

## SIMULATION-BASED INNOVIZATION

This section describes a technique for extracting knowledge from simulation-based optimization, hence called

simulation-based innovization.

Simulation-based innovization (SBI) (Ng et al., 2009) uses simulation-based optimization (SBO) as a method for retrieving the optimization data for analysis. The second component in SBI is data mining which extracts knowledge from the results from SBO, figure 1.

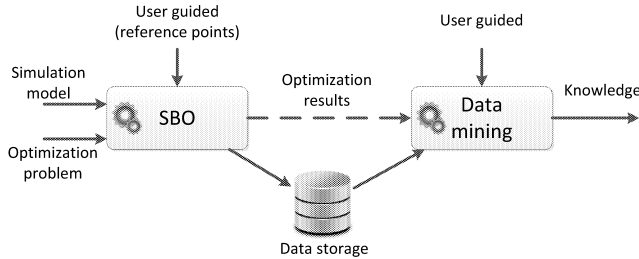


Figure 1: Simulation-Based Innovization Process

Simulation-based optimization is an iterative process where an optimization algorithm uses a simulation model to evaluate solutions. Inputs are generated in the optimization algorithm that are fed into a simulation model. The response from the model is then used by the optimization algorithm for generating a new set of inputs. The optimization algorithm might perform many of these iterations to get a solution that meets the decision maker's requirements. The evaluation time depends on the type of simulation and the complexity of the model. If the simulation model is stochastic, each evaluation of the simulation needs to be replicated so that statistically reliable output data can be generated. Such a resampling also means that the optimization will take even longer time to complete.

A company's use of this kind of software will vary over time. The need increases when the company plans for new products and/or new manufacturing lines. Another common use case scenario will be the continuous performance improvement of an existing manufacturing line. Because of all these different requirements, the utilization of a computer cluster for SBO will also vary. This makes it hard to decide how large the cluster should be. In some situations, the cluster may be too small and at other times it may be unused or with very low utilization. If the cluster is too small, it may hinder the company to make decisions in time and on the other hand an over-dimensioned cluster means the company has to maintain unnecessarily costly computing resources. In the next section, cloud computing is introduced as a possible solution to this problem of computing resource allocation.

## CLOUD COMPUTING

Cloud computing has a broad meaning. In some definitions, both the applications delivered through the cloud and the services and hardware behind it can be referred to as cloud computing (Armbrust et al., 2010). NIST

(National Institute of Standards and Technology) (Mell and Grance, 2011, p. 2) defines cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction".

Cloud computing with its scalable and elastic features has opened up for new flexible and cost-effective web applications that have not been possible before. The move from physical servers to virtual servers has led to the development of different cloud services. A physical server can host several virtual servers where the exact number is limited by the resources on the host computer. This technique gives organizations the possibility to increase and decrease their computational resources which results in flexibility that is not possible with a traditional data center. Such kind of scalability and elasticity is unique to cloud computing, which opens up for new applications that rely on distributed computation.

## Cloud computing for SBI

By utilizing the concept of cloud computing, the problems with a SBI platform can be mitigated. There are many advantages offered by cloud computing, but the bottom line is, it lets a customer to hire in only the required computational capacity at the required time frame. If the SBO platform is run in a cloud environment, the company does not have to own any cluster hardware and in that way save money, electricity and other resources. Cloud computing usually has a "pay as you go" model (Furht and Escalante, 2010, p. 69) so that the customer only pays for the services they use. Because of this, it is possible to start only the number of virtual computers that are needed at the required moment, which makes the solution "elastic" but still highly scalable.

The next chapter describes some of the related work that has been done in the area of simulation and optimization in the cloud.

## RELATED WORK

A survey of related work has been performed and there are attempts to migrate simulation and SBO to the cloud.

Fujimoto et al. (2010) discusses the potential benefits and challenges with running Parallel Discrete Event Simulation in a cloud environment and proposes an architecture they call Aurora. Aurora is based on a master/worker paradigm where a controller sends work packages, called Simulation Packages, to workers which contain information about what tasks the workers should perform. The controller and workers can run on a cloud environment. The main benefit is that it is pos-



sible to hide the complexities with parallel simulation and make parallel simulation more easily accessible to the simulation community. In terms of challenges, the authors mention that there are no tools available for this kind of simulation and that the security and availability must be further investigated.

Liu et al. (2012) also discuss the possibility to make simulation available in the cloud. The authors focus on how existing simulation software can be made available in the cloud and propose an architecture called CSim. The user accesses the simulation software through a website. The website is connected to a set of services that provides the functionality of the simulation software. The services, as well as the website, are running on virtual computers. To be able to make use of the resources in an effective way the CSim architecture can scale the number of computing resources to fit the actual computing requirements.

Another web-based simulation architecture is proposed by Rossetti and Chen (2012) for simulating supply chains in the cloud. The authors propose an architecture they call Cloud Computing Architecture For Supply Chain Network Simulation (CCAFSCNS). The CCAFSCNS architecture includes a workload management system called Condor for distributing the simulation jobs. This system queues and distributes the jobs while it also manages the computing resources. The jobs are originally created on a local computer and are uploaded to the CCAFSCNS server and then later downloaded for analysis on the local computer as well. For the prototype, the authors are using the research grid FutureGrid project to be able to test how the architecture works. When the authors are simulating large multi-echelon supply chains in their prototype, the simulation time is significantly lower compared to a single desktop computer.

An architecture for putting SBO in the cloud for visualizing and optimizing integrated circuits is described by Zhu et al. (2012). The simulation model consists of a Matlab project simulating a dual die CPU. By applying the multi-objective evolutionary optimization algorithm NSGA-II, which feeds input to the simulation model, the authors can determine the best allocation of conductors. This SBO loop is running on a virtualized environment based on the Oracle Virtual Box and can use several virtualized computers to distribute the simulations. A GUI is built in PHP in form of some web page to let the users access the system. Again, by applying cloud computing techniques, the authors have shown significant reduction in computation time.

A concept for a platform for running the SBI process in the cloud is presented in the next section.

## CLOUD BASED SBI PLATFORM

In principle, the process of SBI fits well in the distributed cloud environment. In its most basic form, both

SBO and data mining can be parallelized and therefore the individual computational work packages can be sent to different computers in a distributed environment.

In the following, we propose a new platform that can perform simulation-based optimization and automatically applies data mining techniques on the optimization results. The platform is also designed to make use of the properties of cloud computing, which makes it different from a regular distributed computing.

## Components

The SBI process needs at least two components, SBO and data mining. The SBO process results in a large amount of data from which the data mining process can be used to generate usable knowledge. In figure 2 the main parts in the concept platform is shown.

### *System control*

This part of the system manages the resources and has two main functions, controlling the computing resources and controlling the optimization and data mining processes. The instance controller launches virtual computer instances based on the computational need from the optimization and data mining processes. These instances are later closed when they are not needed.

### *Computing resources*

The computing resource component assembles all the computational needs and runs them on a virtual computer. There are three functions that need computational resources in the platform: evaluation of simulation models, optimization algorithms and data mining algorithms.

The virtual computers that run the computing resource component may run different operating systems, depending on the need for the specific work package. Depending on the cloud environment, there can be several different instances that the platform can choose from. It is the system control component that initiates the launch of the instances. The number of evaluation threads depends on the virtual computer specifications, i.e. number of virtual cores and memory. Both the optimization and data mining algorithms are also controlled by this component. The idea behind the computing resource component is to easily render a virtual computer in different roles, depending on the situation.

### *Data storage*

All the data that are generated in the platform are stored by the data storage component. This can be, but is not limited to, optimization projects with simulation models, optimization results and rules extracted from the data. To be able to make use of the scalability of the cloud environment, the storage should be able to grow together with the amount of data and number of users as well. By storing the data different users can

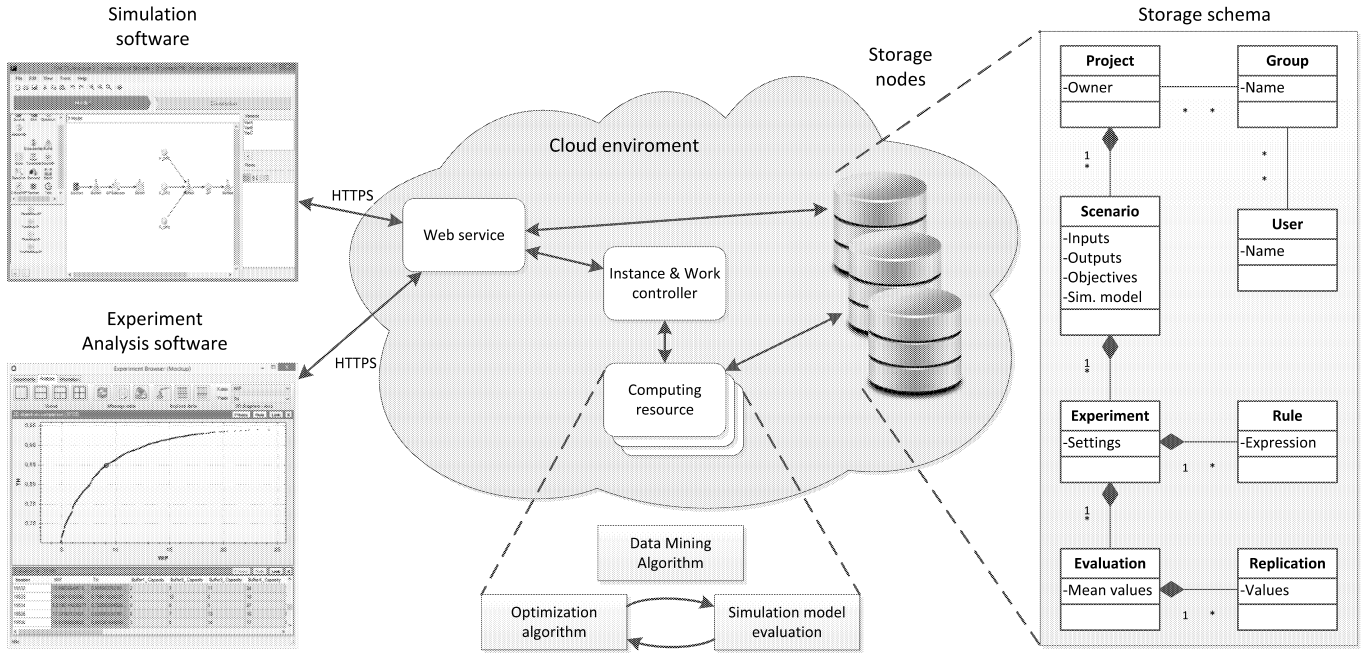


Figure 2: Concept For SBI System

view and modify the same data and because of this collaborate on the same project.

The data storage component can use different technologies for storing the data like NoSQL and relational databases, as well as cloud storage facilities like Amazon S3 (Simple Storage Service). New technologies like NoSQL databases have emerged to handle large data (Nambiar et al., 2014), often called big data. These kind of technologies can handle elastic scaling which is difficult with a traditional storage solution. Therefore the NoSQL databases fit well into the scalable and flexible environment that cloud computing relies on (Pokorny, 2013).

#### Client software and devices

The platform can be interfaced by many types of clients through an API in the system control component. To be able to use the platform, the user needs a simulation model that is supported by the platform. The design of the platform supports all simulation softwares that are interfaced with an API that supports starting simulations, changing setting, as well as retrieving inputs and outputs. The optimization can be initiated directly from a supported simulation software which sends the simulation model together with the optimization problem (e.g. objectives, the selected algorithm and its parameter setting) to the platform.

The experiment analysis software is used to view the results from the optimization as well as performing analysis on the results. This software can also initialize data mining analysis and view the resulting rules extracted by the data mining process. The extracted rules are

stored in the database linked to the experiment.

#### Interactivity

The platform should be able to let the decision maker guide the optimization during its progress. For example, by supplying one or more reference points to a genetic algorithm, it is possible to find better solutions in less time (Siegmund et al., 2012). This can improve the performance of the platform as well as letting the decision maker to point out the preference regions in the objective space to guide the local search. The reference points can be modified throughout the optimization to make sure that the decision maker can provide his/her knowledge during the optimization.

#### Security issues

There are some inherent security issues with using a public cloud computing platform. To many companies and organizations, the major issue is that data are sent to a third party. For example, a simulation model depicting a future manufacturing line can contain a lot of sensitive data like processing times, layouts and production processes. This type of information may be harmful if they are accessible by the competitors. The data locality problem is also one of the main security issues that Subashini and Kavitha (2011) mentions in their extensive review of security issues in cloud computing. One solution to this problem is to encrypt the data stored within the system (Anil and Thanka, 2013). Zissis and Lekkas (2012) also discusses different security issues and

propose the use of encryption both of data and the transportation of the data at a lower level. The traditional security challenges of an IT system also apply to a cloud based system but there are additional challenges as well. Rong et al. (2013) have listed a couple of issues related to resource location and multi-tenancy. The multi-tenancy issue is unique for cloud computing where the cloud providers often use one host computer for several virtual computers that can have different customers. It is a risk that a customer may perform attacks against another customer's virtual computer on the same host.

## APPLICATION IN INDUSTRY

An example of the use of simulation-based innovation in industry can be found in Pehrsson et al. (2013) that has proposed a workflow that supports decision making in the industry with the use of simulation-based innovation.

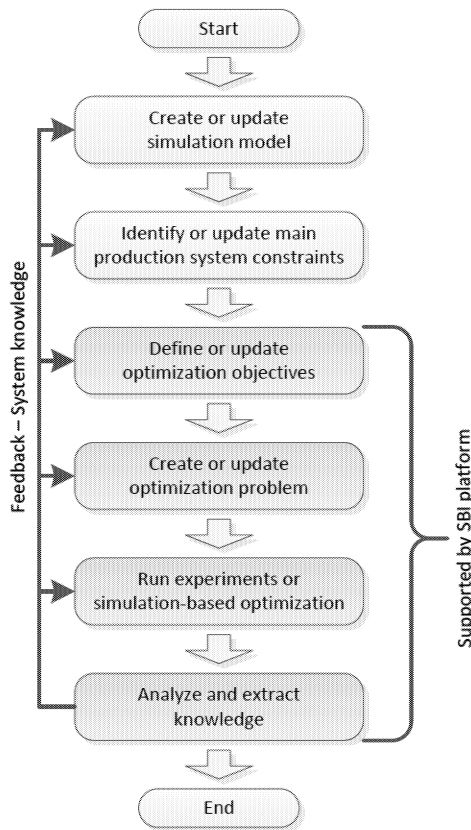


Figure 3: Decision Making Workflow With SBO

The workflow, shown in figure 3, proposes an iterative process in which the decision makers starts with a simulation model that is used for generating data for analysis with the help of simulation-based optimization. The steps in the workflow are refined by the knowledge gained by the subsequent steps in a iterative loop. This leads to decisions of higher quality compared to running the workflow only once (Pehrsson et al., 2013).

This workflow is, in part, supported by the proposed SBI platform. The simulation model is very specific for each problem and is created outside the platform in a simulation software. Identification of system constraints is done by the decision maker but is used as an input to the SBI platform as well as the simulation model. The rest of the steps can be done entirely in the SBI platform to help the decision maker. Results from the steps are stored in the database which can be used for documentation of the decision. This also supports the iterative property of the workflow because earlier used data can be updated and reused. The storage also makes it possible for several users to collaborate on a project which may be difficult if the SBI process is done manually in separate software.

## SUMMARY

In the SBI process, where simulation, optimization and data mining processes are integrated, it can be helpful for supporting decision makers to make decisions for a complex manufacturing environment. Although there are tools and platforms for performing simulation-based optimization, there is no complete tool chain for the complete SBI process. This paper describes the concept of a cloud-based computing platform that is able to assist decision makers in making decisions for designing and improving complex manufacturing systems.

The main feature of the platform is that it has the possibility to adjust the computational resources for the current load. If the platform is running on a public cloud where the users pay for what they are using, this feature will lower the costs of running the platform. It will also give the possibility to run a complex analysis which requires a lot of computing resources. Most of the platforms that have been reviewed in this paper do not provide this capability.

One of the largest challenges is the security aspect of cloud computing in general. The security is very important in this kind of solution because of the sensitivity of the information in the data. This has to be considered throughout the design and implementation of the solution. Encryption of both the communication and the stored data will be required but there are other security issues that encryption cannot alleviate, for example the multi-tenancy issue.

## REFERENCES

- Anil, S. L. and Thanka, R. (2013). A survey on security of data outsourcing in cloud. *International Journal of Scientific and Research Publications (IJSRP)*, 3.
- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., and Zaharia, M. (2010). A view of cloud computing. *Commun. ACM*, 53(4):50–58.
- Bandaru, S. and Deb, K. (2011). Towards automat-

- ing the discovery of certain innovative design principles through a clustering-based optimization technique. *Engineering Optimization*, 43(9):911–941.
- Deb, K. and Srinivasan, A. (2006). Innovization: Innovating design principles through optimization. In *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation*, page 1629–1636, New York. ACM.
- Fujimoto, R. M., Malik, A. W., and Park, A. (2010). Parallel and distributed simulation in the cloud. *SCS M&S Magazine*, 3:1–10.
- Furht, B. and Escalante, A. (2010). *Handbook of Cloud Computing*. Springer Publishing Company, Incorporated, 1st edition.
- Liu, X., He, Q., Qiu, X., Chen, B., and Huang, K. (2012). Cloud-based computer simulation: Towards planting existing simulation software into the cloud. *Simulation Modelling Practice and Theory*, 26:135–150.
- Mell, P. and Grance, T. (2011). The NIST definition of cloud computing—recommendations of the national institute of standards and technology. special publication 800-145, NIST, gaithersburg. *csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf*.
- Nambiar, R., Chitor, R., and Joshi, A. (2014). Data management – a look back and a look ahead. In Rabl, T., Poess, M., Baru, C., and Jacobsen, H.-A., editors, *Specifying Big Data Benchmarks*, number 8163 in Lecture Notes in Computer Science, pages 11–19. Springer Berlin Heidelberg.
- Ng, A., Deb, K., and Dudas, C. (2009). Simulation-based innovization for production systems improvement: an industrial case study. In *The International 3rd Swedish Production Symposium*, page 278–286.
- Ng, A. H., Dudas, C., Boström, H., and Deb, K. (2013). Interleaving innovization with evolutionary multi-objective optimization in production system simulation for faster convergence. In *The 7th International Conference on Learning in Intelligent Optimization (Lion7)*, 7-11 Januari, 2013, Catania, Italy.
- Ng, A. H., Dudas, C., Nießen, J., and Deb, K. (2011). Simulation-based innovization using data mining for production systems analysis. In *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*, page 401–429. Springer.
- Pehrsson, L., Ng, A. H. C., and Stockton, D. (2013). Industrial cost modelling and multi-objective optimisation for decision support in production systems development. *Computers & Industrial Engineering*, 66(4):1036–1048.
- Pokorny, J. (2013). NoSQL databases: a step to database scalability in web environment. *International Journal of Web Information Systems*, 9(1):69–82.
- Rong, C., Nguyen, S. T., and Jaatun, M. G. (2013). Beyond lightning: A survey on security challenges in cloud computing. *Computers & Electrical Engineering*, 39(1):47–54.
- Rossetti, M. D. and Chen, Y. (2012). A cloud computing architecture for supply chain network simulation. In *Simulation Conference (WSC), Proceedings of the 2012 Winter*, pages 1–12.
- Siegmund, F., Bernedixen, J., Pehrsson, L., Ng, A. H., and Deb, K. (2012). Reference point-based evolutionary multi-objective optimization for industrial systems simulation. In *Proceedings of the Winter Simulation Conference*, page 130. Winter Simulation Conference.
- Subashini, S. and Kavitha, V. (2011). A survey on security issues in service delivery models of cloud computing. *Journal of Network and Computer Applications*, 34(1):1–11.
- Zhu, B., Li, H., and Lu, J. (2012). A cloud and parallel computation system for IC electromagnetic compatibility modelling and optimization. In *2012 Asia-Pacific Symposium on Electromagnetic Compatibility (APEMC)*, pages 877–880.
- Zissis, D. and Lekkas, D. (2012). Addressing cloud computing security issues. *Future Generation Computer Systems*, 28(3):583–592.

## BIOGRAPHIES

**INGEMAR KARLSSON** is a Ph.D candidate at the University of Skövde. He has a B.Sc. in Computer Science and a M.Sc. in Automation Engineering from University of Skövde. His research interests include simulation-based optimization, cloud technologies and visualization. His email address is <ingemar.karlsson@his.se>.

**AMOS NG** is a Professor of Production and Automation Engineering at the University of Skövde, Sweden. He holds a Ph.D. degree in Computing Sciences and Engineering. His research interests lie in applying multiobjective optimization for production systems design & analysis. His e-mail address is <amos.ng@his.se>.

**ANNA SYBERFELDT** is a senior researcher at the University of Skövde, Sweden. She holds a PhD in Computer Science from the De Montfort University, UK and a Master’s degree in Computer Science from the University of Skövde, Sweden. Her research interests include simulation-based optimization, artificial intelligence, metaheuristics, and advanced information technology with applications in logistics and manufacturing. Her e-mail address is <anna.syberfeldt@his.se>.

## ACKNOWLEDGEMENTS

This work is partially financed by the Knowledge Foundation (KK Stiftelsen), Sweden. The authors gratefully acknowledge them for the provision of research funding.

# FRAMEWORK FOR AN ADAPTIVE DECISION SUPPORT SYSTEM FOR INDUSTRIAL SHOP-FLOOR OPERATORS

Magnus Holm  
Virtual Systems Research Centre  
University of Skövde  
541 28 Skövde  
Sweden  
Email: magnus.holm@his.se

Göran Adamson  
Virtual Systems Research Centre  
University of Skövde  
541 28 Skövde  
Sweden

Lihui Wang  
KTH Royal Institute of  
Technology  
100 44 Stockholm  
Sweden

Philip Moore  
Academy of Innovation & Research  
Falmouth University  
Tremough, Penryn  
Cornwall, TR109EZ  
United Kingdom

## KEYWORDS

Adaptive decision support system, Shop-floor operator, Function blocks

## ABSTRACT

Today's shop-floor operators' working tasks often stretches over a broad spectra of jobs; from ordinary production assignments to handling errors and performing maintenance. Demands for enhanced skills and knowledge are constantly raised to limit the consequences of tool breakage, machine down time and other stochastic events negatively affecting the production.

To be able to meet these increasing demands a framework for a distributed and adaptive decision support system is proposed. It will help the shop-floor operator to distinguish between decision options and minimize time to consider appropriate actions to maximize productivity both during normal production and when facing unexpected or unscheduled events.

*"In what ways is it possible to support operators in making decisions for optimal productivity?"* was the opening question from the industry partner when beginning this research. Targeting this question a novel framework for an adaptive decision support system (DSS) enabled by event-driven function blocks, based on decision logics is proposed. Its ability to adapt to the actual conditions on the shop-floor is compared to a currently used voice message system in a test case.

## INTRODUCTION

Short product life cycles in combination with an increasing number of product variants due to rapidly changing market demands is a normal day at work for most manufacturing companies today. To be competitive in a rapidly changing global market the production system has to be able to rapidly respond to shifting demands. Much research effort and

resources have focused on how to optimize the machines, buffers, processes etc. of the production system. Less research effort has been spent focusing on the shop-floor operator's perspective. The role of the operator has evolved from one being stationed at one machine solving one specific task to an employee having increased responsibilities and scope which demands extensive skills in an environment where the difference between blue and white collar becomes less distinct (Dencker et.al., 2009). The efficiency of production as a whole is essential to the future competitiveness of a sound manufacturing company.

The environment of the shop-floor operators is affected by external variables set by demands from the market in combination with internal stochastic variables, such as, broken or missing tools, fixture shortages, down-time for machines and changes in the number of available operators. Such stochastically changing conditions raise the bar demanding an ability from shop-floor operators to handle and act in an information intensive environment with an increased degree of uncertainty. Neither traditional planning systems nor ordinary control systems usually have the capability of handling such events, which consequently negatively affect the production efficiency (Xu et. al. 2011). The level of adaptability of the production system is a significant indicator of its ability to handle and manage counterproductive situations. A distributed control system in combination with real-time production data can help facilitate adaptive decision-making and a dynamic control capability. It can enable the vital and valuable capacity of the production system's ability to handle uncertainty (Monostori et. al., 2010) (Wang et. al., 2012). Limited time and a lack of transparency of available options and their consequences challenge the operators in considering appropriate actions both during normal production and when facing unexpected or unscheduled events. This indicates a need for adaptive decision support for shop-floor operators. This paper introduces a proposed architecture for such an adaptive and distributed DSS for shop-floor operators working in a demanding production environment. Adaptive decision logic

is tested in a case study simulation of a production line together with our collaborative partner from the automotive industry.

This paper is organised in the following sections. Section 2 presents a literature review focusing on the shop-floor operator and on decision support for shop-floor operators. Section 3 presents the proposed frame work for an adaptive decision support system. In Section 4 a case study comparing the current voice message system and the proposed adaptive decision logic is presented. Section 5 concludes the paper.

## LITERATURE REVIEW

The great potential of operators to adapt to different situations is discussed by Payne et. al. (1993). However, during production with dynamically changing demands and stochastic events, it is unlikely that an operator is able to process and evaluate all available information to make informed decisions and prioritizations at all times. It would be good if the operators could act as rational decision makers, defined by Lee (1971). A rational decision maker should be completely informed, infinitely sensitive and rational, knowing exactly what to do in every situation. No shop-floor operator can hold and process all production information in real-time being a rational decision maker, but a well-designed DSS can help in getting closer. A decision is not an isolated event in time. The current situation of the shop floor operator is affected by previous production status and also impacts the subsequent status. It is therefore essential that the DSS used on the shop floor is developed with the operator in focus. Despite the obvious importance of the narrow time window and connectedness to the whole, it is seldom modelled in research (Hollnagel 2002).

Research focusing on DSS for industry is not an immature academic topic. In 1971 Gorry and Scott Morton presented the concept “decision support system”. In two comprehensive publications, more than 30 years later, Arnott and Pervan (2005 & 2008) reviewed production research concerning decision support systems. They defined seven major DSS sub-fields in the literature reviewed but could also point out a severe and significant lack of case studies, widening a gap between research and practice. An indirect signal of a general absence in DSS research which focuses on the needs of shop-floor operators is their definition of DSS “the area of information systems (IS) discipline that is focused on supporting and improving managerial decision-making”. Though there is some light in the cloudy sky, some research focusing the shop-floor operator and industrial production case studies in DSS for shop-floor operators has been performed.

A research project aimed at the shop-floor operator is ProAct (Dencker et.al 2009), which focuses on the importance of reaching a proactive working environment for shop-floor operators. Dencker et. al. conclude that operators having proactive behaviour will lead to higher competitiveness through flexibility gains and also a reduction of the total lead time for assembly operations. The design of a proactive assembly system is further discussed by Fasth et. al. (2010). The potential of shop-floor operators is utilised using criteria and interaction of the areas; automation, information and competence of the operators’. The “operator of the future” or

at least an approach trying to find its definition and requirements are presented by Grane et. al. (2012) and Berlin et. al. (2012). Through workshops engaging several Swedish process and manufacturing industries and system developers, the tasks and needs of the future operator were investigated. The single most important variable found in their work is information interpretation. Future shop-floor operators have to be able to correctly interpret their working environment in order to act as a part of it. In a production environment where the amount of information will cause overload of any human recipient the need for a DSS is clear.

A DSS for operators at an engine manufacturing unit using a distributed environment is presented by Gertosio and Dussauchoy (2003). The developed DSS communicates with each operator giving a list of next executable actions the operator can act upon in order to increase effectiveness. Swanepol (2004) worked together with a tube mill and developed a DSS for their operators. It assists the operator to, when needed, manually change operating settings of the welding process to return it to stable conditions. The welding process in the tube mill is monitored on-line acknowledging the operator when an alarm occurs. The DSS indicates appropriate action to be taken by the operator. Elghoniemy et al. (2006) presents a simulation-based intelligent decision support system (IDSS) developed in cooperation with a Canadian milling plant manufacturing doors and windows. The IDSS assists operators in decision-making by presenting recommended actions and through a simulator the operators are able to examine the effects of the recommendations before choosing the one to implement. Continued research on the IDSS at the rough mill was presented by Elghoniemy and Gruver (2012) in which an agent ontology was developed, and a prototype system implemented. The IDSS-demonstrator offers the production staff benefits, such as helping operators to make consistent and standardized decisions, the ability to run the production using less experienced personnel and a possibility to train new operators, as well as knowledge acquisition and transfer when experienced operators update the IDSS recommendations. Gebus and Leiviskä (2009) present a DSS for an electronic assembly line using knowledge acquisition. They state that “an effective decision support system is essential to provide workers with information necessary to identify the causes of a problem and take appropriate action to solve it”. The aim of the proposed DSS is to improve product quality by providing to the operators understanding and formalization of the parameters influencing it. Frantzén et. al. (2011) presents a DSS which uses a simulation-based scheduling system for on-line optimization (OPTIMIST). The OPTIMIST system is implemented at an automotive manufacturer giving the shop-floor operators continuously updated and near-optimum scheduling solutions. Simulation Techniques are also used by Ugarte et. al. (2008) when developing a DSS for an aluminium industry. The approach supports a rescheduling process in an enterprise resource planning (ERP) controlled system. Thorvald et. al. (2014) focuses the information context in a DSS for assembly workers. They discuss and investigate in what way the performance of operators is affected by how the needed assembly information is presented.

Research on decision support systems are comprehensive but

as presented in the literature reviews by Arnott and Pervan (2005 & 2008) has the major approach been from the IS point of view. Their reviews also show that when researchers have approached production systems the focus has been on the management level. Few researchers have published work from the industrial shop-floor operators' perspective of decision support and their need of adaptive support in real-time.

## FRAME WORK FOR AN ADAPTIVE DECISION SUPPORT SYSTEM

To optimally run a modern production line with its combination of automatic, semi-automatic and manual stations is a tough task even if every machine and operation is up and running according to plan. The operators facing tool and machine breakdowns, variances in material and cycle-time and on top of that various alarms will strongly impact the production outcome through made action prioritization. To meet the changing conditions and incoming stochastic events the supporting systems must be able to adapt to actual production status and respond in real-time.

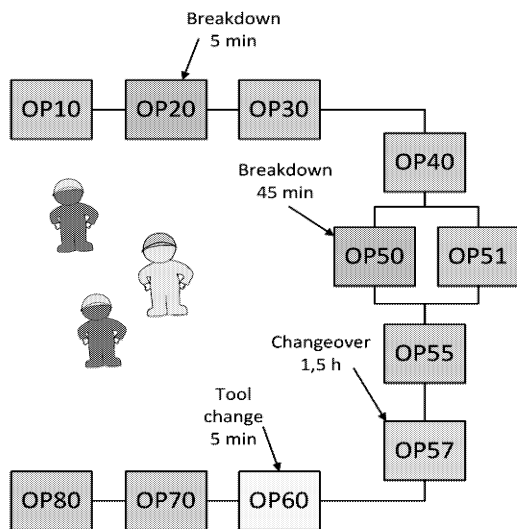


Figure 1. Question initiating project: *In what way is it possible to support operators in making decisions for optimal productivity?*

## Framework for an Adaptive Decision Support System

When initiating the research project Wise-Shopfloor a question from one of the industrial project partners was “In what way is it possible to support operators in making decisions for optimal productivity when the production system calls for action at several locations?” (Figure 1). To meet the demands from the project partner and the shop-floor operators, architecture for an Adaptive Decision Support System (ADSS) is proposed. The aim of the proposed ADSS is to provide the right operator with the right information at the right time in a dynamic production environment with stochastic events calling for actions. It is not enough that the DSS just “reply” with static information. The DSS has to adopt to the actual situation on the shop floor: What operators are available? What other events are calling for action? etc. The DSS should also be possible to customize by

its users as later described in this paper. All together the DSS must be adaptive to different settings and changing conditions – an Adaptive DSS.

In a literature survey by Power (2002) five areas are listed in which an organization can benefit by using a DSS:

- A DSS can enhance productivity.
- A DSS can improve decision quality.
- A DSS can support interpersonal communication.
- A DSS can improve the quality of the made decisions.
- A DSS can increase organizational control through standardization of decisions/actions.

Of all DSS reviewed by Power (2002) none claimed to perform in all five areas. The proposed ADSS focuses the first three areas in Power’s list to facilitate shop-floor operators to make decisions for optimal production output.

The proposed ADSS is divided into four main modules: Production status evaluation, Adaptive decision logic, Dynamic resource position detection and the Operator device (Figure 2). The ADSS also uses some supporting facilities: real-time monitoring of the production line together with process plans and the production status from the production management system (PMS). The status of the production is evaluated and function block based decision logic is used in combination with resource positioning to enable decision support to the shop-floor operators in real-time both during normal production and when incoming events call for action. Supporting modules are marked with dashed lines in Figure 2. The four main modules of the ADSS can be divided into two layers, one server level which holds the three outer modules (marked with \* in Figure 2) and one individual level, each operator’s device, which can be individually configured to some extent.

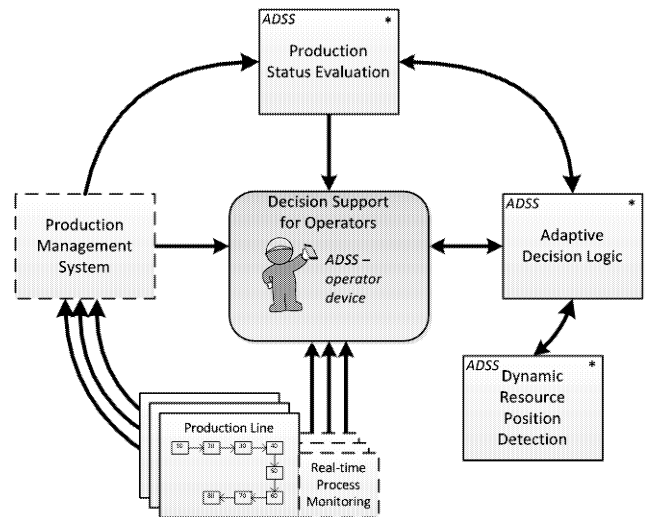


Figure 2. Architecture of proposed ADSS

## Process monitoring and production information

Updated and correct data from the production system is critical for the operators’ ability to correctly understand and interpret the current production status. The proposed ADSS uses real-time monitoring technology as described by Wang et. al. (2011). It uses a web based solution which



considerably reduces network traffic while increasing the controllability of the production system. The web based solution is not limited to shop-floor usage but can also be used from remote locations. The PMS collects and saves information about production rate and status. It is displayed as well as machine availability to both the management and the shop-floor operators. The current production status both in an overall perspective and for the different sectors of the production is evaluated using the outputs from the ADSS adaptive decision logic and the PMS.

### Production evaluation and decision logic

The adaptive decision logic and also parts of the module production status evaluation are built by event-driven function blocks (FB). These FBs are reusable functional modules based on the event-driven models defined in the international standard IEC 61499. The behaviour of the FB is internally controlled by a state machine. Different internal algorithms are executed by the state machine depending on arriving events and the current state. In Wang et. al. (2008) and Holm et. al. (2013) is the functionality and test cases using FBs more thoroughly described.

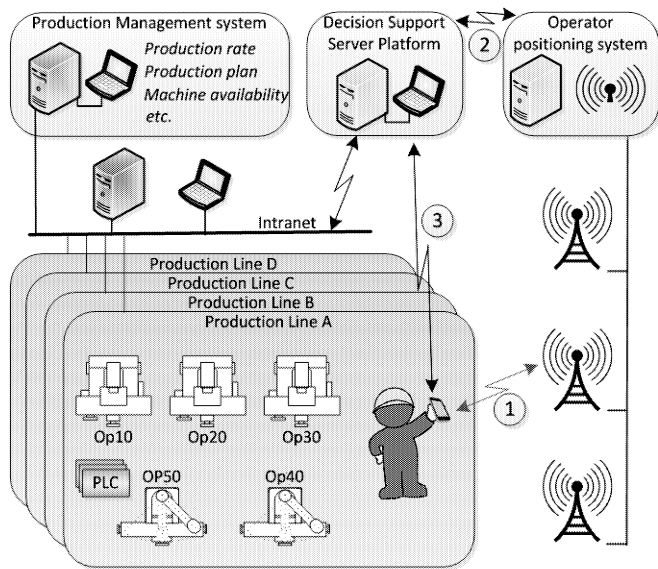


Figure 3. Operator positioning system

Interviews with ten operators at three production lines at Volvo Cars in Skövde were made during 2012. The aim of the interviews was to have their view of the everyday situation and possible needs and requirements of a DSS for shop-floor operators. The most emerging property mentioned by the interviewed operators was that the DSS should be customizable. Each group and in some cases each individual should be able to customize the settings. This requirement is met through the selectable characteristics of the Adaptive Decision Logic by the internal algorithms of the FBs. Run-time output from the other three ADSS-modules are used to control the FBs enabling adaptive decision support in real-time to the shop-floor operators. The algorithms in the FBs can be updated or changed when, for example, a new optimum, due to changing production conditions, is found.

### Dynamic resource location detection

A shop-floor operator of today is usually not stationed at one location but operates over a larger area handling several machines and stations. Also trucks and other joint resources move in the whole production premises and their position is an important variable when giving information and assigning tasks. The proposed ADSS uses low-energy Bluetooth technology for a positioning system that enables tailored information and task allocation depending on position and availability. The positioning system tracks, for example, the position of the operator device ①. It is sent as input ② to the decision support server and the decision logic, affecting the support output ③ to the operators (Figure 3).

### CASE STUDY

In the case study, two different approaches on assigning operators to incoming stochastic events from the production line are evaluated by comparing the throughput per hour. The module Adaptive decision logic is compared to the at site installed voice message system. Online production data in real-time was not available for the test case; instead the production throughput is compared using a simulation model of a production line and its operators. The model has two lines, one of them having 13 stations served by four operators. Between each station is a buffer (triangle) with three places. The other line has eight stations and two operators (Figure 4). This line has two places in each buffer between the stations and the last buffer before it is connected to the longer line has six places. The longer line is divided into three sections and the shorter one into two sections. The shorter line produces products that are used in the longer production line. If necessary the operators can of course help their colleagues both on their own and the other line.

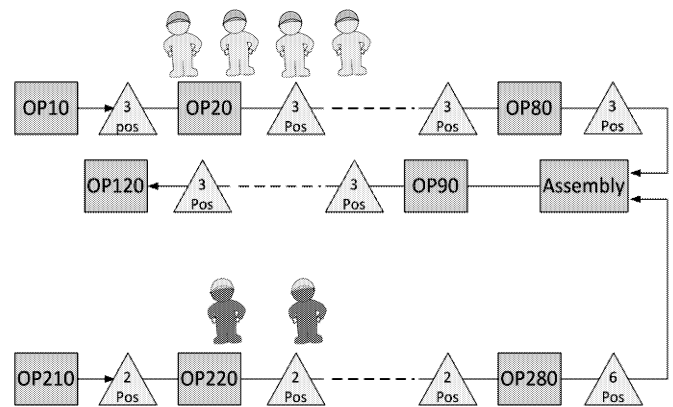


Figure 4. Production system used for the two experiments in the test case.

During the test case two experiments are performed. The first one uses a model of the installed voice message system to acknowledge the shop-floor operators. It polls machine status using the beacon signals of every station. An occurring event is acknowledged to every operator within the sector through a voice message to the operators using handheld receivers. The events can either indicate that a machine is changing



alarm status or that it is time to do a periodic measuring verification. If the operator is occupied, another operator in an adjacent section is called. During working hours the operators rotate the sections in which they work. In Figure 5 is depicted the event flow through the voice message system used in the test case.

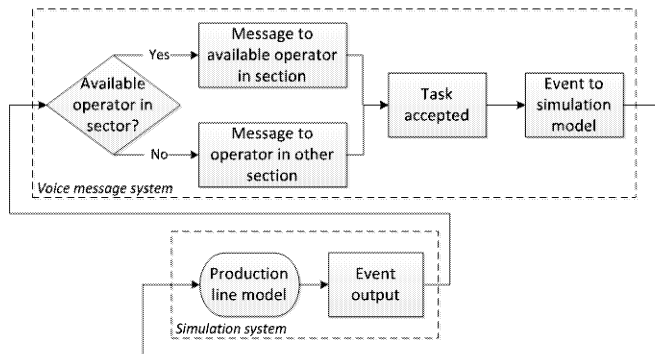


Figure 5. Event flow through the current voice message system.

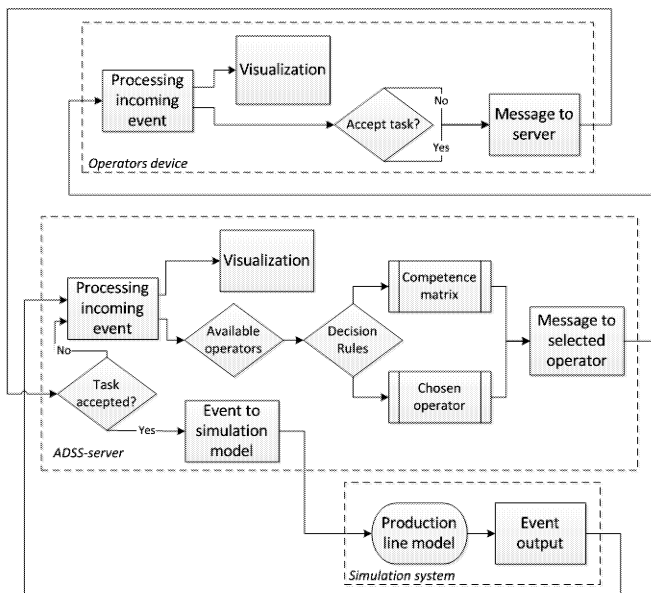


Figure 6. Event flow through the proposed ADSS frame work.

Experiment number two in the test case uses the ADSS and the decision rule “Competence matrix”. The adaptive decision logic uses the competence matrix when assigning an operator to a task. The event flow through the ADSS-framework used in the second experiment is illustrated in Figure 6. Incoming events from the production line are processed by the ADSS-server. The decision logic choses among available operators the one with the most suited competence for the current task. A message is sent to the assigned operator’s device. The assigned task is visualized and the operator has to either accept or refuse the assignment. If accepted the ADSS-server sets the operator’s state to *occupied*. When the task is completed, the ADSS-server is acknowledged and the operator state is reset to *unoccupied*. If the task is denied by the operator or if the limited respond time is exceeded, the ADSS-server assigns the task to another operator.

During the simulation events (errors, tool changes etc) are generated at the stations. The response from each system (VMS and ADSS) giving impact to the model’s output are used for evaluation. During the case study each of the two systems, the voice message system in experiment one and the ADSS in experiment two, was run in the simulation model for 168 hours (seven days) with 50 replications. The voice message system in the first experiment reached an average throughput per hour of 9,739 with a standard deviation of 0,206. The average throughput per hour in the second experiment using the ADSS was increased with almost 9%, reaching 10,608 with a standard deviation of 0,071. The proposed ADSS using the FB-algorithm “Competence matrix” showed a significant better result than the installed voice message system (Figure 7 and Table 1).

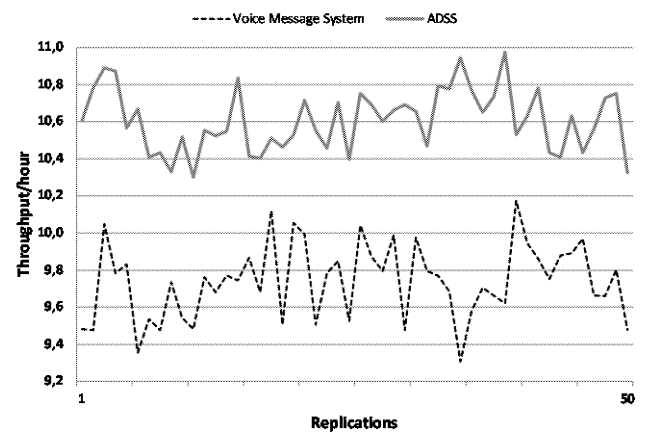


Figure 7. Average throughput/hour for the two experiments in the test case.

Table 1. Statistics from the two experiments in the case study.

Experiment	Tp/h average	Standard Deviation	Minimum	Maximum
Exp 1	9,739	0,206	9,307	10,175
Exp 2	10,608	0,168	10,301	10,976

## CONCLUSIONS

Decision support systems are not newcomers within the manufacturing industry. There are several different systems operating on various organisational levels but few have its base on the shop-floor with the operators in focus. The predominant base for DSS has been a managerial level with a decision time range of at least some hours up to several months. The shop-floor operator cannot spend that much time between incoming event and implemented action. A decision has to be made by the shop-floor operator usually within seconds in a dynamic environment with stochastic events. In this paper a novel DSS aiming the needs of the shop floor operator is presented. The presented scope is within automotive industry but the system can be used in any working environment where operators struggle with stochastic events (errors, tool changes etc) negatively affecting the production.

This paper's contribution against the reported literature can be concluded as follows. 1) The shop-floor operator is in focus of the proposed ADSS-framework. The ADSS gives decision support in real-time compared to managerial decision support systems having a decision time range of several hours or more. 2) The proposed ADSS-framework uses event driven function blocks which enables system adaptability through its encapsulated algorithms. 3) The proposed ADSS-framework facilitates the shop-floor operators to be more like rational decision makers. The software used for simulation in the test case is Plant Simulation.

## ACKNOWLEDGEMENTS

This paper is based on the research project Wise-ShopFloor, which is funded by the Knowledge Foundation in Sweden. The authors would like to thank Per Thim and Per Carlson at Volvo Cars in Skövde for their contribution to the project which made this paper possible. The authors would also like to thank the anonymous reviewers whose comments helped to improve this paper

## REFERENCES

- Arnott D; G. Pervan. 2008. Eight key issues for the decision support systems discipline. *Decision Support Systems* 2008;44:657-672.
- Arnott D; G. Pervan. 2005. A critical analysis of decision support systems research. *Journal of Information Technology* 2005;20(2):67-87.
- Berlin C; J. Andersson; Å. Fasth; C. Grane; L. Abrahamsson; J. Johansson; A. Osvalder; and J. Stahre. 2012. "Keyword mingling workshop – a method for identifying and consolidating industrially perceived needs and requirements of future operators." *Swedish Production Symposium 2012*, ISBN 978-91-7519-752-4-
- Dencker K; Å. Fasth; J. Stahre; L. Mårtensson; T. Lundholm; and H Akillioglu. 2009. "Proactive assembly systems-realizing the potential of human collaboration with automation." *Annual Reviews in Control*, 33:3 IFAC/Pergamon, ISSN: 1367-5788.
- Elghoniemy E; and WA. Gruver. 2012. Agent-based decision support and simulation for wood products manufacturing. *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and reviews*. 2012;42:No 6.
- Elghoniemy E; Ö. Uncu; WA. Gruver; DB. Kotak; and M. Fleetwood. 2006. An intelligent decision-support system for rough mills. *International Journal Manufacturing Technology and Management* 2006;8:203-225.
- Fasth Å; J. Stahre; J. Bruch; K. Dencker; T. Lindholm; and L. Mårtensson. 2010. "Designing proactive assembly systems (ProAct) – Criteria and interaction between automation, information, and competence." *Asian International Journal of Science and Technology in production and manufacturing engineering (AIJSTPME)*, vol 2(4):1-13.
- Frantzén M; AHC. Ng; and P. Moore. 2011. A simulation-based scheduling system for real-time optimization and decision making support. *Robotics and Computer-Integrated Manufacturing* 2011;27:696-705.
- Gebus S; and K. Leiviskä. 2009. Knowledge acquisition for decision support systems on an electronic assembly line. *Expert Systems with Applications* 2009;36: 93-101.
- Gertosio C; and A. Dussauchoy 2003. A distributed decision support system for operational units, application to a manufacturing unit. *Proceedings of IEEE conference on Emerging Technologies and Factory Automation* 2003;2:575-581.
- Gorry GA; and MS. Scott Morgan 1971. A framework for management information systems. *Sloan Management Review* 1971;13(1):1-22.
- Grane C; L. Abrahamsson; J. Andersson; C. Berlin; Å. Fasth; J. Johansson; J. Stahre; and A. Osvalder. 2012. "The operator of the future – a key to competitive industry in a future information society." *Swedish Production Symposium 2012*, ISBN 978-91-7519-752-4.
- Hollnagel E. 2002 Time and time again. *Theoretical Issues in Ergonomics Science* 2002;3:143-158.
- Holm M; G. Adamson; and L. Wang. 2013. Enhancing Adaptive Production using IEC 61499 Event-Driven Function Blocks. *Proceedings of NAMRI/SME* 2013;41.
- Lee W. 1971. Decision theory and human behavior. New York: John Wiley & Sons.
- Monostori L; BCs. Csáji; B. Kádár; A. Pfeiffer; E. Ilie-Zudor; Zs. Kemény; and M. Szathmári. 2010. Towards adaptive and digital manufacturing. *Annual Reviews in Control* 2010;34:118-128.
- Payne JW; JR. Bettman; and EJ. Johnson. 1993. The adaptive decision maker. Cambridge University Press; 1993 ISBN 0-521-41505-5
- Power DJ. 2002. Decision support systems: Concepts and resources for managers. Westport, Connecticut: Quorum Books.
- Saenz de Ugarte B; A. Hajji; R. Pellerin; and A. Artiba. 2008. Development and integration of a reactive real-time decision support system in the aluminium industry. *Engineering Applications of Artificial Intelligence* 2008;22:897-905.
- Swanepol K. 2004. Decision support systems: real-time control of manufacturing processes. *Journal of Manufacturing Technology Management* 2004;15(1):68-75.
- Thorvald P; D. Högberg; and K. Case. 2014. The effect of information mobility on production quality. *International Journal of Computer Integrated Manufacturing* 2014;27(2):120-8
- Wang L; M. Givehchi; B. Schmidt; and G. Adamson 2012. A Function Block Enabled Robotic Assembly Planning and Control System with Enhanced Adaptability. *Proceedings of the 45th CIRP Conference on Manufacturing Systems* 2012. p. 194-201.
- Wang L; M. Givehchi; G. Adamson; and M. Holm. 2011. A sensor-driven 3D model-based approach to remote real-time monitoring. *CIRP Annals – Manufacturing Technology* 2011;60:493-6.
- Wang L; Wise-ShopFloor: An Integrated Approach for Web-Based Collaborative Manufacturing. *IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and reviews* 2008;38(4).
- Xu X; L. Wang; and ST. Newman. 2011. Computer-Aided Process Planning – A Critical Review of Recent Developments and Future Trends. *International Journal of Computer Integrated Manufacturing* 2011;24(1):1-31.

# **MANUFACTURING OPTIMIZATION AND EVALUATION**



# AGGREGATED DISCRETE EVENT MODELLING FOR SIMULATION AND OPTIMISATION OF MANUFACTURING SYSTEMS

Leif Pehrsson<sup>1,2</sup>

Simon Lidberg<sup>1</sup>

<sup>1</sup>Volvo Car Corporation  
SE-405 31, Gothenburg  
Sweden

leif.pehrsson@his.se

Marcus Frantzen<sup>2</sup>

Tehseen Aslam<sup>2</sup>

Amos Ng<sup>2</sup>

<sup>2</sup>University of Skövde  
SE-541 28, Skövde  
Sweden

## KEYWORDS

Simulation, aggregation, discrete event systems, manufacturing systems, multi-objective optimisation.

## ABSTRACT

In many simulation studies for factory analysis, for example, to locate the constraint of an entire factory that consists of multiple production lines, it may not be effective to put every process detail into a single model. Firstly, to develop such a factory-wide model would be very time-consuming. Secondly, it can be very computational costly to run the model, especially if simulation-based optimisation is applied to find the optimal setting from such a complex model that possesses all the details of the processes. In this regard, aggregation, with which multiple process steps are aggregated into some simpler simulation objects, is an effective method to reduce both the development and computational times. On one hand, based on the initial analysis, the simulation expert can pinpoint the sub-system that restrains the performance of the entire factory and decide if a more detailed model is needed. On the other hand, inter-line buffers/storages can be readily optimised by using such an aggregated model. Through an application study with data from a real-world factory, this paper introduces a novel aggregation method and illustrates the potential of the above-said concepts.

## INTRODUCTION

Discrete Event Simulation (DES) techniques have, for many years, successfully been used for modelling and simulation of production lines. In combination with optimisation technology, especially evolutionary multi-objective optimization (MOO), data for decision support can be generated (Pehrsson and Ng, 2011). In recent years, research has shown that the addition of knowledge extraction techniques to MOO can produce very useful decision support (Dudas et al., 2013; Ng et al., 2012) based on the concept of “Innovization”, formed by the expression *Innovation* through *optimisation* (Deb and Srinivasan, 2006). However, aggregated DES modelling techniques for higher level manufacturing systems and supply chain applications have, to the authors’ best knowledge, not been explored to the same extent. Supply chains and logistic systems are often modelled based on mathematical approaches (Aslam et al., 2011). These kinds of models are sometimes considered less

accurate and might be more difficult to understand than DES models. Detailed modelling of complete factories and supply chains tend to be very time-consuming and when considering real-world decision situations, it might not be feasible to meet the deadlines for decision-making when using traditional DES practices. With the application of aggregated DES modelling techniques such shortcomings can potentially be prevented. A combination of detailed and aggregated modelling would open opportunities to explore a new field of multi-level models, enabling the simultaneous optimisation of high-level and low-level production systems parameters.

In this paper, some approaches for aggregated DES modelling of production lines and assembly lines of discrete products (e.g. automotive components) are discussed and then a novel approach is explored, based on a statistical analysis of the characteristics acquired from some detailed models of production lines. The resulting techniques for aggregated modelling will be tested in a simulation model of a complete production system with several inter-connected lines. The simulation model will then be used with multi-objective optimisation in order to generate decision support for improving the production system. However, the main purpose of this paper is to investigate if there is a potential for future development of generally applicable aggregation techniques for DES modelling for higher-level manufacturing systems modelling, such as an entire supply chain.

## AGGREGATED MODELLING

Rooda and Vervoort (2007) point out that manufacturing systems can be viewed and analysed from various levels. They distinguish between four abstraction levels, shown in Figure 1, of a manufacturing system where each level is studied, analysed and viewed upon from a different perspective. The four levels are defined as follows: (1) the *machine* level, where the manufacturing system is viewed upon as individual machines and where the system elements are the components within the machine; (2) the *workstation* level views the system in terms of a group of machines that are scheduled as one entity; (3) the production line level, which Rooda and Vervoort (2007) define as the *sub-network* level, consists of several machines or workstations in an area of the manufacturing plant in which the system elements are the individual machines in this area. The final level as defined by them is (4) the *network* level, which can be seen as the plant or factory level where the production lines and

their group of machines and workstations are connected as the elements of the system.

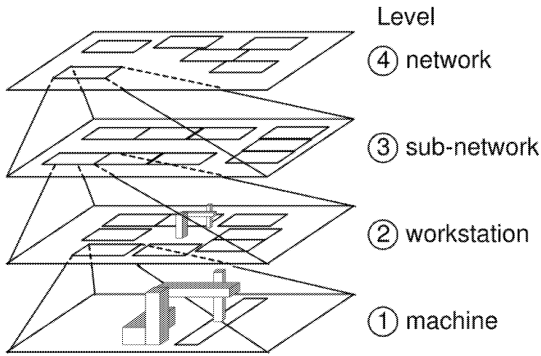


Figure 1: Levels in Manufacturing system, (Rooda and Vervoort, 2007)

In addition to the four abstraction levels for manufacturing systems defined by Rooda and Vervoort (2007), it is possible to add a fifth level, namely the *supply chain level*, in which individual factories or plants are seen as the elements connected in a networked system. Now, modelling and simulation is an effective way of designing, understanding, or analysing these complex systems, i.e. manufacturing systems, in various levels of abstraction and scope (Chwif et al., 2000). DES has over the years become one of the most preferable tools of industrial practitioners for studying and analysing complex systems (Madan et al., 2005). However, as mentioned earlier, detailed modelling of complete factories tend to be very time-consuming, not just in terms of model building, but also in terms of execution time. Particularly, when combining simulation with optimisation, so called simulation-based optimisation, it may require several hundreds or thousands of simulation runs on the simulation model. Chwif et al. (2000) point out that the complexity of a simulation model arises from its level of detail as well as its scope. For instance, when simulating manufacturing, it is possible to model the entire plant or an individual work cell depending on the scope. One way of coping with complexity of a simulation model is through aggregation or sometimes called model abstraction. Frantz (1995) explains that model abstraction or aggregation is an approach through which the modeller is able to reduce the complexity of the simulation model whilst preserving the validity of the simulation results with respect to the scope of the simulation model. He presents a taxonomy of model abstraction techniques which he categorises into three main groups, namely, *model boundary modification*, *modification of behaviours* and *modification of model form*.

There are several potential strategies for aggregated DES modelling of production lines to consider for evaluation. One example is aggregated modelling of a system with Automated Guided Vehicles (Pehrsson and Ng, 2013), in which standard objects such as operations and buffers were used to model a complex system with a reduced number of modelling details. Depending on the required accuracy of the results, the parameters to be altered or measured, and the objective of the simulation study, various trade-offs must be made. For example, a very simple way would be to use operation objects with a single product position, a cycle time parameter, and availability to represent lines and just

evaluate the throughput of the system. Such an approach may give accurate estimations of the throughput of the system, but there will be no accurate information about lead-times and Work In Progress (WIP), rendering such an approach useless for the evaluation of e.g. various production schedules. In order to find a sufficient modelling trade-off to solve the buffering problem of an entire plant with an aggregated DES model, different aggregation strategies have been investigated.

## Modelling principles

The basic requirements on a model for the simulation and optimisation of a manufacturing scheduling problem or the optimised design of a manufacturing system are related to the objectives to be met. At least some guiding information about the lead-time, throughput, and WIP of the system must be obtainable as output variables from the model. Hence there is an analogy with the parameters defining Little's law (Little 1961), in a form commonly defined when applied to manufacturing systems (Hopp and Spearman, 2000) in Equation 1.

$$TH = \frac{WIP}{CT} \quad (1)$$

Where  $TH$ =Throughput (arrival rate),  $WIP$ =Work In Progress, and  $CT$ =the cycle time of the complete system, also defined as the expected time spent in the system, or the lead-time (LT) through the system.

However, the stochastic nature of the system, not covered by Little's law by itself, is an essential factor to consider when evaluating the performance of a complete manufacturing setup and its response to alternative scheduling principles, buffer settings, or designs. One way of including the desired parameters would be to use a buffer object to represent a line. The average lead-time could be represented by the *sojourn time* the products spend in the buffer object and the maximum WIP could be modelled through the maximum buffer capacity. This approach would however be deterministic in its nature if the lead-time is a constant. Varying the lead-time is not easily done, since each individual work piece in the buffer object would be affected by such an operation. Another approach would be to map the behaviour from the first and the last operation of the line and try to mimic the behaviour with a series of objects consisting of an operation, a buffer, and another operation. The distribution in time for work pieces started in the first operation of the line could be mapped to the first operation object in the model, WIP could be handled by the buffer object, and the distribution of work pieces delivered from the last operation of the line could be modelled through the second operation object in the model. With this approach there might be a risk that the input distribution to some extent is a consequence of the output distribution in the real-world line and this might cause an underperforming model due to the fact that the input distribution is disturbing the output distribution and vice versa. Yet another strategy would be to model the line with just a buffer object representing the average WIP and lead-time in combination with an operation

object handling the output distribution. The technique chosen for further analysis is to use two serial operation objects separated by a buffer and to model the stochastic behaviour in the second operation object. The buffer object will hold the WIP of the modelled line and the first operation object will control the input pace.

### Required data and data collection

A number of different production lines have been analysed as input to the selection of an aggregated modelling technique. Several component production lines as well as assembly lines have been covered by this study in an attempt to find system behavioural patterns which can be generalised. One challenge was to capture the pure behaviour of a specific line rather than its behaviour as part of a larger system. The strategy selected to prevent such factors from distorting the data was to measure the required data in decoupled, well validated simulation models of the lines. The time for each work piece finished in the last operation (finish times) was logged. Information about the lead-time through the line was also logged for every produced item.

### Statistical analysis

The statistical analysis is in the greatest extent connected to the data collected and the aggregated modelling technique adopted. Based on the two types of data collected, i.e. (1) the finish times and (2) the lead-time through the line, the following methodology has been used:

1. Select an appropriate aggregated modelling technique.
2. Analyse the finish times and find distributions that fit the data.
3. Test the distributions in a simulation model and generate data to be compared against the original data set.

The first step is to select an aggregated modelling technique which is able to capture the main behaviour of the whole system with as few objects as possible, based on a limited amount of available data. The finish times of the products is the actual output to the other lines, and is the collective product of several disturbances and delays in a production system, such as tool breakage, waiting for operator and setup times. If a distribution is generated that fit the data of the finish times it will only generate correct values if no other disturbances are present.

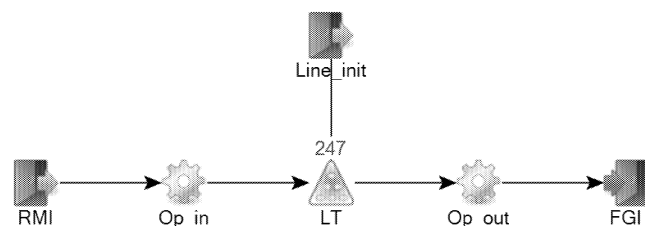


Figure 2: The proposed aggregated modelling technique

Therefore, a simple approach is to have variability in the last station of the production system only, and leave the other objects as deterministic. The production system has a lead-time that needs to be captured by the model as well. An aggregated simulation model of this kind is displayed in Figure 2.

The source object RMI models the unlimited supply of material and the source object Line\_init fills the production system with products at the initialization. The operation object OP\_in models the entrance pace to the system and the operation object OP\_out models the output pace of the system. Furthermore, the buffer object LT models the lead-time of the system and the sink object FGI is used to model the end of the line. LT is set to use the lowest lead-time measured from the data, which allows parts to have a short lead-time if OP\_out have a period of few disturbances. This is possible because of the entrance pace set in OP\_in, which will only feed the system in its own pace. In the example of component line 1 (CL1) demonstrated here, the capacity of the LT was set to 247 because it is the maximum capacity of the system, and the entrance pace of OP\_in was set to the highest processing time possible, whilst still achieving the desired production throughput. The operation objects in the model have a maximum WIP of one part each.

The finish times of CL1 were analysed in the software ExpertFit, and a histogram of the data (zoomed in on the majority of the values) is displayed in Figure 3.

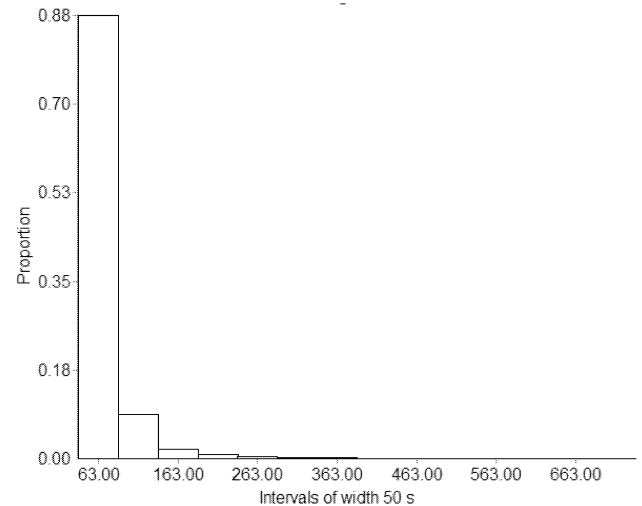


Figure 3: Finish times of CL1

In this case it was not possible to find a single distribution to fit the data because it is the collective result of a vast number of disturbances in the production system. Therefore, the actual processing times (without major disturbances) have been sorted out and displayed in Figure 4. The sorted out processing time related data includes 78% of the complete data set.

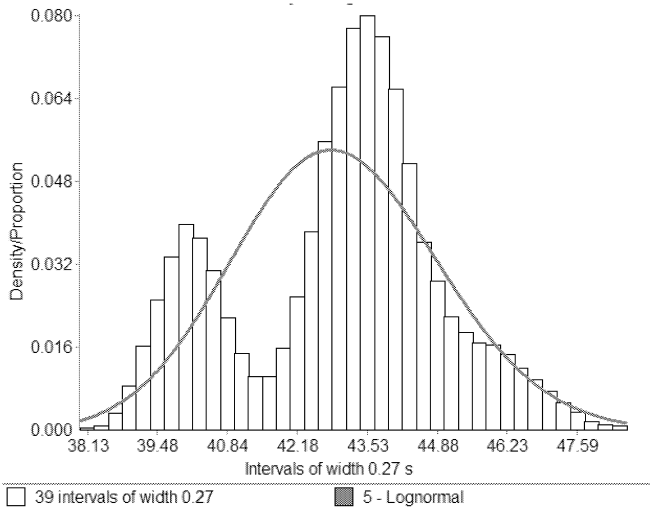


Figure 4: Finish times because of processing times in CL1

The processing time data plot in Figure 4 shows that there are two distinct groups of product variants. An empirical distribution could be used to fit the observed data very accurately. However the objective of this study is to find one or a few distributions to model a variety of production systems, including conceptual systems, on an aggregated level. Therefore, a lognormal distribution has been selected to model all product variants in order to maintain the model simplicity and support general applicability. According to Banks et al. (2010) the lognormal distribution can be used to model a process that is the product of a number of processes, e.g. with different processing times. The remaining 22% of the times, i.e. times above 48.5s, was subtracted with the mean value of processing times distribution ( $\mu_P$ ) and modelled as disturbances. The mean time between failures, i.e. the time from the end of one failure to the start of the next one, was determined to be approximately 195s ( $\mu_P/0.22$ ).

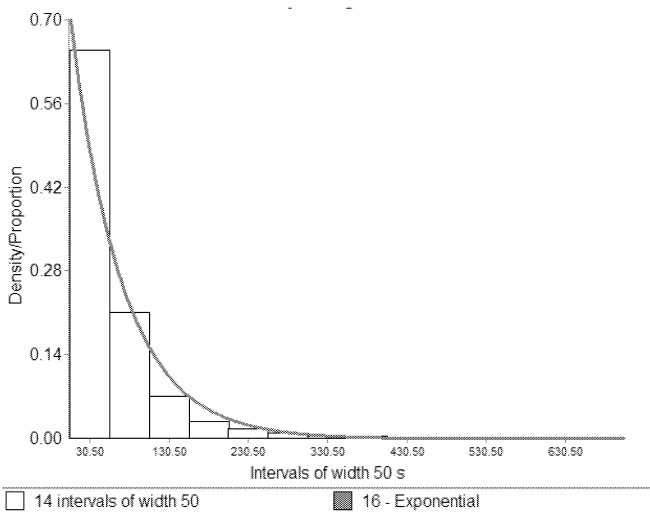


Figure 5: Finish times because of disturbances in CL1

The output data generated from the final aggregated model can be compared with the real-world data as shown in Figure 6.

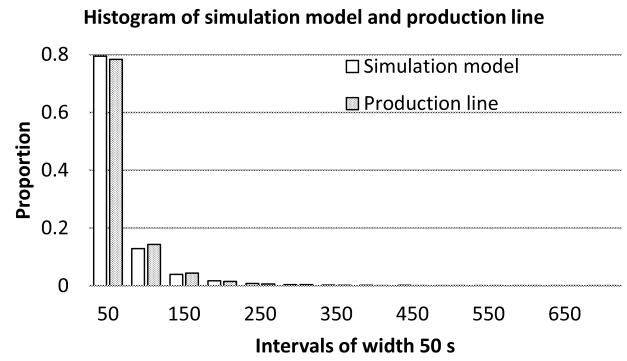


Figure 6: Histogram comparison of finish times

This method uses exponential distribution to model variability of the mean time to repair (MTTR). Because of its memoryless property, exponential distribution is the extreme common probabilistic model for failure time and repair time in manufacturing systems analysis (Curry and Feldman, 2011). Their conformity to the real data set can also be observed in Figure 5. However, the idea is to find reasonably good results with few objects, and the result on other measures will be compared next. Furthermore, not only the finish times have been compared, but also mean values and lead-time variation over time. The lead-time variation over time (days) is shown in Figure 7, in which the black line represents data from the simulation model and the grey line represents data from the real production line.

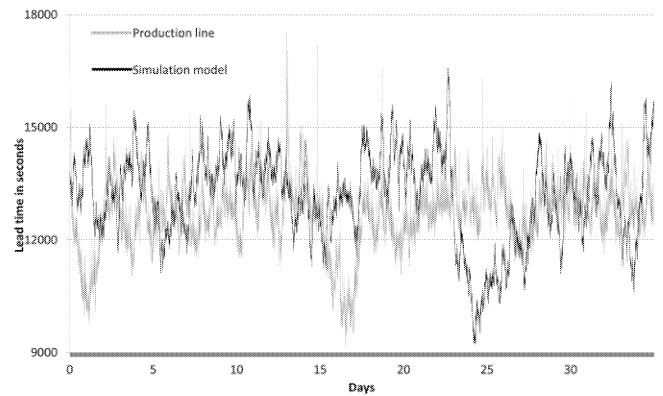


Figure 7: Lead-time variation comparison

The lead-time variation of the simulation model is close to the variation of the component line when considering the variation from one day to another. There is, however, a difference of variance when studying a shorter period. The mean values of the output variables of the simulation model are probably what matter the most, and they are usually used for building the deterministic models of larger systems. The mean values, throughput (TH) in jobs per hour (JPH) and lead-time (LT) in seconds, of the simulation model and the component line are shown in Table 1.



Table 1: Mean values comparison

Output	Simulation model	Component line	Result deviation
TH (JPH)	62.64	62.68	$\sim -0.1\%$
LT (s)	12688.08	12604.04	$\sim +0.7\%$

Consequently, the aggregated modelling technique is considered being able to capture the main behaviour of a whole system with only a few objects, based on a limited amount of available data. The fourth column of Table 1 shows the result deviation, i.e. the mean value of the simulation model divided by the mean value of the component line.

All six lines, three component lines (CL) and three assembly lines (AL), used the same aggregated modelling technique and statistical analysis. The mean values comparison is shown in Table 2.

Table 2: Mean Values Comparison of the Production Lines

Output	CL1	CL2	CL3	AL1	AL2 & AL3
TH	-0.1%	+0.1%	+0.1%	-0.1%	0%
LT	+0.7%	+0.1%	-0.9%	+0.2%	0%

Hence, the simulation models of the different lines generate mean values close to the real-world result and the stochastic of each line resembles quite well to the real-world result.

## AN APPLICATION STUDY

The aggregated modelling technique has been tested on an automotive industry manufacturing system comprising a factory with three component production lines and one sequential assembly line followed by two parallel assembly lines as shown in Figure 8.

## Modelling and simulation

The model was created in FACTS Analyser (Ng et al., 2007), a simulation software well suited for aggregated modelling, with integrated functionality for multi-objective optimisation and data analysis. Each component production line is followed by a finished goods stock with several purposes. One purpose is to enable different shift patterns in component production and assembly, another purpose is to enable batch production of certain variants, and a third purpose is to reduce the effect of availability disturbances in the component production lines. CL1 and CL3 consist of single piece flows and CL2 produces batches of 20 components. These lines originally have 30 to 70 machining cells each, and have automated material handling. The first of the three assembly lines is mainly automated and the two parallel ones are manual lines of walking-worker type (Pehrsson et al., 2013). Each assembly line has originally around 30 stations. The first and the two following assembly lines are separated by a buffer. The component production lines are operated on three shifts and the assembly lines are operated on two shifts.

FACTS Analyser has a function for modelling custom objects, based on any combination of standard objects, which was used to make a template for the aggregated production lines. The template was then used in the model to represent each of the six production lines in the system. Based on the statistical analysis of each line described under the modelling section, the plant model was populated with data. The first test was to run a simulation and log the buffer levels in order to monitor the actual build up of the WIP levels and the variation patterns. The throughput and the lead-time were measured for validation. The results have been found to be fairly accurate corresponding to the throughput of the known bottleneck in the real world system. The maximum throughput was 31.67 JPH, the average lead-time through the whole system was 73.2 hours, and the WIP was 3217 pcs.

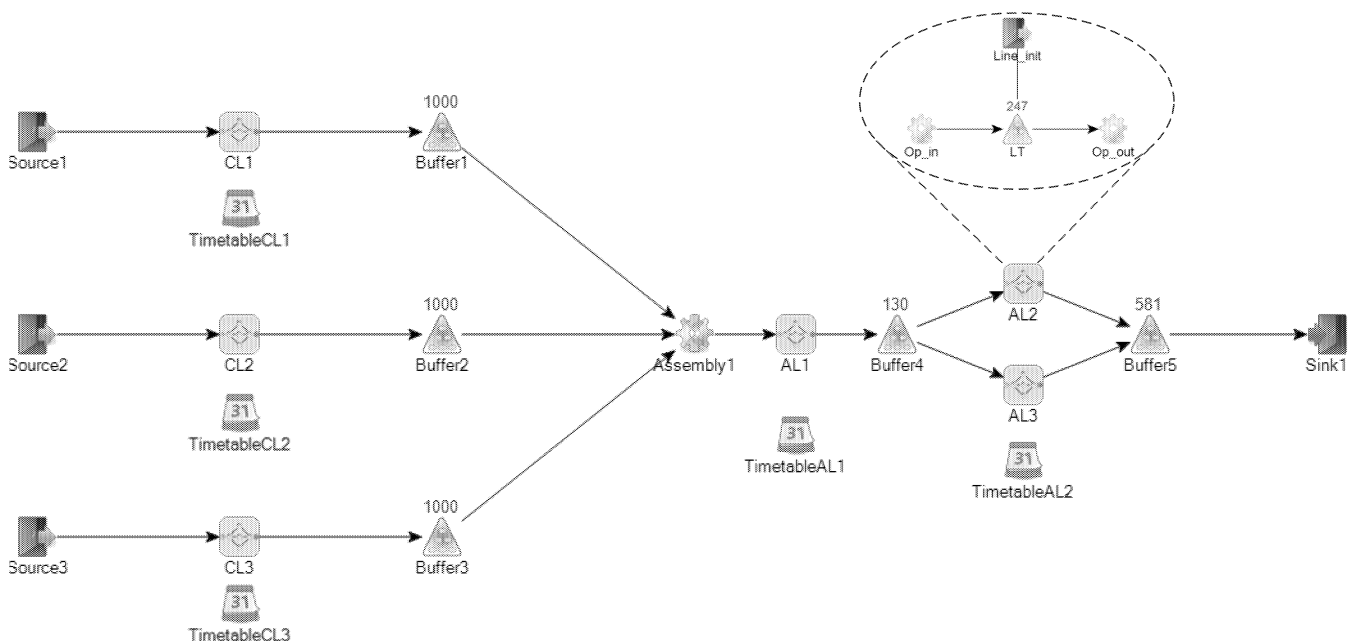


Figure 8: Aggregated plant model in FACTS Analyser with an example of an aggregated line model within the dashed line

This can be compared to the 30.07 JPH, from the constraining lines of the original model when distributed over a calendar week with the shift forms in the model as shown in Table 3.

Table 3: Shift forms and average calendar week TH

Line	Original TH	No of Lines	Hours per day	No of days op/week	Hours per week	Share of a week	Average TH/h during a week
CL1	62.64	1	23	5	115	68.45%	42.88
CL2	63.97	1	23	5	115	68.45%	43.79
CL3	47.38	1	23	5	115	68.45%	32.43
AL1	76.84	1	14.5	5	72.5	43.15%	33.16
AL2&AL3	34.84	2	14.5	5	72.5	43.15%	30.07

## Optimisation

A novel method to test if the aggregated model can capture the behaviour of the model when subjected to parameter changes, by using an optimisation towards some essential manufacturing systems objectives, has been used in this study. The optimisation problem chosen was to investigate the opportunities to search for high performing solutions based on the dimensioning of the buffers in the system. The basic problem can be formulated as minimising the sum of the buffers capacity (Lean buffer sum) and simultaneously minimise the lead-time and maximise the throughput as shown in Equations 2 to 4.

$$f_1 = \min(LT) \quad (2)$$

$$f_2 = \min(\sum_{i=1}^n Buffer_{i_{capacity}}) \quad (3)$$

$$f_3 = \max(TH) \quad (4)$$

where  $f_1$  = objective function for minimisation of lead-time,  $f_2$  = objective function for minimisation of the sum of the capacity in Buffer1 to Buffer5, and  $f_3$  = objective function for maximisation of throughput.

The optimisation was carried out using the NSGA-II algorithm (Deb, 2001) for 5000 iterations with ten replications each. The simulation horizon was 98 days including 28 days for warm-up. The optimisation parameters and the optimisation objective functions were connected to the simulation model through the integrated optimisation platform in FACTS Analyser.

## Results and analysis

According to Weiss and Indurkha (1998) “visual techniques are of primary interest for discovering knowledge” and visual techniques used for exploring multi-dimensional data from optimisations are also suggested by Simpson and Domdelinger (2009). The optimisation results show a natural and expected trade-off between the three objectives throughput, lean buffers, and lead-time when plotted in three dimensions as shown in Figure 9.

The optimal configuration of buffers in the system can be found by plotting the input parameters against the objectives as in the example with throughput, shown in Figure 10 to Figure 13.

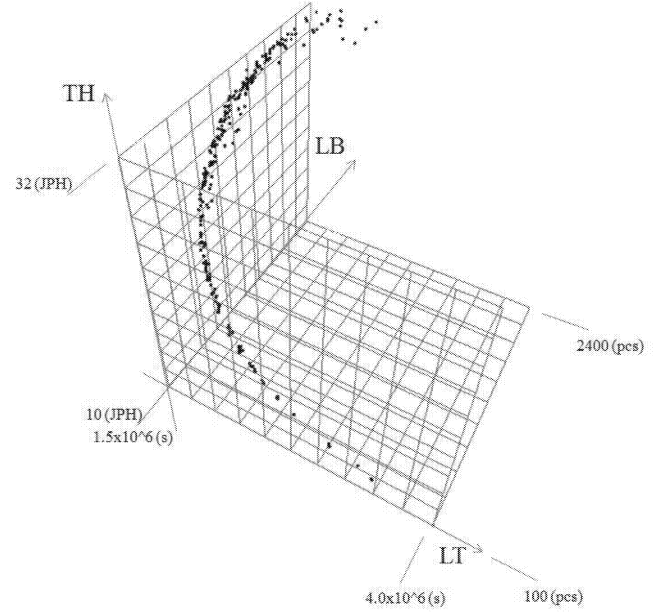


Figure 9: Non-dominated optimisation results for the three objectives

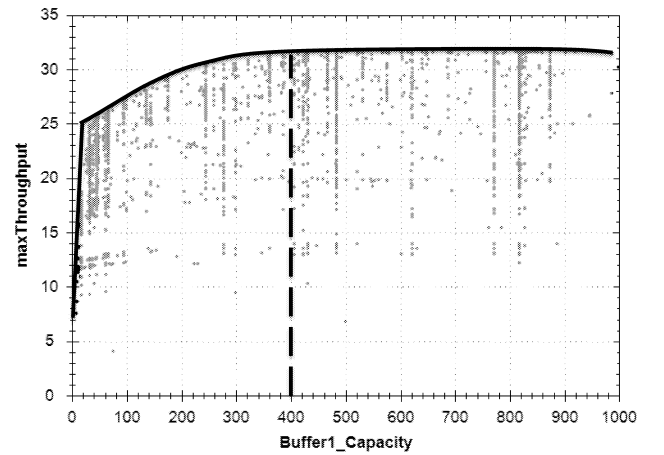


Figure 10: Throughput vs. buffer capacity for Buffer1

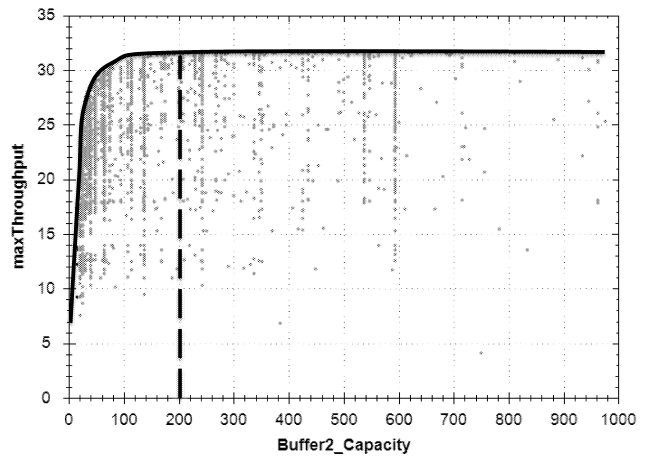


Figure 11: Throughput vs. buffer capacity for Buffer2

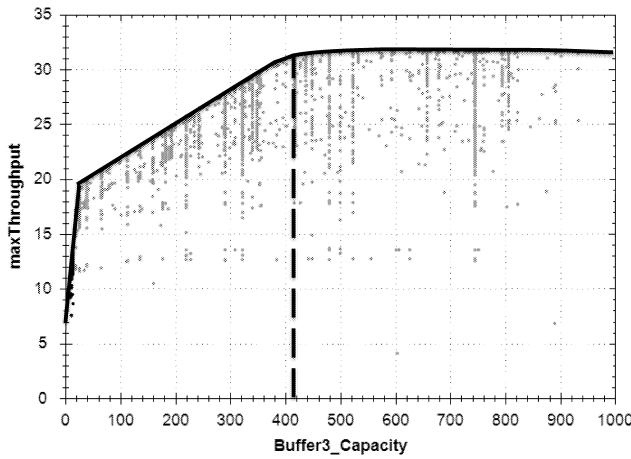


Figure 12: Throughput vs. buffer capacity for Buffer3

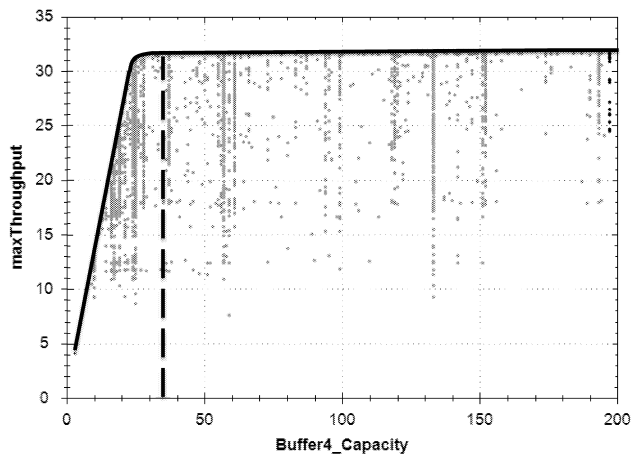


Figure 13: Throughput vs. buffer capacity for Buffer4

In order to achieve the maximum throughput from the system the buffer settings should be approximately: Buffer1 > 500, Buffer2 > 200, Buffer3 > 500, Buffer4 > 30, and Buffer5 > 50. The shortest lead-time for the system would be achieved by keeping the Lean Buffer sum within  $1000 < LB < 2000$ . Filtering the optimisation results accordingly shows that with  $LB < 1800$  and the buffer levels as proposed the throughput can be maintained with a WIP of 1911 pcs. Hence, a significant WIP reduction and lead-time reduction can be achieved while maintaining the throughput of the system, as shown by the parallel coordinate plot with the selected preferable solution highlighted in Figure 14.

By observing the different ramp-up slopes in Figure 10 to Figure 13, it can be seen that throughput is highly dependent on Buffer3, Figure 12, which is the storage situated directly after the component line with the lowest capacity. Below its saturation level at approximately 400 pcs., the system throughput drops almost linearly with reduced buffer capacity. These results, despite being not fully validated, shows an expected behaviour of the system, indicating there is a potential to use aggregation principles such as the one proposed in this paper to model manufacturing systems comprising a number of inter-connected lines in order to analyse and optimise manufacturing systems towards objectives on a system level. The required time for a single simulation was reduced to a few seconds, compared to several minutes for a simulation of just one of the original lines.

## CONCLUSIONS AND FUTURE WORK

Based on the results from the simulations and the optimisation study there seems to be a potential to aggregate line models into more computational efficient models for the use in the analysis and optimisation of systems and supply chains. It might be too early to refer to this study as a proof of concept, but there is a strong indication that the aggregated model can produce useful results for supporting decision-making on a manufacturing systems level. Since aggregation might degrade the accuracy of a model the results must be assessed very carefully. However, the results correspond with the real production system from which the original input data was acquired. The simulation time for the complete system was reduced down to a few seconds, a fraction compared to detailed DES models, approaching the territory of mathematical approaches.

Future work will have to further refine and validate the modelling techniques. The data requirements could potentially be reduced by generalising the principles on how to populate line models with data including common production performance indicators such as, cycle time, up-time, and mean down time. When the behaviour of a line can be generalised based on such measures, the modelling technique can be applied on a wider range of problems with easier access to data, and estimations into models of conceptual systems could become more accurate. The problems to be analysed should include product variants, scheduling, and system bottleneck detection.

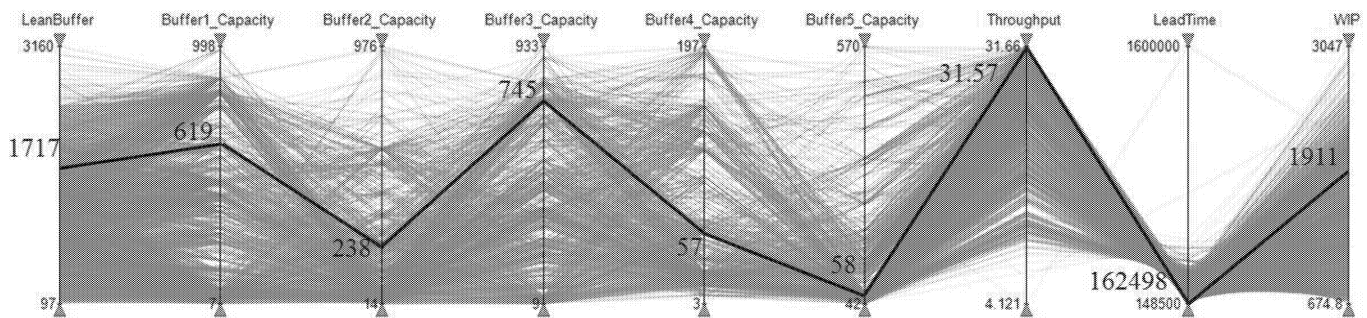


Figure 14: Parallel coordinate plot with the proposed solution in boulder black

## REFERENCES

- Aslam, T., Hedenstierna, P., Ng, A.H.C., Wang, L., 2011. Multi-Objective Optimisation in Manufacturing Supply Chain Systems Design: A Comprehensive Survey and New Directions, in: Wang, L., Ng, A.H.C., Deb, K. (Eds.), *Multi-objective Evolutionary Optimisation for Product Design and Manufacturing*. Springer London, pp. 35–70.
- Banks, J., Carson J. S., Nelson, B. L. and Nicol, D. M. 2010. *Discrete-Event System Simulation*. 5th ed. Upper Saddle River, New Jersey: Prentice-Hall, Inc.
- Chwif, L., Barretto, M.R.P., and Paul, R.J., 2000. On simulation model complexity. In: Joines JA, Barton RR, Kang K and Fishwick PA (eds). *Proceedings of the 2000 Winter Simulation Conference*. IEEE, Piscataway, NJ, pp 449–455.
- Curry, G.L. and Feldman, R.M. 2011. *Manufacturing Systems Modeling and Analysis*, 2<sup>nd</sup> edition, Springer-Verlag Berlin Heidelberg.
- Deb, K., 2001. *Multi-objective optimization using evolutionary algorithms*, 3<sup>rd</sup> ed., Wiley, Wiltshire, UK.
- Deb, K. and Srinivasan, A. 2006. Innovization: Innovating design principles through optimization. In *Proceedings of the Genetic and evolutionary Computation Conference (GECCO-2006)*, The Association of Computing Machinery (ACM), New York, 1629-1636.
- Dudas, C., Ng, A.H.C., Pehrsson, L. and Boström, H. 2013. Integration of data mining and multi-objective optimization for decision support in production system development, *International Journal of Computer Integrated Manufacturing*. DOI:10.1080/0951192X.2013.834481
- Frantz, F. K. 1995. A taxonomy of model abstraction techniques, In: *Proceedings of the 1995 Winter Simulation Conference*, ed. C.Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman., 1413-1420. Institute of Electrical and Electronics Engineers, Piscataway, New Jersey.
- Hopp, W.J. and Spearman, M.L., 2000. *Factory physics: foundations of manufacturing management*, 2nd ed., Irwin McGraw-Hill Higher Education, Burr Ridge, IL.
- Little, J. 1961. A proof of the queueing formula  $L=\lambda W$ . in *Operations Research* 9:383-387, 1961.
- Madan, M., Son, Y.-J., Cho, H. & Kulvatunyou, B., 2005. Determination of efficient simulation model fidelity for flexible manufacturing systems. *International Journal of Computer Integrated Manufacturing*. Vol. 18, No. 2-3, pp. 236-250.
- Ng, A.H.C., Dudas, C., Pehrsson, L. and Deb, K. 2012. Knowledge Discovery in Production Simulation by Interleaving Multi-Objective Optimization and Data Mining. In *Proceedings of the 5<sup>th</sup> Swedish Production Symposium (SPS'12)*, Linköping, Sweden, 6-8 November 2012.
- Ng, A.H.C., Urenda, M., Svensson J., Skoogh, A. And Johansson, B. 2007. FACTS Analyser: An innovative tool for factory conceptual design using simulation, In *Proceedings of Swedish Production Symposium*, Gothenburg 28-30 August 2007.
- Pehrsson L. and Ng A.H.C. 2011. An applied framework for simulation-based multi-objective production system optimisation within production system development, in the *Proceedings of the Industrial Simulation Conference, ISC'2011*, June 6-8, 2011, Venice, Italy
- Pehrsson L. and Ng A.H.C. (2013). Decision-making in conceptual AGV systems design using simulation-based optimisation. In the *Proceedings of the Industrial Simulation Conference, ISC'2013*, May 22-24, 2013, Ghent, Belgium. 171-176.
- Pehrsson, L., Ng, A.H.C., Bernedixen J. and Stockton, D.J. 2013. Sectioned walking-worker lines with loop balancing. In the *Proceedings of the Industrial Simulation Conference, ISC'2013*, May 22-24, 2013, Ghent, Belgium. 187-191.
- Rooda., J.E. and Vervoort, J. 2007. Analysis of Manufacturing Systems using x 1.0. Eindhoven University of Technology,

source:

[http://seweb.se.wtb.tue.nl/education/mis/AMS\\_main.pdf](http://seweb.se.wtb.tue.nl/education/mis/AMS_main.pdf), Accessed March 2014.

Simpson, T.W. and Donndelinger, J.A., (2009). Visual Steering Commands for Trade Space Exploration: User-Guided Sampling With Example. *Journal of Computing and Information Science in Engineering*. Vol. 9.

Weiss S.M. and Indurkha, N., (1998). *Predictive Data Mining: a practical guide*. Morgan Kaufmann Publishers, Inc., San Fransisco.

## BIOGRAPHIES

**LEIF PEHRSSON** holds a Ph.D. degree in Manufacturing Systems from De Montfort University in the UK. He has a M.Sc. degree in Automation Engineering and a B.Sc. degree in Mechanical Engineering both from the University of Skövde, Sweden. He has been working in various engineering and managerial positions within automotive industry for almost 20 years. He is now a Senior Advisor at Volvo Cars Engine and an Affiliated Lecturer at the University of Skövde. His e-mail address is <leif.pehrsson@his.se>.

**SIMON LIDBERG** is a Simulation Analyst at Volvo Cars Engine. He holds a M.Sc. degree in Industrial Informatics from the University of Skövde.

**MARCUS FRANTZÉN** holds a Ph.D. degree in Manufacturing Engineering from De Montfort University in the UK. He has a M.Sc. degree in Manufacturing Management from the Loughborough University and a B.Sc. degree in Automation Engineering from the University of Skövde. He has more than ten years experience from industrial simulation modelling and experimentation. Currently, he is an Assistant Professor at the University of Skövde.

**TEHSEEN ASLAM** holds a Ph.D. in industrial informatics from University of Skövde, Sweden. He has a M.Sc. degree in Manufacturing Management from Loughborough University and a B.Sc. in Automation Engineering from University of Skövde. His research interests include modelling, simulation and multi-objective optimisation for the design and analysis of supply chains.

**AMOS NG** is a Professor of Production and Automation Engineering at the University of Skövde, Sweden. He holds a Ph.D. degree in Computing Sciences and Engineering. His main research interest lies in applying multiobjective optimisation and data mining techniques for production systems design, analysis and improvement. His e-mail address is <amos.ng@his.se>.

# A CASE STUDY OF EVOLUTIONARY SIMULATION-BASED OPTIMIZATION IN AIRCRAFT ENGINE MANUFACTURING

Anna Syberfeldt  
Patrik Gustavsson  
University of Skövde  
Skövde, Sweden  
E-mail: anna.syberfeldt@his.se

Joakim Svantesson  
Torgny Almgren  
GKN Aerospace  
Flygmotorvägen 1  
Trollhättan, Sweden

## KEYWORDS

Simulation-based optimization, aircraft engine manufacturing optimization, evolutionary algorithm, case study.

## ABSTRACT

This paper describes an application of simulation-based optimization within the aircraft engine manufacturing domain. The study is performed at the company GKN Aerospace and considers the optimization of priorities of components in x-ray stations. More specifically, the optimization aims to reduce delays by finding an efficient prioritization of components waiting to be x-rayed. An evolutionary optimization algorithm is being implemented and connected to a discrete-event simulation model for evaluating solutions. For deploying the simulation-based optimization, a web-based system is developed that adopts distributed simulation agents for increased efficiency.

## INTRODUCTION

GKN Aerospace in Sweden develops and manufactures high-technology components for aircraft and gas turbine engines. Components manufactured at GKN Aerospace in Sweden can be found in both commercial and military aircraft markets, as well as in rocket engines. The focus of this study is a newly installed workshop at the company that produces turbine frame structures that demands high quality controls. The workshop includes stations for washing, x-ray, liquid penetrant testing, automatic laser and plasma welding, manual welding, CNC and manual burring. In the study, the three x-ray stations are come under specific focus as these are considered as a critical point in the production flow. Currently, the order in which components are x-rayed in the stations are based on a prioritization done by an operator. For a human to find the best prioritization among components is, however, not trivial due to the complexity of the problem. The problem is both mathematically complex (of size  $(n!)^3$ ) and complex with respect to a fluctuating inflow and variable operation times in the manufacturing process. The complexity has raised a need of replacing the current manual approach of prioritization with an automatic optimization procedure. The main aim of the optimization is to decrease delays, as the components being processed are highly capital-intensive and on-time deliveries are

important. Due to the complexity of the manufacturing process, a simulation-based approach is used for realizing the optimization rather than an analytical method. The simulation model of the workshop, created by simulation experts at the company, is described in the following section.

## SIMULATION MODEL

The simulation model is created using the discrete-event simulation software SIMUL8 (see [www.simul8.com](http://www.simul8.com)). The simulation model (Figure 1) includes all machines, operators, fixtures and buffer zones present in the real workshop. Setup times, breakdown settings, operation times and other variables in the model according to the real manufacturing process. The simulation model has two types of inputs: decision inputs and environment inputs. The decision inputs specify the priority setting for each component, and these inputs are to be generated by the optimization process. The environment inputs specify the current state of the workshop and its settings. This information, including for example components currently in the production flow, number of operators in each work area and processing times, is partly specified manually and partly imported from the company's resource enterprise software SAP.

The outputs measured from the model are the delay respective type of component along with the total delay of all components. These outputs are used in the optimization process to iteratively improve solutions, as further described in the next section.

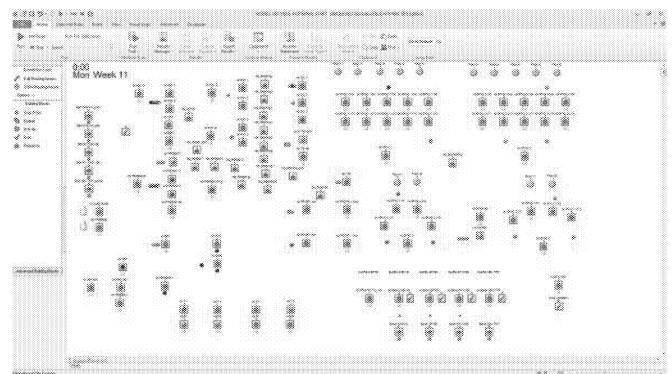


Figure 1. Screenshot from SIMUL8.

## OPTIMIZATION

The optimization of component priorities is a combinatorial problem of complexity  $(n!)^3$ . Using an exact optimization method that evaluates *all* possible solutions in order to find the best one is not possible as this would need a huge computing time. Instead, an optimization technique that is not guaranteed to find the optimal solution, but a sufficiently good one in a short time is appropriate for solving the problem. One such technique that has been proven to successfully optimize complex manufacturing problems is evolutionary algorithms (Boesel et. al, 2001; Deb 2004), and this is also the technique used in this study. Evolutionary algorithms are essentially based on Darwin's theory about "survival of the fittest" (first described in Darwin 1859). According to this theory, in a population of individuals, 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. This same idea is adopted in evolutionary algorithms for iteratively evolving (refining) solutions over time. A solution in this case represents a set of component priorities.

In evolving a population of solutions, evolutionary algorithms apply biologically inspired operations for selection, mating and mutation (Bäck et. al., 2007). The operators are applied in a loop, and an iteration of the loop is called a generation. During each generation, a proportion of the solutions is selected to breed offspring for the next generation of the population (that is, create new solutions). Solutions are selected based on their fitness, representing a quantification of their optimality, in this case the total component delay (which is to be minimized). The fitness is obtained by running the simulation model with the specific solution. From the solutions selected, new solutions are created to form the next generation of the population. The process of evaluating solutions and forming a new generation of the population then starts all over again and continues for a user-defined time period, as shown in Figure 2 and Figure 3.

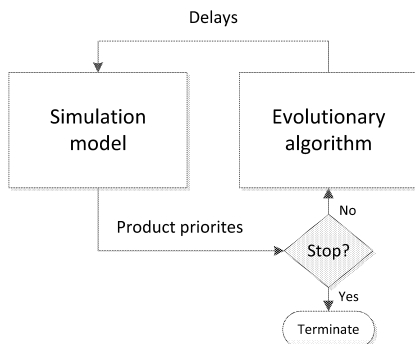


Figure 2. Iterative process.

---

Initialize population  
 Evaluate the fitness of solutions in the population  
 repeat  
   Select solutions to reproduce

Form a new population through mating & mutation  
 Evaluate solutions in the new population  
 until terminating condition

---

Figure 3: Basic steps of an evolutionary algorithm.

In the implementation done in this study, a solution is encoded as a list of unique integers values representing priorities. The first value of the list is assigned to component number one, the second value to component number two, and so on. When the algorithm is initiated, the list is populated with random numbers (priorities). For the creation of new solutions, a so called "order crossover" (Deep and Mebrahtu, 2011) is used due to the strict requirement of uniqueness of values in a solution (no integer can be repeated as this would result in the same priority of two components). In the crossover, two parents,  $p1$  and  $p2$ , are selected and used to create two offspring,  $o1$  and  $o2$  (Figure 4). The crossover operation starts by copying a range of the string from  $p1$  to  $o1$ ; then, the rest of the numbers will be copied from  $p2$  to  $o1$  in the order that they appear and only numbers that have not already been copied. The same process applies to  $o1$  but with  $o2$  as the first parent and  $p1$  as the second. It can be noted that the parents are selected through binary tournaments in which two solutions in the population are randomly picked, and the one with the best quality is selected to become a parent.

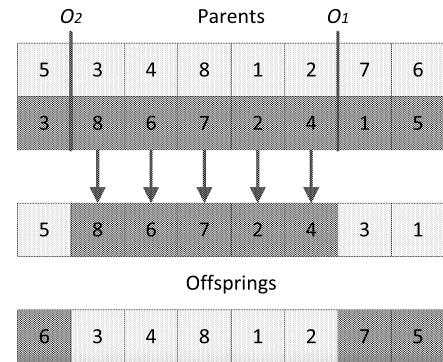


Figure 4. Example of ordered crossover.

Occasionally, the offspring solutions can undergo a small mutation in order to keep the diversity of the population large and avoid local minima. A mutation involves changing an arbitrary part of a solution with a certain probability, and in this case a partial shuffle mutation (Deep and Mebrahtu, 2011) is being used. In this mutation, integers are randomly swapped in a range as shown in Figure 5. First, two offsets are randomly generated, and then all integers in the range between the first and the second offset are swapped randomly.

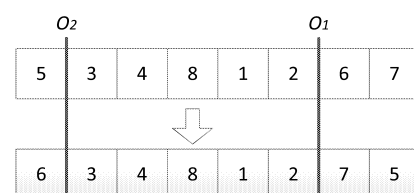


Figure 5. Example of mutation.

The next section continues with describing how the evolutionary, simulation-based optimization procedure is embedded in a software platform for enabling execution and monitoring of optimizations.

## SOFTWARE PLATFORM

The platform is implemented using a web-based approach in which a server holds the simulation-optimization system and a web client shows the user interface (Figure 5). A web-based platform allows for optimizations to be run from anywhere with an internet connection, and not only from the specific computer that has the SIMUL8 software installed. Furthermore, the maintenance of a web-based platform is easier since they do not have to be installed on each client's computer. Updates can instead be made on the server and reach the clients instantly, eliminating virtually all on-site maintenance and allowing for a frequent update scheme.

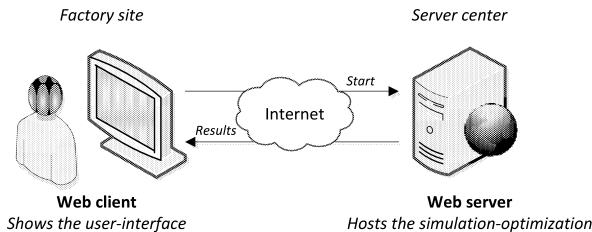


Figure 5. Conceptual picture.

## System architecture

The architectural design of the platform is presented in Figure 6. In order to achieve a manageable complexity of the software, the system is divided into four components: a) user interface, b) web server, c) database server, and d) simulation agents. An optimization is initiated by a start command sent from the user interface to the web server. The web server was implemented using the software Apache Tomcat (see [tomcat.apache.org](http://tomcat.apache.org)), mainly because it supports multiple operating systems and is free of charge. Upon receiving a start command, the web server starts a simulation-optimization process and eventually returns the best priority list found. To reduce the optimization time, a distributed approach with multiple, parallel simulation agents is used. Distributing simulation evaluations is necessary from a practical perspective due to the fact that a single simulation run takes about 1 seconds, and in total about 10 000 simulations are needed for an optimization.

The optimization results are stored in a database from which they can be retrieved over time. In an industrial environment, viewing historical results is often of interest with regard to analyzing the past and making predictions for the future. PostgreSQL was chosen for the database server on the basis of its well-known stability and the fact that the use of the software is free of charge.

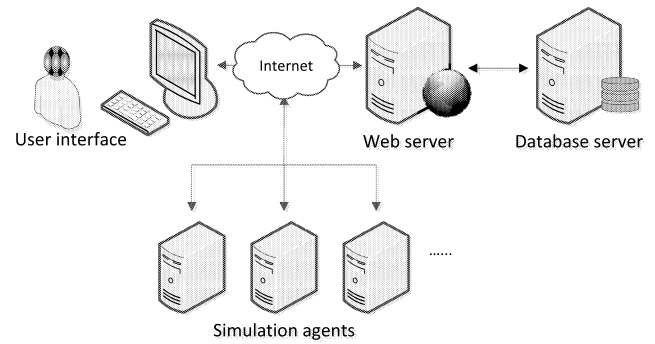


Figure 6. System architecture.

## User interface

The user interface is in the form of a web page reachable from anywhere with an internet connection. The web page was created using PHP, JavaScript and HTML5. The content of the user interface was developed using an iterative process undertaken in cooperation between the company and the software developer at the university. A screenshot of one part of the user interface is presented in Figure 7. In this screenshot, the panel to the left shows the menu from which the user can access various functions in the platform. The user starts by specifying the current status and configuration of the manufacturing process. This is done using the “INDATA” menu alternative (shown in Figure 7). The required data includes shift period, status of components currently in the workshop, list of components planned to enter to workshop, availability of fixtures, availability of operators, and scheduled maintenance. All this information is needed for the simulation model to be able to mimic the operations of the real manufacturing process in a realistic manner. When the user has specified the necessary input, the simulation-optimization is started using the “OPTIMIZATION” menu alternative.

When an optimization is finished, the results is shown in the “PRIORITY LIST” menu alternative. This page shows all components waiting in the x-ray area in the workshop, and also the priorities of components queuing for a specific x-ray. Working with the priorities is an easy, two-step process:

1. When a component arrives to the x-ray area, the operator selects the component from a list in the user interface and the component is automatically inserted to the x-ray queue in the correct position based on the optimized priority list.
2. When a component is finished in an x-ray station, the operator removes the component from both the real x-ray station and from the virtual x-ray queue in the user interface.

INDATA

OPTIMIZATION

PRIORITY LIST

LOG OUT

## Specify input data

Ange indata för optimeringen i de gula fälten samt ladda in detaljstatus från SAP.  
Tryck sedan på knappen Starta för att starta optimeringen.

Operatörskategori	Antal/skift	Skiftform
Svetsning	9	3-skift
Bearbetning	2	3-skift
Kontroll	2	3-skift
Röntgen	6	3-skift
Penetrant	3	3-skift

Starttid för optimeringsperiod: 2014-03-10 08:00

Sluttid för optimeringsperiod: 2014-03-17 17:00

Optimeringstid (min): 10

Detaljstatus (från SAP):

Valj fil DetailStatus\_AllTypes\_NoEmptyOrderNo.XLS Läs in fil

Starta

Detalj	Materialnummer	Planerad GLT	Veckotakt
XWB R	VOLS:10	2.8	5.2
XWB W	VOLS:101	2.4	5.2
GP700	VOLS:1	2.1	2
PW1	VOLS:9	1.1	11.4
GE	VOLS:90	1.8	3

Figure 7. Screenshot of user interface.

## EVALUATION

Experts at GKN Aerospace have created a typical scenario for the test case that involves 59 components, giving a  $(59!)^3$  complexity of the problem. The settings for the optimization algorithm as used in the test case are given in Table 1.

Table 1. Algorithm settings.

Population size	100
Number of evaluations	10000
Initialization	Random
Selection for mating pool	Binary tournament
Crossover	90% Order crossover
Mutation	10% Shuffle mutation

For evaluating the performance of the evolutionary algorithm, a hill climbing algorithm is also implemented and applied on the same problem. Hill climbing is selected since it is well-known, simple to implement and often used for baseline comparisons. Hill climbing is an iterative algorithm that belongs to the family of local search algorithms (Russell and Norvig, 2003). The algorithm starts with a random solution to the problem and attempts to find a better solution by mutating (changing) the solution. If the mutation produces a better solution, it is kept and the procedure is repeated until no further improvements can be made. The hill climbing is implemented with shuffle mutation and is run for 10 000 evaluation, just as the evolutionary algorithm.

The result from the optimization process is shown in Figure 8. The x-axis shows the mean value of the total component delay (for company integrity reason, the values have been transformed and the unit is not specified) and the y-axis

shows the generation of the evolutionary algorithm. The grey arrow marks the best solution found by the hill climbing algorithm after 10 000 evaluations.

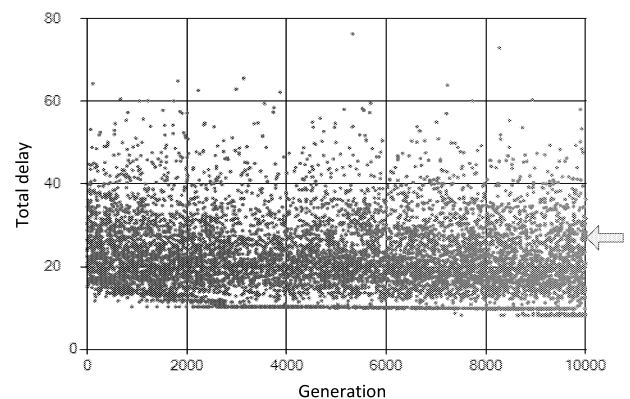


Figure 8. Results from optimizing 59 components.

Figure 8 shows that the evolutionary algorithm successfully optimizes the problem and that it outperforms the hill climbing algorithm. However, the progress of the evolutionary algorithm is relatively slow. To see whether the search performance would improve with a reduced complexity of the problem, an experiment with only one x-ray station where also performed. With only one x-ray station the complexity of the problem is reduced from  $(59!)^3$  to  $59!$ . However, the results of this experiment is similar to the first one, which indicates that the problem has a high complexity also when considering only one x-ray station.

It has not yet been possible to implement and evaluate the optimized solutions in the real workshop. However, compared to the manual approach of handling prioritization of components, the simulation-optimization system



implemented has considerable advantages. Since no manual control or intervention in the optimization process is necessary, a lot of time and effort are saved for the operator responsible for prioritizing components in the x-ray stations. This advantage, in combination with the potential of significantly increasing the efficiency of the workshop, can hopefully motivate a deployment of the simulation-optimization system in the future.

## SUMMARY

This paper presented a real-world case study of simulation-based optimization within the aircraft engine manufacturing domain. The study was performed at the company GKN Aerospace and considers the optimization of priorities of components in three x-ray stations in one of the company's workshops. The aim of the optimization was to reduce delays by finding an efficient prioritization of components waiting to be x-rayed. For performing the optimization, an evolutionary algorithm was implemented and connected to a discrete-event simulation model. The simulation model was created using SIMUL8 and designed to closely mimic the real manufacturing process.

For deploying the simulation-based optimization, a web-based system was developed. A web-based system was considered advantageous as it allows for easy access from anywhere with an internet connection and without the need of installing SIMUL8. The web-based approach also simplifies

maintenance since updates can be made through a server and on-site maintenance is not needed. The graphical user interface of the system is in the form of a web page where the user can execute optimizations and watch results. For increased efficiency, the simulation-optimization system adopts distributed simulation agents. Using a distributed approach means that several simulations can be performed at the same time and the total execution time of the optimization can thereby be significantly reduced. From a practical perspective, the distributed approach is necessary considering that a single simulation run takes about 1 second and about 10 000 simulations are needed for an optimization.

Results from running the simulation-based optimization system showed that the evolutionary algorithm was able to successfully optimize the problem. Compared to a hill climbing algorithm, the evolutionary algorithm found significantly better solutions. However, the search progress of the evolutionary algorithm turned out to be relatively slow, which is probably due to the high complexity of the problem. An important aspect to consider in the future for improving the search progress is to elaborate on the crossover operator used in the algorithm. The crossover is a vital part of an evolutionary algorithm (Goldberg, 1989) and it might be that the implemented operator was not the most efficient one. This aspect needs further research and several crossover operators need to be compared. It would probably also be of advantage if some meta-knowledge of

input/output correlations could be found and used in the crossover process.

## REFERENCES

- Boesel, J., R. Bowden, F. Glover, J. Kelly and E. Westwig. 2001. "Future of simulation optimization". *Proceedings of the 2001 Winter Simulation Conference*, Arlington, VA, pp. 1466–1470.
- Bäck, T., Fogel, D. and Michalewicz, Z. (eds) 1997. "Handbook of Evolutionary Computation". Oxford University Press.
- Darwin, C. 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. 2004. "Multi-objective optimization using evolutionary algorithms". Chichester: John Wiley & sons, Ltd.
- Deep, K. and H. H. Mebrahtu. 2010. "New Variations of Order Crossover for Travelling Salesman Problem". *International Journal of Combinatorial Optimization Problems and Informatics*, Vol 2, No 1.
- K. Deep and H. 2011. "Mebrahtu Combined Mutation Operators of Genetic Algorithm for the Travelling Salesman problem". *International Journal of Combinatorial Optimization Problems and Informatics*, Vol. 2, No.3, 1-23.
- D. Goldberg. 1989. "Genetic algorithms in search, optimization and machine learning". Addison-Wesley Publishing Co, Boston, Massachusetts.
- S.J. Russell, P. Norvig. 2003. "Artificial Intelligence: A Modern Approach". 2<sup>nd</sup> ed, Upper Saddle River, New Jersey: Prentice Hall, 111–114.

## BIOGRAPHIES

**ANNA SYBERFELDT** is a senior researcher at the University of Skövde, Sweden. She holds a PhD in Computer Science from the De Montfort University, UK and a Master's degree in Computer Science from the University of Skövde, Sweden. Her research interests include simulation-based optimization, artificial intelligence, metaheuristics, and advanced information technology with applications in logistics and manufacturing. Her email address is [anna.syberfeldt@his.se](mailto:anna.syberfeldt@his.se).

**PATRIK GUSTAVSSON** is a research engineer at the University of Skövde, Sweden. He holds a MSc in Automation Engineering from the University of Skövde, Sweden. His research interests include simulation, evolutionary optimization, and software engineering with applications in manufacturing. His email address is [patrik.gustavsson@his.se](mailto:patrik.gustavsson@his.se).

**JOAKIM SVANTESSON** is working at GKN Aerospace with logistics development. He has extensive experience in analysis of production systems and supply chains using traditional techniques as well as discrete-event simulation. He has also developed tools for improved processes for resource and flow control, amongst others.

**TORGNY ALMGREN** is working at GKN Aerospace as a specialist within logistics. He is also an associated professor within mathematical science at Chalmers University, Sweden. He is especially interested in developing the possibilities of using operations research in the logistics field.

#### **ACKNOWLEDGEMENTS**

This work has been carried out within the SOL project which is partially financed by the Knowledge Foundation (KK Stiftelsen), Sweden.

# PRACTICAL PRODUCTION SYSTEMS OPTIMIZATION USING MULTIPLE-CHOICE SETS AND MANHATTAN DISTANCE BASED CONSTRAINTS HANDLING

Jacob Bernedixen

Amos H.C. Ng

Production and Automation Engineering, Department of Engineering Science

University of Skövde

SE-541 28, Skövde,

Sweden

E-mails: jacob.bernedixen@his.se, amos.ng@his.se

## KEYWORDS

Production systems, optimization, simulation-based optimization, multiple choice sets, constraint handling.

## ABSTRACT

Many simulation-based optimization packages provide powerful algorithms to solve large-scale system problems. But most of them fall short to offer their users the techniques to effectively handle decision variables that are of multiple-choice type, as well as equality constraints, which can be found in many real-world industrial system design and improvement problems. Hence, this paper introduces how multiple choice sets and Manhattan-distance-based constraint handling can be effectively embedded into a meta-heuristic algorithm for simulation-based optimization. How these two techniques have been applied together to make the improvement of a complex production system, provided by an automotive manufacturer, possible will also be presented.

## INTRODUCTION

Optimization can be defined as the process of seeking the “best” design/operation of a product /system, with respect to certain objective(s). It is the ultimate goal of many on-going scientific research and industrial production system development projects. For instance, simultaneously maximizing the effectiveness and minimizing variability (Six Sigma), waste (Lean) or more directly the running cost of the available resources on hand is the target for almost all companies to remain competitive. Implicitly, optimization is the keyword for most of the continuous system improvement methodologies, found in the field of Management Science and

Operation Research (OR). Nevertheless, optimization models found in OR are not easy to be applied in practice, especially in industry. Firstly, many of these models require high analytical skills to formulate the problems into some mathematical models. Secondly, very often they require unrealistic assumptions, e.g. deterministic process times, hence fall short to represent real-world system complexity. It is therefore not difficult to understand that simulation, particularly discrete-event simulation, is still the most practical choice for industrial practitioners (Tempelmeier 2003). This is not a surprise when considering that any system complexity and stochastic variability can be easily modeled using almost any simulation software. On the other hand, simulation is not an optimization tool when used standalone. The connection of meta-heuristic search methods, such as Genetic Algorithms (GA) or Tabu Search (TS), to simulation models, referred to as Simulation-based Optimization (SBO), can be used for finding the optimal setting of many decision variables in a design optimization problem (Law and McComas, 2002). Extensive research focus has been paid to this area since mid-1990s, when the first commercial SBO package was launched. The basic mechanism of these packages is the integration of meta-heuristic search algorithms, with Discrete-Event Simulation (DES) software so that “optimal” or close to optimal solutions can be found automatically, as illustrated in Figure 1.

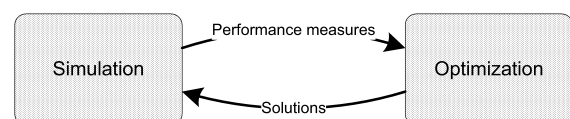


Figure 1: Simulation-based Optimization

SBO has opened a new era for applying optimization to solve real-world problems in industry because: (1) meta-heuristic search algorithms in general treat a target simulation model as a black box, therefore the same generic algorithm can be used to solve a range of problems without any re-configuration; (2) meta-heuristic search algorithms, like GA and TS, are able to handle a large set of decision variables with mixed data type (e.g. continuous, discrete and combinational). Many industrial application studies for addressing production systems design and analysis problems have been reported. Nevertheless, despite the huge potential that can be offered by SBO, there are some deficiencies that still prohibit the effective use of general SBO packages for production systems optimization in practice. Specifically, we are referring to two issues that have to be resolved if SBO is to be used effectively for the continuous improvement of production systems: (1) decision variable ranges that cannot be well modeled by using standard integer or continuous optimization variables with a lower and an upper bound; (2) constraint handling to deal with practical constrained optimization problems. The aim of this paper is to introduce how these two specific issues can be resolved by the use of multiple choice sets as input variables and Manhattan distance based constraint handling respectively. These two techniques have been implemented into FACTS Analyzer (Ng et al. 2007; 2011), and then applied together to make an improvement project of a complex production system, provided by an automotive manufacturer, possible. This application study and its results are included with some details in this paper.

## MULTIPLE-CHOICE SETS AS INPUTS

Parameters commonly used as inputs in optimization problems of production systems include, processing times, availability, mean time to repair, buffer capacities, etc. In practice these parameters are not well modeled using standard integer or continuous optimization variables with a lower and an upper bound, see 1 and 2 in Figure 2. For instance, reducing the processing time of a machine is not done in a linear fashion nor is it necessarily restricted to only integer values, e.g. it can be improved from 32.1 seconds to 24.3 seconds. The same is true for availability and mean time to repair. Even though buffer capacities usually are represented by integers, they need not

be increased/decreased by one when changed, e.g. adding a level to a rack might result in several more storage places in the rack based on the amount of parts that fits on one level. This calls for a more versatile type of optimization input similar to what Tomlin (1988) referred to as a *multiple-choice set* (MCS).

A MCS is a set consisting of a finite number of unique integer and/or continuous values (obviously integrality must be enforced if the underlying variable is an integer). A special notation is introduced to simplify the definition of large sets. First, the standard definition of a bounded optimization variable with a *lower*- and *upper* bound is extended to include an additional parameter *base*. It represents the distance between consecutive values between the *lower*- and *upper* bound, possibly except for the last two values of the set where the distance might be shorter. The notation used for this is  $\{lower\_bound;upper\_bound|base\}$  and it is called a standard set (SS), e.g.  $\{3;4.2|0.5\}$  would represent the values 3, 3.5, 4, and 4.2. It is worth to note that the upper bound is included even though it is in violation with the specified base. The reason for this is that the upper bound probably is of some significance if specified in such a manner. This kind of SS would represent a valid MCS, as illustrated by case 3 in Figure 2. However, it can also be extended to form more complex MCSs by allowing scalar values and/or multiple SSs to be combined. The notation for this is  $\{*,*,...\}$  where  $*$  = a scalar value or a SS, e.g.  $\{3;4.2|0.5\}, 6.5, 8, \{99;105|3\}$  would represent the values 3, 3.5, 4, 4.2, 6.5, 8, 99, 102, and 105 or illustration 4 in Figure 2 is another example.

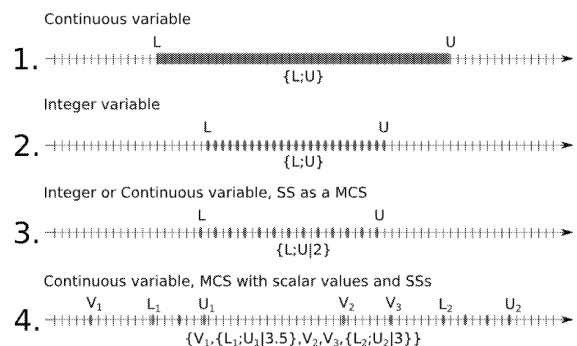


Figure 2: Illustration of standard integer and continuous variables (1-2) and MCS variables (3-4)

These MCSs have been implemented in a generic fashion into FACTS Analyzer that supports

multiple evolutionary optimization algorithms, by treating the MCSs as special cases in the crossover and mutation operators. The MCSs are represented as a list of values in which each value can be accessed by its zero-based index in the list. Using this representation, specific crossover and mutation operators need not be implemented for the MCSs, even though the authors acknowledge that such specific operators are likely to be more efficient. Instead some real-parameter crossover and mutation operators (for instance some described by Deb 2001) can be used on the indexes before the actual value of the MCS is returned using the computed index. The procedure is illustrated in Figure 3 using the random mutation operator as an example. The procedure is analogous for other real-parameter mutation and crossover operators.

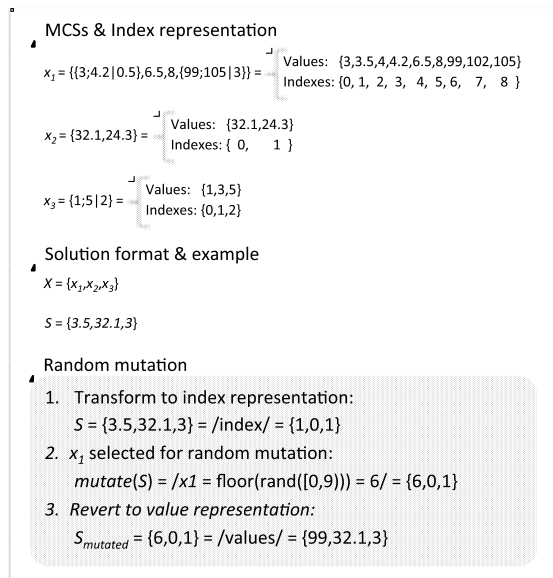


Figure 3: Example of mutation of MCSs using index representation

This representation of the MCSs and the fact that only the actual value (not the corresponding index) of a MCS is stored during the optimization imposes some restrictions on the MCSs. There need to be a one-to-one mapping from actual value to its corresponding index. This means that duplicate values are not allowed in a MCS and SSs within a MCS are not allowed to overlap.

## CONSTRAINTS HANDLING USING MANHATTAN DISTANCE

It is not uncommon that optimization problems of production systems are constrained in some way or another. For instance, space restrictions might limit

the total number of storage places (total buffer capacity) or for line-balancing the total processing time might be constant even though it may be allowed to vary between individual stations in the production system. The former represents an inequality constraint (less than) whereas the latter is an equality constraint. Any combination of these types of constraints is allowed as long as they are linear (even though integrality constraints are allowed on optimization variables). Linearity is important since it will enable an infeasible solution  $S$  to be replaced by a feasible solution  $X^*$  that is the closest feasible one to  $S$  by solving a mixed-integer programming (MIP) problem (or possibly a linear programming problem if all optimization variables are continuous). Avoiding evaluation of infeasible solutions is, as state by Laguna and Rafael (2002), important for these types of complex systems where evaluations are performed by running time-consuming simulation models.

Infeasible solutions with regards to the set of constraints might result from the random generation of the first population or the creation of a new population (the crossover and mutation operations). Regardless of the cause of the infeasibility the same MIP problem formulation can be used to find a feasible solution. The MIP problem formulation used is an extended version of the one suggested by Laguna and Rafael (2002) that can also handle the MCS optimization variables. The extension draws upon what Healy (1964) referred to as Multiple Choice Programming. In order to formulate the MIP problem some notations and definitions are presented in Table 1.

Table 1: Notations and definitions for the MIP problem

Notation	Definition
$X$	The original optimization variables $\{x_1, \dots, x_n\}$ where all MCS variables are represented as standard variables bounded between each MCS's smallest and largest value.
$L$	Lower bounds of all variables in $X$ .
$U$	Upper bounds of all variables in $X$ .
$J$	Set of all $\{j \subseteq \{1, \dots, n\} : x_j \in X \wedge x_j \text{ represents a MCS}\}$ .
$AX \leq B$	Notation used for original set of constraints, even though equality constraints are allowed.

$S$	Infeasible solution $\{s_1, \dots, s_n\}$ .
$X^*$	Feasible solution $\{x_1^*, \dots, x_n^*\}$ that minimizes the Manhattan distance between $S$ and $X$ .
$P$	Positive deviation of $X$ from $S$ $\{p_1, \dots, p_n\}$ where $p_i \geq 0, \forall i \in \{1, \dots, n\}$ .
$M$	Negative deviation of $X$ from $S$ $\{m_1, \dots, m_n\}$ where $m_i \geq 0, \forall i \in \{1, \dots, n\}$ .
$X_{MCS}$	Subset of $X$ , $\{x_j : j \in J\}$ .
$V_j$	Vector containing all allowed values of MCS variable corresponding to $x_j$ where $j \in J$ , $\{v_{j1}, \dots, v_{jk(j)}\}$ and $k(j)$ is the number of values in the MCS.
$H_j$	Set of binary variables belonging to $x_j$ where $j \in J$ , $\{h_{j1}, \dots, h_{jk(j)}\}$ and $k(j)$ is the number of values in the MCS.

The mathematical formulation of the MIP problem looks like:

$$\begin{aligned}
&\text{minimize} && \sum_{i=1}^n p_i + m_i && (1) \\
&\text{s.t.} && AX \leq B && (2) \\
&&& s_i - x_i - m_i + p_i = 0, \forall i \in \{1, \dots, n\} && (3) \\
&&& -x_i + \sum_{l=1}^{k(j)} v_{jl} * h_{jl} = 0, \forall j \in J && (4) \\
&&& \sum_{l=1}^{k(j)} h_{jl} = 1, \forall j \in J && (5) \\
&&& L \leq X \leq U && (6) \\
&&& P, M \geq 0 && (7) \\
&&& H_j \in \{0, 1\}, \forall j \in J && (8)
\end{aligned}$$

where (1) represents the minimum Manhattan distance, (2) are the original constraints of the problem, (3) set of constraints that minimizes the Manhattan distance, (4) set of constraints that together with what Healy (1964) called integer sums (5) forces the MCS variables in  $X$  to values contained in the MCSs, and (6)-(8) bounds on the variables.

This MIP problem finds the closest (based on the Manhattan distance metric) feasible solution  $X^*$  to an infeasible solution  $S$ , illustrated in Figure 4. The use of the Manhattan distance metric might seem convenient since it is a linear metric contrary to the widely used quadratic Euclidian distance metric. However, as pointed out by Aggarwal and Hinneburg (2001) it is actually preferable for high-dimensional space. The probability that the Manhattan distance metric provides a significantly higher relative contrast is very high for dimensionalities of 20 or larger, but not insignificant even for lower dimensionalities. In this case it is the number of optimization variables that dictates the dimensionality and for complex production systems it is not uncommon that it

exceeds 20, see the section A Complex Industrial Optimization Problem.

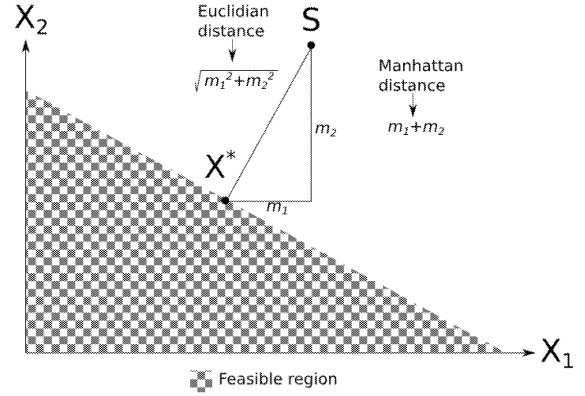


Figure 4: 2D-illustration of Manhattan distance and Euclidian distance

## A COMPLEX INDUSTRIAL OPTIMIZATION PROBLEM

This section describes how FACTS Analyzer with MCSs and the described constraint handling can be used to optimize the performance of a complex real-world production line. The production line is the same one first presented in Siegmund et al. (2012), but in the current study the simulation model is built with FACTS Analyzer (Ng et al. 2007; 2011), or simply FACTS hereafter. There is a commercial version of FACTS distributed by Evoma AB ([www.evoma.se](http://www.evoma.se)) but the current work uses the academic version of FACTS developed at the University of Skövde. There are two main features that distinguish FACTS from other commercial production system simulation software: (1) rapid modelling, no programming, easy to learn and use; (2) tightly integrated with powerful multi-objective optimization algorithms, making optimization of production systems straightforward, when compared to other products. This latter feature renders any FACTS models to flexibly cope with a list of generic optimization parameters, instead of like the case in many commercial discrete-event simulation software where tailored optimization parameters have to be added (usually in form of programming code) to the models. The snapshot of the FACTS model in Figure 5 illustrates the complexity of the production system. The production line includes several parallel sections, assembly stations, machines with several operations, portal cranes, variants, etc. The size and complexity of the line makes it extremely hard to locate what to improve, let alone the effect of such an improvement. With the objectives to

simultaneously maximize the system throughput as well as to minimize the number of improvements, to apply simulation-based multi-objective optimization is in order.

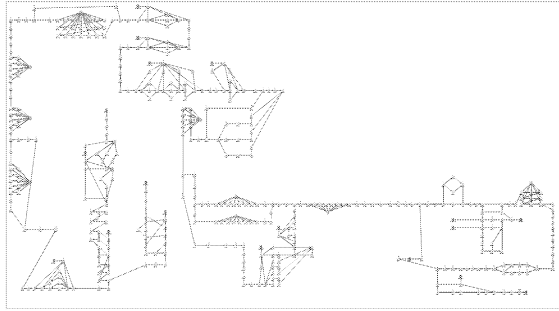


Figure 5: FACTS Analyzer model of a large complex real-world production line

Industrial engineers in charge of the production line provided improvement options like reduced processing times (per variant where applicable), increased availabilities, and reduced mean times to repair. The level of improvement (from the original value) varied between different stations. Table 2 lists the number of improvement variables of each type and also the range of the improvements of that type, e.g. the processing time reduction ranges from only 0.2 % for one station to 41.8 % for another station.

Table 2: Improvement details

Type of improvement variable	Number of variables of this type	Range (min-max)
Processing time	317	-0.2% – -41.8%
Availability	82	+0.1% – +23.8%
Mean time to repair	65	-5.5% – -92.1%

That sums up to 464 improvement alternatives represented in the optimization problem as binary MCS variables in the format {*original value*, *improved value*}. That is equivalent to  $2^{464} \approx 4.8E+139$  possible improvement combinations, which obviously makes it infeasible to try all of them. Again, an indication that it's a problem well suited for an SBO approach.

On several locations this production line has stations with machines in which several operations are performed in sequence. Using FACTS this is modeled with a serial line of operations. However,

improvements that are made to such a machine affect all operations simultaneously, i.e. one such operation cannot be improved without all other operations of that machine also being improved. In the optimization problem, this is handled by adding equality constraints that ensure that either all operations are improved or all remain not improved. To illustrate this, consider a sequence of three operations OP1, OP2, and OP3; with original processing times  $PT_{OP1\_orig}=PT_{OP2\_orig}=PT_{OP3\_orig}$  and improved processing times  $PT_{OP1\_imp}=PT_{OP2\_imp}=PT_{OP3\_imp}$ . This would be modeled by MCSs  $X_1=\{PT_{OP1\_orig},PT_{OP1\_imp}\}$ ,  $X_2=\{PT_{OP2\_orig},PT_{OP2\_imp}\}$ , and  $X_3=\{PT_{OP3\_orig},PT_{OP3\_imp}\}$  in the optimization problem. Then two (the number of operations minus one) equality constraints are needed to ensure that they are either all being improved at the same time or all remain not improved at all. For instance  $X_1=X_2$  and  $X_1=X_3$ . With this arrangement, this model actually contains 153 such equality constraints.

The result from this SBO problem is shown in Figure 6. The objective was to maximize the throughput while keeping the number of improvements to a minimum. Here the throughput is expressed as relative change in percent from the initial state with no improvements. It can be seen that these improvements can improve the throughput significantly (over 80 %) and a substantial improvement (about 50 %) can be achieved with only 7 improvements.

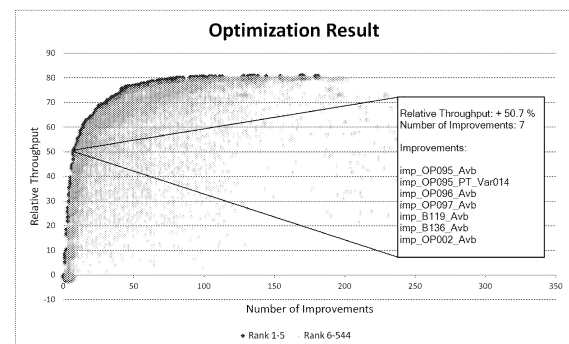


Figure 6: Results from the improvement optimization

In order to determine the importance of the different improvements among the best solutions (black ones in Figure 6) a simple frequency analysis was done on them. In this case the best solutions have been selected as the rank 1 through 5 solutions, using the non-domination rank described by Deb et al. (2002). The frequency analysis

calculates the frequency of each improvement among these best solutions. The 20 most frequent once are shown in Figure 7. Here operation OP095 stands out as very important. Five out of these top 20 are improvements related to OP095, one for the availability (Avb) and the other 4 for the processing times (PT) of different variants.

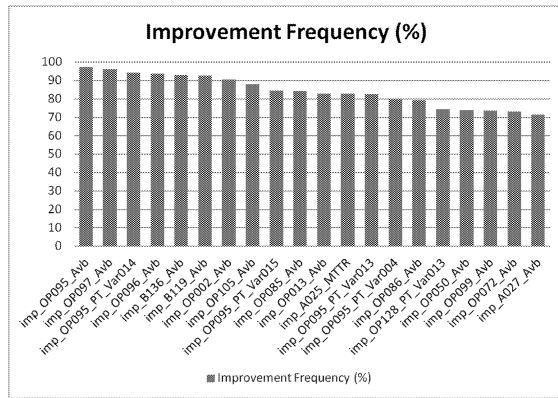


Figure 7: The 20 most frequent improvement alternatives among the best solutions (rank 1 to 5)

## CONCLUSIONS

In this paper, we have argued the importance of offering the techniques for handling multiple-choice type of input variables as well as equality constraints to solve real-world industrial systems design and improvement problems. Specifically, how multiple-choice sets and Manhattan distance-based constraint handling can be effectively embedded into a meta-heuristic algorithm for simulation-based optimization has been addressed. A complex production system with 464 input variables and 153 equality constraints have been solved by the implemented techniques in FACTS Analyzer. The study can conclude that only 7 improvements, out of the 464 improvement alternatives, are enough to significantly increase the throughput of the entire system by 50 %. The techniques for handling multiple-choice type of input variables and equality constraints have therefore been proved to be highly applicable in solving real-world industrial problems. Future research will expand this work to more complex real-world problems that involve more than one level of improvements, i.e., by utilizing MCS for improvement alternatives that consist of multiple levels.

## REFERENCES

- Aggarwal C.C. and A. Hinneburg. 2001. "On the Surprising Behavior of Distance Metric in High-Dimensional Space". In *Proceedings of the 8<sup>th</sup> International Conference of Database Theory* (London, UK, Jan.4-6). Springer, 420-434.
- Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, 2001.
- Deb, K., Agrawal, S., Pratap, A. and Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, Vol. 6, No.2, 182–197.
- Healy, W.C. 1964. "Multiple Choice Programming (A Procedure for Linear Programming with Zero-One Variables)." In *Operations Research* 12, No.1 (Feb), 122-138.
- Laguna, M. and M. Rafael. 2002. "The OptQuest Callable Library". In *Optimization Software Class Libraries*, S. Voß and D.L. Woodruff (Eds.). Springer, US, 193-218.
- Law, A.M. and McComas, M.G. (2002). Simulation Based Optimization. *Proceedings of the 2002 Winter Simulation Conference*, WSC2002, San Diego, California, USA, December 8-11, 2002, 41-44.
- Ng, A.H.C., Urenda, M., Svensson J., Skoogh, A. and Johansson, B. (2007). FACTS Analyser: An innovative tool for factory conceptual design using simulation. In *Proceedings of Swedish Production Symposium*, Gothenburg 28-30 August 2007.
- Ng, A.H.C., Bernedixen, J, Urenda, M. and Jägstam, M. (2011). Factory Flow Design and Analysis Using Internet-enabled Simulation-based Optimization and Automatic Model Generation. In *Proceedings of the 2011 Winter Simulation Conference*, S. Jain, R.R. Creasey, J. Himmelsbach, K.P. White, and M. Fu, eds. 2181-2193.
- Siegmund F., J. Bernedixen, L. Pehrsson, A.H.C. Ng, and K. Deb. 2012. "Reference Point-Based Evolutionary Multi-Objective Optimization for Industrial Systems Simulation". In *Proceedings of the 2012 Winter Simulation Conference* (Berlin, Germany, Dec.9-12). IEEE, 1-11.
- Tempelmeier, H. (2003). Practical considerations in the optimization of flow production systems. *International Journal of Production Research*, 41, No. 1, 149-170.
- Tomlin, J.A. 1988. "Special ordered sets and an application to gas supply operations planning". *Mathematical programming* 42, No.1-3, 69-84.



## **AUTHOR BIOGRAPHIES**

**JACOB BERNEDIXEN** is a system developer and PhD student at University of Skövde. He holds a M.Sc. degree in Industrial Engineering and Management from the University of Linköping, Sweden. His research interests include production system improvement using multi-objective optimization and simulation modeling and development.

**AMOS H.C. NG** is a Professor of Production and Automation Engineering at the University of Skövde, Sweden. He holds a Ph.D. degree in Computing Sciences and Engineering. His main research interest lies in applying multi-objective optimization and data mining techniques for production systems design, analysis and improvement.

## **ACKNOWLEDGEMENTS**

This work is partially funded by the Knowledge Foundation (KK Stiftelsen), Sweden. The authors gratefully acknowledge them for the provision of research funding.

# STUDY OF THE IMPACT OF PRODUCTION QUANTITY LOADED TO A PRODUCTION LINE ON APPAREL PRODUCTION SYSTEM PERFORMANCE USING SIMULATION

Lottie L.C. Mak, W.K. Wong, S.Y.S. Leung  
Institute of Textile and Clothing  
The Hong Kong Polytechnic University  
Hung Hom, Kowloon,  
Hong Kong  
E-mail: [calvin.wong@polyu.edu.hk](mailto:calvin.wong@polyu.edu.hk)

## KEYWORDS

Computer simulation, production loading, operator learning effect

## ABSTRACT

In apparel manufacturing process, one of the functions in production planning process is to define the production quantity to be loaded to the production line. It involves splitting customer order into lots of different sizes before it is loaded. The splitting decision is mainly based on the primary concerns of avoiding late delivery and making the most out of the production capacity; it has received minimal attention from both industry and academic research. In this paper, computer simulation method is used to explore the impact of different lot sizes to be loaded in production lines and its interaction with operator competence on production system performance. The simulation results confirm the significant effect of production quantity to be loaded to the production line and its interaction effect with operator competence on all performance measures.

## INTRODUCTION

In the current manufacturing world which stresses short lead time and Quick Response supply, planners split the customer order into lots of different sizes before loading them to the production lines to enhance responsiveness and flexibility. The decision regarding the production quantity to be loaded is mainly based on the primary concerns of avoiding late delivery and making the most out of production capacity; minimal concern has been placed on the impact of the defined quantity on the production system performance. To improve the production loading process, computer simulation is applied to explore the impact of the production quantity to be loaded to the production line and its interaction with operator competence on the production system performance in terms of aspects include Work-In-

Progress (WIP) level, flow time, machine and operator utilization rates.

## BACKGROUND

In common practices when planners split the customer order into lots of different sizes, the impact of the lot size on production system performance is beyond their consideration. In fact, especially in labor-intensive industry, such as apparel and footwear manufacturing, there is a close relationship between the production quantity loaded to the production line and the production line efficiency. Many studies have proved the presence of learning effect associated with an increase in cumulative unit of practices which contributes to shortening of production time (Janiak and Rudek 2009; Lee and Chin 2009; Mosheiov 2001). Leung et al. (2006) attempted to integrate customer order size as a factor in the simulation model of a Lean Production line and it showed that order size indeed affected machine utilization rate. However, little research has extended the investigation to explore the impact of the production quantity loaded to the production line on the production system performance. A research to further explore the mentioned issue would be helpful in providing insights into the production loading process.

## FRAMEWORK OF THE SIMULATION MODEL

The framework of the simulation model consists of three phases (Figure 1). In phase 1, based on real production data including operation sequences, the Standard Allowed Minutes (SAM) of each operation and operator competence records, the allocation of machines and operators are computed. A simulation experiment is thus configured to imitate a specific production system. The experiment is verified with the use of animation and debugger tool and validated with the use of hypothesis test based on cycle time. In phase 2, two selected factors of interest which are operator competence level and production quantity loaded to the production line are

incorporated to the simulation experiment. In phase 3, simulation experimental output is generated and analyzed to study the mentioned research issue.

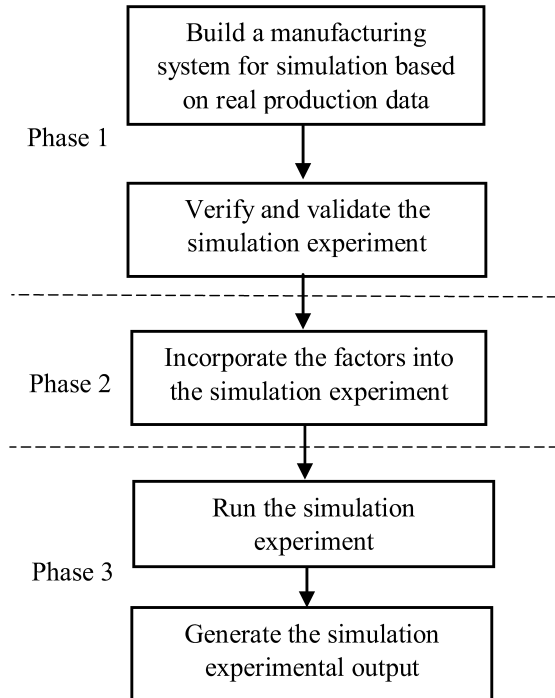


Figure 1: Simulation Model Framework

A simulation model which involves human components with dynamic behavior is very complex (Sirovetnukul 2010). To control the level of model details, project objectives have to be formulated carefully in the initial stage. As this paper aims to explore the impact of production quantity and operator competence on production system performance, the simulation model is constructed under the following assumptions which are also commonly adopted by simulation practitioners:

- The queuing discipline is First-In-First-Out
- Raw materials are always available
- Operators are multi-skilled
- Machine downtime has no impact on the production system performance
- The production system works eight hours a day

## APPLICATION OF THE FRAMEWORK

To illustrate the application of the proposed simulation mode, a U-shaped one-piece flow production system used in the apparel industry is selected in this paper to demonstrate the proposed framework. The operators are interchangeable among operations and are multi-skilled so they can pick up each other's work along the production line. They are grouped in a module to work on machines

which are arranged in a U-shaped line. An operator move clockwise with the raw material to the next machine and perform the operation until he reaches another operator. The raw material is then passed to the operator he meets and the operator who is out of work moves anti-clockwise and interrupts the first operator he meets for more additional work (Tang & Hammond 1994; Wang et al. 1991).

### Phase 1: Simulation Modelling of a Production System

Based on real-life production data collected from factory, a simulation experiment is configured to simulate a U-shaped one piece flow production system. The production data concerning operation sequence, work measurement and operation complexity is showed in Table 1. The production consists of 12 operations and involves 14 operators. Since the U-shaped one piece flow production system generally contains fewer operators, 14 operators are split into two groups and arranged to work in two production lines with the same machine layout. The layout of the production system is presented in Figure 2.

Table 1: Real Production Data

Operation	Work measurement (minutes)	Complexity
1	21.9	Medium
2	2.0	Medium
3	28.2	Complex
4	17.7	Easy
5	25.1	Medium
6	15.1	Easy
7	25.2	Easy
8	24.4	Medium
9	1.4	Medium
10	1.7	Medium
11	22.8	Medium
12	8.4	Easy

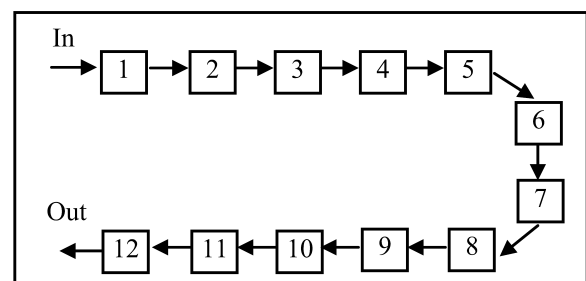


Figure 2: Layout of the Production System

Next, verification is carried out with the debugger tool to ensure the programming errors are eliminated and

validation is accomplished by using hypothesis test based on mean cycle time with a 95% confidence interval (Guner & Unal 2008). The hypothesis are:

$$\begin{aligned} H_0: \mu_F - \mu_E &= 0 \\ H_1: \mu_F - \mu_E &\neq 0 \end{aligned}$$

where  $\mu_F$  is the mean cycle time from field and  $\mu_E$  is the mean cycle time from the simulation experiment. The hypothesis test is computed as follow in Equation (1):

$$t = \frac{(\bar{x}_f - \bar{x}_e)}{\left( \frac{(n_f - 1)S_f^2 + (n_e - 1)S_e^2}{n_f + n_e - 2} \right) \sqrt{\frac{1}{n_f} + \frac{1}{n_e}}} \quad (1)$$

where

$\bar{x}_f$  is the mean cycle time from the field

$\bar{x}_e$  is the mean cycle time from the simulation experiment

$n_f$  is the number of samples collected from field

$n_e$  is the number of samples collected from simulation experiment

$S_f^2$  is the cycle time variance from field

$S_e^2$  is the t cycle time variance from simulation experiment

From the equation, it proves that there is no significant difference between the mean cycle time from the field and simulation experiment ( $t = 0.046$ ); therefore, the simulation experiment constructed in phase 1 is validated.

## Phase 2: Incorporating Factors into the Simulation Experiment

The factors of operator competence level and production quantity loaded to the production line are incorporated into the constructed simulation experiment. To simulate the operator competence level, with reference to the classifications of operators and task difficulty proposed by Tyler & Fozzard (1987), operations are classified as Easy (1), Medium (2) and complex (3) and operators are classified into Experienced (A), Moderate (B) and Inexperienced (C); nine learning curves are computed to reflect the performance variations of three

groups of operators in handling three difficulty levels of operations (Figure 3). The 14 operators in the production system are classified into different levels (A, B or C) and the operations in the production line are classified into different levels of difficulty (1, 2 or 3). The factor is manipulated to form three factor levels; each level consists of 14 operators with unique combination of operator classifications (Table 2).

Table 2: Factor Levels of Operator Competence

Factor levels	Operator classification			Total
	A	B	C	
1 (Most experienced line)	8	2	4	14
2 (Moderate line)	4	8	2	14
3 (Least experienced line)	2	4	8	14

The second factor is the production quantity loaded to the production line. This factor refers to the quantity planner loads to the production line when they split the customer order into lots of different sizes. 16 factor levels are derived from the current daily output of the production line to reflect different industrial situations as production may take a day or even months to complete (Table 3).

Table 3: Factor Levels of Quantity Loaded to the Production Line

Factor levels	Quantity loaded	Factor levels	Quantity loaded
1	175	9	1575
2	350	10	1750
3	525	11	1925
4	700	12	2100
5	875	13	2275
6	1050	14	2450
7	1225	15	2625
8	1400	16	2800

Based on the factor settings, 48 factorial-designed experiments are formulated.

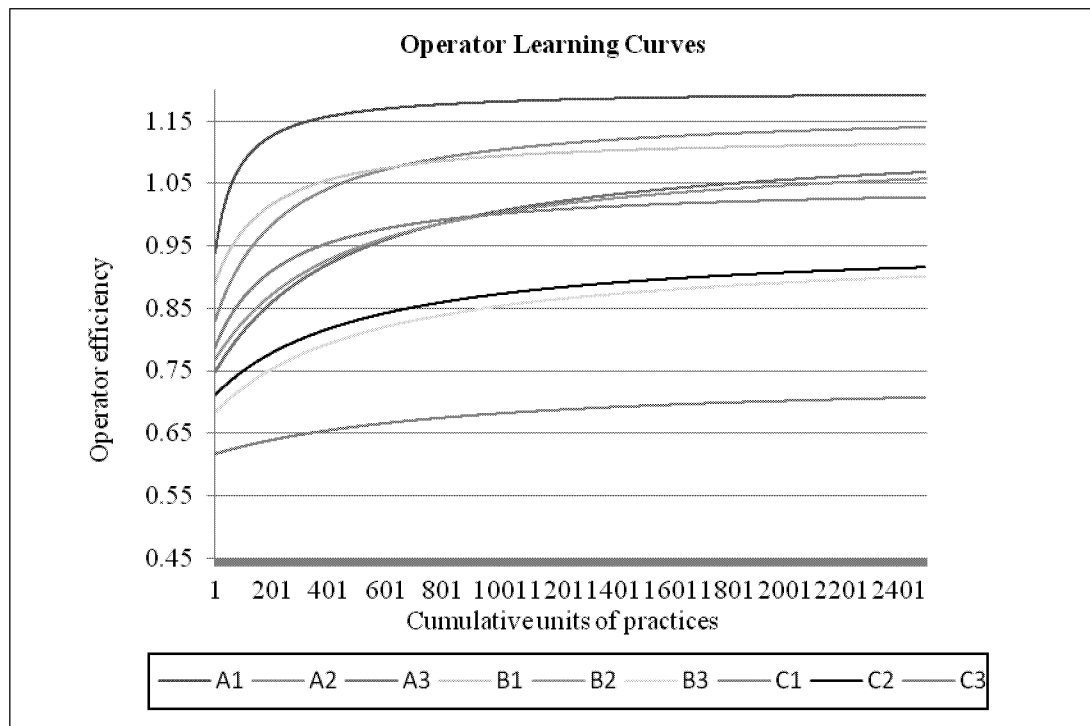


Figure 3: Operator Learning Curves

### Phase 3: Result Analysis and Discussion

The simulation experimental results are analyzed with the use of two-way Analysis of Variance (ANOVA) at 95% confidence interval to examine the main effects and interaction effect of the factors on the manufacturing system performance in terms of machine utilization rate, operator utilization rate, flow time and WIP level.

In Table 4, P value indicates whether the factor significantly affects the dependent variable; a P value which is less than 0.05 confirms the importance of the factors in affecting the dependent variable. Therefore, all main effects of the two factors are confirmed to be significant on all performance measures as all P value is less than 0.05. Table 5 shows the two-way interaction effect between quantity loaded to the production line and operator competence on the performance measures. All P values are less than .05 which means the two-way interactions are statistically significant on all performance measures.

Table 4: Main Effects of the Factors

Performance measures	<i>F</i>	<i>P</i>
Machine utilization rate	558.98	.00
Operator utilization rate	2218.82	.00
Flow time	691.37	.00
WIP level	142.15	.00

Table 5: Two-way Interactions of the Factors

Factors	Performance measures	<i>F</i>	<i>P</i>
Operator competence	Machine utilization rate	2492.33	.00
	Operator utilization rate	83206.11	.00
	Flow time	45848.58	.00
	WIP level	4004.77	.00
Quantity loaded to the production line	Machine utilization rate	2845.95	.00
	Operator utilization rate	12609.74	.00
	Flow time	10452.19	.00
	WIP level	89.77	.00

Figures 5 and 6 further reveal the importance of quantity loaded to the production line in all production lines with different competence levels in the aspects of machine and operator utilization. The average machine and operator utilization rates versus the 48 experiments respectively are shown. Despite the difference in operator competence, there is an abrupt drop in both machine and operator utilization rates (see experiments 12 through 16 and experiments 41 through 48). This is mainly due to the phenomenon that the gaps among operators' learning curves widen further as the cumulative units of practices increases (Figure 3), and it affects the balance of the

production line. Therefore, we conclude that an increase in the quantity loaded to the production line leads to a decrease in both machine and operator utilization rates.

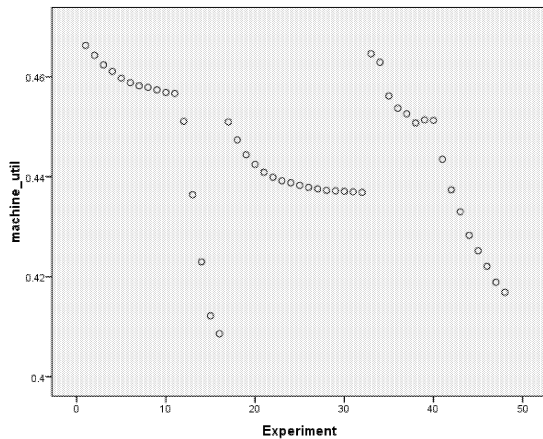


Figure 5: Average machine utilization rates versus experiment (10 replications)

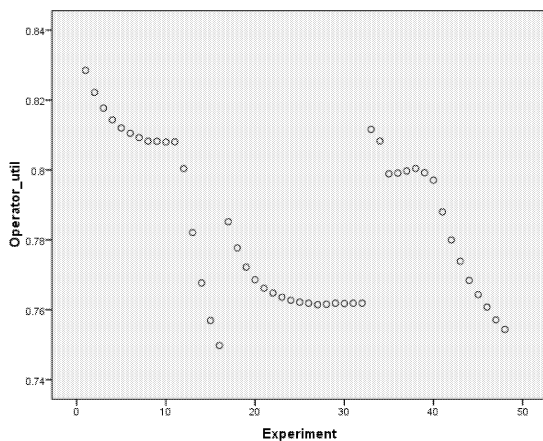


Figure 6: Average operator utilization rates versus experiment (10 replications)

## CONCLUSION

This paper attempts to draw the attention to the impact of quantity loaded to a production line on the manufacturing system performance in the hope of improving the production loading process. The experimental results prove the significant effect of both factors on the apparel manufacturing system performance in terms of machine and operator utilization rates, WIP level and flow time; it is also proven that the two-way interaction effect between quantity loaded to a production line and operator competence is statistically significant on all performance measures. In addition, the simulation results reveal that an increase in production quantity loaded to a production line associated with a decline in

machine and operator utilization rates. The results provide an insights to the planners and researchers when they encounter production loading problem; when deciding the quantity to be loaded to a production line, it has to be strategically selected to avoid an abrupt decline in the manufacturing system performance.

This research has provided an initial investigation on the impact of quantity loaded to a production line and its interaction with operator competence on the manufacturing system performance. It aims to draw the attention from the industry and researchers to the importance of these factors in the production loading process by proving the presence of the significant effect of these factors on the manufacturing system performance. Future research could focus on further investigation of the significance of these factors on the manufacturing system performance in order to optimize the decision of selecting production quantity to be loaded to the production line.

## REFERENCES

- Guner M.G. and C. Unal. 2008. "Line Balancing in the Apparel Industry Using Simulation Techniques." *Fibres & Textiles in Eastern Europe* 16, 75-78.
- Janiak A. and R. Rudek. 2009. "Experience-based Approach to Scheduling Problems with the Learning Effect, Systems, Man and Cybernetics, Part A: Systems and Humans." *IEEE Transactions* 39, 344-357.
- Lee W.C. and C.C. Wu. 2009. "A Note on Single-machine Group Scheduling Problems with Position-based Learning Effect." *Applied Mathematical Modelling* 33, No.4, 2159-2163.
- Leung Y.S.; W.K. Wong; P.Y. Mok; and S.F. Chan. 2006. "An Evaluation of the Adoption of Lean Production as a Supply Chain Management Strategy in the Apparel Industry: a Case Study of a Swimwear Manufacturing Company Using Computer Simulation Techniques." *Proceedings in IEEE International Conference on Management of Innovation and Technology* 2, 862-866.
- Mosheiov G. 2001. "Scheduling Problems with a Learning Effect." *European Journal of Operational Research* 132, 687-693.
- Sirovetnukul R. and P. Chutima. 2010. "The Impact of Walking Time on U-shaped Assembly Line Worker Allocation Problems." *Engineering Journal* 14, 53-78.
- Tang Z.B.; J.H. Hammond; and F.H. Abernathy. 1994. "Design and Scheduling of Apparel Manufacturing Systems with both Slow and Quick Production Lines." *Proceedings of the 33rd Conference on Decision and Control*, 1603-1608.
- Tyler D.J. and G.J.W. Fozzard. 1987. "Simulation Incorporating Characteristics of Manual Skill." *Proceedings 3rd International Conference Simulation in Manufacturing*, 95-106.
- Wang J.; B.J. Schroer; and M.C. Ziemke. 1991. "Understanding Modular Manufacturing in the Apparel Industry Using Simulation." *Proceedings 1991 Winter Simulation Conference*, 441-447.

# **AUTHOR LISTING**





## AUTHOR LISTING

Adamson G. .... 75  
 Almgren T. .... 91  
 Andersson M. .... 61  
 Aslam T. .... 83  
 Atorf L. .... 24  
  
 Bakır M. .... 14  
 Bengtsson V. .... 61  
 Bernedixen J. .... 97  
  
 Camilleri L. .... 56  
 Canan U. .... 14  
 Çelebi S. .... 14  
 Conley W. .... 45  
  
 Dias L. .... 5  
  
 Fajdek B. .... 38  
 Frantzén M. .... 83  
  
 Gustavsson P. .... 91  
  
 Hekimoğlu Ş B. .... 14  
 Holm M. .... 75  
 Huitema G.B. .... 53  
  
 Karlsson I. .... 69  
 Kaya M. Ş. .... 14  
 Krehel M. .... 24  
 Kurowski M. .... 33  
 Kuşoğlu G. .... 14

Leung S.Y.S. .... 104  
 Lia K. .... 56  
 Lidberg S. .... 83  
 Lokshina I. .... 17  
  
 Mak L.L.C. .... 104  
 Moore P. .... 75  
  
 Ng A.H.C. .... 61/69/83/97  
  
 Oliveira J. .... 5  
  
 Pehrsson L. .... 83  
 Pereira G. .... 5  
  
 Rossmann J. .... 24  
  
 Schiele E. .... 17  
 Stachura M. .... 38  
 Studzinski J. .... 33/38  
 Sukhatme G.S. .... 24  
 Svantesson J. .... 91  
 Syberfeldt A. .... 61/69/91  
 Szirbik N.B. .... 53  
  
 van der Burg R.H. .... 53  
 Vieira A. .... 5  
  
 Wang L. .... 75  
 Wong W.K. .... 104  
 Wortmann J.C. .... 53