9^{th} INTERNATIONAL INDUSTRIAL SIMULATION CONFERENCE

2011

ISC'2011

EDITED BY

Simonetta Balsamo

and

Andrea Marin

JUNE 6-8, 2011

VENICE, ITALY

A Publication of EUROSIS-ETI

9th Industrial Simulation Conference 2011

VENICE, ITALY

JUNE 6-8, 2011

Organised by

ETI- The European Technology Institute

Sponsored by

EUROSIS - The European Simulation Society

In Cooperation
With

Universitá Ca' Foscari Venezia

Co-Sponsored by

AEKI

BITE

Ghent University

Higher Technological Institute

University of Skovde

UPV

Hosted by

Centro Culturale Don Orione
Venice, Italy

EXECUTIVE EDITOR

PHILIPPE GERIL
(BELGIUM)

EDITORS

General Conference Chair
Prof. Simonetta Balsamo, Universita Ca'Foscari di Venezia, Venice, Italy

General Program Chair
Andrea Marin, Universita Ca'Foscari di Venezia, Venice, Italy

Post Conference Chair
Gábor Lencse, Szechenyi Istvan University, Gyor, Hungary

Publications Chair
Yan Luo, NIST, Gaithersburg, USA

INTERNATIONAL PROGRAMME COMMITTEE

Discrete Simulation Methodology, Languages and Tools
Jon Andersson, Chalmers University of Technology, Gothenburg, Sweden
Matthias Becker, University Hannover, Hannover, Germany
Helge Hagenauer, Universitaet Salzburg, Salzburg, Austria
Sophie Hennequin, ENIM, Metz Cedex, France
Mauro Iacono, SUN II, Universita di Napoli, Naples, Italy
Bjorn Johansson, Chalmers University of Technology, Gothenburg, Sweden
Panagiotis Katsaros, Aristotle University of Thessaloniki, Thessaloniki, Greece.
Sarka Kvetonova, Brno University of Technology, Brno, Czech Republic
Jan Lemeire, VUB, Brussels, Belgium
Goreti Marreiros, Polytechnic of Porto, Porto, Portugal
Stefano Marrone, Seconda Universita di Napoli. Naples, Italy
Maria do Rosario Moreira, University of Porto, Porto, Portugal
Aziz Naamane, ISIS-EPUM, Marseille, France
Werner Sandmann, TU-Clausthal, Clausthal, Germany
Antonella Zanzi, Institute for the Protection and Security of the Citizen, JRC - European Commission, Ispra Italy

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

Simulation in Manufacturing
Pascal Berruet, Universite Bretagne Sud, Lorient, France
Peter Byrne, Dublin City University, Dublin, Ireland
Ana Camacho, UNED, Madrid, Spain
Luis Camarinha-Matos, Universidade Nova de Lisboa, Lisbon, Portugal
Eduardo Castellano, IKERLAN Technol. Res. Centre, Mondragon-Arrasate, Spain
Remy Dupas, Universite Bordeaux 1, Bordeaux, France
Alexander Felfernig, University of Klagensfurt, Klagensfurt, Austria
Michel Gourgand, Universite Blaise Pascal, Clermont-Ferrand, France
Pouria Homayonifar, REC Scanwafer, Prosgrunn, Norway
Imed Kacem, Université Paul Verlaine Metz, Metz, France
Habtom Mebrahtu, APU, Chelmsford, United Kingdom
Carlo Meloni, Politecnico di Bari - DEE, Bari, Italy
Pascal Meyer, Forschungszentrum Karlsruhe, Karlsruhe, Germany
Marina Valles, UPV, Valencia, Spain
Markus Vorderwinkler, PROFACTOR GmbH, Steyr, Austria
Roland Wischnewski, (RIF) e.V., Dortmund, Germany

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

INTERNATIONAL PROGRAMME COMMITTEE

Simulation in Automotive Systems
Track Chair: Steffen Strassburger, Technical University of Ilmenau, Ilmenau, Germany
Naoufel Cheikhrouhou, EPFL, Lausanne, Switzerland
Julien Richert, Daimler AG, GR/PAA, Sindelfingen, Böblingen, Germany
Rob Walker, Anglia Ruskin University, Chelmsford, United Kingdom

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

Simulation in Electronics, Computer and Telecommunications
Teresa Alvarez, University of Valladolid, Valladolid Spain
Christos Bouras, University of Patras, Patras, Greece
Boguslaw Butrylo, Bialystok Technical University, Bialystok, Poland
Adrian Florea, University "Lucian Blaga" of Sibiu, Sibiu, Romania
Domenico Giunta, ESA-ESTEC, Noordwijk, The Netherlands
Paulo Maciel, University of Pernambuco, Recife, Brazil
Silvia Mirri, University of Bologna, Bologna, Italy
Maurizio Palesi,Universita di Catania, Catania, Italy
Marco Roccetti, University of Bologna, Bologna, Italy
Fernando Boronat Segui, UPV, Valencia, Spain
Renate Sitte, Griffith University, Gold Coast, Australia

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

Simulation in Computer Science
Paul Chung, Loughborough University, Loughborough, United Kingdom
Lipika Deka, Tezpur University and IIT Guwahati, India
Ernst Kesseler, NLR, Amsterdam, The Netherlands
Wolfgang Kreutzer, University of Canterbury, Christchurch, New Zealand

Simulation in Logistics, Transport and Harbour Simulation
El-Houssaine Aghezzaf, Ghent University, Ghent, Belgium
Isabel Garcia Guttierez, Univ. Carlos III de Madrid, Madrid, Spain
Martin Grunow, Technical University of Denmark, Lyngby, Denmark
Thomas Hanne, Univ. of Applied Sciences Northwestern Switzerland, Olten, Switzerland
Dmitry Ivanov, Chemnitz University of Technology, Chemnitz, Germany
Peter Lawrence, Australian Catholic University, Fitzroy, 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
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

Hospital Logistics Simulation
Track Chair: Giorgio Romanin-Jacur, University of Padova, Vicenza, Italy
Antonio Abelha, Universidade do Minho, Braga, Portugal
Joel Colloc, Universite 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
Christophe Claramunt, Naval Academy Research Institute, Brest, France
Gabriella Dellino, University of Siena, Siena, Italy
Werner Dubitzky, University of Ulster, Coleraine, United Kingdom
Krzysztof Fujarewicz, Silesian Technical University, Gliwice, Poland
Laszlo Gulyas, AITIA International Inc., Budapest, Hungary
Eugene Kindler, Ostrava University, Ostrava, Czech Republic
Alexandre Nketsa, LAAS, Toulouse, France
Michal Pechoucek, Czech Technical University, Prague, Czech Republic
Miguel Rocha, University do Minho, Braga, Portugal
Alfonso Urquia, UNED, Madrid, Spain

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

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
Blazej Balasz, Technical University of Koszalin, Koszalin, Poland
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

Simulation in Energy and Power Systems
Saleh Al-Jufout, Tafila Technical University. Tafila, Jordan
Jean-Frederic Charpentier, Ecole Navale, Brest, France
Manuel Cordeiro, CETAV/UTAD, Vila Real, Portugal
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
Rene Banares-Alcantara, Oxford University, Oxford, United Kingdom
Diganta Das, Loughborough University, Loughborough, United Kingdom
Vahid Nassehi, Loughborough University, Loughborough, United Kingdom
Mohamad R. Riazi, Kuwait University, Kuwait

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

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

INTERNATIONAL PROGRAMME COMMITTEE

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

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

Workshops

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

NANOSIM
Clemens Heitzinger, University of Vienna, Vienna, Austria
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 Optimization in Industry
Amos H.C. Ng, University of Skovde, Skovde, Sweden
Anna Syberfeldt, University of Skovde, Skovde, Sweden

INDUSTRIAL SIMULATION 2011

All articles published in these Proceedings have been peer reviewed

EUROSIS-ETI Publications are Thomson Reuters and INSPEC referenced

Selected papers of this conference are published in scientific journals.

A CIP Catalogue record for this book is available from the Royal Library of Belgium under nr 12620

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

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

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

EUROSIS-ETI Publication
ISBN: 978-90-77381-63-2
EAN: **978-90-77381-63-2**

PREFACE

Dear participants,

it is with the greatest of pleasure we welcome you to the 9th EUROSIS Annual Industrial Simulation Conference in Venice. The conference is held in a wonderful location namely the Centro Culturale Don Orione facing the Giudecca channel and close to San Marco square.

This year the ISC program is very rich and puts together research papers coming both from academic institutions and industry. Accepted papers come from many European countries (Austria, Belgium, Czech Republic, Estonia, Germany, Greece, Ireland, Italy, Romania, Russia, Sweden, Switzerland, The Netherlands and Turkey) but a great contribution comes also from research groups outside of Europe.(Australia, Cuba, Colombia, Thailand, USA).

Traditionally, Venice is a city where the mix of cultures coming from all over the world has lead to great cooperative results -S. Marco Cathedral is a wonderful example-and we are sure that this Conference will not be an exception!

The topics addressed from the conference papers are very interesting for the simulation community and range from the specification and analysis of simulation tools to applications of simulation in various contexts such as electronics, computer systems, logistics, economic markets, manufacturing, multimodal transport systems and the environment.

We would like to thank the keynote and invited speaker who contribute to the effectiveness of the conference. We thank the authors of all submitted papers for their contributions and the paper reviewers for reviewing papers and providing valuable feedback to the authors.

Last but not least, we would like to express special thanks to Philippe Geril from EUROSIS, for managing the multi-facetted processes that have gone into the preparation of this conference.

Simonetta Balsamo
General Chair

Andrea Marin
Program Chair

Universita' Ca' Foscari Venezia
Dipartimento di Scienze Ambientali, Informatica, Statistica

CONTENTS

CONTENTS

PRODUCTION OPTIMIZATION

INDUSTRIAL PLANT SIMULATION

MULTIMODAL TRANSPORT SIMULATION

CONTENTS

CONTAINER TERMINAL SIMULATION

IN-HOUSE TRANSPORT LOGISTICS

MATERIALS HANDLING

ENVIRONMENTAL SIMULATION

LATE PAPERS

SCIENTIFIC PROGRAMME

SIMULATION TOOLS

DISCRETE SIMULATION TOOLS RANKING –
a Commercial Software Packages comparison based on popularity

Luís M. S. Dias; Guilherme A. B. Pereira; Pavel Vik; José A. Oliveira
Department of Production and Systems,
University of Minho
Braga, 4710-057, PORTUGAL
http://www.dps.uminho.pt < {lsd,gui,zan}@dps.uminho.pt >
Phone: +351253604764/+351936271733 Fax: +351253604741

KEYWORDS
Simulation languages, Survey, Review, Web-Presence

ABSTRACT

This paper documents a work on all-purpose discrete event simulation tools evaluation. Selected tools must be suitable for process design (e.g. manufacturing or services industries). Rather than making specific judgements of the tools, authors tried to measure the intensity of usage or presence in different sources, which they called "popularity". It was performed in several different ways, including occurrences in the WWW and scientific publications with tool name and vendor name. This work is an upgrade to the same study issued 5 years ago (in 2006). It is obvious that more popularity does not mean necessarily more quality or being better to the purpose of a simulation tool; however a positive correlation may exist between them. The result of this work is a short list of 19 simulation tools.

INTRODUCTION

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.
Simulation languages have been replaced by simulation packages/tools.
High market prices of simulation tools in the past decades, added to other factors like: ease of construction of a simulation tool; the emerging graphics facilities; the wide field of applications and the absence of strong standards or languages; lead to a large, or may be too large, tools offer (Dias, 2005).
Thus, for instance, in the Industrial Engineering Magazine (1993/July) there is a list of 45 commercial simulation software products. The sixth biannual edition of simulation software compiled by James J. Swain in 2003 identifies about 60 commercial simulation products, 55 in 2005 and 48 in 2009 (Swain, 1991-2009). The annual 2004 SCS edition - "M&S Resource Directory" lists 60 simulation products (Klee, 2004). In the "Simulation Education Homepage" (Simulation tools list by William Yurcik) there were more than 200 simulation products, incl. non commercial tools.
This work started with Swain´s list, removing non discrete event simulation environments, and adding some tools found in more than one list sources. Some other relevant simulation tools like SIMPRO don´t appear in our list due to lower web-presence (see Table 2).
This tools comparison was performed previously in 2006, and is here extended with more parameters and relevant changes are discussed.

Product names in this paper are trademarks or registered trademarks of their respective owners.

MOTIVATION – WHY MEASURING *POPULARITY*?

In this scenario of such a large simulation tools' offer it is unfeasible to perform a consistent experiment. The comparison, based on features or characteristics is also very difficult or non conclusive because most of them have similar features lists.
The measure here called "popularity" was the way that we found to overcome those difficulties identifying the tools that are potentially the best or most used.
To choose a popular simulation tool is positive in two *ways*:

- If you are a company, it is easier to find simulation specialists with know-how on a popular tool;
- If you are a simulation specialist, it is easier to find companies working with a popular tool.

The second *way* includes educational purposes because students should be the future simulation specialists.
Nevertheless, popularity should never be used as a unique parameter for simulation tools selection. If so, new tools, would never gain market share - and this is a generic risk, not a simulation particularity.
So, the popularity may be seen as a significant "blind" factor to be used in conjunction with direct evaluation mechanisms like features comparison and experimentation

DEVELOPMENT – *POPULARITY* EVALUATION

Our evaluation method, in order to identify a short list containing the most popular or important tools, was essentially based in the intensity or level of presence on:

- WWW (Internet);
- Winter Simulation Conference scientific publications.
- Document database oriented sites (*new*)
- Social networks (*new*)
- Selected set of sources (e.g. scientific surveys, lists and homepages).

TECHNIQUE

For the purpose of measuring the web-presence, the **Google searching engine** was used. The reasons are:

- It is the most-used search engine on the Web (http://searchenginewatch.com/3630718) – around 61% of all searching actions in 2008
- Google *owns* different sources of relevant information (books, youtube, synonyms, maps, translator, etc.)
- It supports a function for getting an approximated number of results (for this project we developed also a function for automatically updating data in an Excel sheet)
- It supports restricted search to specific domains (e.g. scribd.com, books.google.com, linkedin.com, facebook.com).

FACTORS DESCRIPTION AND *TUNING*

We used **around 40 parameters/factors** for evaluating each simulation tool, listed in Table 1. For each tool we defined the two following labels:

- **"Tool"** represents the search string containing the name of the simulation tool, the word "simulation" and some additional words to avoid finding pages out of the topic due to common English words used as tools names. (e.g. "Arena", "Witness", "Extend", "Quest" etc)
- **"Vendor"** represents the search string containing the name of the simulation tool vendor.

"T" is also used as abbreviation of "Tool" and **"TV"** as "Tool"+"Vendor".

In the **factors** where the results represent the number of occurrences, the values may vary from units to millions. The sum of all of them together would lead to irrelevant factors mixed with absorbent factors. To reduce the impact of different orders of magnitude, the uses of mathematical functions were studied in order to "control" big numbers, although keeping relative differences. Square and cubic root, Natural and ten base logarithms were the evaluated possibilities.
After an extensive iterative process, the **cubic root** was chosen once it was proven to consider both small and big numbers adequately - see Figure 1 (cubic root (x) = x^1/3).
The use of a cubic root of a number in place of the number it-self, is the same as comparing the volume of cubes, using only the value of the their width.

	^1/2	^1/3	LN	LOG10
1000000	1000	100	14	6
100000	316	46	12	5
10000	100	22	9	4
1000	32	10	7	3
100	10	5	5	2
10	3	2	2	1
1	1	1	0	0

Figure 1 Possible Functions to Factors Adjustment

Almost all factors between #1 and #20, were adjusted using the cubic root of the number of occurrences, multiplied by "***Factor Weight***" (indicated under the label of each column).

In each of those columns in the datasheet is a pair of values:
-The right sided values represent the "raw" number of occurrences;
-The left sided values represent the result of the cubic root of those values, multiplied by the "*Factor Weight*". Those values are then directly added to the respective tool scoring.
Above the left side of each column is the average of the adjusted values. Each of those averages shows up the real influence of each factor in the tool's score. Those values are named as "**Average Effective Factor Weight**".

The complete list with factors description (in Table 1) is organized in **two groups**.

The first group includes all factors that are calculated based on Google search results (approximate number of results). This group is split into two tables: Figure 2 (with Winter Simulation Conference, Documentation sites and Social Networks Scoring) (#1→#10) and Figure 3 (with general searches in WWW, including tools URL web-presence and Google's and Yahoo page rankings) (#11→#18).

	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
Tool(s)	WSC "T"	WSC "TV"	**amazon**.com "TV"	**Scholar.Google** "T"	**Scholar.Google**.com	**scribd**.com	**docstoc**.com	**youtube**.com	**linkedin**.com	**facebook**.com
	0,8	1,6	1	0,05	0,2	0,3	0,4	0,3	0,6	1
	3,1	[illegible]	1,2	0,5	1,3	1,1	1,1	1,0	2,6	2,3
	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3	1/3
Arena	5,8 380	9,5 210	3,1 30	0,9 5100	2,7 2.500	3,5 1.530	1,8 96	2,2 390	4,8 500	6,7 300
Simul8	3,9 120	6,2 57	2,8 23	0,5 1100	1,5 460	1,2 72	2,1 137	1,0 41	4,3 380	5,9 210
WITNESS	4,5 180	6,0 54	1,6 4	0,7 2300	1,8 680	1,3 77	2,1 144	2,7 750	3,6 220	5,7 190
ProModel (Service	5,0 250	7,9 120	2,9 24	0,7 2500	2,1 1.100	0,9 25	2,0 124	0,9 29	4,8 500	1,9 7
ExtendSim	3,5 81	6,3 62	1,3 2	0,6 1900	1,8 750	1,3 74	1,0 18	1,2 57	1,2 8	1,8 6
AnyLogic (eXperimental (	3,4 75	5,7 45	1,9 7	0,5 1000	1,5 440	2,0 46	0,8 9	0,5 6	2,2 50	2,0 8
FlexSim	2,2 22	4,5 22	0,0 0	0,5 870	1,3 300	0,6 10	0,8 8	2,9 930	3,4 190	5,9 210
AutoMod	5,1 260	7,0 84	1,7 5	0,6 1800	1,4 360	0,6 8	0,8 8	1,1 50	3,1 140	1,0 1
Plant Simulation -	2,2 21	3,1 7	1,0 1	0,4 358	0,8 70	1,3 89	1,9 99	1,4 110	2,8 100	3,1 29
QUEST; DPM POWERTR	2,4 27	4,6 23	1,0 1	0,5 900	1,3 290	0,8 22	2,1 142	1,8 210	4,1 310	1,9 7
Enterprise Dynamics	2,7 39	4,0 16	0,0 0	0,3 250	0,9 95	0,6 7	0,5 2	0,5 4	3,7 230	0,0 0
SIMPROCESS (SIMSCRI	3,9 120	4,5 22	1,4 3	0,7 2500	1,2 220	0,8 16	1,0 16	0,0 0	0,9 3	0,0 0
ProcessModel	2,1 18	4,2 18	1,3 2	0,4 410	1,5 420	1,6 156	1,5 55	0,9 26	2,6 80	3,4 38
Simio *NEW*	2,3 24	4,1 17	1,0 1	0,3 140	0,6 23	0,6 9	0,4 1	1,0 33	2,7 92	3,8 54
Micro Saint + IPME	3,1 57	3,4 10	0,0 0	0,4 650	1,0 120	0,6 7	0,7 5	0,0 0	1,1 6	0,0 0
SimCAD Pro	1,0 2	2,0 2	0,0 0	0,1 15	0,5 14	0,4 3	0,6 3	0,5 6	1,0 4	1,0 1
SLX + Proof 3D + Proof 5	3,1 58	6,1 55	0,0 0	0,3 150	1,0 143	0,6 9	0,7 5	0,0 0	1,1 6	1,0 1
ShowFlow (based on Tay	1,2 3	2,3 3	0,0 0	0,2 44	0,4 9	0,5 4	0,0	0,0 0	0,6 1	0,0 0
GPSS World for Windows	1,6 8	2,9 6	1,0 1	0,3 320	1,2 210	1,2 57	0,7 6	0,0 0	0,0 0	1,0 1

Figure 2 WSC, DOCs and Social

	#11	#12	#13	#14	#15	#16		#17	#18
Tool(s)	**WWW** "T"	**WWW** "TV"	N. of **Links** to "Site"	**"Site"** in **WWW**	Domain age, in years	Google PageRank 2006		**Google PageRank** 2011	**Yahoo** (new!) **checkpagerank**.net
	0,02	0,1	0,4	0,06	0	0,1		0,5	0,2
	1,0	2,8	1,3	1,4		0,5		2,4	1,7
Arena	2,4 1800000	5,6 180000	1,4 46	1,8 28000	10	0,6 6	0	3,0 6	1,8 743
Simul8	0,9 99000	3,2 32000	1,5 57	1,9 33000	14	0,6 6	-1	2,5 5	2,9 2 995
WITNESS	1,6 490000	3,6 48000	1,5 54	1,7 22000	14	0,5 5	0	2,5 5	2,2 1 393
ProModel (Service	1,3 250000	3,0 26000	1,8 93	1,5 15000	16	0,6 6	-1	2,5 5	2,3 1 503
ExtendSim	1,5 460000	4,3 80000	1,4 39	2,0 38000	06	0,6 6	0	3,0 6	1,8 717
AnyLogic (eXperimental (	1,1 190000	2,8 22000	2,1 142	1,9 34000	13	0,6 6	0	3,0 6	2,9 2 918
FlexSim	0,6 33000	2,6 18000	1,2 24	2,0 39000	10	0,6 6	-1	2,5 5	3,6 5 956
AutoMod	0,7 38000	2,8 23000	1,4 47	0,8 2800	16	0,6 6	-3	1,5 3	1,1 161
Plant Simulation -	1,0 120000	4,8 110000	0,7 6	0,8 2500	11	0,4 4	1	2,5 5	0,8 63
QUEST; DPM POWERTF	1,5 430000	4,2 72000	1,9 114	1,1 6400	11	0,6 6	0	3,0 6	2,0 1 016
Enterprise Dynamics	0,5 14000	1,9 6400	1,6 70	1,2 8000	13	0,3 3	1	2,0 4	0,6 27
SIMPROCESS (SIMSCRI	0,7 36000	1,7 4700	1,4 44	0,9 3700	13	0,5 5	-1	2,0 4	1,2 235
ProcessModel	0,7 46000	3,7 50000	0,8 8	1,0 4200	15	0,6 6	-2	2,0 4	1,7 668
Simio *NEW*	0,9 80000	2,0 8000	1,4 47	1,4 14000	06	- -		2,0 5	1,7 607
Micro Saint + IPME	0,6 27000	1,8 6000	1,1 20	1,9 32000	14	0,6 6	-1	2,5 5	1,2 238
SimCAD Pro	0,4 11000	1,4 2900	0,8 7	1,1 6700	14	0,5 5	-1	2,0 4	1,8 715
SLX + Proof 3D + Proof 5	0,7 48000	1,3 2000	0,6 4	1,0 4000	14	0,5 5	-1	2,0 4	1,0 128
ShowFlow (based on Ta	0,5 16000	0,7 320	0,8 7	0,7 1900	11	0,5 5	-1	2,0 4	0,9 101
GPSS World for Windows	0,4 6760	1,3 2300	0,7 6	0,9 3000	12	0,4 4	0	2,0 4	1,0 126

Figure 3 WWW Searches and Site ranking

The second group is based on a selection of scientific works of review/survey, software lists and conference activities and sponsorships. The 20 factors used in this group are in one table in Figure 4 (The first three columns refer to relevant reviews with some kind of tools evaluation and their results were used here with proportional scoring. All the others are just binary scores when the tool name is referenced in the specified sources) (#24→#44).

Tool(s)	#24 Mustafee 2007 "A	#25 Abu-Taieh, 2007 CSP	#26 VIVACE review 2004	#27 SimulationTools.bib,	#28 ORMS Survey 2009	#29 WSC 2010	#30 Systemflow list 2009	#31 Google Directory	#32 Wikipedia - List of	#33 ORMS Survey 2003	#34 PMC short list (2010)	#35 www.averill	#36 SimServ WhitePaper	#37 IIE Exhibitors (2011)	#38 Simul8Site (2006)	#39 WSC 2005	#40 Solution Simulation	#41 Hlupic, 2000	#42 Babulak 2006	#43 Pemberton Cyrus	#44 Edwin Valentin (2002)	#45 tot. selected sources
	3	2	2	3	4	3	2	2	2	1	2	1	1	2	1	1	1	1	2	1	1	0,8
	[illegible]	[illegible]	[illegible]	2,7	2,5	1,6	1,5	1,3	0,8	0,8	0,7	0,7	0,6	0,5	0,5	0,5	0,4	0,4	0,4	0,4	0,3	
Arena	2,2	2,8	3,6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	50
Simul8	4,1	2,7	2,2	1	1	1	1	1	1	1	1	1	1	-	-	1	1	1	1	1		49
WITNESS	3,2	2,3	3,3	1	-	1	1	1			1	1	1		1			1	1	1	1	40
ProModel (Service	2,1	2,3	3,1	1	1	1	1	1		1	1	1	1	1	1	1		1		1		42
ExtendSim	2,6	2,0	3,2	1	1	1	1	1		1		1	1	1	1	1						39
AnyLogic (eXperimental	2,7	3,2	2,1	1	1	1	1	1	1	1						1	1					38
FlexSim	0,6	2,9	2,5	1	1	1	1	-		1		1	1	1		1	1	1		1		33
AutoMod	1,3	3,2	2,4	1	-	1	1	-		1	1	1	1		1	1	1		1	1	1	34
Plant Simulation -	1,9	2,0	1,0	1	1	1	1	-	1	1		1	1				1				1	30
QUEST; DPM POWERTF			3,4	1	-		1	-			1	1	1		1		1					17
Enterprise Dynamics	2,1	2,3	2,3	1	-		1	1	1	1	1		1		1	1					1	31
SIMPROCESS (SIMSCR	2,7	2,3		1	-		1	1		1		1			1			1				23
ProcessModel	0,6		1,0	-	-			-		1		1			1							6,2
Simio NEW				1	1	1		1	1													14
Micro Saint + IPME	1,3	1,2	1,0	1	1		1	-		1		1	1			1				1		22
SimCAD Pro	1,2	2,7	1,0	1	1			1	1	1				1			1					26
SLX + Proof 3D + Proof 5			1,0	1	1		1	1	1			1				1						17
ShowFlow (based on Ta	0,6	2,1	1,0	1	1			1		1			1					1			1	21
GPSS World for Windows		2,3		-	-			-	1	1					1			1				8,9

Figure 4 Selected Sources

Tool(s)	#19 WSC	#20 DOCS	#21 SOCIAL	#22 WWW	#23 tot. (WSC docs social WWW)	#45 tot. selected sources	#46 Total Score	#47 Price (Swain, 2009 or 2003) K$ Oth	<	~	>
					1	0,8			<	~	>
Arena	15	14	11	17	58	50	9,9		0,5	??	25
Simul8	10	9	10	14	43	49	9,0		1,5		5
WITNESS	11	10	9	14	44	40	8,7				
ProModel (Service	13	9	7	13	42	42	8,6		3,5	20	40
ExtendSim	10	7	3	15	35	39	8,1		1,8	2,5	5
AnyLogic (eXperimental (	9	7	4	14	35	38	8,0		6		16
FlexSim	7	6	9	13	35	33	7,8		15	20	65
AutoMod	12	6	4	9	31	34	7,6			24	
Plant Simulation -	5	7	6	11	29	30	7,3			20	30
QUEST; DPM POWERTF	7	8	6	14	35	17	7,0				
Enterprise Dynamics	7	3	4	8	21	31	6,8		3,5		
SIMPROCESS (SIMSCRI	8	5	1	8	23	23	6,4		0,8		10
ProcessModel	10	7	6	11	34	6,2	6,3			2,5	
Simio NEW	6	4	6	9	26	14	6,1			10	14
Micro Saint + IPME	7	3	1	10	20	22	6,1		9		
SimCAD Pro	3	2	2	8	15	26	6,0			1	
SLX + Proof 3D + Proof 5	9	3	2	7	21	17	5,8		1,8		9
ShowFlow (based on Ta	3	1	1	6	11	21	5,2			1,5	
GPSS World for Windows	5	4	1	7	17	8,9	4,9		4,5		

Figure 5 Ranking Total Scores and Prices

Figure 5 (and Figure 7) contains the table with Total Scoring. Columns #19 to #22 have the scores of the first group (Winter Simulation Conference, Documentation sites, Social Networks Scoring and WWW). Columns #23 and #45 have the sub-totals of the two Groups, where the #45 is the scoring of the second group - Selected Sources. The column #46 has the Final Score of this Paper Research, adjusted to a 0-10 scale, using the square root mathematical function.

Table 1 Factors Description

#	Factor Name	Description
1	WSC "only Tool"	Occurrences of **"Tools" in** www.informs-sim.org. (Institute for Operations Research and Management Science - Simulation Society) (includes all **Winter Simulation Conference** – papers 1997-2011April)
2	WSC "TV"= "Tool+Vendor"	Occurrences of **"Tools" + "Vendor" in** www.informs-sim.org (same source as **#1**))
3	amazon.com	Occurrences of "Tools" + "Vendor" in site:amazon.com
4	Scholar.Google "T"	Occurrences of **"Tools"** in site:Scholar.Google
5	Scholar.Google	Occurrences of "Tools" + "Vendor" in site:Scholar.Google
6	scribd.com	Occurrences of "Tools" + "Vendor" in site:scribd.com
7	docstoc.com	Occurrences of "Tools" + "Vendor" in site:docstoc.com
8	youtube.com	Occurrences of "Tools" + "Vendor" in site:youtube.com
9	linkedin.com	Occurrences of "Tools" + "Vendor" in site:linkedin.com
10	facebook.com	Occurrences of "Tools" + "Vendor" in site:facebook.com
11	WWW only "Tool"	Number of web pages with **"Tools"**+"simulation"(the "simulation" string was used to count only the internet pages in the simulation area)
12	WWW "TV"	Number of web pages with "Tools"+"Vendor"+"simulation"
. 13	N. of WWW Links to "Site"	Number of web pages pointing with a link to the "Site" of the "Vendor" (in Google)
14	"Site" in WWW	Occurrences of vendor's site address in WWW
15	Domain age,	Vendor's site years old (factor not used, just information)
16	Google PageRank'06	Google "PageRank" (Google automatic evaluation about page importance). Record from 2006.
17	Google PageRank'11	Google "PageRank" (Google automatic evaluation about page importance). Current value (2011).
18	Yahoo (new!)	Yahoo "PageRank" Yahoo evaluation about page importance. *checkpagerank.net*. Current value (2011).
19	WSC	Sum of WSC related factors: #1 + #2 (Factors adjusted with cubic root function)
20	DOCS	Sum of Documents Repositorium related factors: from #3 to #8 (Factors are adjusted with cubic root function)
21	SOCIAL	Sum of Social Networks Activity related factors: #8 + #9 + #10 (Factors are adjusted with cubic root function)
22	WWW	Sum of general WWW webpresence factors: from #11 to #18 (Factors are adjusted with cubic root function)
23	Tot.(WSC docs social WWW)	Total Sum of Factors related to Google Search engine used for measuring webpresence (from #19 to #22)
24	Mustafee 2009	Mustafee N. 2007 "A Grid Computing Framework For Commercial Simulation Packages". Brunel University, West London, PhD Thesis. bura.brunel.ac.uk/ bitstream/2438/4009/1/Fulltext(Thesis).pdf
25	Abu-Taieh, 2007	Abu-Taieh. 2007. Commercial Simulation Packages - CSP. I.J. of SIMULATION Vol. 8 No 2. ISSN 1473-804x (http://ducati.doc.ntu.ac.uk/uksim/journal/Vol-8/No-2/paper-7.pdf)
26	VIVACE review 2004	VIVACE review: "Techniques to Model the Supply Chain in an Extended Enterprise", Kim et.al, 2004.
27	SimulationTools.bib 2010	List with Simulation Tools with Short Description. By Andrea Emilio Rizzoli. SimulationTools.bib, 2010 http://www.idsia.ch/~andrea/sim/simlang.html
28	ORMS Survey 2009	Swain J. 2009. Simulation Software Survey. OR/MS Today magazine, Institute for Operations Research and the Management Sciences (INFORMS). Lionheart Publishing. 1991-2009. www.lionhrtpub.com /orms/surveys/Simulation/Simulation.html
29	WSC 2010 sponsorship	Sponsors of the Winter Simulation Conference 2010 (Memory registered in year 2011)

#	Factor Name	Description
30	Systemflow list 2009	Simulation Software List – System flow Simulations, Inc. (2005-2009)http://www.systemflow.com/software_list.htm
31	Google's Simul. S/W	Google Directory of Simulation Software www.google.com/Top/Science/Software/Simulation/
32	Wikipedia - List of Simul.S/W	Wikipedia - List of discrete event simulation software http://en.wikipedia.org/wiki/List_of_discrete_event_simulation_software
33	ORMS Survey'03	Swain J. 2003. (See Factor #28)
34	PMC short list (2010)	List of the simulation tools where the PMC Company have competency (http://www.pmcorp.com/sim_services.shtm)
35	www.averill-law.com	Averill-law list of simulation training software: (www.averill-law.com/simulation-training-software.htm)
36	SimServ WhitePaper (2004)	Sim-Serv organization white paper about simulation tools. Jaroslaw Chrobot. 2004. (http://www.sim-serv.com/wg_doc/WG1_White_Paper_discussion.pdf)
37	IIE Exhibitors (2011)	Exhibitors of the IIE Conference (2011) (Institute of Industrial Engineers) (http://www.iienet2.org/annual2/details.aspx?id=6790)
38	Simul8Site (2006)	Brooks homepage (Simul8) identification of concurrency (www.simul8.com/products/webdemo.htm)
39	WSC 2005	Sponsors of the Winter Simulation Conference 2005 (Memory registered in year 2006)
40	Solution Simulation 2004	Sponsors of the conference "Solution Simulation 2004". http://www.simsol.org/2004%20files/SimSol%20onsite%202004%20revised.pdf
41	Hlupic, 2000	Hlupic V. 2000. Simulation software: an operational research society survey of academic and industrial users. In (J. Joines et. al., eds.) Proc. WSC 2000. (Piscataway, New Jersey), IEEE, 1676-1683.
42	Babulak 2008	Babulak B and Wang M. 2008. "Discrete Event Simulation: State of the Art" International Journal of Online Engineering (iJOE), Vol 4, No 2 (2008) ISSN: 1861-2121
43	P. Cyrus 2004 Sim. S/W	Simulation Software List by Pemberton Cyrus, 2004 http://pt.scribd.com/doc/38056975/Simulation-Software-2004-05-28
44	Edwin Valentin (2002)	Tools systematic evaluation based on experimentation (Valentin, 2002). (http://www.tbm.tudelft.nl/webstaf/edwinv/SimulationSoftware/index.htm)
45	Selected Sources Total	Sum of Factors related to the 20 selected sources, from #24 to #44.
46	Total Score	Total sum of all factors. Calculated as SQRT(#23+0,8*#45)/K. The 0,8 coefficient was defined to balance the relative weight between web presence and selected sources. The K divisor was used to adjust scale to 0-10.
47	Price	Software tools prices from (Swain, 2009 or 2003) in thousands of Dollars (K$). Minimum, Average and Maximum prices mentioned, when available. Not used for scoring.
48	Ranking 2006	Ranking of our 2006 ranking
48		Ranking changes analysis for each tool
50	Ranking 2011	Current Ranking.

COMPARISON WITH FORMER 2006 RANKING

In this ranking we introduced the **social networks** communities presence that are used every day (e.g. Facebook, Linkedin, youtube), and show recent activity.
Another set of factors was introduced – the number of occurrences on **documents** database oriented sites (e.g. amazon.com, Scholar.Google, scribd.com, docstoc.com, youtube.com).
The two new sets of factors are basically fed by people in the spirit of web 2.0. It is remarkable that a **page rank reduction** has occurred on many vendors' sites. This could mean that users share more among them rather than being connected to vendor's websites.

Figure 6 shows tools rankings evolution between 2006 and 2011 (Dias et.al 2007).
The four most popular tools are Arena, Simul8, Witness, and Promodel. Simul8 registered a significant increase of its popularity level, as well as AnyLogic, Plant Simulation and Enterprise Dynamics.
A tool like Siemens Tecnomatix – Plant Simulation is more popular now than 5 years ago basically as a result of the distribution of Digital Factories Packages by big automotive companies to their suppliers.
There is one new "player" (Simio) that climbed directly to a noticeable position.
Simprocess, SLX and Automod registered a significant popularity loss.

Tool(s)	Site	#48 Ranking 2006	#49 R2006 - R2011 (5 years)	#50 Ranking 2011
Arena	www.arenasimulation.com	1	0	1
Simul8	www.simul8.com	7	5	2
WITNESS	www.lanner.com	5	2	3
ProModel (Service	www.promodel.com	2	-2	4
ExtendSim	www.extendsim.com	6	1	5
AnyLogic (eXperimental	www.xjtek.com	15	9	6
FlexSim	www.flexsim.com	10	3	7
AutoMod	www.appliedmaterials.com/services-software/libr	3	-5	8
Plant Simulation -	www.plm.automation.siemens.com/en_us/produc	17	8	9
QUEST; DPM POWERTI	www.delmia.com	8	-2	10
Enterprise Dynamics	www.incontrolsim.com	16	5	11
SIMPROCESS (SIMSCR	www.simprocess.com	4	-8	12
ProcessModel	www.processmodel.com	12	-1	13
Simio *NEW*	www.simio.com		NEW	14
Micro Saint + IPME	www.maad.com	14	-1	15
SimCAD Pro	www.createasoft.com	-	-	16
SLX + Proof 3D + Proof !	www.wolverinesoftware.com	11	-6	17
ShowFlow (based on Ta	www.showflow.com	-	-	18
GPSS World for Window:	www.minutemansoftware.com	18	-1	19

Figure 6 Ranking comparison 2006-2011 and Tool's Site

LIST OF OTHER SIMULATION TOOLS

The process of tools selection, lead to the exclusion of many simulation tools. They were not suited for discrete event simulation or because of their lower popularity score. In the following Table 2 is the list of such software tools.

Table 2 List of other simulation tools

AP3, Capstone, COCODRIS (realistic 3D), COOPS, Crystal Ball, CSIM-19 (c++,c), DecisionPro, DESMO , Factory Explorer, G.R.A.S.P., GAUSS, HighMast, HOCUS, iGrafx, INSIGHT , INSTRATA, IRT_PETRINWZ, KanbanSIM, Lean Modeler, MAST, ModSim, NET, NETWORK II.5 (CACI), OMNEST, OPTIMA, PACE, PCModel, PIMSS, Process Charter, Proplanner Manufacturing Process Management Software, QGERT, Resource Manager, SDI Supply Chain (Supply Chain Builder), SIGMA, Siman/Cinema, SIMFACTORY II.5 (CACI), SIMPRO, SimPRO (other), DOSIMIS-3, SIMULA, SLAM, SLIM, SLOOP/TERMINAL, TOMASWeb, VISSIM (traffic), VS7; VS6, VSE Visual Simulation Environment, WebGPSS, WORKSPACE, XCELL+

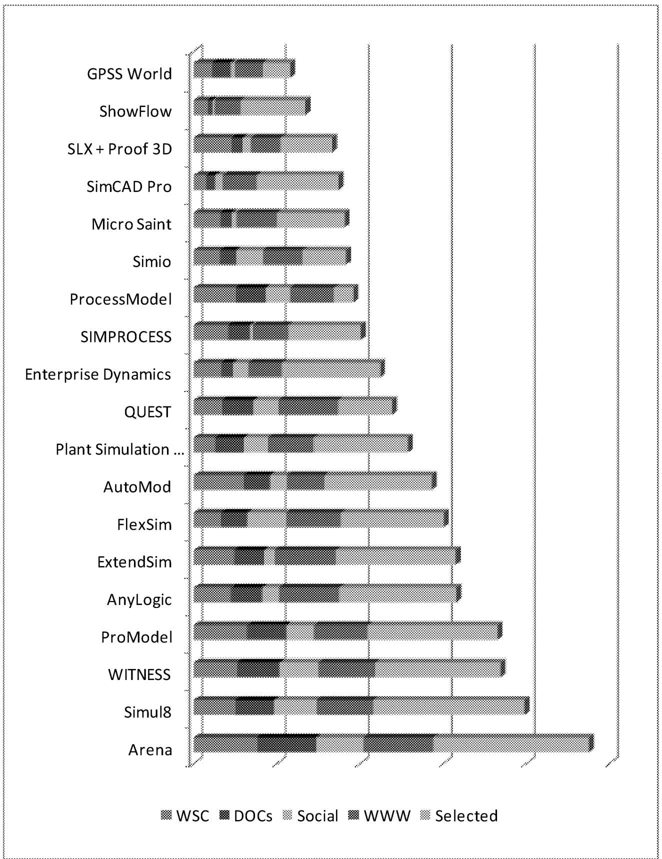


Figure 7 Scoring distribution

CONCLUSION

This list was created based on the subjective evaluation of a parameters set. Different parameters may be used alternatively with different weights producing other results. Anyway, even with subjectivity, we believe that the **Top 10** "popular" simulation commercial tools are included in this list (of 19). As well as it is most probable that this list includes the top 10 "most used" and "best" contemporary simulation tools.

The chart in Figure 7, can help to visualize the strengths and weaknesses of each tool, in a comparative analysis.

In measuring popularity some other relevant parameters could be considered like the number of sold licences in the industry area (with a company size factor) or used at universities for education purposes. Although it is quite difficult to reliably collect these types of data.

One relevant improvement to this study may consist on giving more weight to recent references, using some **time-line** approach to analyse trends. Some effort have been made using Google but searching with multiple keywords was not yet successful in searching the historical data of the search engine.

REFERENCES

Abu-Taieh, 2007: CSP - COMMERCIAL SIMULATION PACKAGES. I.J. of SIMULATION Vol. 8 No 2. ISSN 1473-804x (http://ducati.doc.ntu.ac.uk/uksim/journal/Vol-8/No-2/paper-7.pdf)

Babulak B and Wang M. 2008. Discrete Event Simulation: State of the Art. International Journal of Online Engineering (iJOE), Vol 4, No 2 (2008) ISSN: 1861-212

Dias, L.; Rodrigues, A.; Pereira, G. 2005. "An activity oriented visual modelling language with automatic translation to different paradigms", In *Proceedings of the 19th European Conference On Modelling And Simulation* (ECMS 2005), Riga, Letónia. Ed. Yury Mercuryev et al. Junho de 2005. pp. 452-461

Dias, Luís S. 2005. Automatic Interactive Modelling of Simulation. Ph.D. Thesis. (written in Portuguese - Modelação Automática Interactiva de Simulação). Department of Production and Systems Engineering University of Minho, Portugal.

Dias, L., Pereira, G., Rodrigues, A., 2006. A Shortlist of the Most 'Popular' Discrete Simulation Tools, In (M. Becker, H. Szczerbicka, eds.) Proc. ASIM 2006 - 19th Symposium on Simulation Technique, SCS.(Hanover, Germany) pp. 159-163. 12-14

Dias, L.; Pereira, G.; Rodrigues, A. 2007. A Shortlist of the Most "Popular" Discrete Simulation Tools. Journal SNE Simulation News Europe., April 2007, pp 33-36. ISSN 0929-2268.

Hlupic, Vlatka and Ray J. Paul. 1999. Guidelines for selection of manufacturing simulation software. In *IIE Transactions on Design and Manufacturing*, 31(1). 21-29.

Hlupic, Vlatka. 2000. Simulation software: an operational research society survey of academic and industrial users. In *Proceedings of the 2000 Winter Simulation Conference*, ed. J. A. Joines, R. R. Barton, K. Kang e P. A. Fishwick. 1676-1683. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Klee, Harold (ed.). 2004. Modeling and Simulation Resource Directory. In *Modeling and Simulation*, Directory Issue 2004, vol.3 nº4, Oct-Dec. The Society for Modeling and Simulation International (SCS), San Diego, USA. ISBN 1-56555-297-0 ISSN 1537-7733

Law, A. M. e McComas, M. G. 2002. Simulation optimization: simulation-based optimization. In *Proceedings of the 34th Conference on Winter Simulation*: Exploring New Frontiers (San Diego, California, December 08 - 11, 2002). Winter Simulation Conference. 41-44. portal.acm.org/citation.cfm?id=1030461

Mustafee N. 2007. A grid computing framework for commercial simulation packages. Brunel University, West London, PhD Thesis.bura.brunel.ac.uk/bitstream/2438/4009/1/Fulltext(Thesis).pdf

Swain, James J. 2009. Simulation Software Survey. OR/MS Today magazine from Institute for Operations Research and the Management Sciences (INFORMS). Lionheart Publishing. 1991-2009. www.lionhrtpub.com/orms/surveys/Simulation/Simulation.html

Solution Simulation 2004, Site of the conference - the sponsors list: (http://www.simsol.org/2004%20files /SimSol%20onsite%202004%20revised.pdf).

Recent Sponsors of the Institute of Industrial Engineers Conference (www.iieannual.org/IndustryTracks.htm)

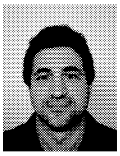
Sponsors of the Winter Simulation Conference 2005 (www.wintersim.org/exhibits.htm)

ORMS Simulation Tools List (www.lionhrtpub.com/orms/surveys/Simulation/Simulation.html)

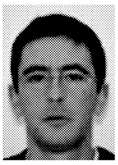
Site of the INFORMS simulation society (www.informs-sim.org)

Site with scientific simulation publications of the Institute for Operations Research and the Management Science - Simulation Society. Contains the Winter Simulation Conference Publications (www.informs-sim.org)

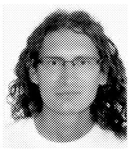
BIOGRAPHY



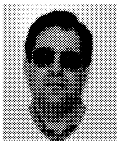
Luís M S Dias was born in 1970 in Vila Nova de Foz Côa, Portugal. He graduated in Computer Science and Systems Engineering in the University of Minho, Portugal. He holds an MSc degree in Informatics Engineering and a PhD degree in Production and Systems Engineering from the University of Minho, Portugal. His main research interests are Simulation, Systems Performance, Operational Research and Systems Visual Modeling.



Guilherme A B Pereira was born in 1961 in Porto, Portugal. He graduated in Industrial Engineering and Management in the University of Minho, Portugal. He holds an MSc degree in Operational Research and a PhD degree in Manufacturing and Mechanical Engineering from the University of Birmingham, UK. His main research interests are Operational Research and Simulation.



Pavel Vik was born in 1982 in Mladá Boleslav, Czech Republic. He studied at the Technical University of Liberec where he obtained his MSc degree in Manufacturing Systems, in 2005. He is making doctoral studies in Manufacturing Systems and Processes and he is currently at the University of Minho. His main research interest is integration of computer simulation and CAD system in area of production systems design. His email address is <vikpavel@seznam.cz>.



José A Oliveira was born 1966 in Matosinhos, Portugal. He studied Mechanical Engineering at the University of Porto, Portugal. He graduated with a Ph.D. in Production and Systems Engineering at University of Minho, Portugal. His main research interests are Optimization with Heuristic Methods in Systems Engineering.

Appendix

Tool(s)	"Tools" in WSC	"Tools" + "Vendor" in WSC	N. of WWW Links to "Site"	"Tools" + "simulation" in WWW	"Tools" "Vendor" "simulation" in WWW	Google PageRank	tot. (WSC+WWW)	ORMS Survey 2003	Hlupic, 2000	Simul8Site	Edwin Valentin	SimServ WhitePaper	PMC short list	Law+McComas(2002)	Solution Simulation 2004	IIE Conference (Institute	WSC 2005	tot. selected sources	t) ~Total
	2	3	2	0,5	1	0,5		2,0	2,0	0,5	1,0	0,5	0,5	0,5	1,5	1,0	2,0		6,0
Arena	382	55	67	999 999	45 000	6	22,1	X	X	X	X	X	X	X	X	X	X	13	19
ProModel (Service Model, MedM	181	181	60	60 000	50 000	6	22,3	X	X	X		X	X			X	X	11	19
Automod (Brooks Automation -	146	40	35	20 000	700	6	17,6	X		X	X	X	X		X	X	X	11	18
SIMPROCESS, SIMSCRIPT	79	26	96	50 000	20 000	5	19,0	X	X	X		X				X		9	18
WITNESS	108	36	106	900 000	10 000	6	20,2		X	X	X	X	X			X		7	18
Extend	150	38	68	999 999	700	6	19,0	X		X		X				X	X	7	18
Simul8 (Brooks Automation - Aut	48	6	96	30 000	350	6	14,8	X	X			X	X		X	X	X	11	18
QUEST; DPM POWERTRAIN; R	104	19	105	20 000	10 000	6	18,5			X		X	X		X	X		5	17
Crystal Ball	30	8	300	100 000	*20 000*	6	17,8	X									X	4	17
Flexsim (Taylor II, até 2002)	12	3	47	20 000	1 000	6	12,5	X	X			X			X	X	X	10	17
GPSS/H + SLX	103	44	16	15 000	1 700	5	17,0					X					X	4	17
ProcessModel	11	11	44	10 000	*10 000*	6	14,9	X		X								4	16
iGrafx (2005)	4	4	400	25 000	*25 000*	7	15,2			X		X			X			3	16
Micro Saint	40	1	49	15 000	600	6	11,8	X				X				X	X	7	16
AnyLogic (eXperimental obJect T	4	1	115	10 000	500	6	10,4	X							X	X	X	7	15
Enterprise Dynamics - (Taylor ED	17	5	2	900	400	3	9,5	X		X	X	X	X				X	7	15
eM-Plant (ex:SIMPLE++)	46	2	3	15 000	30	4	9,0	X			X	X			X			6	15
GPSS World for Windows	3	3	11	10 000	300	4	9,2	X	X	X								6	15
TaskResourceSIM,DistributionSI	11	10	16	150	150	5	11,1	X								X		3	15
Visual Simulation Environment	14	8	6	700	150	4	10,5	X			X							3	14

Figure 8 Table from 2006 Ranking (Dias et al. 2006)

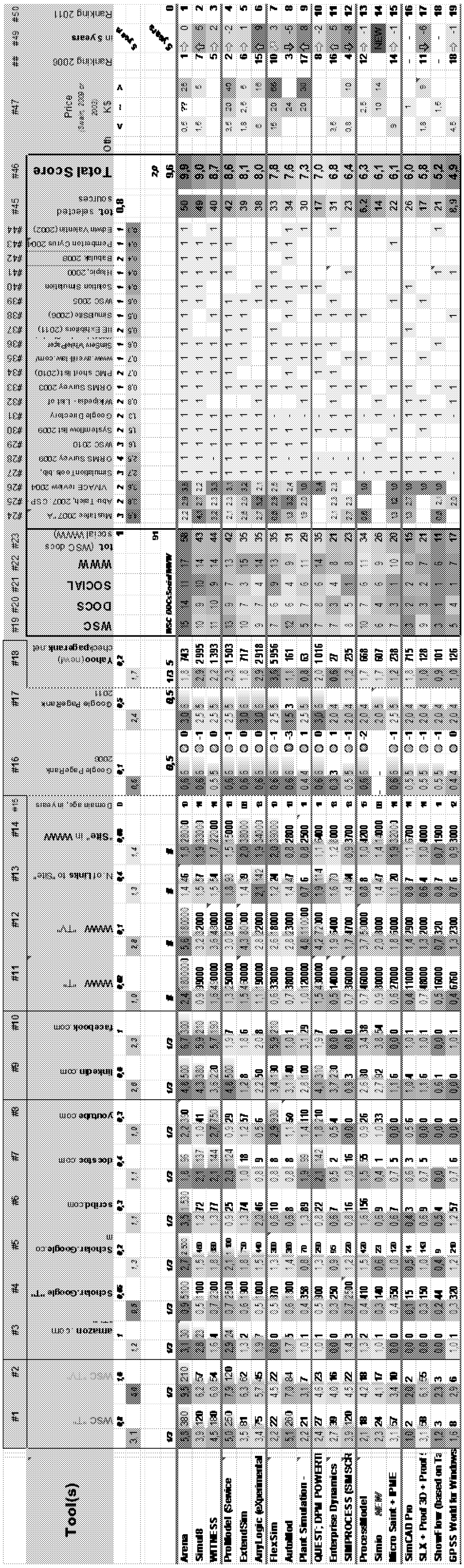


Figure 9 Full Table With 2011 Tools Comparison

THE SIMSED FRAMEWORK FOR MODELLING AND SIMULATION OF TRANSITIC SYSTEMS UNDER UNCERTAIN ENVIRONMENT

Romain Bévan
Jean Louis Lallican
SYDEL
ZI de Keryado
Rue du Gaillec – BP 2934
56312 Lorient Cedex, France
romain.bevan, jean-louis.lallican
@sydel.fr

Willy Allègre
Pascal Berruet
Lab-STICC
Université de Bretagne-Sud
Centre de Recherche – BP 92116
56321 Lorient Cedex, France
willy.allegre, pascal.berruet
@univ-ubs.fr

KEYWORDS
Design, simulation, transitic systems, dynamic engine, virtual environment, hazards of functioning, scenario.

ABSTRACT

This paper presents a virtual reality environment used for the design and the simulation of transitic systems (conveying systems). This framework, called SimSED, enables testing and validating the control code of the system before on-site implementation. The virtual reality environment provides a tool for representing three-dimensionally a transitic system, which encapsulates the realistic behaviour of the system including failure occurrence. The 3D animation improves graphical display and the use of a physical engine called ODE provides a technical solution for realistic and precise material part behaviour. The control code is validated by joint simulation of the transitic system's material part and control part. The tool enables the control behaviour to be tested in case of material breakdown. SimSED has been used to design and to simulate a complete industrial application. This framework has then been validated by tests.

INTRODUCTION

Within the framework of production systems, as well as the design of systems, designers try to find a correct solution as quickly as possible, which meets the requirements in terms of production time, robustness, flexibility … They frequently test the automation of their equipment directly on-site before the commissioning of facilities. This testing phase, which comes very late in the design process, is inadequate (Lallican 2007). In fact, we need to ensure, at the earliest design phase of the control system, that no faults leading to non compliance with safety constraints and requirements set by the specifications were committed.

One solution consists in using synthesis (Ramadge & Wonham 1987, Hietter et al. 2008) that guarantees the obtained solution. As this kind of work is promising, it has to overcome the major difficulty of scalability for common industrial uses. Another solution consists in checking the solution after elaboration. Simulation is largely used for this in industry. Its major advantage is that it is quite easy to carry out compared to proof techniques. This is particularly true for analyzing complex systems that cannot be easily modelled using mathematical or analytical models. The drawback is that the system is only validated according to a set of scenarios. Thus scenarios may include failure occurrences in order to take degraded situations into account.

The objective of the presented work is to ensure that the control is adequately designed and to test the impact of a failure occurrence on the control code. In order to reduce the number of experiments carried out on the real system that may induce material damages, this phase should be performed before on site implementation. Essentially due to a time to market reduction, the time for testing may be short in industrial project. We propose to use the simulation of control code coupled to a virtual material part of the system and have therefore developed a tool in collaboration with the Sydel company to help its engineers in this task.
As the systems evolve in an uncertain environment, it has to be noticed that we need to test the system in both normal and degraded conditions. Indeed, the simulation has to take into account component wear present in a system and the various issues that may affect their proper functioning. The remark is important for the requirements of the future testing tool.

The paper is organised as follows: the next section summarises the state of the art in terms of simulation and tools for control code simulation. This both enables to establish our contribution and to deduce some interesting requirements for the simulation tool. Section 3 presents the environment, how it interacts in a global design flow. Some specific features, that enable the capturing of some uncertainties of the environment, are delineated in section 4 and an industrial application is presented before discussion and conclusion.

STATE OF THE ART

Simulation is classically divided into three categories: static simulation, continuous simulation, and discrete event simulation.
Static or Monte Carlo simulation enables solving stochastic problems without needing explicit time representation (Page Jr 1994). The last two kinds of simulations concern more dynamic systems.

In continuous simulation, state variables evolve with time without any interruption. In discrete event simulation they change according to event occurrences (Ray & Claramunt 2003).

In manufacturing systems, simulation software tools can be classified into two kinds: discrete event simulation and geometric simulation (Klingstam & Gullander 1999).

Discrete event simulation (also called flow simulation) is suitable for an analyzing system and its performances. This kind of simulation often expresses flows and has the advantage of rapidly providing results based on large simulated periods. But as the control behaviour is embedded in the model, it is not relevant for control code testing. Arena, Extend, Cadence and Quest are examples of flow oriented simulators. Others are based on simulation languages such as Simula, Simscript.

In contrast, geometric simulation simulates the geometry of a part, or the whole manufacturing system. Geometric simulation often refers to continuous simulation. Generally this kind of simulation allows for the testing of the control code of the system. System evolution can be displayed through two-dimension or three-dimension techniques. An example of a two-dimension display simulator is ControlBuild (Gros et al. 2006). This tool uses an interesting reuse-based approach, which consists in reusing components (elementary building blocks) present in a library, to improve the quality and productivity of a complex systems design. This software is configurable (cycle time management, actuator configuration, user settings) unfortunately it does not take the interactions between actuators into account.
Three-dimension display simulation applications are as follows: Virtual Factory (Wenbin et al. 2002), Robotic (Ju-Yeon et al. 1997) and ITS PLC (Real Games Lda. 2008). In Robotic applications, simulation often concerns a small part of the system. ITS PLC tool uses a quite realistic environment that can be directly connected with an industrial PLC, but the application is currently limited to five existing systems and not configurable (no cycle time management, only sixteen inputs / outputs).

Previous works have led to a virtual reality environment for the design and simulation of transitic systems (conveying systems). This framework called SimSED (Simulation of Discrete Event Systems) is the result of a joint effort by the Lab-STICC laboratory (ex LESTER laboratory) and Sydel, a company specialized in process engineering and automation, for fourteen years.

The tool developed in Java enables the modelling of the operative part. It is based on a component-based approach (association of elements that can be parameterised) which increases the productivity and the reliability of the design. The system simulator used is a three-dimension display simulator. Operating part simulation is performed simultaneously with control code simulation, taking into account the PLC cycle. Products and parcels are simulated as individual entities, which allow for a precise simulation taking into account collision problems (J. L. Lallican et al. 2005).
As the prototype was used by the Sydel company, some possible adjustments were pointed out, such as for time simulation that remained relatively long for complex systems, or for the copy/past mechanism that was not easy to use. Moreover, the library had to include more elements, such as elevators which were missing to represent more complex systems, only normal behaviour was possible and the user could not interact with the system (parcels, actuators, sensors).
The simulator has thus been improved to take into account these new requirements.

SIMSED FRAMEWORK

This section first recalls the flow and the simulation principle.

Global flow

The global process is part of a usual flow based on simulation to validate or modify the design parameters. It also integrates a component-based approach facilitating and increasing the design process. Simulation concerns both operating and control parts, the control program being associated with the operating part.

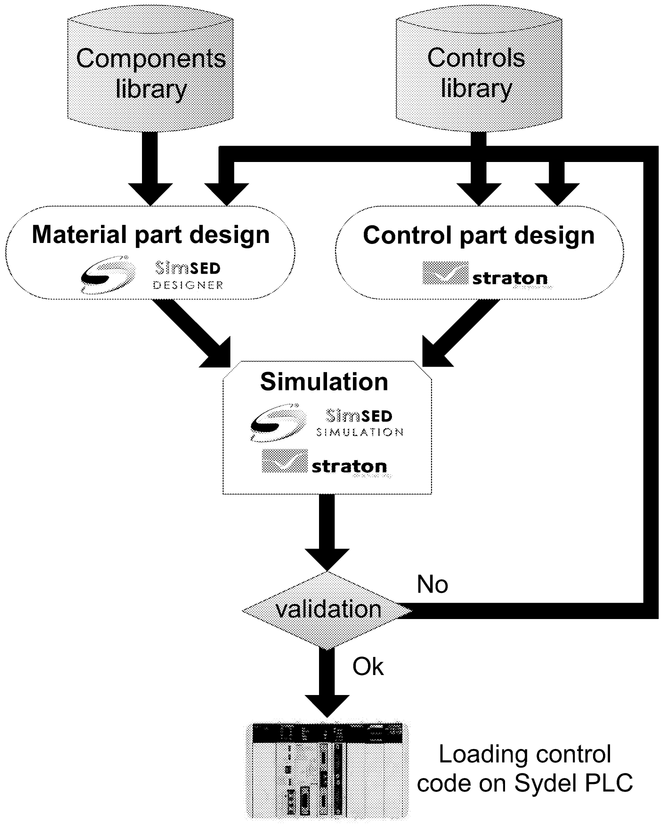


Figure 1: Global process design

Figure 1 illustrates the global process composed of three steps: the material part design, the control part design and the simulation.

The material and control part designs are performed using libraries. After validation, the control program can be loaded in a PLC. If simulation does not fit the requirements, the operating part and/or the control part are modified to achieve objectives.

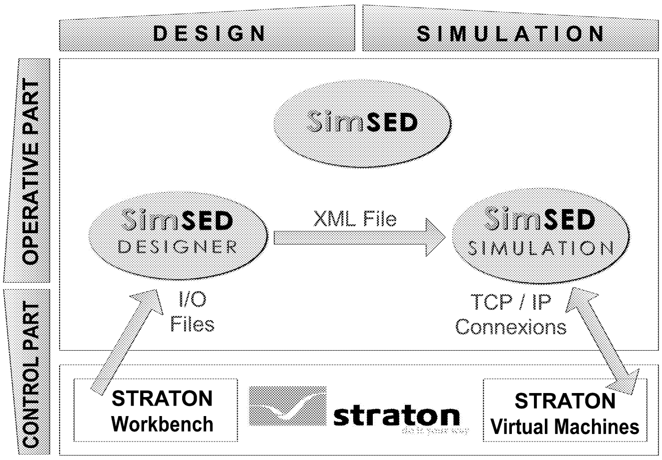


Figure 2: SimSED Framework

As depicted by figure 2, the SimSED framework is composed of two tools: SimSED DESIGNER for the material design part and SimSED SIMULATION for the simulation. As for the control part, it is written with an IEC 61131-3 standard-compliant software: STRATON.

The following sections introduce the different parts of SimSED framework: the operating part, the control part and the simulation.

Operating part design

The material part design is performed using the SimSED DESIGNER tool.
A component-based model approach has been adopted to provide an easy way to reuse previously modelled elements. The complete model of a system is seen as an assembling of components. All components include parameters: static parameters such as position, orientation in the 3D environment and dynamic parameters, for example the speed for a motor. Components are stored in a library. This software is seen as an ergonomic interface for the 3D simulator. To design a system it is sufficient to select components from the library and to parameterize them according to system features.

Control part design

The control part is written using a software compatible with the IEC 61131-3 standard. STRATON software is used to write the different controls and to download them to a virtual machine for the simulation (Copalp 2004).
The control part implementation is realized with a STRATON Workbench and its simulation uses a STRATON Virtual Machine. The interest is to simulate and test controls that will be really implemented, without any transcription. This tool has been chosen because our partner, the Sydel Company, has developed a PLC based on a Vx Works operating system and a STRATON Virtual Machine.

Analysis using joint simulation

Dynamic analysis enables the study of the behaviour that the system is going to have in operation, in order to validate the control part of the system and to check whether the chosen parameters will enable the obtaining of the requested performances.
The validation of the control part is performed using simulation of control code coupled with operating part one. The dedicated simulator is called SimSED SIMULATION.
This method uses continuous simulation respecting the PLC cycle. Synchronization management between the two kinds of software is dedicated to SimSED SIMULATION. It is also meant to control STRATON simulation execution.

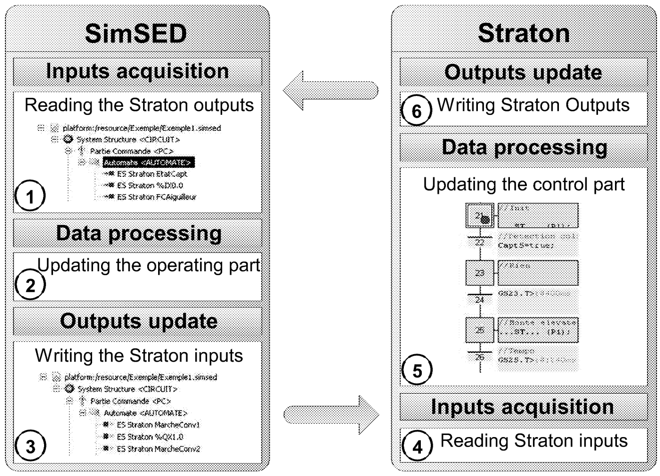


Figure 3: Description of the simulation cycle

Figure 3 described the simulation cycle that is divided into 2 parts. The goal of the first part, carried out by SimSED, is first to memorize STRATON output values. Then components execute one simulation cycle. After STRATON inputs are updated. The second part details the evolution of the control program according to the new values of these inputs. When the control program simulation cycle is performed, STRATON outputs are updated.

SimSED Features

To be relevant, a simulation has to be as close as possible to the real system. As the control is checked, it can be seen as a parameter of the other part of the system that is the operating part. To provide a realistic behaviour of the material part, SimSED SIMULATION (Figure 4), developed in C++ language, integrates a physics engine called Open Dynamic Engine (Smith 2008). This open source library, also written in C++ language, enables the simulation of rigid body dynamics. It has advanced joint types and integrated collision detection with friction. Problems like critical speed or acceleration, low sensor tolerance, and parcel collision can be pointed out.

Animation is also an important feature. 3D animation helps to visualize clearly the behaviour of the simulated workshop and emphasizes the understanding of the system. Coupling the 3D environment with the physics engine ODE guarantees 3D realistic graphics and real-time physics that closely report what happens in real life. The proposed 3D animation enables the designer to zoom in on a specific part of the workshop to watch it in detail, or to zoom out for overall review. It is also possible to follow an object moving around the virtual system. Designers can also change viewpoints as desired. All these features enable easier detection of critical points. Finally, SimSED SIMULATION provides a clear interface which allows for cycle time management, STRATON Virtual Machine connexions for centralised or distributed architecture connexions and user settings for a perfect control of simulation parameters.

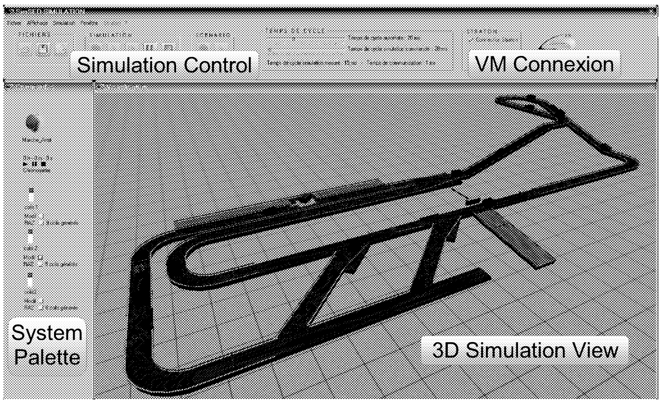


Figure 4: SimSED SIMULATION interface

IMPROVEMENTS

The prototype used by Sydel pointed out some previously cited weaknesses. SimSED has thus been improved to take into account these new requirements.
The former editor SimSED DESIGNER was initially developed, "by hand", in Java. With hindsight, we realize that it was adapted to its time but nowadays, it lacks in scalability and convenience that recent technologies could allow it to acquire.
To capitalize on our expertise and ensure the continued existence of our product, we have followed a model-driven engineering (MDE) approach. This methodology focuses on creating models, or abstractions, closer to our domain concepts than computing (or algorithmic) concepts. Based on the principle that our business logic is stable, we have made our business models to represent our systems regardless of implementation platform.
The Eclipse Modelling Framework (EMF) (Budinsky et al. 2003) is a modelling framework for Eclipse. EMF includes a metametamodel Ecore, close to the MOF metametamodel, for describing (meta) models and runtime support for the models. After transcription of our business model as a business metamodel within EMF which conforms to Ecore, we have generated a complete product (application-specific branding) for operating part design.
The new SimSED DESIGNER provides a clear and user-friendly interface with native operations of Eclipse, such as the "Copy/Paste" functionality, thus improving the system design.
The complete model of a system is seen as an assembling of components (Berruet et al. 2001). All components include parameters providing adaptability to different designs. Several components of level L-1 can be grouped into an aggregate component of level L. As components have the same structure at any abstraction level, the hierarchical components can be easily stored and reused for future workshop designs.

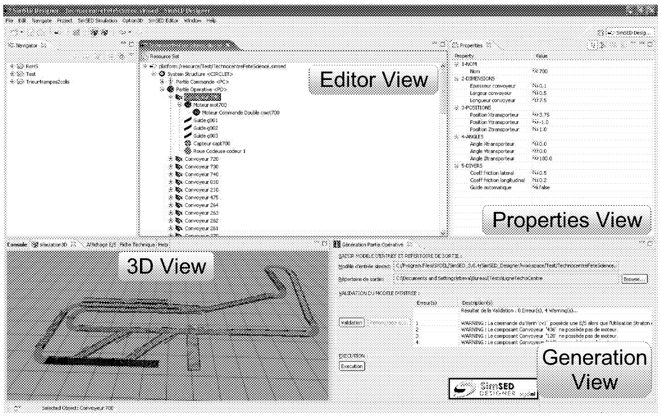


Figure 5: SimSED DESIGNER interface

SimSED DESIGNER modular interface (Figure 5) is composed as follows:

- The *Editor View*, generated by EMF from our business metamodel, offers a hierarchical view for the design of a transitic system, compliant to our requirements through its construction.
- The *Properties View*, related to the *Editor View*, enables the configuration of the components (static parameters, dynamic parameters and input/output connexions of PLC's).
- The *Generation View* forwards the validation of the transitic system model and the generation of the simulation file (XML file conforms to SimSED SIMULATION metamodel) using model transformations described with Atlas Transformation Language (ATL) (Jouault & Kurtev 2006).
- The *3D View*, developed with Java OpenGL-bindings (JOGL), displays the transitic system edited in the *Editing View*. This view is particularly useful for the designer-user to check the correct positioning of the components.

To simplify component settings, a parent / child structure (Figure 6) has been adopted and two abstraction levels have been defined. Each level provides its own reference. There are two parent components: the "normal" conveyor and the turn one. The other components, such as jacks, sensors, and stoppers are considered as children because they are necessarily associated with a parent component. The parent components are arranged according to site level (absolute reference of the workshop), whereas the child components are arranged according to their parent component levels. This structure makes the modelling of the workshop easier.

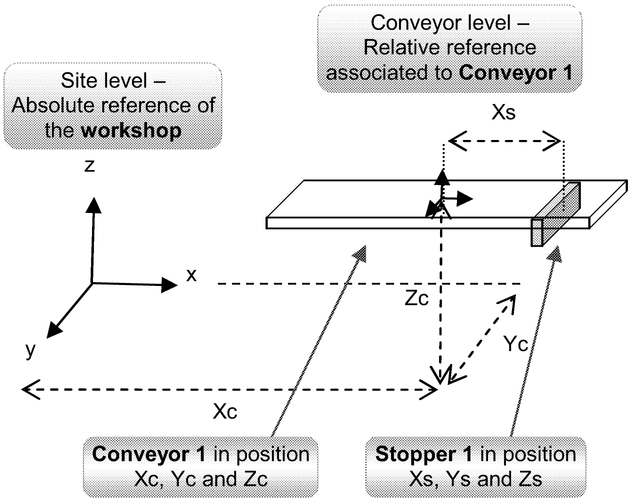


Figure 6: Parent / Child structure

The library was initially constituted of two different types of components: the support component and the basic component. A support component, such as a conveyor, is composed of basic components, such as jacks, sensors, and stoppers. To represent more complex systems, the library had to include more missing elements. For this purpose, the library has been enriched. Among the additions, the elevator has enabled the introduction of a new component typology: the handling component. Thus, a support component can be composed of both basic components and handling components to address a broader array of complex systems.

HAZARD INTEGRATION / HANDLING

Simulation, the goal of which is to get as close as possible to reality, must take into account unexpected events related to the wear or malfunctioning of components in a transitic system.

Indeed, after several months or years, a component may fail or become faulty. This material defect, called internal hazards, may first have important implications on the production line, until sometimes stopping it. This translates into a potentially large economic impact for the company concerned. In this case, the chain is forced to stop until a maintenance team arrives on site.

Thus, these material defects should be considered for the design of the control part, as well as the operative part, which is why we propose to integrate and handle hazards within the simulation.

Hazard integration

Simulation considers an abstraction of the physical system, on which the command control code is tested, and the functioning of the operative part can be visualized. To get as close as possible to reality, the transitic system model integrates hazards (Figure 7).

Each component is associated with one or many types of hazards, defined accordingly to the needs and feedback of control engineer experts. It is almost always elementary components situated on a conveyor, such as a stopper, a jack, a motor or a sensor. Two types of hazard are defined in this work: there is "blocking at a given value" for the sensor, whereas there is "speed reduction" for the others. The list of components concerned by hazards, as well as the nature of the latter, can be expanded to meet specific needs.

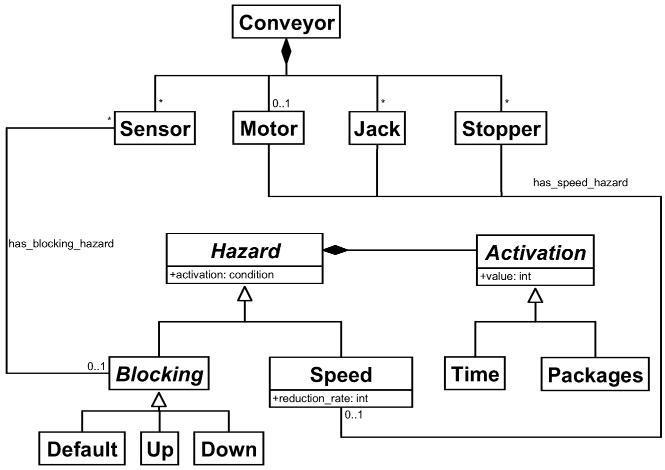


Figure 7: Class Diagram of components affected by hazards

Hazard integration in SimSED SIMULATION is done through an XML script. The latter can be edited by the user, handling an adapted GUI before the simulation. At this level, the user can:

- Select one component in which hazards have to be integrated. It is possible to let the system choose randomly between all instances that are on the site.
- Set hazards according to the type of component. This choice can also be done randomly.
- Select and set the hazard activation condition: it can be time (simulated or PLC cycle time) or the number of generated parcels. Once again, a pseudo random value (bounded value) associated with this condition can be generated.

At each simulation step, the program checks and updates component properties from the physics engine (e.g. collision test). Furthermore, PLC commands are checked and updated at each controller cycle step. Time and generated parcel numbers are tested for all components that are subject to a defect. When the event occurs, the physical properties of the component are changed according to the nature of the hazard that has to be simulated. The faulty component flashes black and white to give visual feedback to the user.

Finally, the integration of hazards allows users to realize that there might be errors in the control part of their system, and thus change the PLC program accordingly. But the static integration alone is not enough: the possibility to add/remove hazards dynamically during the simulation and the possibility to play back scenarios are essential features to obtain a more fault tolerant system.

Handling hazards

Recording / Playback scenarios

Once the hazard has been taken into account in the PLC program (e.g. output jack on a tempo rather than sensor information when the later failed), the user must be able to replay the same scenario, which has been previously recorded, to test the modified program.

Recording a complete scenario requires significant memory resources. The backup of all physical parameters (related to ODE physics engine) associated to all components at each execution step is too costly. The video capture meets neither the joint criteria for CPU occupancy, nor for video quality. Thus scenarios are based on an XML structure containing specifications (time, size, references, etc.) of the generated parcels sequences, as well as previously programmed hazard specifications (component, time, and parcel).

SimSED SIMULATION, through the PLC program, reacts logically in the same way to an identical sequence of hazards and generated parcels. The user can, in this case, test and change his program recursively until obtaining one that can fit his requirements.

Dynamic interaction with the environment
OpenGL, an open-source graphics library, is widely used to design applications that produce 3D images. It uses representations of projective geometry to avoid any situation involving the infinite. A picking function allows one to detect which objects are below the mouse or in a square region of the OpenGL window. It is implemented through the use of the OpenGL selection mode. This method is retained in this work because it offers a fast and efficient hardware selection management as managed by the GPU (Graphics Processing Unit).
From this picking function, features of hazard management have been developed. The user may, during the simulation, choose to remove the programmed hazard component, or on the contrary, to add some at runtime in order to do a robust system control. But conveyor components are not the only ones to cause disturbances on the functioning of the system: parcels can sometimes block and cause an untimely pile. Picking allows one to move or remove the latter without restarting the simulation from the beginning.
Beyond the application to hazards, interaction with the 3D environment is very useful for moving parcels, as well as manual and intuitive control components. This makes it possible to move towards more real conditions in the testing phase that can be later verified on the prototype or site.

ILLUSTRATION

The basics of our approach have been validated by comparing a real system, which exists at Sydel, to its simulated one on SimSED (J. L. Lallican et al. 2005). The behaviour provided by simulation was similar to the behaviour of the real system, controlled by the same control code. The principle of control code validation using the 3D virtual simulation provided correct and promising results.
Different huge complex systems were simulated using the new simulator version. The new enhancements of SimSED make it possible to take into account operating hazards, such as sensors and actuators malfunctioning which cause error situations or system jams. We can thus test, improve and validate our control code to be as robust and effective as possible before the commissioning of facilities, in order to prevent risks of human injury or machine damage.

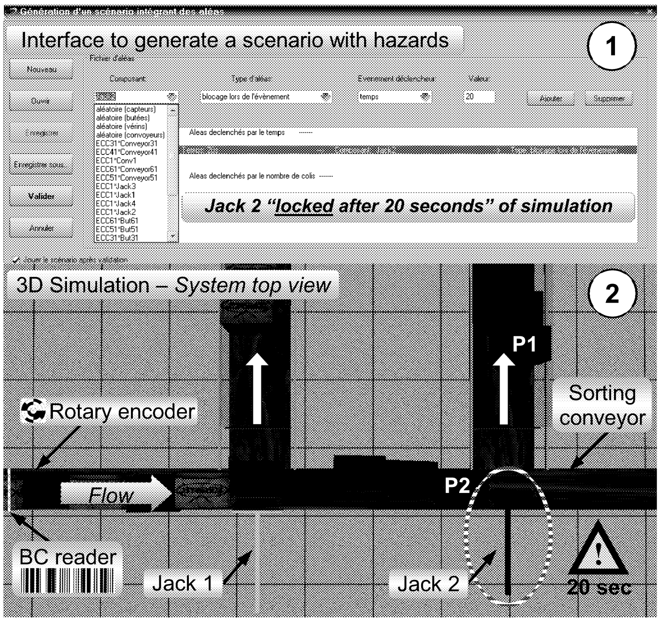


Figure 8: Interface to generate a scenario with hazards (1) and simulation results on a sorting conveyor example (2)

Figure 8 illustrates, on a sorting conveyor example, the use of the generation with hazards. A sorting conveyor is often deployed to sort products (parcels) by size or by bar code. In our case, the sorting conveyor is composed of a main motorized conveyor, a rotary encoder placed on the main conveyor motor, a bar-code reader (BC reader), two jacks, and two outputs conveyors (ramps).
When operating normally, the upstream conveyor spaces out the products and sends them to the main conveyor from which they are scanned by the BC reader. Then the PLC connected to the system begins to measure the distance covered by the parcel with the help of the rotary encoder. It then decides to perform a transfer with a jack when the product is in front of its output conveyor.
Figure 8 (1) presents the interface to generate a scenario with hazards. The scenario we have decided to play is to lock *Jack2* in its current position after twenty seconds of simulation ("blocking at a given value" hazard). The result, Figure 8 (2), is that after twenty seconds of simulation, *Jack 2* which just transferred product *P1*, locks while it was coming back (jammed halfway). Visually, the three-dimensional representation of J*ack2* flashes in black and white: this makes it possible to easily locate the component affected by malfunctioning in a huge system. The *Jack2* problem leads to the jamming of the system because product *P2* abuts on it and then blocks the product flow. It can be noticed that our control code is clearly not adapted to this situation: we should, for example, transfer all the new products onto the first ramp through *Jack 1* or at worst, stop the main conveyor to avoid a build up and an amplification of this blocking problem.
During the simulation, picking can allow us to unlock *Jack2* and to delete the desired parcels in order to unblock the system. However, that does not solve our control problem. Therefore, SimSED offers the possibility to replay this

scenario an unlimited amount of times. We can thus adapt and refine our control code and retest it to improve system operations and to avoid this blocking situation.

SimSED is also used for the development / enhancement of the control program of *Technocentre Sydel*. *Technocentre Sydel* is a platform for tests and customer demonstrations of the expertise of the company. This system has, among other elements, a sorting conveyor, a collecting conveyor, and a converging conveyor. Figure 5 presents the design part on SimSED DESIGNER. Figure 4 shows the global view of the *Technocentre* on SimSED SIMULATION and Figure 9 offers comparative views of the real system and the simulated one. The simulation results closely emulate what happens in a real system. We were able to refine our control code to obtain a quality code checked by performance indicators present in SimSED (like flowmeters including graphical representations) without any physical damages.

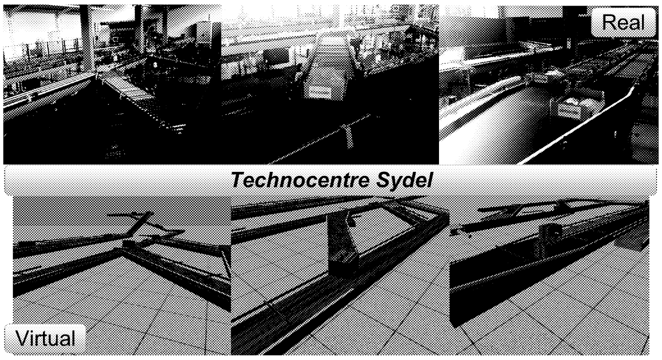


Figure 9: Comparative views of *Technocentre Sydel* system

DISCUSSION AND FUTURE WORKS

A framework for testing and validating control programs of transitic systems using operative part simulation has been presented. A same program developed using the virtual environment can be used on a real one. The 3D environment and the use of a physical engine guarantee a technical solution to obtain realistic and precise material part behaviour that closely emulates what happens in real life. The new enhancements of the framework integrate operating hazards (failure occurrences) and scenarios for refining, improving, and "robustifying" control programs and thereby reducing the risks of damage during on-site implementation. These additional features are also used after implementation to recreate and then solve critical situations which happen in real systems. It should also be noted that the tool has a significant educational aspect. In fact, students can learn and practice PLC programming without any risks of human injury or machine damage.
Future works are planned to comply to the framework with new hardware targets, such as Siemens or Schneider PLCs in order to expand the range of action of the framework.

Finally, further works focus on obtaining a reconfigurable control program with the operating part design. For this, the description of the system is based on components that include several views with partial models. Promising results for nominal control program generation from a conveying system modelling have been obtained. We are currently working on the generation of multiple versions of control code to include at least a nominal and a degraded control program by component.
Thus, we will be able to test the system response upon the occurrence of a failure, such as a sensor or a jack becoming faulty. The control code will switch between control versions (normal/degraded). The SimSED interface to generate a scenario with hazards is suitable for this purpose and it will enable us to introduce these types of hazards that can be replayed as desired as scenarios (Cf. Section 4 *Hazard integration / Handling*).
We will also be able to test the reliability of the control code against the insertion/suppression of parcels by human operators. Again, the SimSED framework and its picking functionality will be particularly relevant to carry out these tests and validating our works.

REFERENCES

Berruet, P. et al., 2001. Modeling and validation of transitic systems based on reusable components. In WMSCI 2001, IIIS Conference. Orlando, Florida, USA, pp. 286-291.
Budinsky, F. et al., 2003. *Eclipse Modeling Framework*, Addison-Wesley Professional, 1st edition.
Copalp, 2004. STRATON handbook.
Gros, R. et al., 2006. Fault monitoring & fault tolerant control in Controlbuild software platform. In *6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, SafeProcess 2006. Beijing, China, 641-645.
Hietter, Y., Roussel, J. & Lesage, J., 2008. Algebraic synthesis of dependable logic controllers. In *17th IFAC World Congress*. Seoul, Korea, p. 4132-4137.
Jouault, F. & Kurtev, I., 2006. Transforming models with ATL. In *Satellite Events at the MoDELS 2005 Conference*. Montego Bay, Jamaica, 128–138.
Ju-Yeon, J. et al., 1997. Virtual Testing Of Agile Manufacturing Software Using 3D Graphical Simulation. In *proceedings of the 1997 IEEE International Conference on Robotics and Automation.* New Mexico. p. 1223–1228.
Klingstam, P. & Gullander, P., 1999. Overview of simulation tools for computer-aided production engineering. *Computers in Industry*, 38(2), 173-186.
Lallican, J.L., Berruet, P. & Philippe, J.L., 2005. SimSED: a tool for modeling and Simulating Transitic Systems. In *International Mediteranean Modeling Multiconference,Conceptual Modeling and Simulation Conference*, I3M-CMS'05. Marseille, France.
Lallican, J.L., 2007. *Proposition d'une approche composant pour la conception de la commande des systèmes transitiques*. Thèse d'Etat. [S.l.]: Université de Bretagne-Sud.
Page Jr, E.H., 1994. *Simulation modeling methodology: principles and etiology of decision support*. Virginia Polytechnic Institute and State University.
Ramadge, P.J. & Wonham, W.M., 1987. Supervisory Control of a Class of Discrete Event Processes. *SIAM Journal on Control and Optimization*, 25(1), 206.
Ray, C. & Claramunt, C., 2003. A distributed computing system for the simulation of disaggregated data flows. *Knowledge-Based Systems*, 16(4), 191–203.
Real Games Lda., 2008. ITS PLC Professional Edition User Guide.
Smith, R., 2008. Open Dynamics Engine User Guide.
Wenbin, Z. et al., 2002. An integrated simulation method to support virtual factory engineering. *International Journal of CAD/CAM*, 2(1).

A SIMULATION PACKAGE FOR AN ENERGY-AWARE COMPARISON OF ARQ PROTOCOLS

Gian-Luca Dei Rossi Andrea Marin Matteo Rosati
Simonetta Balsamo
Università Ca' Foscari di Venezia, Dipartimento di Scienze Ambientali, Informatica e Statistica
Via Torino 155, Venezia
email: {deirossi,marin,mrosati,balsamo}@dsi.unive.it

KEYWORDS
Simulation,Computer Networks, Energy Management, Go-Back-N

ABSTRACT

In this paper we consider the ARQ Go-Back-N protocol, using both simulation and theoretical analysis in order to derive some performance indices, like the throughput and the energy efficiency, given some channel error distributions. We then propose a heuristic for the optimisation of energy consumption when errors are strongly correlated, and compare our approach with other solutions in literature.

INTRODUCTION

Automatic Repeat Request (ARQ) protocols are widely used as methods to detect and correct errors in data transmission through the use of acknowledgements and timeouts in order to achieve reliability on a intrinsically unreliable channel. Three basic types of ARQ protocols are traditionally used, namely Stop-and-Wait (SAW), Go-Back-N (GBN) and Selective Repeat (SR). Although those protocols are widely analysed in literature, the recent widespread use of wireless networks and mobile agents has posed new challenges to implementers of those protocols, especially regarding energy consumption minimisation and performances in very error-prone channels.
In this paper we focus on the performance evaluation of the Go-Back-N ARQ protocol, and in particular on two performance indices, i.e., the throughput and the energy efficiency, proposing an heuristic to increase the energy efficiency of the protocol implementations and using theoretical analysis and simulation to support our claims.

Related Works In Chockalingam and Zorzi (2008) an approach to energy minimisation in Go-Back-N ARQ protocol is proposed. Although the aim of the paper is similar to ours, the proposed heuristic and its performances and properties are really different, as can be seen in the comparison made in the rest of the paper.
In Chlamtac et al. (1998) some energy minimisation techniques for Go-Back-N protocols are proposed. However, the approach of the authors differs from ours in the fact that they actually change the communication protocol in order to achieve better results. Altering the protocol requires that both parties in the communication are aware of those modification, i.e., their protocol implementation has to be modified in order to interoperate with the other. In our approach, only the sender implementation has to be changed, and, because the protocol is left untouched, it can communicate with every other party in the system.

Paper Structure The paper is structured as follows: first we recall some background notions needed to maintain the paper self-contained, then we introduce an analytical Semi-Markov model for Go-Back-N protocols. In the next section we present a simulator created in order to do experiments with Go-Back-N performances, and then we propose an heuristic for energy efficiency optimization in those kind of protocols. In the conclusive section we analyse the results of the simulation and we compare our approach with another heuristic proposed in the literature. Some final remarks are then presented.

THEORETICAL BACKGROUND

In this section we recall some well-known theoretical results that are needed in order to understand the remaining part of the paper.

The Go-Back-N Protocol

Go-Back-N (GBN) is a sliding window protocol that uses an Automatic Repeat Request (ARQ) mechanism. In this kind of protocol, each packet (or *frame*) transmitted has an associated sequence number (SN). The receiver memorizes in a variable the next expected SN for an incoming packet. If a packet has the right number, the variable is incremented and an acknowledgement with the SN of the received packet is sent back to the sender. If a received packet has a SN that is not expected, the receiver simply discards the packet and sends back an

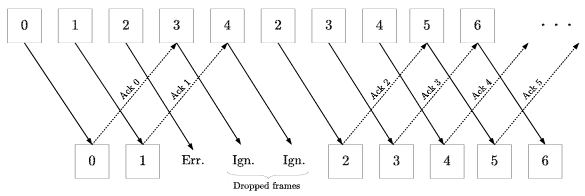


Figure 1: Go-Back-N Protocol Example.

acknowledgement with the SN of the last correctly received packet.

The sender, on the other hand, maintains a sequence number for the next packet to be sent S_n, the sequence number of the last acknowledged packet S_a and a *window* size N. A packet can be sent if $S_n - S_a < N$. At each packet transmission the value of S_n is updated, and at each acknowledgement reception if its sequence number $S_r > S_a$, S_a is update accordingly.

The protocol behaviour is illustrated in Figure 1.

A more detailed description of Go-Back-N and of other ARQ protocols can be found in Tanenbaum (2002).

Renewal Reward Processes

A renewal process is a generalization of a Poisson process where the times between consecutive events are not necessarily exponentially distributed. Formally, let $\{N(t), t \geq 0\}$ be a counting process, and let X_n be the non-negative random variable that denotes the time between events $(n-1)$ and n, then, if the random variables in the sequence $\{X_1, X_2, \ldots\}$ are independent and identically distributed (IID), the counting process is said to be a *renewal process*. Whenever an event occurs, the time between that and the next event is independent from the time between the previous ones and it has the same probability distribution. We can say that at each event a *renewal* occurs.

Let us consider a renewal process $\{N(t), t \geq 0\}$ as described before, and suppose that whenever a renewal occurs, i.e., an event happens, there is a *reward*. We denote by R_n, $n \geq 1$ the reward for the nth renewal, and we assume that each reward is independent and identically distributed. An usual choice for the reward R_n is the time X_n between the renewal n and the previous one. Let $R(t)$ be the total reward received by time t, then:

$$R(t) = \sum_{n=1}^{N(t)} R_n \ . \tag{1}$$

A detailed description of Renewal Reward Processes can be found in Ross (2006).

Semi-Markov Processes Let a stochastic process have N states $1, 2, \ldots, N$. Whenever the process enters in state i it remains in the state for a random amount of time with mean μ_i and then has a transition to state j with probability p_{ij}. Such a process is called a *Semi-Markov Process*). Let X_n be the state of the process after the nth transition, then $\{X_n, n \geq 0\}$ is a Discrete-Time Markov Chain (DTMC) with transition probability matrix $\mathbf{P} = [p_{ij}]$ with $i, j \in \{1, \ldots, N\}$, i.e., the same transition probabilities of the original Semi-Markov process.

Let $\boldsymbol{\pi}$ be the stationary probability vector for the DTMC, i.e., the solution for the linear system

$$\begin{cases} \boldsymbol{\pi}\mathbf{P} & = & \boldsymbol{\pi} \ , \\ \boldsymbol{\pi}\mathbf{1} & = & 1 \ , \end{cases} \tag{2}$$

then the stationary probabilities for the Semi-Markov Processes can be computed using the formula:

$$P_i = \frac{\pi_i \mu_i}{\sum_{j=1}^{N} \pi_j \mu_j} \ , \tag{3}$$

where π_i is the ith component of $\boldsymbol{\pi}$, i.e. the stationary probability of being in the state i of the DTMC.

THEORETICAL MODEL

Model Assumptions

According to the analysis method for the performances of the Go-Back-N ARQ Protocol proposed in Chockalingam and Zorzi (2008), we consider a model for the sender behaviour, using the following assumption:

- There are two agents in the system, namely *sender* S and a *receiver* R;
- S and R communicates using a noisy channel C, whose bandwidth is b and propagation delay is d;
- The sender S is always ready to send packets, unless the window is full;
- The packets have size s, and therefore the time to send a single packet is $f = \frac{s}{b}$;
- The window of the protocol is fixed, and has the optimal size $N = \frac{f+2d}{f}$;
- Acknowledgements have a negligible size (modelled as size 0), and are never lost;
- The time-out t for the reception of an acknowledgement has to be at least equal to round-trip time plus the time to send a packet, then $t > 2d+f$, i.e., $t = 2d + f + \varepsilon$ with $\varepsilon > 0$. However, in order to maintain efficiency, ε should be as small as possible, and therefore in our model we will assume that $t = 2d + f$;

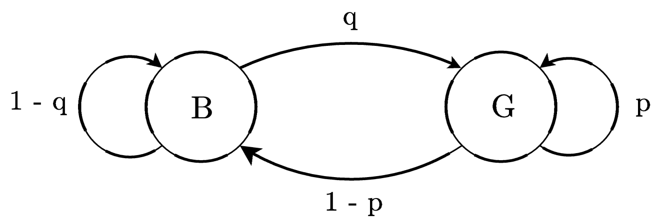


Figure 2: Channel State Markov Chain.

- At every packet transmission, if the channel is in a good state, i.e., there is no error, it remains in this state with probability p or it switches to bad state with probability $1 - p$.
- Whenever the channel is in bad state, it returns to good state with probability q or it remains in bad state with probability $1 - q$.

The last two assumptions characterise the behaviour of errors in the channel at each packet transmission, and could be represented by a DTMC as in Figure 2, where the state B represents the bad (error) state, and the state G the good (absence of error) one, and whose transition probability matrix P_e is

$$\begin{pmatrix} 1-q & q \\ 1-p & p \end{pmatrix} . \tag{4}$$

The stationary probability distribution for this chain, computed using Equation (2) , is given by:

$$\pi(B) = \frac{1-p}{1-p+q} \qquad \pi(G) = \frac{q}{1-p+q} . \tag{5}$$

A model with such assumptions is clearly limited in its scope, but it is analytically tractable and useful for the validation of the simulator.

Semi-Markov Model

Given the above assumption, we can easily build a Semi-Markov Process that models the behaviour of the sender, where each state $1 \dots N$ represents how many packets are waiting to be acknowledged in the window. Moreover, there is a state E that represents an error condition detected by the sender. The resulting process is shown in Figure 3. Transitions are labelled using the following schema: first there is the transition probability, second is the reward, third is the number of packets sent, and last is the number of packets that reach the destination correctly and are acknowledged.
Note that the probabilities of going from the error state E to the state 2 and to remain in the state E are respectively q^* and $1 - q^*$, and not q and $1 - q$. The reason is that we defined q as the probability of returning to a good state after a single packet transmission, i.e., after a time f, but in the model we have to consider the probability that the channel is in a good state after a timeout t is elapsed, i.e., after N attempts to send a packet were done. Given the transition matrix P_e defined by the Equation (4), if the one-step-probability q of remaining in state B starting from the same state is $P_e(1, 1)$, then the probability of being in the state B starting from B after N steps is

$$q^* = P_e^N(1, 1) . \tag{6}$$

Notice also that the state 1 is transient, because, after the very beginning of the communication, it will never be visited again, therefore it can be omitted, as could be seen in Figure 4.

Lumping The model just described could be used directly to analyse some performance indices of the Go-Back-N protocol, like the throughput and the energy efficiency, but it is complex and the number of its states depends on the window size. It is possible to further simplify the underlying stochastic process using a technique called *lumping* (see Kemeny and Snell (1976)) that reduces the space state of a Markov Chain partitioning the process according to some constraints. In Figure 5 is shown the lumped Semi-Markov model, where all good states $2, \dots, N$ are collapsed into one, named T (for *transmitting*), i.e., the state in which the transmission activity of the sender is effective. The transition matrix for the model is given by

$$\begin{pmatrix} 1-q^* & q^* \\ 1-p & p \end{pmatrix} , \tag{7}$$

and, using Equation (2), we can compute the stationary probabilities, that are:

$$\pi(E) = \frac{1-p}{1-p+q^*} \qquad \pi(T) = \frac{q^*}{1-p+q^*} . \tag{8}$$

Using Equation 3 combined with Equation (8) we can compute the stationary probability P_T of being in the transmitting state T, i.e., the fraction of time used to transmit data on the whole elapsed time:

$$P_T = \frac{f(\pi(E)q^* + \pi(T)p)}{\pi(E)[fq^* + t(1-q^*)] + \pi(T)[fp + t(1-p)]} . \tag{9}$$

When the protocol doesn't use any mechanism in order to avoid unnecessary transmission, and all of them requires, on average, the same amount of energy, this probability is also what is called the *energy efficiency* of the protocol, i.e., the fraction of transmission that are effective in communicating data to the receiver:

$$E_{\text{eff}} = P_T . \tag{10}$$

From the energy efficiency is also possible to derive the *throughput* of the system, i.e., how many packets are

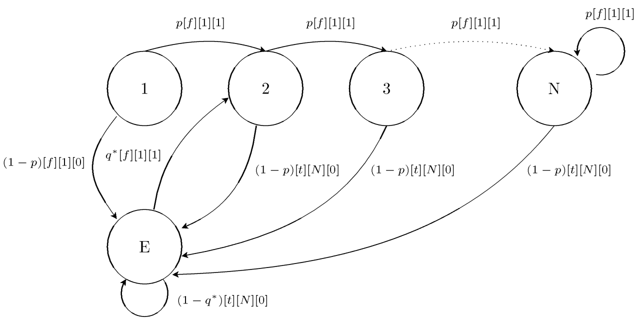


Figure 3: Semi-Markov Model for Go-Back-N, transient state 1 included.

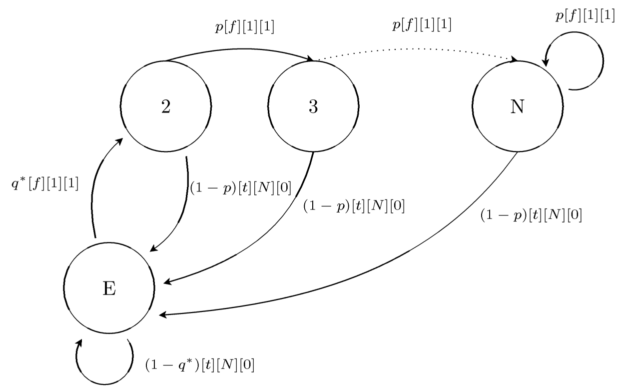


Figure 4: Semi-Markov Model for Go-Back-N without transient states.

effectively sent, and thus received on the other end of the channel, in the chosen unit of time. As stated in the model assumptions, the system is continuously trying to send packets on the channel whenever it is ready, i.e., if the time to send a packet is f, in the chosen unit of time $1/f$ packet transmissions are attempted. Of those packets, only a fraction E_{eff} is effectively transmitted, and then the throughput is:

$$\text{Th} = \frac{1}{f} E_{\text{eff}}. \tag{11}$$

Error Correlation

In our analytical model for the state of the channel, the presence of two different probabilities p and q for the transition to the state for the next transmission, as seen in Figure 2, influences the distribution of errors over time. Suppose that the error distribution is correlation-free: then $\Pr\{X_{n+1} = B | X_n = B\} = \Pr\{X_{n+1} = B\}$, i.e., the probability that, in the next transmission, the process would be in the error state is the same whether the process is already in an error state or not. In our model, this condition is satisfied only when $p = q$. In

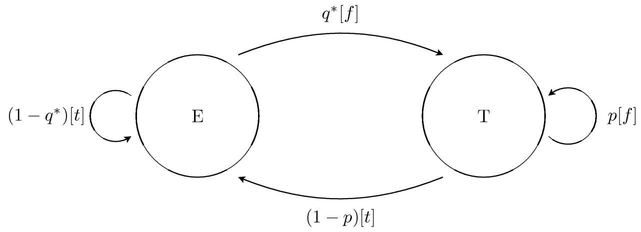


Figure 5: Lumping of the Semi-Markov Model for Go-Back-N.

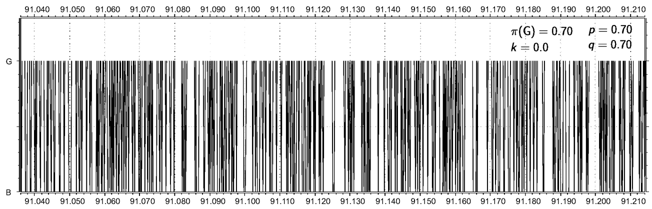


Figure 6: Uncorrelated errors in the channel.

all the other cases we have a *correlation* of errors over time.

Analysing the model, it would be useful to express the error distribution on the channel in terms of the overall probability of being in a good state $\pi(G)$ and a correlation parameter k. After some algebraic manipulation of Equation (5), imposing the constraint that $k = 0$ when $p = q$, we could rewrite p and q as:

$$p = (1 - \pi(G))k + \pi(G) \qquad q = \frac{\pi(G)(p-1)}{\pi(G) - 1} \ . \tag{12}$$

In Figure 6 the error distribution in a channel without correlation is shown: errors are independent and uniformly distributed. In figure 7 a channel with strong error correlation but with the same stationary probabilities is depicted: the errors tend to concentrate in some regions, leaving the channel in a uninterrupted good state for longer intervals of time.

Notice that, for $k = 1$, the DTMC for errors is no more ergodic, because $p = 1$ and $q = 0$, independently from the desired $\pi(G)$. Moreover, in our study we will not consider negative correlation parameters, so we will assume that $k \in [0, 1)$.

SIMULATOR DESCRIPTION

In this section we introduce our simulator for Go-Back-N ARQ protocols and how it has been used in order to achieve some experimental results. Our software is based on the OMNeT++ framework, described in Varga (2010), that provides object-oriented libraries to write discrete-event simulators focusing on the domain logic of the problem.

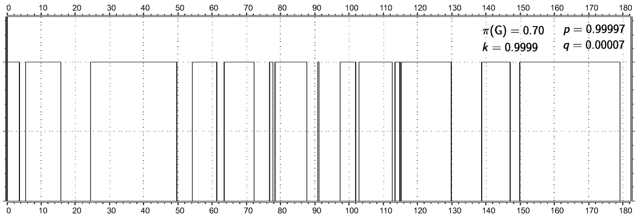


Figure 7: Strongly correlated errors in the channel.

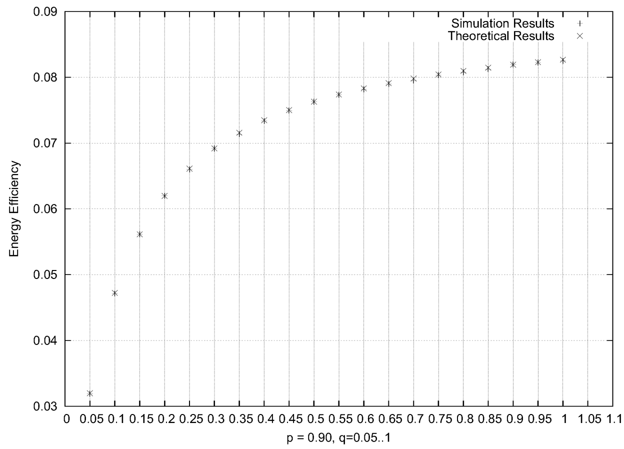


Figure 8: Energy Efficiency for correlated errors: theoretical vs. simulation results.

Our simulator models a packet sender, a receiver and a channel between them. The simulation parameters are:

- The maximum bandwidth of the channel, in our examples it is assumed to be 54Mbps, like in a 802.11g wireless network);
- The delay of the channel, i.e., how much time is spent between the transmission of the first bit of a message and its reception on the other side of the channel;
- The size of the packets transmitted;
- The timeout delay;
- The window size, N;
- The probability p of correctly transmitting a frame, given that the previous one was correctly transmitted;
- The probability q of correctly transmitting a frame, given that the previous one was lost.

The simulator behaves as the Go-Back-N protocol prescribes for sender and receiver. The channel state at each packet transmission is modelled according to the chain of Figure 2.

For the pseudo-random number generation we have chosen a Mersenne Twister RNG (see Matsumoto and Nishimura (1998)). In order to eliminate the effects of initial transient in the steady state simulation, we use the techniques described in Welch (1981) and thus discarding the results of the simulator during the *warmup* time, estimated using pilot runs.

Validation

In order to validate our simulator we compared its results with the ones obtained analytically using the model described in the previous sections. Clearly, the application scope of the simulator is much broader than the one of the analytical model, e.g., it is possible to observe the behaviour of the system when the window size is not optimal or when acknowledgements could be lost, but nevertheless a validation is a useful aid to verify the correctness of a simulator.

The validation process is complex and lengthy, thus in this section we will only show some highlights of the tests that were actually done. In Figure 8 we can see a comparison between simulation and theoretical results for the energy efficiency of a system parametrised with a bandwidth of 54Mbps, a packet size of 1492 bytes, a delay of 11ms, a timeout of 22.3ms and a window of size 100 when the correlation parameter k varies. All the experimental measurements shown are actually the mean of the energy efficiency values for 100 runs with a simulated time interval of 300s, without resetting the RNG seed. Those 100 measurements were then considered to be IID, and mean, variance and confidence intervals for $c = 0.95$ were computed. It was verified that the theoretical results obtained from the model were always inside the aforementioned confidence intervals, whose maximum width was less than $3 \cdot 10^{-4}$.

It should be noted that the same procedure is used for every validation test that we have done, e.g., the throughput and the energy efficiency for different correlation indexes, different stationary probabilities for the error state, different delay times (and thus different timeouts and optimal window sizes).

A HEURISTIC FOR ENERGY MINIMISATION

In this section we present a novel heuristic for the minimisation of the energy consumption of the sender, i.e., for the maximisation of the energy efficiency. Our main observation here is that the knowledge of the error correlation parameter can lead to better energy efficiency without degrading performance sensibly.

In Figure 7 we have seen that strong correlation parameters lead to error concentration in some time intervals. This is quite realistic, thinking of typical sources of noise or signal loss in wireless networks, e.g., an engine in its starting phase or an agent entering and going through a tunnel. In those cases, trying to send packets continuously until an acknowledgement is received is not an optimal strategy. On the other hand, stopping transmissions for a pre-determined time or slowing the communication reverting to a Stop-And-Wait behaviour doesn't exploit the knowledge of the error correlation parameter and leads to poor performances. An optimal behaviour should be skipping the transmission that, given the probability distribution of errors, are likely to be ineffective, and resume normal operations whenever there is an high probability to be in a good state.

In order to understand how this behaviour could be achieved, we have to consider the model of Figure 2 and recall that in a DTMC the residence time in a state is modelled as a geometrically distributed random variable X. Given the fact that the process is in the state B (that models the bad channel state), the probability that the process will exit the state *after* the kth step is:

$$\Pr\{X = k\} = q(1-q)^k \ . \tag{13}$$

We recall also that the expected value for that random variable is:

$$E[X] = \frac{1}{q} \ . \tag{14}$$

This result means that, on average, if the channel is detected to be in a bad state, it will remain in that state for the next q^{-1} steps. An ideal heuristic for Go-Back-N, then, should be, whenever a packet is lost, to skip the next q^{-1} transmission, i.e., for a time fq^{-1}. The problem with that solution is that the sender is not able to directly and instantaneously detect packet losses. The error detection is done only *a posteriori*, when a timeout has occurred. Under the assumption of optimality of the window size and of the timeout value, when a packet loss is detected, other N packets have already been sent. It is clear that it is not possible to avoid those transmissions, but if the error correlation parameter is high, it is likely that q^{-1} is higher than N, and thus some energy could be saved avoiding to send the remaining packets until the first value is reached.

We can model the amount of packet transmission slots that should be skipped using the variable s:

$$s = \max\left(\left\lfloor \frac{1}{q} \right\rfloor - N, 0\right) . \tag{15}$$

If the windows size is not optimal, we cannot assume that, waiting the timeout, exactly N packets are sent, thus we have to use a more general formula, where the number of already sent packets is tf^{-1}:

$$s = \max\left(\left\lfloor \frac{1}{q} \right\rfloor - \frac{t}{f}, 0\right) . \tag{16}$$

Using the results in Equation (15) or (16) we can derive a simple heuristic for optimising the energy efficiency of the Go-Back-N protocol, i.e., whenever the sender is in a sending state, and a timeout occurs, the next s transmission could be skipped.

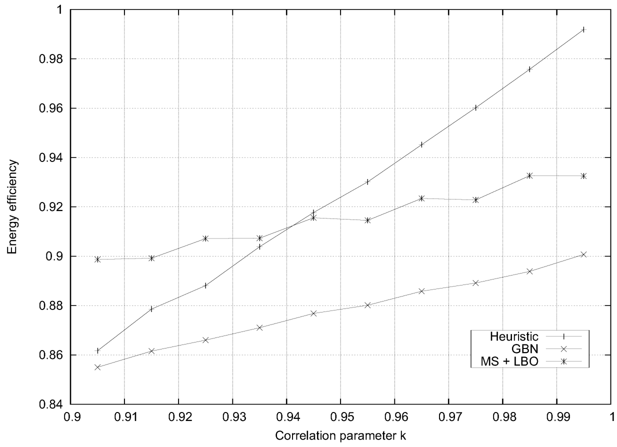


Figure 9: Comparison between energy efficiencies.

Simulation Results

Using the simulator that we have built, we can now analyse the behaviour of the protocol when it uses our proposed heuristic and compare the performances with the plain Go-Back-N and the heuristic proposed in Chockalingam and Zorzi (2008). All the simulations were done adopting the same criteria used for validation (Initial transient elimination and repeated runs). In this section we briefly present the salient results of our experiments, using the set of parameters shown in Table 1.

In Figure 9 is shown a comparison between the energy efficiency of our proposed heuristic, the one proposed in Chockalingam and Zorzi (2008) (labeled MS+LBO) and a naive implementation of Go-Back-N (labeled GBN). We can see that the naive implementation has the worst efficiency, while the MS+LBO approach behaves well when the correlation is not too strong. Our heuristic performs as the naive GBN when the correlation is low, but has an significant advantage over the other approaches when the correlation is strong.

In Figure 10, on the other hand, is shown a comparison between the throughputs of the three approaches. In this case the better results are achieved by the naive GBN, because the protocol continuously attempts to send packets. Our approach has a lower throughput, that decreases slightly as the correlation parameter increases. The MS+LBO approach has the worst performances, but they get slightly betters with high error correlation.

Note that, for $k > 0.94$, our heuristic has both a better energy efficiency and a better throughput than the MS+LBO approach.

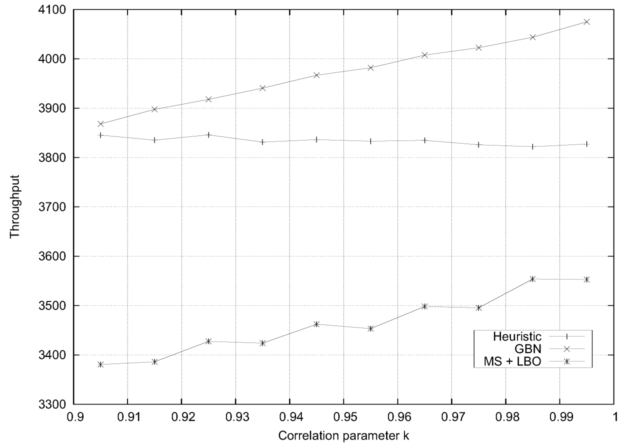


Figure 10: Comparison between throughputs.

Parameter	Value
Bandwidth	54Mbps
Delay	0.99ms
Frame size	1492 B
Timeout	2.21ms
$\pi(G)$	0.9

Table 1: Parameter values for the simulation.

CONCLUSION

In this paper, we have presented a simulation package to analyse energy efficiency and other performance indices for ARQ protocols, focusing on Go-Back-N and its variants. We have then proposed a novel heuristic for energy efficiency optimisation, using the results provided by a theoretical model.
Using the simulation package that we have developed, we have analysed and compared the energy efficiency and the throughput of a system adopting our heuristic, pointing out the advantages of our approach in the strong error correlation case.

Future Works

Future developments of this research could be manifold. From a theoretical point of view, an extension of the analytical model to cases not currently tractable, e.g., when the window size is not optimal or when acknowledgement can be lost, would be desirable. From the tool development perspective, the simulator could be extended in order to allow a more precise description of errors behaviour, e.g., specifying signal fading and other physical characteristics.
The proposed heuristic could be modified in order to achieve better results when the correlation index is low, maybe combining techniques already used in other approaches.

For the implementation of the heuristic in real devices, techniques to efficiently guess values for $\pi(G)$ and k from statistics on the previous behaviour of the channel should be considered.

REFERENCES

Chlamtac I.; Petrioli C.; and Redi J., 1998. *Energy-conserving go-back-N ARQ protocols for wireless data networks.* In *Universal Personal Communications, 1998. ICUPC'98. IEEE 1998 International Conference on.* IEEE, vol. 2. ISBN 0780351061, 1259–1263.

Chockalingam A. and Zorzi M., 2008. *Adaptive ARQ with energy efficient backoff on Markov fading links. Wireless Communications, IEEE Transactions on*, 7, no. 5, 1445–1449. ISSN 1536-1276.

Kemeny J.G. and Snell J.L., 1976. *Finite Markov Chains.* Springer, second ed. ISBN 978-0-387-90192-3.

Matsumoto M. and Nishimura T., 1998. *Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 8, no. 1, 3–30. ISSN 1049-3301.

Ross S.M., 2006. *Introduction to Probability Models.* Academic Press, ninth ed. ISBN 978-0-125-98062-3.

Tanenbaum A., 2002. *Computer Networks.* Prentice Hall Professional Technical Reference, 4th ed. ISBN 0130661023.

Varga A., 2010. *OMNeT++.* In K. Wehrle; M. Günes; and J. Gross (Eds.), *Modeling and Tools for Network Simulation*, Springer. ISBN 978-3-642-12330-6.

Welch P.D., 1981. *On the problem of the initial transient in steady-state simulations.* Tech. rep., IBM Watson Research Center, Yorktown Heights, NY.

BIOGRAPHY

SIMONETTA BALSAMO is a full professor of Computer Science at the University Ca' Foscari of Venice, Italy. Her research interests include performance and reliability modelling and analysis of computer and communication systems, parallel and distributed processing, distributed simulation, quantitative analysis of software architectures and integration of specification languages and performance models. She has published several papers in international journals, conference proceedings, books and special editions. has served as general chair, program chair and program committee member for several international conferences, is associated editor of Performance Evaluation Journal.

GIAN-LUCA DEI ROSSI received his M.Sc. degree in Computer Science from the University of Venice in 2010. He is currently a Ph.D. student at the same University. His research area is on stochastic modelling and performance evaluation.

ANDREA MARIN received his degree in Computer Science from the University of Venice in 2002, and the Ph.D. in Computer Science from the same University in 2009. Most of his research is focused on the analysis of compositionality of Markovian stochastic models, in particular in stochastic Petri net domain. Some novel results about product-form solutions have been derived in this context. He now works as a post-doctoral researcher at the University of Venice

MATTEO ROSATI received his B.Sc. degree in Computer Science from the University of Venice in 2011. He is currently a M.Sc. student at the same University.

A TOOL TO VISUALIZE RESPONSE SELECTION AND EXECUTION ON PACE RATING

Peeradaech Suwittayaruk
Dirk Van Goubergen
Ghent University
Technologiepark 903
B-9052 Ghent-Zwijnaarde, Belgium
E-mail: peeradaech.suwittayaruk@ugent.be,
dirk@vangoubergen.com

KEYWORDS
Pace rating, a visualizing tool, response selection, response execution, decision making pattern.

ABSTRACT

Pace rating has always been one of the main responsibilities for industrial engineers (IEs) when manufacturing companies need standard times to use for planning, scheduling, cost efficiency and budget calculations. Although computer technologies have been applied to model, analyze, and visualize complex systems to help engineers solve specific issues, a simulation or visualization tool on pace rating that reveals and visualizes decision making patterns has as yet not been developed to help IEs perform worker's pace rating. This paper presents the development of a tool that can visualize response selection and execution, showing decision making patterns and help IEs to perform pace rating.

What is pace rating?

Pace rating is a kind of performance rating. Performance rating is defined as the process of rating work's performance by comparing the actual task being performed and evaluating the observed performance quantitatively with the industrial engineer's concept of normal performance (Aft 2000). Performance rating methods can be categorized in 6 types (Barnes 1980) consisting of (1)Skill and effort rating or the Bedaux system, (2)Westinghouse system of rating, (3)Synthetic rating, (4) Objective rating, (5) Physiological evaluation of performance level, and (6)Pace rating. Barnes (1980) also reports an investigation that 90% of 72 US companies at that time used the pace rating system in time study practices while the rest used others. Among the different ways, pace rating is considered to be the fastest method only the operator's speed of movements is rated but it still has some unresolved issues (see further). Pace rating is also a part of the objective rating method (Mundel and Danner 1994). In this method speed of motions is evaluated together with six factors of job difficulty containing the amount of body used, foot pedals, ambidexterity, eye-hand coordination, handing requirement and weight are then rated with their respective tables of percentages.

Going through the most popular literature on work measurement, we observe most authors write about pace rating in terms of "the most difficult and controversial step"(Groover 2007), "the most important step" (Niebel and Freivalds 2003), "the most challenging aspect"(Meyers and stewart 2002),"two most controversial aspects" (Kanawaty 1992), "the most important and the most difficult part" (Barnes 1980), "very difficult part", and "the most subjective aspect" (Miller and Schmidt 1990). Pace rating has always been treated as difficult, subjective and even controversial. This research is investigating ways to overcome these issues.

Pace rating (PR) is an important element to compute standard time. Direct time study is the most widely spread work measurement technique. It is used when manufacturing companies need standard times with accuracy and high confidence. The standard time is composed of two elements as in Equation(1) below.

Standard time = normal time + allowances (1)

Normal time = observed time x (PR/100) (2)

Allowances are additional times to compensate for non optimal work circumstances while normal time can be calculated as in Equation(2) above.

The fundamental purpose of determining PR is to normalize or adjust the mean observed time for each element being performed in order to become the normal time that is applicable to every worker. Standard times are used for planning, scheduling, cost efficiency and budget calculations as well as for process design and improvement. Reliable standard times are needed for all activities. The quality of the standard time is considerably dependant on the quality of the pace rating process.

Rating scales are needed in order to have an effective way for quantifying the observed pace of working in comparison with the standard pace. There are several pace rating scales in use, the most frequently used scales are (1) Scale A - 100 percent equals normal performance or 100-133 MTM scale, (2) Scale B - 60 points equals normal performance, (3)Scale C - 125 percent equals incentive performance, and (4) Scale D - 100 percent equals incentive performance. The 100-133 MTM scale is easy to use and to understand as the 100% pace corresponds to walking 3 miles per hour (Barnes 1980) or dealing 52 cards into four equal stacks around 30"x30" in 0.5 minute.(Nadler 1955; Watmough 1975; Barnes 1980;

Mundel and Danner 1994; Meyers and Stewart 2002; Niebel and Freivalds 2003; Groover 2007)

Pace rating in traditional way

Traditional direct time study is carried out by using a stopwatch or other timekeeping device to time and record a task to obtain an observed time. To normalize the observed time in order to obtain a normal time, pace rating is typically done in the traditional way. Eventually, allowances are added for converting the normal time into the final standard time. Figure 1 illustrates a model showing pace rating in the traditional way. To perform pace rating, industrial engineers (IEs) are trained to remember a speed of motion at 100% standard pace like dealing 52 cards into four equal stacks. Vision, one of the five basic human senses, is then used in the process of pace rating on an actual work method. The IE sees a worker performing a task at a workstation. Next, this information is processed in the brain by comparing the actual work rate with the 100% standard pace. Finally, the rated pace is estimated and documented.

In the traditional pace rating way, three body parts of the IE are involved: eyes, the brain and hands. The most significant step is the information processing in the brain, including hidden factors that affect response selection and execution. Until now, this has never been studied in detail. The whole process can be considered as a black box which, until now, never has been modeled, simulated or visualized in pace rating. Hence, the traditional pace rating method can be considered to be very subjective and most challenging.

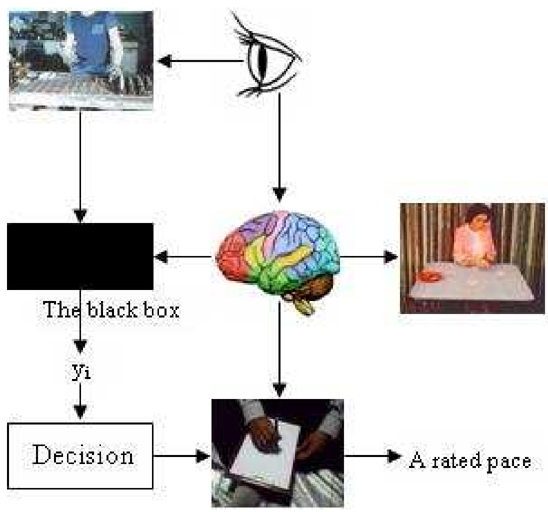


Figure 1: Pace rating in the traditional way

Pace rating using video technology

Based on video technology, Van Goubergen and Vancauwenberghe (2006) proposed a new alternative method for pace rating of work being performed. As shown in Figure 2, an actual video of the task being studied (the right side) and a reference video calibrated at 100% pace (the left side) are displayed at the same time. By adjusting the speed or pace of the reference video, IEs can synchronize and match the motion patterns in both videos and hence, quantify the actual pace of the method under study. Therefore, it appears that video technology can help IEs to rate the real pace of a worker performing a task. However, only the basic idea of synchronizing videos has been presented in literature. A tool that visualizes response selection and execution, with particularly more on decision making patterns for pace rating, needs to be developed as part of the validation of this new pace rating method.

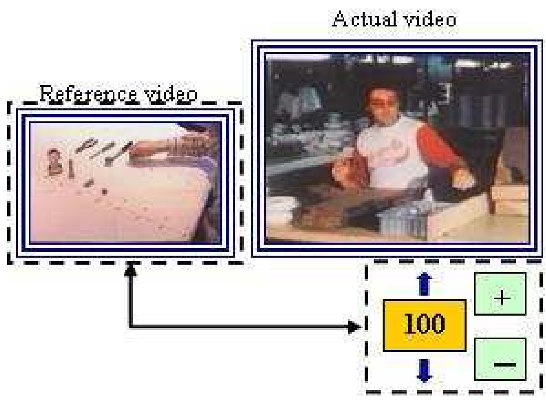


Figure 2: Pace rating using video technology (Van Goubergen and Vancauwenberghe 2006)

The need for a visualizing tool

Since computer and simulation technologies have been used to model, analyze and visualize complex systems, many simulation and visualization packages have been developed as tools to help engineers and decision makers solve specific issues in both manufacturing and service systems. The results of simulation based on certain inputs and operational characteristics can be different alternatives available to the decision makers. However, no simulation or visualization package on pace rating has been developed yet, that reveals and visualizes the decision making pattern that are involved as a tool to help IEs better to determine worker's pace rating. This paper presents a tool that can visualize response selection and execution, showing decision making patterns and help IEs to perform pace rating.

HOW TO MODEL AND VISUALIZE RESPONSE SELECTION AND EXECUTION ON PACE RATING

Figure 3 illustrates the basic idea to model and visualize response selection and execution on pace rating. Our idea is associated with observing events of a system as they evolve over time. Each event influences the state of a system at discrete points in time.

As can be seen from Figure 3, the variables we defined are :
e_i = each event
s_1 = tempo of RV (Reference Video) is less than AV (Actual Video)
s_2 = tempo of RV is more than AV
s_3 = tempo of RV is equal AV

a_1 = clicking the ⇑ button
a_2 = clicking the ⇓ button
a_3 = clicking the stop button
t_i = clock time recoded on e_i
ET_i = execution time
TET = total execution time
ET_i and TET can be calculated by
$ET_i = t_i - t_{i-1}$
$TET = \Sigma\, ET_i$

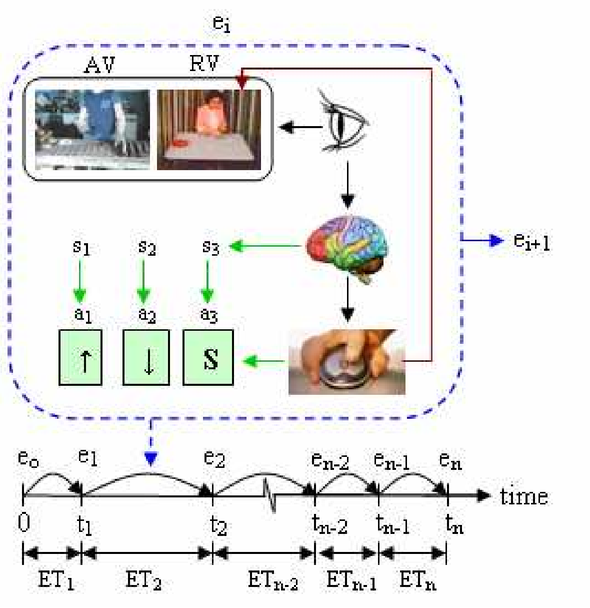


Figure 3: How to model and visualize response selection and execution on pace rating

An event is defined as a state when eyes, the brain and a hand are 'connected' to perform the process of pace rating. Initially, eyes look at the actual video (AV) and the reference video (RV). The final response required is s_3 that means a_3 ('STOP') is activated. To obtain a_3, however, we rate the process under study by comparing the actual video and the reference video, starting at a significantly different tempo. In our research, we set the starting point of the reference video at a minimum tempo difference corresponding to 20 % and a maximum tempo difference at 180%. The time elapsed between two clicks (time factor) is recorded in order to compute and visualize the data series of rated paces. After clicking the mouse, this causes the system state to change instantaneously. The desired outcome of a pace rating method is a figure indicating the actual pace of a work method. In case of the newly proposed method, this corresponds to the speed of the video when the stop button (a_3) is pushed.

The concept of discrete-event dynamic systems was chosen to model and visualize the decision making patterns, as updating the clock time and recording the events is done at discrete moments in time. In this first case, Figure 4 shows an event graph that displays a behavior of visual response selection and execution on pace rating with three zones : a_1, a_2 and a_3 activated. The time is plotted along the X-axis and pace level is indicated on the Y-axis.

In contrast, if the initial point of the reference video is set at a maximum tempo difference of 180 %, the decision making pattern is generated in the opposite way as displayed in Figure 5.

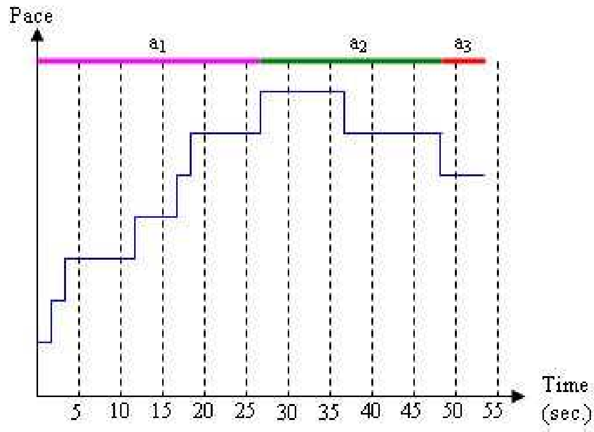


Figure 4: The event graph of a behavior of response selection and execution on pace rating (min)

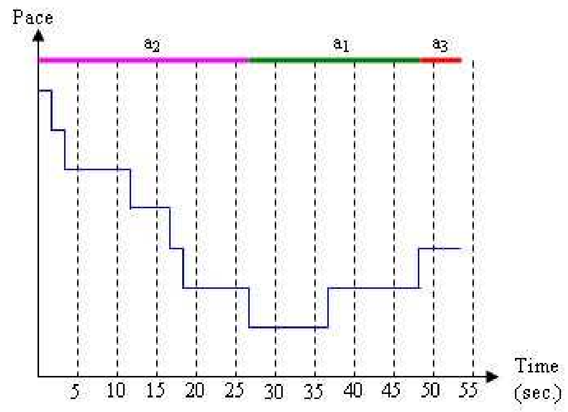


Figure 5: The event graph of a behavior of response selection and execution on pace rating (max)

The basic idea on which the software is based, is how to connect response selection and each execution time. Adjusting the speed of the reference video was programmed so it can be changed with 5 % increments and 5 % decrements. Figure 5 shows a screenshot of the software developed using JAVA programming, suitable for generating decision making patterns.

Figure 6: Screenshot of the visualization tool for pace rating

CASE APPLICATION

In order to visualize response selection and execution on pace rating based on comparing two videos, a reference video, analyzed and normalized with MTM-1 at 100% pace, was chosen to be used in the experiment as displayed in Figure 7 (the right side). MTM-1 is a predetermined time standards technique that provides the highest accuracy compared to other similar systems. It gives a normal time based on predetermined time values at pace 100% for fundamental motion such as : reaching, moving, turning, apply pressure, grasping, positioning, releasing, disengaging, eye movements, and body movement (Karger&Bayha 1987; Salvendy 2001).

Figure 7: A case application

The actual video (the left side of Figure 7) showing hand movements rated at 100% based on the traditional pace rating way was selected from the TMI pace rating videos collection (Watmough 1975). For conducting the experiments, a master student in Industrial Engineering and Operations Research at Ghent University was selected as a participant.

The following instructions were given to the participant for conducting the experiment: Reference and actual videos are shown at the same time without giving any quantitative information on the pace. The participant is asked to compare both videos with regard to pace of working. As a starting point, the video under study is shown at a lower pace (20%). When the pace is perceived by the participant as being lower, the ⇑ button (a_1) needs to be clicked. The speed of the reference video will now increase. The same question is asked repeatedly until the pace is judged as being higher. The ⇓ button (a_2) can be then clicked until the participant is satisfied with the resemblance. Finally, the stop button (a_3) needs to be pushed in order to end the experiment. After that, the two videos are shown again but the starting point of the video under study is shown at a maximum pace (180%). The instruction to perform the experiment is given in procedure similar way.

After the experiment, we obtain two decision making patterns as depicted in Figure 8 and 9. These graphs depict the evolution of the perceived pace in function of time, visualizing the phenomenon of pace rating in the human brain of the participant.

Alternatively, Figure 10 displays two decision making patterns so they can be compared and analyzed at the same time. Two vertical lines drawn on the chart divide the pattern into two important zones: an easy decision zone and a difficult decision zone. The left part is the easy decision zone: as apparently the difference in pace between the two videos is significant, the human being will click relatively fast to approach a value that is closer in the range of the real pace value (in this experiment 100%). From this point on we observe a more ‘difficult’ decision zone: both ⇑ and ⇓ buttons are used and it takes longer for the human being to decide to click a button. Analyzing the decision making patterns shows that especially in the range of the real pace it is more difficult to rate. Based on the two final responses obtained : 95% and 100%, the average value(97.5%) can be used as the rated pace.

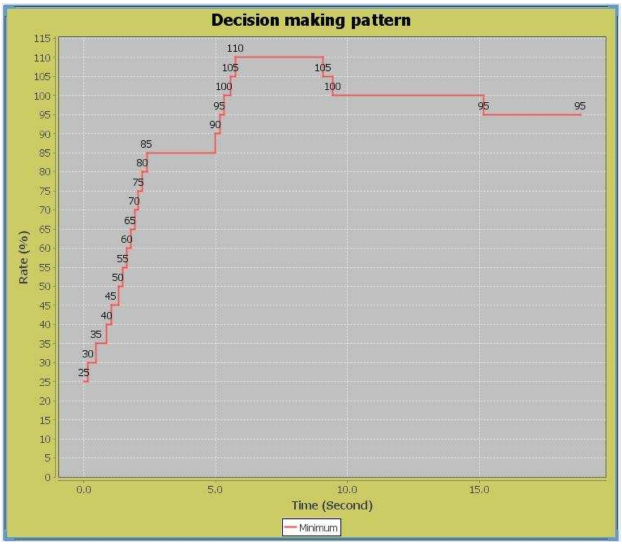


Figure 8: A decision making pattern(minimum)

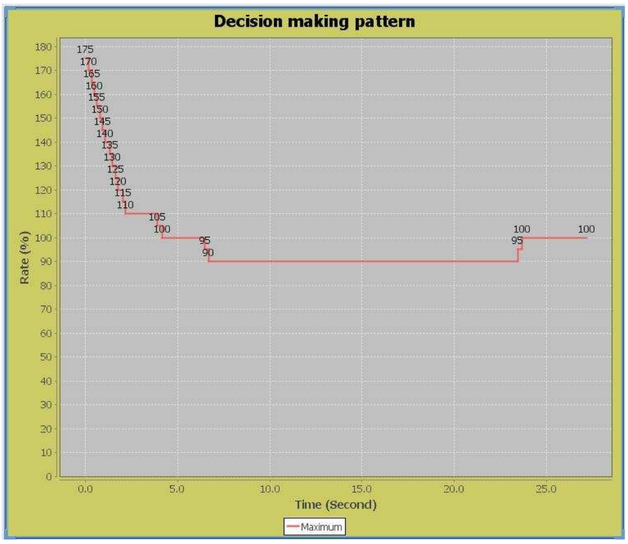


Figure 9: A decision making pattern(maximum)

CONCLUSIONS AND FURTHER RESEARCH

The purpose of this paper is to present a tool that can visualize response selection and execution for pace rating. This research attempts to reveal and generate decision making patterns based concepts of discrete-event dynamic systems. Paces sequentially evaluated with this visual and reasonable procedure can model and visualize how pace rating is processed in the human brain of the rater. The average value computed from the two final responses is used as the rated pace. In addition, documentation of a decision making pattern can serve as an aid to communicate between management and union. Until now, we obtained the rated pace value based on a case with only one person. Additional research is needed to investigate further how to use the proposed lean tool in different cases and when multiple people are involved. And finally, we are confident that our research will serve as a base for future studies on signal detection theory on pace rating, synchronizing two videos.

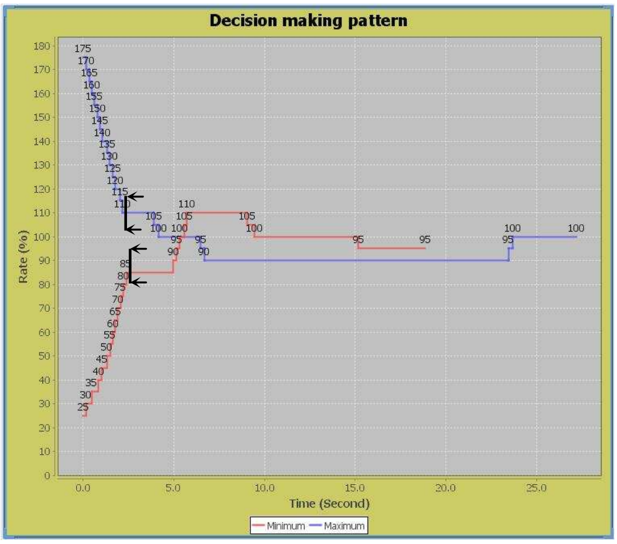


Figure 10: Two decision making patterns(min&max)

REFFERENCES

Aft, L. S. 2000. *Work Measurement and Methods Improvement*. John Wiley and Sons, Inc New York.

Barnes, R. M. 1980. *Motion and Time Study Design and Measurement of Work (7th ed.)*. John Wiley and Sons, Inc, New York.

Groove, M. P. 2007. *Work Systems and the Methods, Measurement,and Management of work*. Pearson Prentice Hall, Inc, New Jersey.

Kanawaty, G. 1992. *Introduction to Work Study (4th ed.)*. International Labor office, Geneva.

Karger, D. and Bayha, F. 1987. *Engineered Work measurement (4th ed.)*. Industrial Press, Inc, New York.

Meyers, F. E. and Stewart, J. R. 2002. *Motion and Time Study for Lean Manufacturing (3rd ed.)*. Prentice Hall, Inc, New Jersey.

Miller, D. M. and Schmidt, J. W. 1990. *Industrial Engineering and Operations Research*. John Wiley and Sons, Inc,Singapore.

Mudel, M. E. and Danner, D. L. 1994 *Motion and Time study: Improving Productivity(7th ed.)*. Prentice Hall, Inc, New Jersey.

Nadler, G. 1955. *Motion and Time Study*. McGraw-Hill, New York.

Niebel, B. and Freivalds, A. 2003. *Methods standard and work design (11th ed.)*. McGraw-hill, Inc, Boston.

Salvendy, G. 2001. *Handbook of Industrial Engineering (3rd ed)*. John Wiley and Sons, Inc, New York.

Van Goubergen. D. and Vancauwenberghe, F. 2006. *4 Mudas of Productivity : Quantifying waste in Work Methods*. IIE Annual Conference, Florida.

Watmough, E. B. 1975. *TMI Pace Rating Videos*, Tempo Manufacturing Institute.

BIOGRAPHY

PEERADAECH SUWITTAYARUK is a PhD student in Industrial Engineering and Operations Research at the department of Industrial Management, Ghent University(Belgium). His research interests lie in work measurememt and motion study, ergonomics and lean manufacturing. In his PhD he is researching on improving work measurement practice for method improvement and process design using video analysis.

DIRK VAN GOUBERGEN is a professor at the department of Industrial Management at Ghent University. He is an expert in the area of work measurement and method engineering, productivity improvement and lean manufacturing. Additionally, he has more than fifteen years of international experience as a trainer/consultant in the area of lean manufacturing within different industrials throughout Europe. He has a M.S. degree in Mechanical Engineering from the Royal Military Academy (Brussels-Belgium), a M.S. in Industrial Management from Ghent University and a PhD in Industrial Engineering from Ghent University. He is a senior member of the Institute of Industrial Engineering(IIE).

ENGINEERING SIMULATION

RISK MANAGEMENT BASED ON FUZZY LOGIC FOR A FRANCIS TURBINE

Diego Tobon Mejia*
Sophie Hennequin*
Francisco Botero**
Nidhal Rezg***
Leonel Castaneda**

*Ecole Nationale d'Ingénieurs de Metz
1, route d'Ars Laquenexy
F-57078 Metz cedex 3
**EAFIT University
Carrera 49 N° 7 Sur - 50
Medellín - Colombia
***UFR MIM - Université Paul Verlaine de Metz
Ile du Saulcy
F-57045 Metz
E-mail: dtobonme@gmail.com, hennequin@enim.fr, fboteroh@eafit.edu.co, rezg@univ-metz.fr, lcasta@eafit.edu.co,

KEYWORDS
Maintenance, risk evaluation and assessment, fuzzy logic, Francis turbine.

ABSTRACT

To facilitate the maintenance of hydroelectric groups and ensure the production of electrical energy, we defined a decision making tool to help the construction of maintenance plans by integrating the concept of risk for a Francis turbine currently operating in Colombia. As the risks are generally expressed in terms of human knowledge, it seems very interesting to use fuzzy logic to solve this kind of problems. Moreover, it can also be applied in case of detectable failures (via sensors) or undetectable failures. This work has been done in collaboration with EAFIT University in Colombia.

INTRODUCTION

To facilitate maintenance of hydroelectric groups and ensure the electric power production, we conducted the study and assessment of potential risks for a Francis turbine currently operating in hydro plants in Colombia. These hydro plants, La Herradura and La Vuelta have been operational in October 2004 in Frontino (Antioquia) Colombia and developed by the company Empresas Publicas of Medellin (EPM). Currently these two plants work continuously, but climate change, which currently generates the seasons of heavy rain and drought is beginning to cause problems. In fact, during periods of water shortage, the station must continue its production of electrical energy, despite a lower nominal level than expected, and that is why the operating speed has to be reduced. This reducing generates vibration and wear. Moreover, due to a design error, the recovery systems of sand were undersized and therefore fail to recover all the sand that damages components. Replacing these components is very difficult because it requires stopping the installation (and therefore stopping the production of energy). It is however imperative for the region to maintain its production of energy for these plants, and, in addition to supplying energy, generate jobs for the region. A maintenance plan must be defined to minimize the risk of failures.

Maintenance of industrial systems has becomed a key issue in design and operations, both for reasons of security and availability (Al-Najjar, 2003). Until recently, techniques of corrective maintenance (Zwingelstein, 1996) or preventive maintenance (Valdez-Flores, 1989) were the only ones used. The second is not always applicable because of the number of human resources and equipment needed, while the former involves extra holding costs for the company. In response to this problem appeared the predictive maintenance, or conditional maintenance (Mobley, 2002). The principle is to simultaneously measure several parameters of the system to monitor, to analyze data to detect the emergence of a potential failure, and to respond with the correctest and fastest way in case of possible problem. Our work concerns decision tools for predictive and preventive maintenance and definition of maintenance policies to ensure the availability of components of La Herradura and La Vuelta hydro plants. In this article, our work is limited to a pre-study of risks management for the maintenance of the Francis turbine used in the both hydropower plants. Research works are carried out through cooperation between the University EAFIT (especially the laboratory GEMI) and the laboratory LGIPM. For risks management and assessment of failures, we propose to use the failure mode and effects analysis (FMEA) method (Garin, 1994). This method is used as a technique for evaluating failures to determine the reliability of equipment and system. Unfortunately, in an industrial context, it is sometimes very difficult to quantify exactly the effects of failures, in case of lack of information (no diagnosis, for example) and / or lack of resources. In this case, fuzzy logic can be used (Yager, 1994). This theory helps to define concepts that are subjective, imprecise, vague, and uncertain. Indeed, fuzzy logic can be defined as Boolean logic which has been extended to meet the concept of partial truth. In Boolean logic, truth values are defined as "completely false" and "completely true" (0 or 1). With fuzzy logic, all possible values between 0 and 1 could be used (Zimmermann, 1990, Vasant, 2004). Several studies have used fuzzy logic to manage some stages of maintenance. Fuzzy logic has for example been used for decision support during the integration maintenance policies with the production management (Gomez et al., 2007), but also to model the human factor which is difficult to quantify (Hennequin and al., 2009). Fuzzy logic is also used for the definition of failures in diagnosis. We can, for example, cite the work developed by (Wang et al., 2008) in which a model is defined to determine the criticality based on fuzzy weights to define the input variables. These weights have led to better estimate the criticality and the proposed model is realistic and simple

In this article, fuzzy logic is used to represent risks in a FMEA model. The studied industrial system is presented in section 2. In a third section, we present our methodology based on fuzzy logic. The fourth part presents numerical results of our fuzzy models. Finally, we finish this article with some conclusions and perspectives.

INDUSTRIAL SYSTEM

In this work, we consider two hydroelectric plants currently operating in Colombia: La Herradura (Figure 1) and La Vuelta. These plants include two conventional dams, with two pressure tubes, two vacuum-tube diffusers, two sand traps and two engine rooms with Francis turbine (Figure 2). The Francis turbine is a jet engine, which means that the inlet pressure of the wheel is greater than the output pressure of the wheel (Varlet, 1964). In that sense it uses both reaction and impulse. The reaction part is the part where the static pressure of the fluid is transmitted to the wheel while the impulse part is the part where the kinetic energy of the fluid is transmitted to the wheel. Francis turbines are used for medium falls and can develop very significant powers.

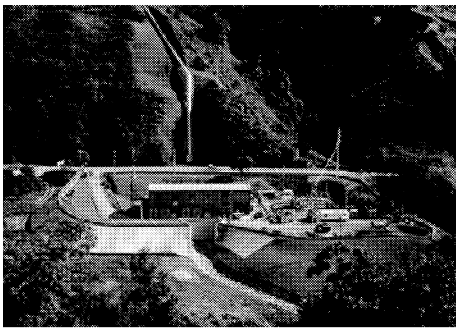


Figure 1: La Herradura plant (EPM)

A hydropower group is a very complex system with many variables, and therefore, is difficult to model and study. In this study, we limit the study to components of the Francis turbine. To make this choice, we use the information found in Guarnizo Luisa Fernanda Gomez's thesis (Gomez Guarnizo, 2007), in which all the key variables for the diagnosis of the turbine are presented. In her work, the most common failures for this type of systems and components are detailed. We defined a model of risks management based on fuzzy logic for certain components of the Francis turbine.

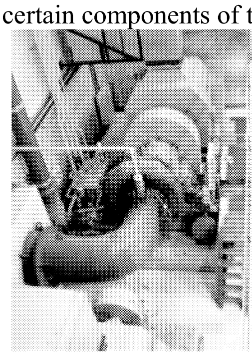


Figure 2: Francis turbine in La Herradura plant (EPM)

For our proposed method, we used the work given in [Faisal I. Khan, 2003]. This method is described by 4 steps and a looping process allows a continuous feedback for risks management. The steps are the following:

- System analysis: In this first step, an analysis of the system is made to understand its operation and identify what are the basic components, and among them, the critical components that must be evaluated and monitored.
- Risk estimation: This step involves the identification of failure modes of critical parts, found in the earlier step using an FMEA sheet. These failure modes are quantified in relation of detection, occurrence and gravity.
- Risk assessment: In this step, we take the values defined in the previous step to calculate the criticality of each failure mode; we use a model based on fuzzy logic.
- Maintenance Schedule: when the criticality of failure modes has been calculated, we must act on the modes and critical parts to reduce their criticality. The reduction of criticality allows the development and implementation of maintenance activities. When these actions are implemented, a new estimate of risk is made.

METHODOLOGY

System analysis

The generator group is composed of various sub-systems; among them, the most critical and complex are the turbine (which performs the conversion of hydraulic energy into mechanical energy) and generator (which makes the conversion of mechanical energy into electricity). In this study, we looked at the Francis turbine because it is affected by factors which are not easily modeled (such as cavitations, wear, leaks, etc.). To better understand these phenomena, we examined their effects on different components of the turbine, especially whose affected by the phenomenon of degradation due to the blaster water, and we also evaluated the indicators to measure the risks associated with these phenomena. The water that circulates throughout, and that brings a large amount of dust, can cause serious failures that may affect the continuous production of energy. The detection of these types of failures is very difficult to implement, firstly because it is costly and sometimes impossible to place sensors on each critical element, and secondly, there is at our knowledge no diagnostic method that ensures to take into account the interaction of different components with each other and back reliable information about the failures of the Francis turbine. Therefore, we initially conducted a special study of certain parts which enabled us to establish the failure modes of these components and a tool to determine their criticality.

Risk estimation

To estimate risk, we used the FMEA method combines with fuzzy logic in order to identify failure modes of the basic components. Firstly, for a period of study of two years, the number of failures for each failure types in function of components are determined (Figure 3).

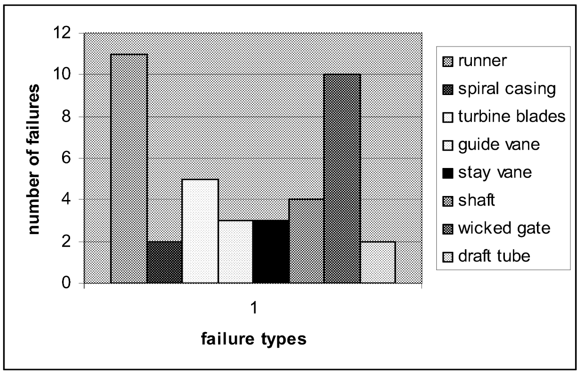


Figure 3: Failures mode vs components

Then, the most recurring failures are those due to the blaster water (Figure 4).

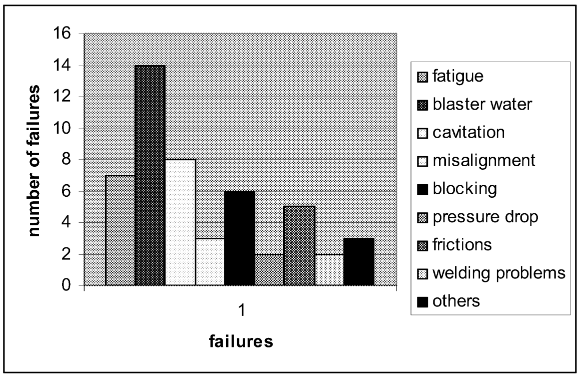


Figure 4: Failures mode vs causes

This cause is very difficult to detect because we do not have a history of dust concentration in the water to connect with failures. On the other hand, the establishment of a monitoring system for this variable can be very costly and complex that is why in this article, we present this particular case, to try to build a system able to determine the criticality of components affected by this phenomenon. To implement the FMEA methodology three variables must be defined to determine the criticality of each component. These variables are: occurrence, gravity and detection.
Detection is defined as the ability to see or detect changes on the monitoring system. This variable is defined using fuzzy logic (Figure 5) and fuzzy sets.

We chose the form of a bell to define the fuzzy sets as considering the Gauss curve allows us to incorporate the probabilistic aspect of variables.

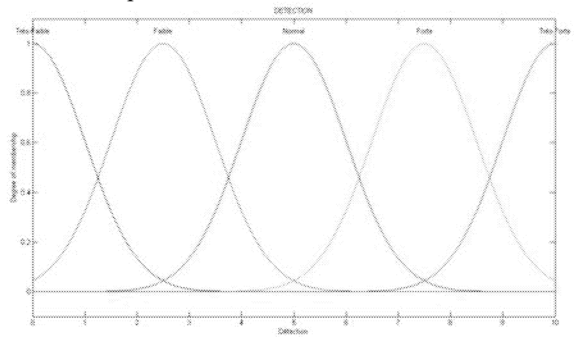


Figure 5: Fuzzy sets for the detection variable

Similarly to detection, the occurrence must be defined in terms of values of FMEA and we defined the corresponding fuzzy sets (Figure 6).

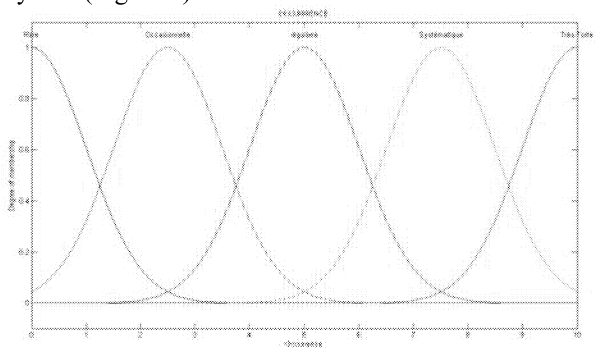


Figure 6: Fuzzy sets for the occurrence variable

The third variable to consider is gravity, which is defined from the effects of failure on the system (Figure 7).

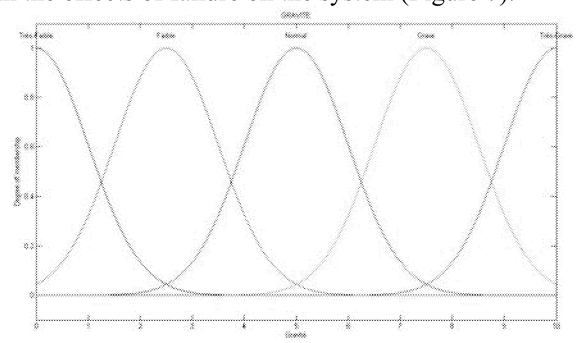


Figure 7: Fuzzy sets for the gravity variable

The criticality is the output variable (Figure 8). To calculate the criticality, we can use the multiplication of occurrence time detection time gravity or using our model derived from fuzzy logic.

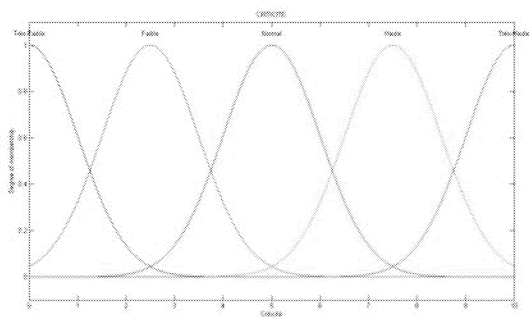


Figure 8: Fuzzy sets for the gravity variable

Risk assessment

The risk assessment is conducted using an inference system, derived from fuzzy logic, which takes into account all the knowledge of experts and maintenance personnel.

From this knowledge, we creat a rule base to compute the criticality of potential risks while reducing the effects of uncertainty on the assessment of FMEA variables (occurrence, gravity and detection). Thanks to studies done by the University EAFIT and information provided by EPM, we could build two models to determine the criticality, a model for all detectable failures and another model for failure due to undetectable phenomena.

For this case, we have identified five potential failure modes, the most critical failure mode is the blocking blades. It is quite obvious that this failure is the most critical since the onset of this failure involves a cessation of production of energy. Furthermore, to replace the part located inside the system, it is necessary to remove the turbine totally. It is then necessary to define a very good maintenance schedule to ensure that the criticality of this failure mode is reduced.

MAINTENANCE SCHEDULE

To reduce the failures on different components, it should provide and implement appropriate actions and measures for each component and failure mode and this, according to its criticality. When we know the critical components, we could act directly on these components to increase system availability and reduce the likelihood of a shutdown.

For our case of bloking blades (due to blaster water), we have to act preemptively against this component in order to reduce the interval of occurrence of this failure, and then reduce the criticality and swing to a normal level.

Until now, only some preventive maintenance actions (mainly cleaning and upkeep) have been implemented. Indeed, the enterprise EPM has decided to investigate and seek new technologies to lessen more the effects.

RESULTS

To create our fuzzy models, we used the module Fuzzylogic software Matlab. Operation "and" and implication use the minimum operator, and the operation "or" and aggregation are defined using the maximum operator.

For our fuzzy model for detectable failures, we obtain the following surfaces (Figure 9). We can note that in this model, the symmetry is fully respected, the only defect that we find are peaks.

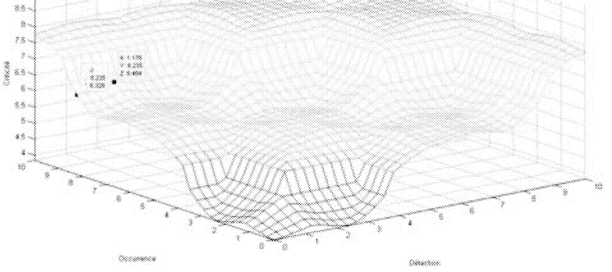


Figure 9: Occurrence vs Detection vs Criticality for the model with 125 rules

For the fuzzy model for undetectable failures, we did not simplify the number of rules because cases were assessed to be manageable (25 possible cases). So we obtain the following surface (Figure 10).

The results obtained using our method seems very interesting and should be modified and adjusted based on the expertise of EAFIT.

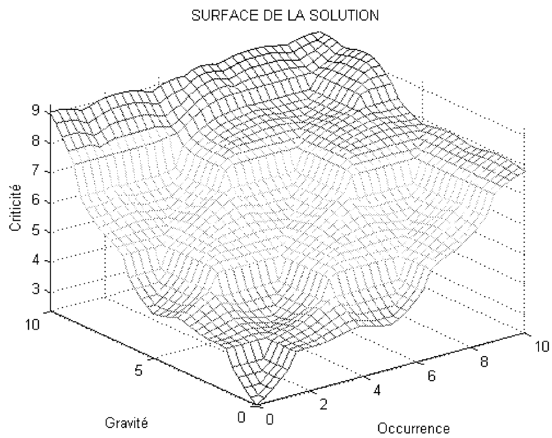


Figure 10: Occurrence vs Gravity vs Criticality for the model with 25 rules

CONCLUSION

We defined a tool for decision support for the definition of maintenance plans by incorporating the concept of risk. The major difficulty in risk management is, of course, as we have seen, the quantification of risks against each other, quantification which is essential to verify that the maintenance actions put in place ensure both the availability of the system and security of all. As the risks are generally expressed in terms of human knowledge, it seemed very interesting to use fuzzy logic to solve this problem. In addition, it may effectively be applied in the case of detectable failures (using sensors) or in a more difficult case to estimate where the failures are not detectable. This case is very important for the industrial system studied in this paper because the implementation of sensors and expertise of these sensors are very expensive, and on the other hand, require time and human resources. The integration of tools from artificial intelligence will thus create relays which can be placed in all production sites of the company EPM to help staff make sound decisions during maintenance, while limiting costs. Moreover, the method we proposed, based on a decomposition of the system into subsystems and individual parts associated with the FMEA method to identify parts and critical failure modes, allowed us to be familiar with the operation and behavior of the system and thus to highlight the key components to be priority given (high research topic currently in the laboratory GEMI) with the system analysis step. We were able to demonstrate that all the cases studied were not well monitored and instrumented. The proposed method is very interesting because it does not require a significant financial contribution. It is of course obvious that our tool will not be used for maintenance of very expensive systems such as satellites, but it remains very promising for many systems and many other countries. The interest of our approach is to propose a model which does not require the definition of complex physical models that are sometimes inaccurate (because of different assumptions made) and difficult to define. Similarly, we have created a method of risk management performance which, because of looping, allows a continuous assessment of risks to redefine a more appropriate and efficient operations and increase the availability and reliability in production energy.

To improve our system, we consider various ways:

- In the inference step, we could identify the rules or inputs whose membership degree is important and integrate them into the calculation of the defuzzification.
- We could establish a system of neuro-fuzzy learning to allow the model to automatically rebalance and adjust to small changes on information provided by the sensors.
- When the study of the laboratory GEMI will be over, a step of statistical analysis to find the best way to define the fuzzy variables associated with inputs and outputs would significantly improve the accuracy of our model.

The fuzzy rules must also be reworked and adapted according to the results provided by EAFIT.

ACKNOWLEDGMENTS

The authors wish to thank the EPM company without whom this work would not have been initiated or developed.

REFERENCES

Al-Najjar, B., and I. Alsyouf. 2003. Selecting the most efficient approach using fuzzy multiple criteria decision making, *Int. Journal of Production Economics,* Volume 84, no 1, 85-100 (ISSN: 0925-5273).

Fernandez Gomez, R., S. Hennequin, and N. Rezg. 2007. Decision Making based on Fuzzy Logic for Products Subcontracting taking into account maintenance actions, *in the 3rd annual IEEE Conference on Automation Science and Engineering*, Arizona, USA, 22-25 september, 2007.

Garin, H. 1994. *AMDEC, AMDE, AEEL: l'essentiel de la méthode* - Paris : AFNOR, 1994 (in French).

Guarnizo Gomez, L. F. 2007. *Conceptos Basicos Para Un Diagnostico Técnico en una,* Ph.D. thesis, MEDELLIN : EAFIT, 2007 (in Spanish).

Hennequin, S., G. Arango and N. Rezg,. 2008. Optimization of imperfect maintenance based on fuzzy logic, in *Journal of Quality in Maintenance Engineering,* Vol. 15, N. 4, pp 412-429, 2009, ISSN 1355-2511.

Khan, F. I. and M. M. Haddara. 2003. Risk-based maintenance (RBM): a quantitative approach for maintenance/inspection scheduling and planning, *Journal of Loss Prevention in the Process Industries,* Vol. 16, Issue 6, November 2003, p. 561-573.

Mobley, K. R. 2002. *An Introduction to Predictive Maintenance*, Plant Engineering, Elsevier Science, second edition, USA, 2002.

Varlet, H. 1964. *Turbines hydrauliques et groupes hydroélectriques,* Paris v : Eyrolles, 1964. - Vol. I (in French).

Wang, Y.-M., K.-S. Chin, Gary Ka Kwai Poon, J.-B. Yang. 2008. Risk evaluation in failure mode and effects analysis using fuzzy, *Expert Systems with Applications,* article in press.

Yager, R., D. Filev. 1994. Essentials of Fuzzy modeling and control. *John Wiley and Sons.* USA: 1994 (ISBN: 0471017612).

Zimmermann, H. 1990. Fuzzy Set. *Theory and Its Applications*. Second Revised Edition. Kluwer Academic Publishers. ISBN 0-7923-9075-X, 1990.

Zwingelstein, G. 1996. Reliability centered maintenance - *Editions Hermès*. ISBN 2-86601-545-2, 1996.

BIOGRAPHY

SOPHIE HENNEQUIN is Associate Professor at the National Engineering school of Metz (France) since 2002. She has obtained her PhD degree at the University of Franche-Comté (France) in 2000. She is currently member of the Laboratory of Industrial Management and Production of Metz (LGIPM) and member of the INRIA CosTeam team. Her field of interest is performance evaluation and optimisation of stochastic systems and definition of maintenance policies.

Preparing a representative dataset by using a workplace for experimental brake systems testing

Tomáš Kohlschütter, Pavel Herout
Department of Computer Science and Engineering
University of West Bohemia
Univerzitní 8, 306 14 Plzeň, Czech Republic
E-mail: kolsi@kiv.zcu.cz, herout@kiv.zcu.cz

ABSTRACT

The important request for simulations is to have a large quantity of data from real systems. This paper describes a methodology and practical experience with a workplace for measurement, evaluation and storage of data acquired during extensive measurement of brake systems on cars or special device called brake stand. The workplace provides now over one hundred data sets per a few days which represent a huge data volume. Software system used for the controlling whole processing cycle is a modular complex system designed to simplify particular stages of measurement, data processing, storage evaluation and results archiving. Nowadays, the workplace is used as a reference one for industrial partners, but existing data set has a great value for a future research. Selected data from whole dataset will be used for the future simulation of brake system in several directions – for validations of simulation model, as real examples of possible failure states or for creating simulation scenarios.

INTRODUCTION

The safety is the most important part of automotive industry – not only safety of car's crew but of all participants in traffic operations. One of key elements of the safety is brakes' behaviour while driving, slowing down a vehicle or complete stopping when a driver puts the brakes on. While car is braking, the places with significantly higher temperature – hotspots – can appear on the brake disc as a result of thermo-mechanical instabilities. They can cause an unstable contact between the disc and brake pads which can result in undesirable vibrations, noise and brakes efficiency reduction. Since this and no other possible failure is acceptable, neither in minimal rate, the car brake system must be completely examined before the vehicle is able to be used in normal traffic.

It has been two years now since Department of New Technology (research centre of University of West Bohemia) finished building up two special devices designed for brake system characteristics measurements – one for measuring car's built-in brakes and one for separate brake components. These devices are called the automobile and the brake stand respectively and they allow measurements of large amount of characteristics like pressure evolved on the brake disc during deceleration, temperature distribution on the brake disc, brake moment, RPM, acceleration etc. The other individual components of the devices are able to measure e.g. noise or vibrations produced by braking process and a lot of other parameters. All characteristics are scanned by a number of sensors and are transmitted as an analog signal into a control computer. This data acquisition process results in a huge number of data values so further processing is necessary to get human-readable results which could be used for diagnostics, simulations and future research of brake systems.

Our first step was to develop stand-alone independent software system which provides a complex fully-automated solution as an addition to specialized software connected to the measuring interface. The basic requirements are:

- full automation,
- high performance,
- fast access to evaluated results,
- storage of already measured data that can be used for another research in the future,
- modularity to allow easy extensibility to satisfy urgent needs on brake system measurement, especially for already mentioned brake system simulations,
- visualisation of results captured from data evaluation

These measurements are considerably specific and they mostly take place in development centres of car factories only. Only a few relevant references can be found in available literature, for example (Yong Fu Zhan et al. 2010) talks about designing and implementing a data acquisition and processing system for testing vehicle's braking performance which has the function of storage, real time display and analysis. Functional samples of the automobile and the brake stands are described in (Honner et al. 2009) and (Lang et al. 2009) respectively.

DATA MEASUREMENT

As was written above, the experimental testing of brake system can proceed either on the brake stand or the automobile stand. Figure 1 shows an illustration of the brake stand which is able to simulate either static tests on constant speed, e.g. driving down the hill, or dynamic tests with varying speed such as stopping. The brake stand control allows testing different braking modes – constant pressure in the braking system, constant moment, constant speed or defined changes of named parameters. (Šroub and Lang 2008)

Many sensors are located on the stand and they sample values at very high sample rate. Information from one sensor is called a channel. It is possible that each channel is sampled at different rate. The channels with single value can exist too, e.g. some statistical averages. Scanned information is passed into the measuring card connected into the control computer via PCI interface. This card is license-bundled with the software DEWESoft (http://www.dewesoft.com) which is responsible for an acquisition and storing all data. It is also possible to install measuring system in the car directly for measuring in real conditions.

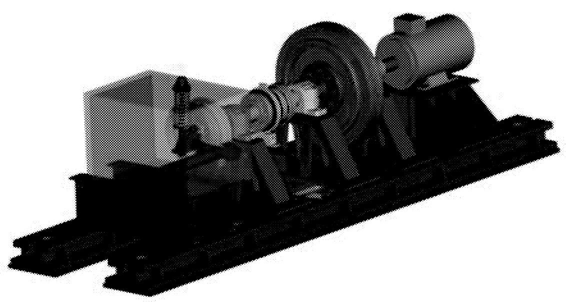


Figure 1: The illustration of the brake stand

It is important to highlight that DEWESoft provides basic means for acquisition and storage of measured data. Since each measured application is unique on principle, nobody can expect this software to be able to evaluate the arbitrary application. It is obvious that experimental evaluation can be tried by tools like Matlab, but everyday practice usage must be performed with dedicated software.

DATA EVALUATION

Up to now, all data acquired from measurements were evaluated by Matlab software which was supported by the fact that DEWESoft allows data exporting to .mat format. But since Matlab is a purely computational environment, it would not allow easy automated file exports of measured data. It is not an optimal way to get evaluation results either – exporting 1.94 GB data file resulted in 4.07 GB Matlab file and it took 20 minutes to finish exporting. During that time the application was completely unresponsive, so another measurement was not possible. Another drawback is the need to load whole data file although only some chunk is going to be evaluated. There was also a problem that we were not able to evaluate data in the environments where Matlab was not installed for whatever reason. These all disadvantages resulted in decision to code a stand-alone application in C++.

Hotspots evaluation

There are several possibilities for data evaluation in prepared software. The most important is an evaluation of temperatures on the brake disc. The wheel rotations must be detected in an incoming signal at first. The brake stand provides measurement data for one wheel only. The automobile stand or direct car measurement logically provide data for two wheels, left one and right one, hence whole evaluation process must be executed twice. The inner and outer surfaces are evaluated separately. Individual rotations are denoted with amplitude peaks (eventually with troughs for an opposite wheel) in the signal. They can be easily detected and separated by setting amplitude threshold to the half of the maximum value in the scanned signal segment. Problem is that peaks with very small amplitude cannot be detected with this method, so we are still in work to improve it.

Temperatures are computed for each rotation by mapping voltage to temperature using sensors' calibration curves. These are obtained from a theoretical calibration exponential curve and constants c_1 and c_2, which must be defined for each measurement, as:

$$U_s = c_1 \cdot U_t + c_2$$

where U_s and U_t are the diode voltage and the theoretical voltage respectively. Computed temperatures can be utilised for later visualisation in the form of temperature maps showing temperature distribution on the brake disc as shown in Figure 2. The performance of plotting is very important, because animations of the temperature distributions during a whole measurement process will be one part of prepared simulations.

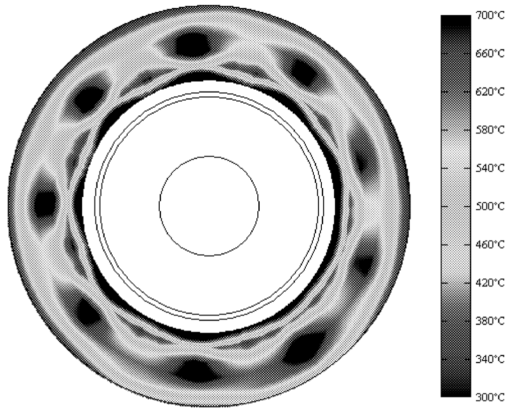


Figure 2: Visualisation of temperature evaluation with the visible hotspots

Noise detection

Next step of evaluation is noise detection during braking. One measurement of noise contains a lot of data (as speed, rotations, vibrations, sound, moment, brake pressure etc.) for many cycles of starting, accelerating, driving, decelerating and complete stopping. Each cycle and its parts must be identified and it is divided into (possibly overlapping) segments of N samples length. Each segment is windowed, using Hanning window, and then transformed via fast Fourier Transform to N spectral samples. Many statistics are obtained during the evaluation process too. The measurement is executed many times for different speeds, pressures and other parameters to simulate different conditions of real usage. Batch processing of more data sets is supported.

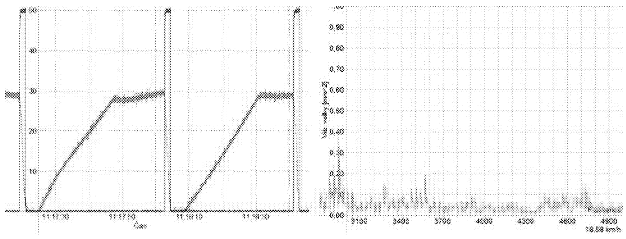


Figure 3: View at noise detection signals. Left graph shows speed (green) and brake pressure (red), right one shows FFT of vibrations.

Other types of evaluations are also possible. Majority of results is planned to be incorporated into the prepared simulations.

SYSTEM ARCHITECTURE

Whole software system is divided into several parts where each one is responsible for different tasks. The core is the most important part which represents the heart of the application. The base of it is built up on plug-in systems that are responsible for data processing operations and storing them into data manager in the core. Data can be obtained from several sources as DEWESoft data files or running DEWESoft instance via DCOM (Distributed Component Object Model) interface, exported text files, our binary files etc. Each data source must have its own plug-in else it is not possible to read from or write to it. Multiple sources per one plug-in are also supported.

DATA STORAGE AND ARCHIVING

All measured data and evaluated results must be stored for future usage. DEWESoft data file were saved immediately after acquisition process in the measuring computer which has a limited disc capacity. Only relative small number of files could be kept there and it was inevitable to either delete old files with time or move them to external medium. The storage of complete data files was necessary to be solved. There was also a need to archive evaluated results without necessity to go through whole evaluation process again.

Data storage

All data captured by DEWESoft application are stored in measuring computer's hard disk in its native binary format. Its biggest disadvantage is that it has very dynamic structure changing all the time with each release as new functions are added. It results in many files stored with one version of this software cannot be loaded with different version. This makes the future research with formerly measured data very difficult, maybe nearly impossible. Installing all possible software versions and guessing the version of the stored file is not a suitable solution. The lack of non-existing documentation makes it impossible to load into any external application.

Due to the reasons mentioned above there was a need to find a better solution than storing data in DEWESoft's native format. We were discussing many already existing data formats, for example well known XML format. Its drawback is that it is not suitable for huge data sets – files need to be read sequentially and they would grow to enormous sizes as measurement requests increase. The same pays for plain text or Matlab format. Therefore we ended up with the opinion that creating our own format will be the best solution and we will be able to design it directly for our purposes.

We have developed a new file format for storing all data acquired during measurements. The designed binary structure makes it well-extensible with preservation of backward compatibility when some change needs to be

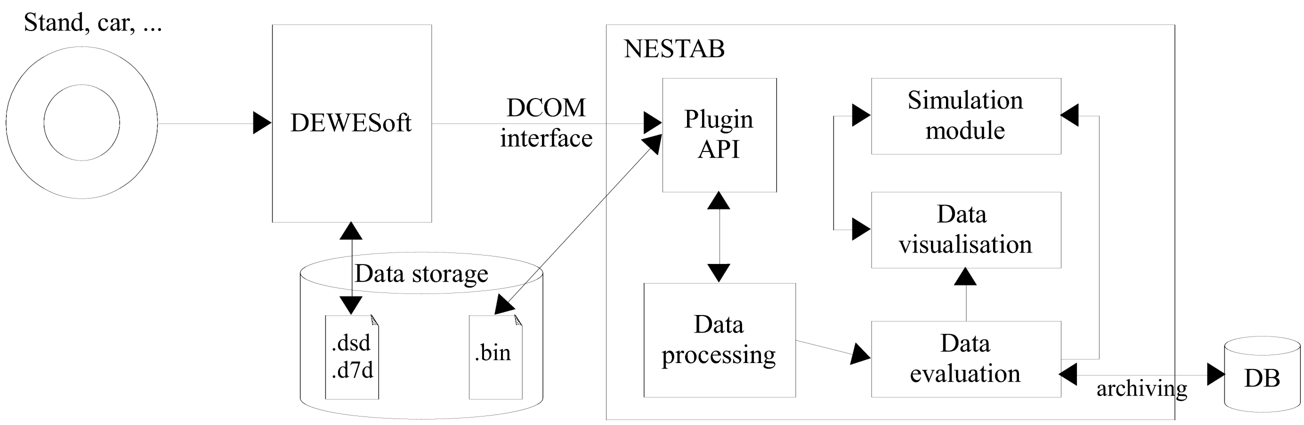


Figure 4: Data flow in our software solution

suitable compressing algorithm because file sizes must retain as small as possible – or at least no more than the original DEWESoft data files. We have managed to reduce the file size up to a half of original files in some cases. Any requested information is loaded from disk almost instantly, because there is no need to load whole file in memory which is the most important property of this format. Table 1 shows size comparison of exported files in formats supported by DEWESoft. The last columns shows file size of our data format.

Table 1: Examples of file size comparison of several data formats

DEWESoft data files	Matlab format	Plain text format	Our binary format
1.94 GB	4.07 GB	7.47 GB	1.44 GB
455 MB	1.71 GB	1.69 GB	339 MB
343 MB	1.98 GB	2.21 GB	155 MB

Our file format has an easy binary structure arranged into the blocks as scheme below shows. Since all data are referenced by the block index and its size, it is very easy to append any new kind of information to either any channel data block or at the end of the file and backward compatibility will still be maintained. This is controlled by adding new flag in the header of file. New versions of application will be able to read older format and vice versa. It is obvious that application will ignore fields which it does not understand. The block structure also makes loading individual channel data very fast by avoiding sequential reading of whole file.

<identificator><flags><channels><data_count>
<sample_rate><start_time>[<channel_name_length>
<channel_name_string><channel_min><channel_max>
<async_data_count>]^n*<indexes><channels_data>*

The identificator is used for checking whether loaded file is really in our format. Flags are there as an option for any future extension. Next values denote a number of available channels, how many values were captured, their sample rate and time of acquisition process start. Individual channels data can be compressed (controlled by flags) using more algorithms to minimize storage space usage. We performed benchmarks of common compression algorithms but none of them satisfied us. Either compression ratio was bad (e.g. ZLIB) or performance was very poor (e.g. BZip2 or LZMA - Lempel-Ziv-Markov-Chain Algorithm). After several studies, we came across a sequence of pre-processing and compression algorithms which provides very good compression ratio with minimal impact on compressing and decompressing performance. At first, the channel data block is pre-processed with Schindler's Sort transformation (Schindler 2002-2004) and then it is transformed using Move-To-Front algorithm (Bentley et al. 1986) which improves the speed of following compression. It consists of common RLE (Run-Length-Encoding) compression encoded into 3-byte packets followed by range encoding (Schindler 2002-2004). Decompression is performed in reverse order and is very fast.

As was already mentioned, the file format description for reading DEWESoft files does not exist (this fact was also confirmed by DEWESoft's support team), hence it was necessary to find out any way to load data from them. First, we utilized data exporting from DEWESoft in plain text format and importing it into our application. Such way was very inefficient because text files were really huge so we had to help it with developing special decimal number format. This gained at least some speed and improved memory usage a lot. Also a lot of information (especially info about asynchronous and single channels) was lost as it was not saved in exported files. Later we developed plug-in for our software application which allows communicating with the measuring computer over network. This way quickly downloads demanded data from running DEWESoft instance using provided DCOM interface. Since different computers (mostly used by different persons) can be responsible for each kind of evaluation (temperatures distribution, noise detection etc.), the network communication makes possible to distribute captured data between them. This data distribution makes much easier finding requested files than searching in tons of external mediums. Our designed format is used for this data storage so also less disc space is required.

The only downside of using different format for data storing and archiving is the necessity of file conversion that can take some time. Currently, operating personnel must do it manually but we put all effort to do it automatically because the goal is to have fully automated system so we could target on brake simulations only. The question is whether to execute conversion after measurement is finished or do it simultaneously when measurement is in process. The first option would require additional time after the measurement; the second option could influence the measurement performance because direct co-operation with DEWESoft application (which is busy with data acquisition in that time) is required. Several tests must be done to choose the most suitable solution. All formerly measured data stored in DEWESoft native format can be easily converted into our format using implemented batch processing (of course, as was already mentioned, appropriate DEWESoft versions must be installed but this is avoided for any future measurement).

Data loading and storing speed is the most critical part of data processing. Experiments show that the best loading performance is gained by using the operating system's native functions executed asynchronously where you can process (evaluate, compress etc.) one data segment while another one is still being loaded. Also native functions know better how the operating system works with file system and they do not perform so much additional checking as universal functions do. Data storing is performed using said method. Data loading is done via memory-mapping, because it is the ideal choice for processing files with block structure.

Results archiving

It was also necessary to archive individual results of evaluation process, such as statistical parameters, pre-processed data needed for visualisation of hotspots or visualised graphs. This step is very important to avoid re-evaluation due to every triviality. Such results are archived in collective MySQL database placed on a remote server. Every member of team has an access to it. Any requested information can be downloaded from there by one click without searching for necessary data files. The database ERA model is shown in Figure 6**.** Since this archiving is applied on data evaluation only, the complete data set is not placed in the database. Those are needed only when some new kind of evaluation must be done or evaluation process must be re-executed for serious reason. Measurement description and identification of the external medium (CD, DVD, HDD etc.) where complete data file can be found are also saved to the database.

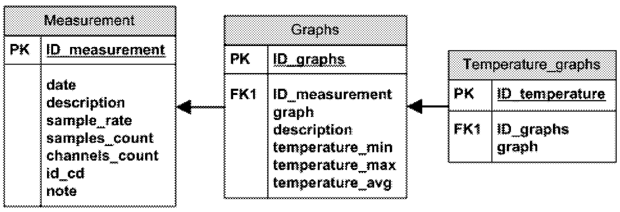


Figure 5: Database ERA model

CONCLUSION

In this paper, the methodology of data acquisition and processing essential for experimental testing of the brake system was introduced. Some types of evaluation were outlined. We coded our own software solution which represents comprehensive system for complex processing, evaluation, storage and archiving of all possible values captured during referenced measurement. We were successful in accomplishing all requirements written in the introduction of this paper. Designing our own data format specialized for this kind of work provides efficient usage of data storage and almost instant access to the measured data with regard to the original file format. The application has been fully optimized for the best performance and everything has been made fully automated by one click. All performed steps led to the time reduction of data evaluation. Up to now, we have already gone through approximately five hundreds measurements which resulted in the extreme amount of huge data files. As was mentioned before, collected data are going to be utilized for simulations of braking systems and all obtained experience employed in the future research. We are going to use this data in prepared simulation models in several ways for validations of simulation model, as real examples of possible failure states or for creating simulation scenarios. The main value of this paper and described research is in fact that it introduced quite practical experience with industrial measurement and processing of huge datasets acquired during this measurement. This activity precedes preparation of many simulation models.

ACKNOWLEDGMENT

The work was supported by the UWB grant SGS-2010-028 Advanced Computer and Information Systems.

REFERENCES

Bentley, J. L., Sleator, D. D., Tarjan, R. E., Wei, V.K. 1986. “A Locally Adaptive Data Compression Scheme.” *Communications of the ACM-Vol. 29.*

Bžoch, P. 2010. *Data visualisation from physical measuring of brake system*. ZČU Plzeň. [in Czech].

Honner, M., Šroub, J., Voldřich, J. 2009. “Frictionally excited thermoelastic instability and suppression of its exponential rising in disc brakes.” *Thermal Stresses 2009.*

Honner M., Lang V., Klepáček. J., Šroub, J., Mertl, J. 2009. “Function sample of an automobile brake testing rig.” Plzeň.

Lang, V., Mertl, J., Klepáček, J., Šroub, J., Honner, M., 2009. “Function sample of a brake testing rig . ” Plzeň.

Litoš, Pe., Honner, M., Lang, V., Bartík. J., Hynek, M. 2006. “The measuring system for the experimental research of thermo-mechanical instabilities of disc brakes.” *Modelování a měření nelineárních jevů v mechanice.*, pp. 117-126. [in Czech].

Šroub, J. and Lang, V. 2008. “Evaluation of measurement of thermo-mechanical instabilities on disc brakes.” November 11th. [in Czech].

Yong Fu Zhan et al., 2010. “Data Acquisition and Processing System for Vehicle's Braking Performance Based on Visual C++.” *Applied Mechanics and Materials (Volumes 44-47)*, pp. 3627-3631.

WEB REFERENCES

Kath, R. 1993. “Managing Memory-Mapped Files in Win32.” MSDN Library. Microsoft Developer Network Technology Group. (http://msdn.microsoft.com/en-us/library/ms810613.aspx).

Schindler, M. 2002-2004. “Compression consulting Schindler.” (http://www.compressconsult.com).

AUTHOR BIOGRAPHY

TOMÁŠ KOHLSCHÜTTER was born in Teplice, the Czech Republic and went to the University of West Bohemia in Pilsen where he studied software engineering. He obtained his master degree in 2010. Now he partly works for a research centre Department of New Technology where he is responsible for software programming and data evaluation. Simultaneously, he studies Ph.D. at the Department of Computer Science and Engineering.

TUNING OF ELECTROHYDRAULIC ROTATIONAL SPEED REGULATORS BY MEANS OF REAL-TIME SIMULATION TECHNIQUES

Dragos Daniel Ion Guta [1)], Teodor C. Popescu [1)],
Petrin Drumea[1)], Constantin Chirita[2)]

[1)] Hydraulics and Pneumatics Research Institute – INOE 2000 – IHP Bucharest, Romania
[2)] Technical University "Gheorghe Asachi" of Iasi, Romania

E-mail: dragos.ihp@fluidas.ro, popescu.ihp@fluidas.ro
E-mail: ihp@fluidas.ro, constantin.chirita@disahp.net

KEYWORDS

Real-time simulation, virtual instrument, electrohydraulic servomechanism

ABSTRACT

This paper presents a real-time simulation system developed for providing automatic tuning and adjustment of electrohydraulic systems in the energy field. The system developed allows modeling of control upon the rotational speed of a hydro unit, identification/ parameterization of model coefficients based on experimental data and developing of a process control algorithm. The system was developed because of the necessity of tuning and off-site testing of the automatic rotational speed electrohydraulic regulators within the structure of hydro units.

INTRODUCTION

In the global energy system balance between production and consumption is achieved with speed and power regulators. Through the joint action of the interconnected generators is ensured a high safety level for consumers because it may be set a stationary value of system frequency, less affected by disturbances. In power systems can be found two control strategies: primary adjustment and secondary adjustment. Both strategies require speed and power regulators for all energy units but the dominant part is played by hydro units which can react quickly in case of local or global load variation in the system [1]. Conventional speed and power regulators used to control hydro units are made in two versions: mechano-hydraulic and electrohydraulic regulators.

The rapid evolution of digital industrial computers and major progress achieved in the field of electrohydraulic systems have imposed as the natural solution the use of electrohydraulic equipment in development or upgrading of speed and power regulators in the energy field. Thus the special static and dynamic performances of hydraulic actuation elements are well mixed with the flexibility and reliability of digital control systems.
Currently, at national and international level, there is a constant concern for the modernization of the automatic speed regulators of hydro units. By introducing the new types of automatic speed regulators substantial improvements have been achieved in the following directions:

- increased safety in the operation of hydro units;
- increasing the reliability of automatic speed regulators;
- increased flexibility in the operation of automatic speed regulators;
- increasing the annual production of electricity.

PROBLEM FORMULATION

The subject of this paper is that category of automated systems using amplification by hydraulic means of the control value up to the level required for the execution value. Usually, applications wherein automatic hydraulic systems are found are characterized by a very high amplification ratio between the signal of automation action execution and the control signal. In combination with electronics, and especially with digital electronics (characterized by abstractization of signal processing in a specialized firmware/ software), hydraulics is associated with the idea of precision.

The objective of this work was to develop a computer system used for tuning and automatic adjustment of electrohydraulic systems in the field of energy. Automatic adjustment means all actions performed on a process for it to behave in a desired manner. For practical application of automatic adjustment it takes up a few basic steps, namely:

- modeling of processes;
- identification of processes based on experimental data and estimation of parameters;
- processing of signals by filtering, prediction, state estimation;
- design of control signals for automatic management

The system developed for tuning and automatic adjustment of electrohydraulic systems must allow running of the afore-mentioned steps. By means of this system it must be possible to be identified the adjustment law and compliance parameters which provide for the whole system a certain unit step response, i.e. satisfying certain required transient and stationary performances. Also the system should allow the selection of a desired response in relation to the actuation of the disturbance values, preserving at the same time certain performances relative to the input value.

The system must provide capabilities for interfacing with physical processes and obviously for driving them in real time. Based on information gathered from the process models of the process can be created to be used during the stage of off-line testing on the adjustment laws adopted. Off-line testing will involve real-time simulation of the process in parallel with execution of the adjustment law adopted.

Real-time simulation of systems has obvious advantages. One can test, with a low risk level, different methods of adjustment and control. Although tuning 'on site' of complex systems is unavoidable, real-time analysis method allows reducing the time needed for tests and adjustments during processes [2]. Reduction of the time required to achieve these objectives minimizes cost and at the same time decreases the possibility of damage for the real systems modelled

The system that adjusts the speed of the hydro unit must provide a set of performance requirements. In the case of the analyzed problem can be identified three important elements [3]:

- the fixed element: hydro unit, electrohydraulic servo motor of the driving device and rotational speed transducer;
- objective of designing the regulator: is to determine the adjustment laws and compliance parameters of the regulator so that to be provided a certain unit step response of the system, thus ensuring transient and stationary performance requirements;
- method of adjustment.

REAL-TIME SIMULATION OF ELECTROHYDRAULIC SERVOMECHANISMS WITH POSITION REACTION

It is hard to imagine now the analysis of a complex dynamic system without the benefit of being able to model and simulate the system. Dynamic systems modeling and simulation are techniques widely used in computer-assisted analysis of systems, also representing an important step in the design (synthesis) assisted by computer systems.

Numerical simulation of dynamic systems is the process whereby information is obtained about the evolution over time of the characteristic parameters of systems by means of the digital computer.

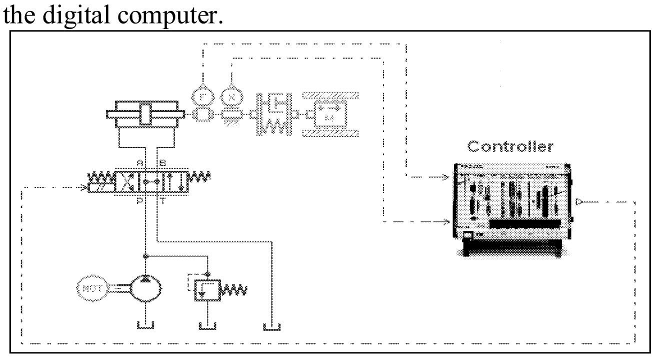


Fig. 1. Real-time simulation of an electrohydraulic adjustment system

Real-time simulation of systems corresponds to the capacity of some computer systems to perform numerical simulation over determined time intervals. An immediate advantage of these types of simulations is the possibility of *interfacing* the computer system performing numerical integration with the investigated physical systems or subsystems [4].

Numerical simulation of dynamical systems allows getting the necessary information on their behavior based on mathematical models that describe these systems. The mathematical models used to develop real-time simulation networks must enable their development in the time range that confers the real-time character to the simulation. The mathematical models used may be linear or nonlinear models. If there is noticed execution of a computation loop outside the allowed time rage, models must be simplified so that the time assigned for execution to be satisfied. For an electrohydraulic servomechanism with position reaction (Fig. 1.) the linear mathematical model can be used:

$$\frac{z(s)}{\varepsilon(s)} = \frac{\mu \dfrac{K_{Qx}}{A_p}}{s\left(\dfrac{m}{R_h}s^2 + \dfrac{mK_P}{A_p^2}s + 1\right)} \qquad (1)$$

Hydraulic natural pulsation,

$$\omega_h = \sqrt{\frac{R_h}{m}} \qquad (2)$$

Damping factor,

$$\zeta = \frac{K_P}{2A_p^2}\sqrt{mR_h} \qquad (3)$$

Speed amplification factor,

$$K_v = \frac{\mu K_{Qx}}{A_p} \qquad (4)$$

Direct path transfer function:

$$H(s) = \frac{z}{y - z} = \frac{K_v}{s\left(\dfrac{s^2}{\omega_h^2} + \dfrac{2\zeta}{\omega_h}s + 1\right)} \qquad (5)$$

Servomechanism transfer function:

$$H_0(s) = \frac{z(s)}{y(s)} = \frac{K_v\omega_h^2}{s^3 + 2\zeta\omega_h^2 s^2 + \omega_h^2 s + K_v\omega_h^2} \qquad (6)$$

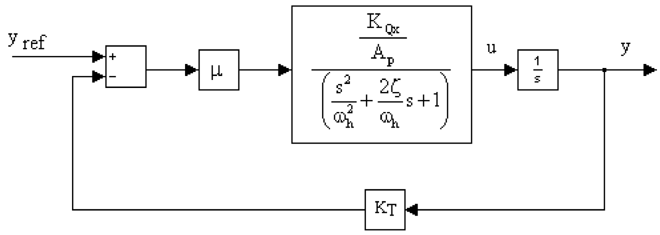
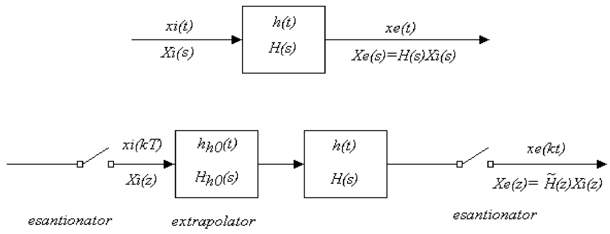


Fig. 2. Linearized simulation model of the electrohydraulic servomechanism

Using a numerical system for interfacing with the physical part requires sampling the analog signal acquired from the process. The mathematical model must be transformed from the complex space s into z, determining the equivalent transfer function in z. If it is intended for the output of the continuous system to coincide with the output of the system with sampling during sampling moments there is introduced an extrapolator for signal reconstruction



In the case of a continuous-time signal *f (t)*, having *Laplace* image F(s)=L [f (t)], by digitization there is obtained a signal whose transform Z is:

$$F(z)=Z[F(s)] \tag{7}$$

For a continuous-time process, having transfer function H (s), the discrete equivalent of transfer function H(z) must also include the transfer function of the zero order extrapolator (ZOE).

The zero-order extrapolator translates an unitary impulse into a constant string along duration T - sampling step. This string can be expressed as the sum of two step signals. Extrapolator transfer function is:

$$H_{EOZ}(s) = \frac{1}{s} - \frac{e^{-sT}}{s} \tag{8}$$

Discrete equivalent H(z) of the transfer function H(s) of a process is obtained applying transform Z to the product $H_{EOZ}(s)H(s)$.

$$\begin{aligned} H(z) &= Z[H_{EOZ}(s)H(s)] = \\ &= Z\left[\left(1-e^{-st}\right)\frac{H(s)}{s}\right] == \frac{z-1}{z} Z\left[\frac{H(s)}{s}\right] \end{aligned} \tag{9}$$

Calculation of transfer function H(z) from H(s) according to $H(z) = \frac{z-1}{z} Z\left[\frac{H(s)}{s}\right]$ can be performed analytically by rests or there can be used the functions c2d or c2dm (*continuous - to discrete - time models*) of Matlab.

It is considered:

$$b_0 = \mu \frac{K_{Qx}}{A_p}; \quad a_3 = \frac{m}{R_h}; \quad a_2 = m\frac{K_P}{A_p{}^2}; \tag{10}$$
$$a_1 = 1;\ a_0 = 0$$

$$\frac{z(s)}{\varepsilon(s)} = \frac{b_0}{a_3 s^3 + a_2 s^2 + a_1 s + a_0} \tag{11}$$

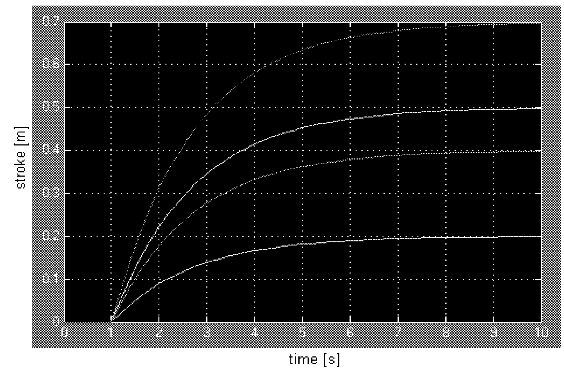
The calculated coefficients of the linear model of electrohydraulic system can be found in Table 1. Furthermore there are presented the coefficients of the discrete model obtained in Matlab using the conversion function c2dm. Comparison between the response to step-type excitation signals of various amplitudes, for the linear model, and the discretized model is shown in Figure 3, a - response of the servomechanism in space s and b - response of the system in z.

Table 1: Coefficients of the electrohydraulic system

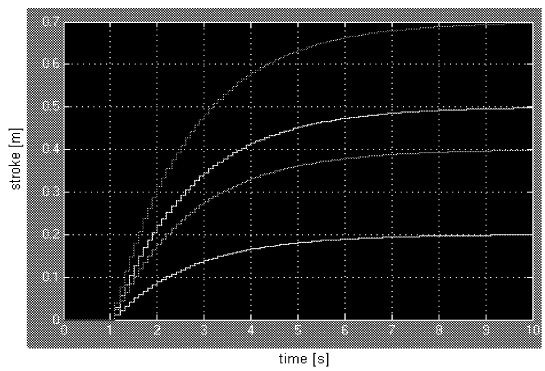
m – 3e4 kg	a_3	a_2	a_1	a_0	b_0
(I -3.8)	7.89e-4	1.23e-2	1	0	0.335

Example of digitization in MATLAB
Sampling step $T = 0.1$ *s*

```
>> T=0.1;
>> num=[0.335];
>> den=[7.89e-4 1.23e-2 1 0];
>> [x,y]=c2dm(num,den,T,'zoh')
x =      0    0.0289    0.0297    0.0110
y =   1.0000   -0.1328   -0.6568   -0.2104
```



(a) Response of the servomechanism in s



(b) Response of the servomechanism in z

Fig. 3. Response over time to step-type excitation signals

Using the zero-order extrapolator generates a dead time delay equal to half the sampling period:

$$H(s) = \frac{1-e^{-sT_e}}{s} \cong \frac{1-1+sT_e-(sT_e)^2/2+\ldots}{s} \tag{12}$$
$$= T_e e^{-s\frac{T_e}{2}}$$

SIMULATION NETWORK AND VIRTUAL INTERFACE OF THE ADJUSTMENT MODEL

To identify the optimal system adjustment process an application for co-simulating the process has been developed using AMESim and MATLAB simulation environments. In its structure there are found the models of the hydro unit, of the electrohydraulic drive system and the software component for automation/ adjustment of the process.

Hydraulic drive systems have been modelated in AMESim, while the adjustment algorithm has been developed using the Simulink toolbox, part of MATLAB software.

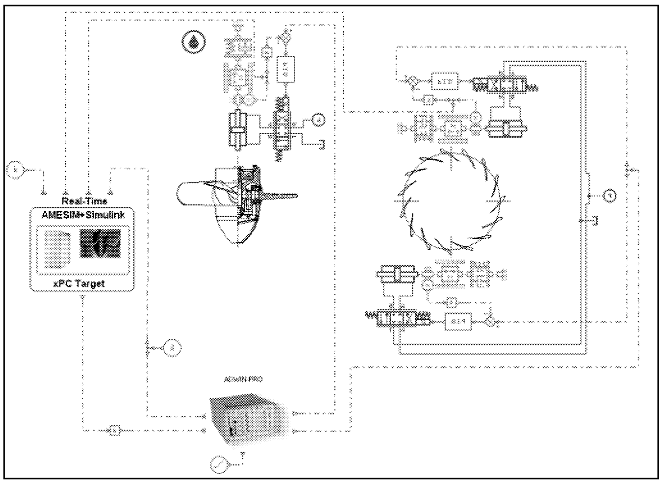


Fig. 4. Software simulation network (co-simulation AMESim - MATLAB)

Real-time simulation has been performed using in the first stage the mathematical model of the system (fig. 4) consisting of:
a) model of the hydro unit
b) model of the speed regulator (process computer)
c) model of the actuation system of the driving device
d) model of the actuation system of rotor blades.

The mathematical model of hydro unit has been experimentally identified by a set of data acquired from the unit of interest. Identification has ben performed using the IDENT toolbox of Matlab. Data aquired experimentally has been filtered (Fig. 5.) to remove the electrical nature noises, disturbances, and it has been divided into two sets (one used for identification itself and the other for verifying the mathematical model chosen).

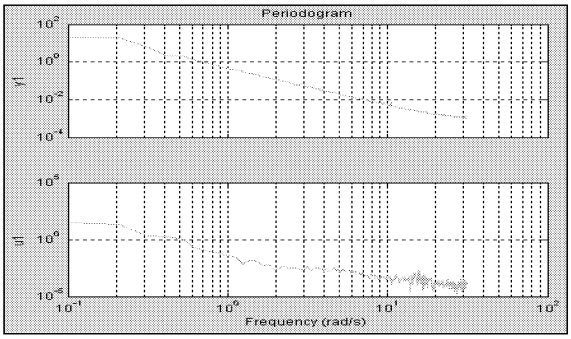


Fig. 5. Frequency spectrum of experimental data range

To remove unwanted components of the signal used to identify the parameters of a mathematical model there has been used a low-pass filter with the upper limit of 10 rad/s.
Identification, based on experimental data, of parameters of hydro unit mathematical model involves four steps:
a) acquisition of input / output data (Fig. 6.)
b) choice of model structure
c) assessment of model parameters
d) validation of the identified model (validation of structure and parameter values - Fig. 7.)

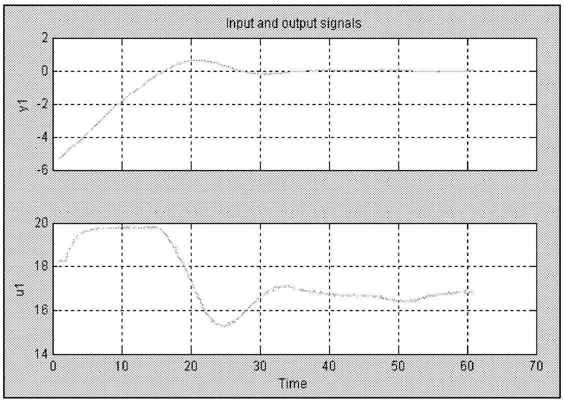


Fig. 6. Experimental data set used in the identification procedure.
u1 – position of driving device blades (%)
y1 - difference between prescribed rotational speed and rotational speed of the turbine shaft (%)

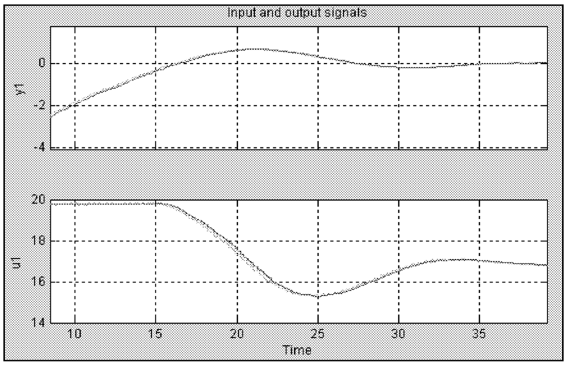


Fig. 7. Comparison of the signals used for the identification procedure, before and after applying the correction filter

Following experimental data processing by means of toolbox Ident threre was obtained a transfer function with the following coefficients:
>> ident
Opening System Identification Tool .... done.
Process model with transfer function

$$G(s) = K * \frac{1+Tz*s}{(1+Tp1*s)(Tp2*s)}$$

with K = 0.6256
Tp1 = 0.1254
Tp2 = 6.0856
Tz = -0.20365

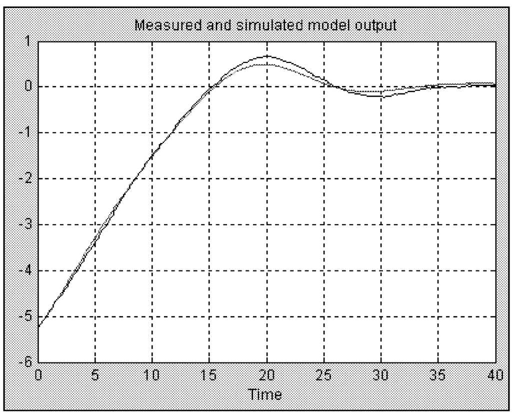


Fig.8. Comparison between the graph obtained with experimental data and the model identified (similar input data)

PID controller studied results from discretizing a continuous PID controller with the independent actions P, I and D.
It is considered the transfer function of the continuous controller [5]:

$$H_{PID}(s) = K\left[1 + \frac{1}{T_i s} + \frac{T_d s}{1 + \frac{T_d}{N} s}\right] \quad (13)$$

Parameters:
K – proportional amplification
Ti – integral action
Td – derivative action
Td/N – derivative action filtering
For digitization there is approximated derivation s through $(1\text{-}q^{1})$/Te and integration 1/s through Te / $(1\text{-}q^{-1})$. It results:

$$\frac{1}{T_i s} = \frac{T_e}{T_i} \cdot \frac{1}{1-q^{-1}}$$

$$T_d s = \frac{T_d}{T_e} \cdot \left(1-q^{-1}\right)$$

$$1 + \frac{T_d}{N} s = \frac{1}{1 + \frac{T_d}{NT_e} \cdot \left(1-q^{-1}\right)} = \frac{\frac{NT_e}{T_d + NT_e}}{1 - \frac{T_d}{T_d + NT_e} q^{-1}} \quad (14)$$

Entering these expressions in the transfer function of the continuous controller there is obtained:

$$H_{PID}(q^{-1}) = \frac{R(q^{-1})}{S(q^{-1})} = K\left[\begin{array}{c} 1 + \frac{T_e}{T_i} \cdot \frac{1}{1-q^{-1}} + \\ \frac{\frac{NT_e}{T_d + NT_e}(1-q^{-1})}{1 - \frac{T_d}{T_d + NT_e} q^{-1}} \end{array}\right] \quad (15)$$

Polynominals R(q^{-1}) and S(q^{-1}) have the form:

$$R(q^{-1}) = r_0 + r_1 q^{-1} + r_2 q^{-2}$$
$$S(q^{-1}) = (1-q^{-1})(1+s_1 q^{-1}) \quad (16)$$

where:

$$s_1 = -\frac{T_d}{T_d + NT_e}$$
$$r_0 = K\left(1 + \frac{T_e}{T_i} - Ns_1\right)$$
$$r_1 = K\left[s_1\left(1 + \frac{Te}{T_i} + 2N\right) - 1\right] \quad (17)$$
$$r_2 = -Ks_1(1+N)$$

To determine the numerical controller parameters:
- there is determined the digitized process
- performances are specified
- there are determined the coefficients of polynominals *R(q^{-1})* and *S(q^{-1})*

After verifying functionality of the real-time "software" model, there have been replaced in the model the software used to simulate the process computer and the model of the blade actuation system of driving device with their physical equivalent.

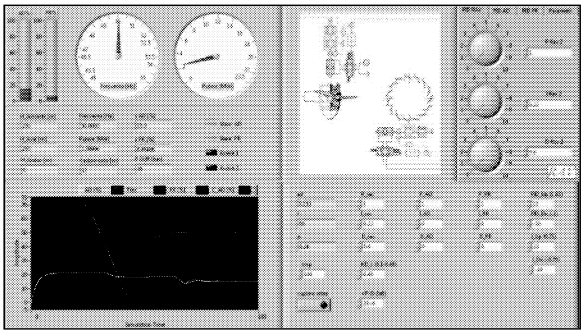


Fig. 9. Control Panel

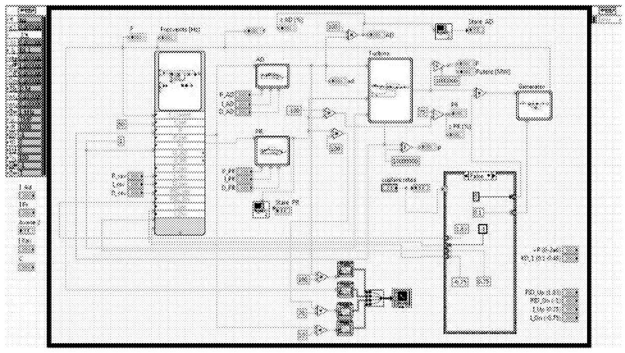


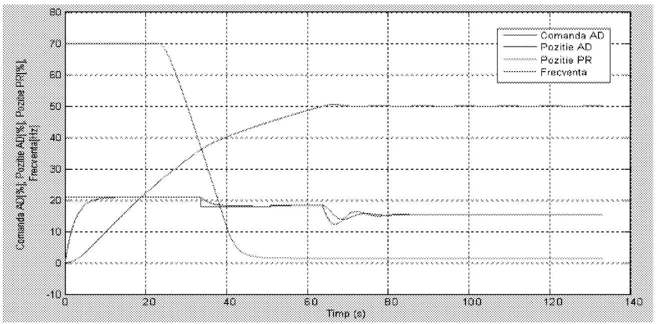
Fig. 10. Application Block Diagram

The objective of these tests was to validate the proposed model in order to use it for developing an "embedded" simulator of the hydro unit. The utility of this simulator is derived from the necessity for preliminary tuning of speed regulator coefficients. To allow easy modification of regulator parameters and recording of the obtained results a monitoring and control application has been developed using the graphical programming language LabVIEW (fig. 9, fig. 10).
Tuning of the regulator parameters was performed taking into consideration the system stability and restrictions related to its dynamic behaviour. Objectives targeted were to

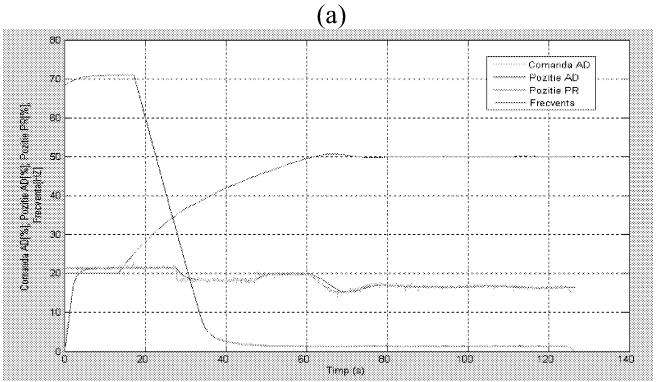
improve the dynamic characteristics and response to disturbances. Analysis criteria used were chosen taking into consideration the possibility to obtain information on system stability, and also on influence of the parameters which are to be optimized.

DYNAMIC CHARACTERISTIC OF THE SYSTEM

To perform simulations there has been developed the modeling network of the hydro unit, comprising these subsystems: turbine model, synchronous generator model, electrohydraulic driving subsystems and digital regulator - ASR. The results were compared with a set of experimental data acquired from a real system (fig.11.)



(a)



(b)

Fig. 11. Start with continuance at idling
(a) - simulated, (b) – experimental

CONCLUSION

As it can be seen in figure 11, results obtained by numerical simulation are comparable with experimental results. This confirms the possibility of using the system for preliminary "laboratory" tuning of such systems. The developed system can also be used for off-line testing/debugging of the firmware required by the equipment for automation/adjustment of these types of processes.

REFERENCES

1. *And Experimental Researches On A New Type Of Digital Electrohydraulic Speed Governor For Hydraulic Turbines,* Vasiliu, N., Călinoiu, C., Vasiliu, D., Ofrim, D., Manea, F. Theoretical 1st International Conference on Computational Methods in Fluid Power Technology, Methods for solving Practical Problems in Design and Control, Melbourne, 2003.

2.*Experimental identification of electrohydraulic servomechanism with virtual instruments technique*, Ion Guta Dragos, Catalin Dumitrescu, Ioan Lepadatu, Corneliu Cristescu, *Hidraulica nr.3 / 2010*, ISSN 1453-7303, pag.49-56.

3.*Using electrohydraulic drive systems in the execution of hydroelectric facilities*, PUB Teodor Costinel POPESCU, Dragos Daniel ION GUTA, Constantin CALINOIU, P.U.B. Dorin Pavel 2010.

4. *Research activities regarding energetic and functional advantages of hydraulic transmissions* P. Drumea, T.C. Popescu, Ion Guta Dragos, 10th International multidisciplinar scientific geoconference - SGEM 2010, ISSN 1314-2704, pag.323-330.

5. *System Identification and Control Design,* Landau, I.D. Prentice-Hall, 1990.

BIOGRAPHY

Dragos Daniel ION GUTA, Ph.D. graduated Mecanical Engineering Faculty from University Politehnica of Bucharest in 2002. He works as scientific researcher in fluid power systems at field at Hydraulics and Pneumatics Research Institute – INOE 2000 – IHP Bucharest, Romania; he is working also in the field of modeling and real-time simulation of the fluid control systems.

Teodor Costinel POPESCU, Ph.D., graduated Institute of Civil Engineering – Faculty of Fluid Systems in 1978, and Polytechnic Institute of Bucharest – Faculty of Mechanichal Engineerig in 1989, Fluid Machines Faculty. Since 1983 he is working as senior scientific resercher in fluid power systems field at Fluid Power Systems Research Institute – INOE 2000 – IHP Bucharest, Romania.

Petrin DRUMEA, Ph.D. in fluid power systems; he graduated Mechanics Faculty at Polytechnic Institute of Bucharest, senior scientific resercher, manager of Hydraulics and Pneumatics Research Institute / INOE 2000 – IHP Bucharest, Romania, President of the National Professional Association of Hydraulics and Pneumatics FLUIDAS, Romania.

Constantin CHIRITA, Assoc. Prof., Ph.D. in fluid power systems; he graduated of Machine Manufacturing and Industrial Management Faculty from Technical University "Gheorghe Asachi" of Iasi, manager of Hydraulics and Pneumatics Systems Engineering Department, Romania, manager of HYDRAMOLD, Romania, VicePresident of the National Professional Association of Hydraulics and Pneumatics FLUIDAS, Romania.

Self-tuning of PID Controllers Model and Simulation Based on Neuro-Predictive Control

Raaed K. Ibrahem
College of Electrical and Electronics Techniques
Foundation of Technical Education
Baghdad/Iraq
Email:raid_khalid2000@yahoo.com

KEYWORDS

Neural Network, Controller predictive, simulation, magnetic bearing

ABSTRACT

In this paper, have introduced a different self-tuning process for PID controllers based on neuro-predictive control. A bounded horizon optimal control problem will solve on-line, authorizing to analyze the tuning parameters of the PID controller. The suggested technique can realized on a PM electromagnets active magnetic bearing process and a comparison with traditional auto-tuning techniques assumed.

1. Introduction

Although most of the industrial processes are complex nonlinear systems, they are still controlled with classical PID control structures, which are tuned to give good results only around a fixed operating point. Under these circumstances, in order to obtain the optimal response over the entire operating range, on-line adaptation or self tuning of the controller is required, and several methods have been proposed in the last decade, e.g. [1], [2], [3]. In [1], the existent types of adaptive techniques are classified based on the fact that if the process dynamics are varying, then the controller hold compensate these variations by adapting its parameters. There are two types of process dynamics variations: predictable and unpredictable. The predictable ones are typically caused by nonlinearities and can be handled using a gain schedule, which means that the controller parameters are found for different operating conditions with an auto-tuning procedure that is employed thereafter to build a schedule. In this paper, a new self-tuning method for PID controllers designed to control processes with predictable dynamics variations is presented. The gain scheduling principle replaced by using a neural network based model that is capable to capture the predictable dynamics variations of the process. The neural network model is also used to develop a neural structure that predicts the future control error caused by process dynamics variations. The controller tuning parameters are calculated solving a finite horizon optimal control problem that minimizes the predicted control error. Real-life experimental results are given for a PM electromagnets active magnetic bearing plant, which demonstrate the practical benefits of this self-tuning method[3].

2. Description of the self-tuning procedure

The proposed self-tuning approach is based on two parallel control structures (see Fig. 1) that are synchronized with the reference clock of the predictable dynamics process in closed-loop with a PID controller. The upper structure uses a predictive control loop consisting of a neural predictor and a PID controller with adaptive tuning parameters. The predictive structure, with the sampling rate *Tp*, works faster than the real-time control loop supplying the predicted control error over a finite future time horizon. The tuning parameters are calculated at each sample time instant through the minimization of the predicted control error and the obtained values are used to update the tuning parameters of the real-time control loop. Thus, the controller parameters are adapted based on the Predictive optimization of the control system behavior and the desired performances can be achieved over the entire operating range.

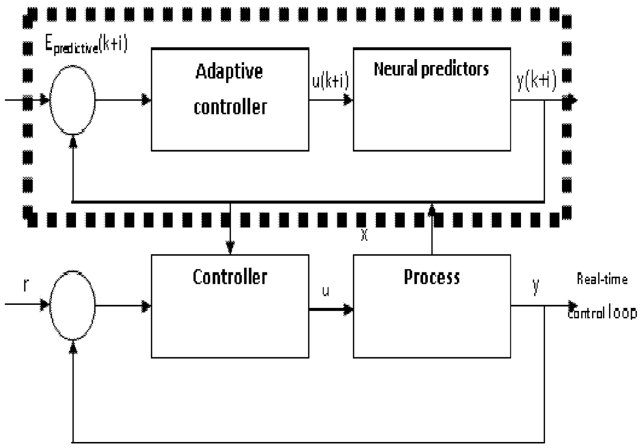


Fig.1. Neuro-predictive structure

2.1 Process neural model

Neuro-predictive control loop contains a neural network model, which models the real process with predictable dynamic variations. The use of neural networks for nonlinear process modeling and identification is justified by their capability to approximate unknown non-linear systems. A nonlinear model that includes a large class of non-linear processes is the following NARMAX model

$$y(k) = f[y(k-1),\ldots,y(k-n),u(k-d-1),\ldots,u(k-d-m)] \ldots (1)$$

Where, $f(.)$ is some nonlinear function, d is the dead time in T/sample, n and m are the orders of the nonlinear system model, u and y being the input and the output of the process with order n and m respectively. A neural network based model, NNARMAX, corresponding to the NARMAX model, may be obtained by adjusting the weights of multi-layer perceptron architecture with adequately delayed inputs [4].

Where, $f(.)$ is some nonlinear function, d is the dead time in T/sample, n and m are the orders of the nonlinear system model, u and y being the input and the output of the process with order n&m respectively. A neural network based model, NNARMAX, corresponding to the NARMAX model, may be obtained by adjusting the weights of Where f^N denotes the input-output transfer function of the neural network, which replaces the non-linear function f in (1) and, **u**$(k\text{-}d\text{-}1)$ and **y**$(k\text{-}1)$ have the following structure

$$y(k) = f^N[u(k-d-1), y(k-1)]$$

Where f^N denotes the input-output transfer function of the neural network, which replaces the non-linear function f in (1) and, **u**$(k\text{-}d\text{-}1)$ and **y**$(k\text{-}1)$ have the following structure

$$u(k-d-1) = [u(k-d-1)u(k-d-2)\ldots\ldots u(k-d-m)]^T$$

$$y(k-1) = [y(k-1)y(k-2)\ldots\ldots\ldots y(k-n)]^T$$

(3)

For a two layer network, the following expression is obtained from eq.(2)

$$y(k) = \sum_{j=1}^{p} W_j \sigma_j (W_j^u u(k-d-1) + W_j^y (k-1) + b_j) + b$$

(4)

Where p is the number of neurons in the hidden layer, σ_j is the activation function for the j-th neuron from the hidden layer, u_j^w the weight vector for the j-th neuron with respect to the inputs stored in **u**$(k\text{-}d\text{-}1)$, y_j^w the weight vector for the j-th neuron with respect to the outputs stored in **y**$(k\text{-}1)$, b_j the bias for the j-th neuron from the hidden layer, w_j the weight for the output layer corresponding to the j^{th} neuron from the hidden layer and b the bias for the output layer. Such structures with a single hidden layer are considered satisfactory for most of the cases. Since all the industrial processes are working in closed-loop, a closed-loop identification method has been used to obtain the neural model of the process. In order to capture all the nonlinear dynamics of the process, the training data had to be attained around several different operating points such that the entire variation range of the process output to be covered. For this reason, a stepwise reference was chosen and then summed with a pseudo random binary signal generated with a shifting register [5].

2.2 Neuro-predictive control loop

In order to obtain the predictable dynamics variations at the time instants k, a neural predictor based on the neural-based model of the process was used. A sequential algorithm based on the knowledge of current values of u and y together with the neural network system model gives the i-step ahead neural predictor

$$y(k) = \sum_{j=1}^{n} W_j \sigma_j (W_j^u u(k-d+i-1) + W_j^y (k+i-1) + b_j) + b \quad (5)$$

The future control **u**$(k\text{-}d\text{+}i\text{-}1)$ from (5) is obtained running the neuro-predictive control loop. Thus, at time instant k, the predicted output $y(k+i)$ is determined, for $i = N_1, N_2$ where N_1 and N_2 are the prediction horizons. If T_p is the sampling time with which the predictive control loop operates, this must satisfy: $(N_2 - N_1)T_p << T$ .Placing the neural model of the process to operate in the neuro-predictive control loop allows for transferring the current state $\boldsymbol{x}$ of the process to the neural predictor Figure (1) at each time instant k. Thus, at each time instant k, the predicted behavior of the process is obtained in the vector form

$$y_{predictive} = [y(k+N_1)y(k+N_1+1)\ldots\ldots\ldots y(k+N_2)]^T$$

(6)

The process output $\mathbf{y}_{predictive}$, predicted by the neural predictor, is used to calculate the predicted control error based on the controller set-point. Considering the discrete form of a PID controller,[6]

$$u(k) = u(k-1) + q_o e(k) + q_1 e(k-1) + q_2 e(k-2)$$

(7)

and the model (5), yields the following equation for the predicted control error

$$e_{predictive}(k+i) = \left(\begin{array}{l} \sum_{j=1}^{n} W_j \sigma_j W_j^u u(k-d+i-1) + \\ W_j^y y(k+i-1) + b_j + b \end{array} \right)$$

$$-r(k+i) \ldots\ldots (8)$$

where the vector **u**$(k\text{-}d\text{+}i\text{-}1)$ is a function of the tuning parameters vector $\mathbf{q}=[q_0\ q_1\ q_2]$. Minimizing the cost function

$$J = \frac{1}{2} \sum_{i=N1}^{N2} e^2_{predictive}(k+i) \quad (9)$$

3. Analysis of System Dynamic Model

Figure (2) shows the schematics of the active magnetic bearing system. It consists of a levitated object (rotor) and a pair of opposing E-shaped controlled-PM electromagnets with coil winding. An attraction force acts between each pair of hybrid magnet and extremity of the rotor [7]. The attractive force each electromagnet exerts on the levitated object is proportional to the square of the current in each coil and is inversely dependent on the square of the gap. Assuming a minimum distance to the length of the axis, the two attraction forces assure the restriction of radial motions of the axis in a stable way. The rotor position in axial direction is controlled by a closed loop control system, which is composed of a non-contact type gap sensor, a PID controller and an electromagnetic actuator (power amplifier). This control is necessary since it is impossible to reach the equilibrium only by permanent magnets. The rotor with mass m is suspended. Two attraction forces F_1. and F_2 are produced by the hybrid magnets. The applied voltage E

from power amplifier to the coil will generate a current i which is necessary only when the system is subjected to an external disturbance w. Equations governing the dynamics of the system are

$$F_1(y,i) + F_2(y,i) - mg + w = m\frac{d^2y}{dt^2} \quad (10)$$

$$E = Ri + N\frac{d}{dt}(\phi_1(y,i) + \phi_2(y,i)) \quad (11)$$

Under small disturbance, the above equation becomes

$$\Delta E = R\Delta i + N\frac{d}{dt}(\phi_1(y,i) + \phi_2(y,i)) \quad (12)$$

$$\Delta E = R\Delta i + N\left\{\frac{\partial(\Delta\phi_1 + \Delta\phi_2)}{\partial\Delta y}\frac{d\Delta y}{dt} + \frac{\partial(\Delta\phi_1 + \Delta\phi_2)}{\partial\Delta i}\frac{d\Delta i}{dt}\right\} \quad (13)$$

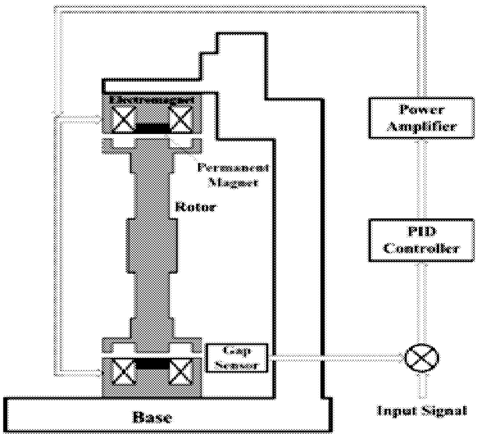


Fig.2. Configuration of PM electromagnets active magnetic bearing

If the weight of rotor is equal to the sum of these two attraction forces, the rotor will rotate on specific gap. According to Eq. (11), the disturbance equation at specific gap is calculated as follows

$$\Delta F_1(\Delta y,\Delta i) + \Delta F_2(\Delta y,\Delta i) + w = m\frac{d^2\Delta y}{dt^2} \quad (14)$$

And

$$\Delta F_1(\Delta y,\Delta i) = \frac{\partial\Delta F_1}{\partial\Delta y}\Delta y + \frac{\partial\Delta F_1}{\partial\Delta i}\Delta i \quad (15)$$

$$\Delta F_2(\Delta y,\Delta i) = \frac{\partial\Delta F_2}{\partial\Delta y}\Delta y + \frac{\partial\Delta F_2}{\partial\Delta i}\Delta i \quad (16)$$

Where y is the distance from gap sensor to bottom of rotor. R and N are the resistance and number of turns of the coil. φ_1 and φ_2 are the flux of the top and bottom air gap, respectively. We denote $\varphi = \varphi_1 + \varphi_2$ and $F = F_1 + F_2$. The system is linearlized at the operation point ($y=yo$, $i=0$) and described as follows

$$\frac{d^2\Delta y}{dt^2} = \frac{1}{m}\frac{\partial\Delta F}{\partial\Delta y}\Delta y + \frac{1}{m}\frac{\partial\Delta F}{\partial\Delta i}\Delta i \quad (17)$$

$$\frac{d\Delta i}{dt} = -\frac{R}{L}\Delta i - \frac{N}{L}\frac{\partial\Delta\phi}{\partial y}\frac{d\Delta y}{dt} + \frac{1}{L}\Delta E \quad (18)$$

$$\frac{d}{dt}\begin{bmatrix}\Delta y\\ \Delta y^{\bullet}\\ \Delta i\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\ a_{21} & 0 & a_{23}\\ 0 & a_{32} & a_{33}\end{bmatrix}\begin{bmatrix}\Delta y\\ \Delta y^{\bullet}\\ \Delta i\end{bmatrix} + \begin{bmatrix}0\\0\\b\end{bmatrix}E + \begin{bmatrix}0\\d\\0\end{bmatrix}w \quad (19)$$

Where

$$a_{21} = \frac{1}{m}\frac{\partial\Delta F}{\partial y} \qquad a_{23} = \frac{1}{m}\frac{\partial\Delta F}{\partial\Delta i}$$

$$a_{32} = -\frac{N}{L}\frac{\partial\Delta\phi}{\partial y} \qquad a_{33} = -\frac{R}{L}$$

$$b = \frac{1}{L} \qquad d = \frac{1}{m} \qquad L = N\frac{\partial\Delta\phi}{\partial\Delta i}$$

The partial derivatives are calculated from the experimental characteristics at the normal equilibrium operating point. The characteristic roots of the system is found.[7] This system has to be stabilized by a PID controller with appropriate controller parameters tuning.

4. Neural model of the plant

In order to estimate the parameters of the neural model, a training sequence was built such that the process output explores its whole operating range. Thus, a stepwise reference summed with a pseudo random binary signal was applied to the real time control loop and, by monitoring the control signal $u1$ and the process output y_1, a training sequence was obtained. Using the training sequence, a two layer neural network was trained off-line.

Model parameters m, L, N and R were estimated based on the collected input-output data and on the physical structure of the process (the inner flow loop and the tank). With a sampling rate of 2sec, it was found that the process has a delay N=2 and m=2, L=2, R=2. For the training and the validation of the neural network that models the process, the software instruments presented in [6] were used.

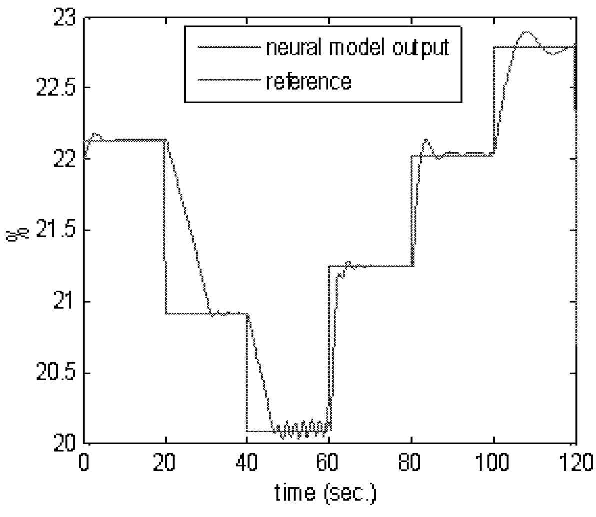


Fig.3.Process neural model validation

In Figure 3, the results of a closed-loop experimental validation of the neural model are plotted.

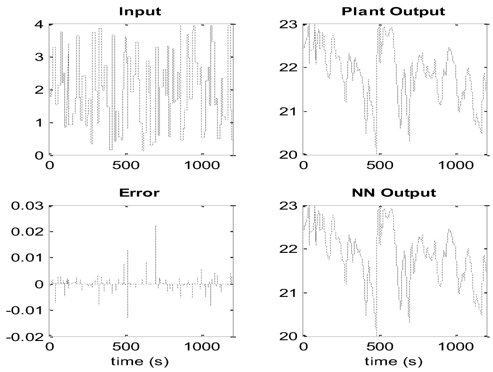


Fig. (4-a) training data for neural network

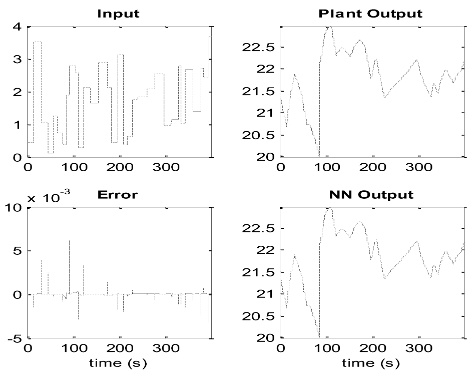


Fig. (4-b) validation data for neural network predictive control

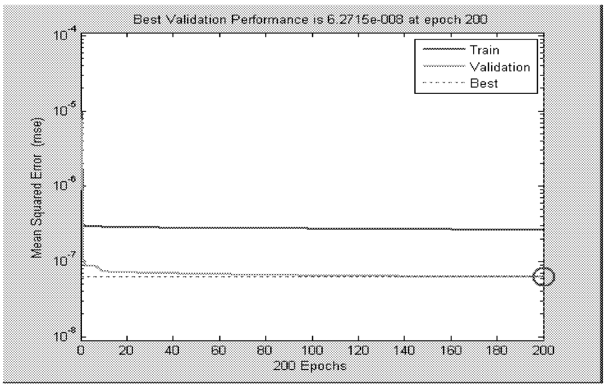


Fig. (4-c) performance for neural network predictive control

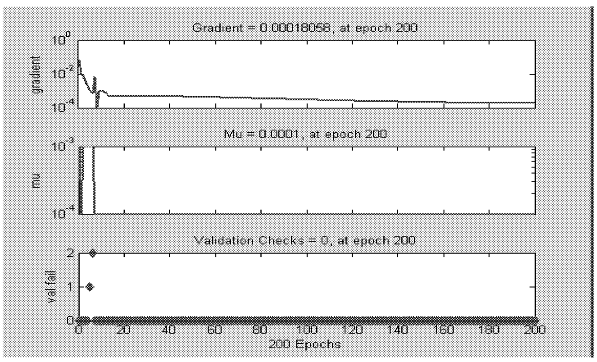


Fig. (4-d) training state for neural network

Predictive control

5. Conclusions

A neuro-predictive control based self-tuning procedure for PID controllers has been developed. The main advantage of the method consists in the on-line adaptation of the controller parameters and in the possibility to track different process operating regimes. The proposed method has been implemented on a benchmark real-life system with good results and a comparison with a *classical* auto-tuning method for PID controllers has been given.

6. References

Hagglund T., Astrom K.J., ***ndustrial Adaptive Controllers Based on Frequency Response Techniques***, *Automatica*, ol.27, no.4, 599-609, 1991.

Cheng-Ching Yu, ***Autotuning of PID Controllers, Springer-Verlag,*** London, 1999.

Astrom K.J. Hagglund T., ***PID Controllers: Theory, Design and Tuning***. New York, Instrument Society of America: Research Triangle Park, 1995, 120-125.

Chen S, Billings S.A. and P.M. Grant, ***Non- linear system identification using neural networks,*** Int. J. Control, vol 49, no.3 ,1013-1032,1990. Landau I. D. System Identification and Control Design, Prentice Hall, 1990.

Norgaard, M, **Neural Network Based System Identification Toolbox,** Technical Report 00-E-891, Department of Automation, Technical University of Denmark, 2000.

Hung C.Chen and Sheng Chang, ***Genetic Algorithms Based Optimization Design of a PID Controller for an Active Magnetic Bearing***, IJCSNS International Journal of Computer Science and Network Security, Vol.6 No.12, December 2006.

EXPERIMENTAL STUDY FOR MODEL-BASED FAULT DETECTION IN FLUID POWER APPLICATIONS

C. Angeli
Department of Mathematics and Computer Science
Technological Institute of Piraeus
Konstantinoupoleos 38, N. Smirni
GR-171 21 Athens, Greece
E-mail: c_angeli@otenet.gr or
c.angeli@computer.org

A. Chatzinikolaou
S. Patsi 62
GR-118 55 Athens, Greece
E-mail: achatzi@otenet.gr

KEYWORDS

Modelling, Simulation, Fluid Power systems, Fault detection

ABSTRACT

Model-based methods improve the reliability of fault diagnosis for technical systems especially when the diagnostic system is able to conclude under dynamic conditions. The diagnostic system presented in this paper is based in inconsistency between the actual process behaviour and its expected behaviour as described by an analytical model. The simulation results and the measured quantities from the actual system are compared on-line. The inconsistency between model and systems is exhibited in residual signals. Decision making about possible faults is performed based on the evaluation of these signals. The experimental results show the effectiveness of the proposed method.

INTRODUCTION

Studies on fault detection and isolation are attracting considerable attention in industrial production world due to the increasing complexity and productivity of modern industrial systems. For the complex highly automated industrial systems it is fundamental to be able to monitor the condition of the installation continually in order to detect faults and to locate deteriorated components.

A lot of industrial monitoring systems rely on the comparison of measured signals to specific thresholds. Many of them do not exploit the correlation existing between the different measured signals. Therefore they only allow the detection of significant deviations from operating conditions and they are not able to detect incipient deviations. These drawbacks motivate the development of diagnostic systems based on analytical models. The main idea behind such systems is to check the consistency between the measurements of different variables of the supervised system and the expected behaviour of this system as described by an analytical model.

In model-based fault detection a model (mathematical or heuristic) is employed to describe the nominal behaviour of the monitored system. The generated residual signals that indicate differences between the model's output and measured process output are interpreted and evaluated to isolate faults.

Model-based fault diagnosis offers a well established approach and many paradigms for engineering systems have been proposed by researchers such as (Patton et al. 2000, Frank et al. 2000, Isermann 2005, Korbicz et al. 2004, Gentil et al. 2004, Medjaher and Zerhouni 2009, Freddi et al. 2009). However the model-based fault diagnosis is build on a number of idealized assumptions for the engineering practice. One of them is that the model of the system is a faithful replica of a system dynamics. Another one is that disturbances and noise acting upon the system are known. In consequence a suitable residual evaluation technique should apply in order to minimise the false alarms and simultaneously maximize the sensitivity of the fault diagnosis system.

Fluid power systems are used in a variety of applications ranging from robotics and aerospace to industrial systems and they are becoming more complex in design and function. An efficient maintenance scheme is necessary for the reliability of the installation. In this paper a model-based approach for interaction of modelling information with relevant measured values of the actual system for the detection and diagnosis of faults in an electro-hydraulic drive system is presented. The dynamic behaviour of the actual system was modelled. Acquired values referring to pressure signals and to the angular velocity signal are compared with the relevant variable values of the simulation process. Previous research work in development of various kind of diagnostic functions for these systems is provided among other researchers by (Angeli 2008, Meuser and Schmidt 2006, Ghoshal and Samanta 2009 Muenchhof and Clever 2009, Kashi et al. 2006, Jelali, and Kroll 2003, Angeli and Chatzinikolaou 2005).

MODELLING PROCESS

The experimental layout used for this work is shown in Figure 1. This system is a typical drive system for a production machine, Figure 2.

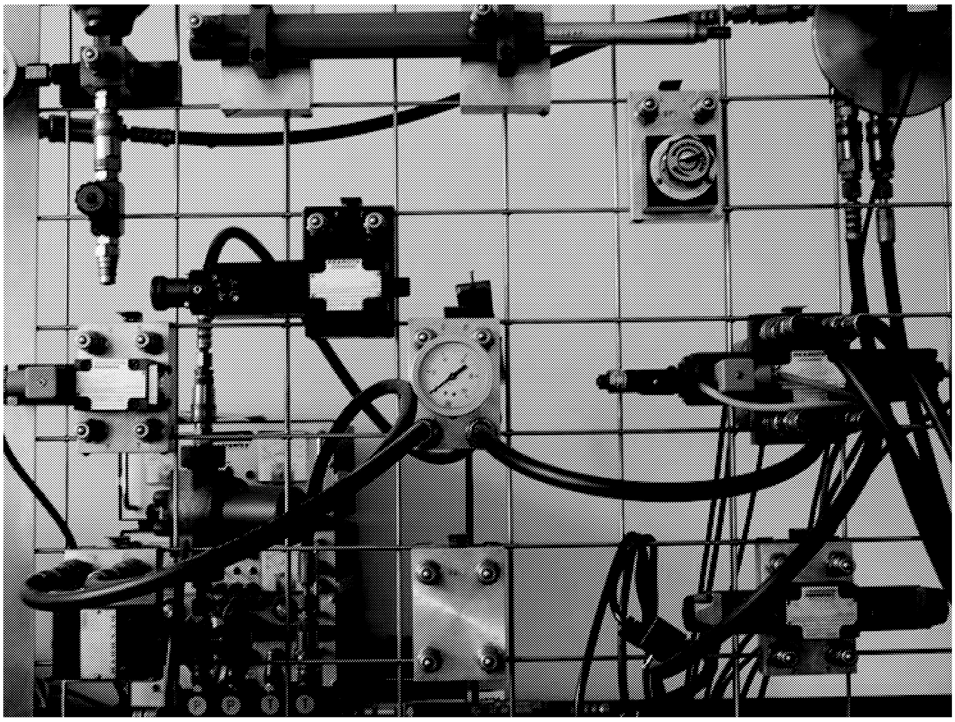


Figure 1: The experimental layout

This system consists basically of a hydraulic motor, a proportional 4-way valve and the connection pipes between them. The hydraulic motor is rotated by means of a hydraulic power unit. The proportional 4-way valve controls the actual flow rate to the hydraulic motor according to an input current and determines so the angular velocity, the acceleration and the deceleration of the hydraulic motor with an attached rotating mass J_m.

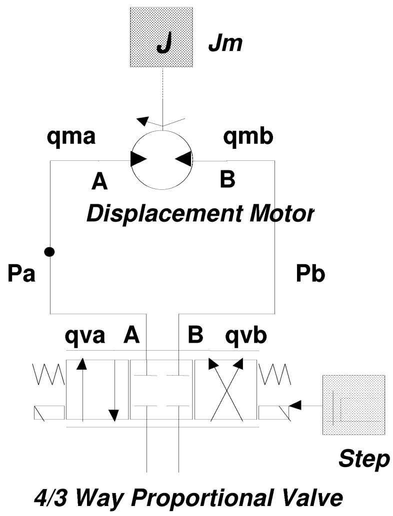


Figure 2: The actual system

This hydraulic system rotates the hydraulic motor and the attached load. The function diagram represents the operating curves of the motor for various values of the operating pressure. From these curves the actual flow in l/min for a demanded rotation speed in min^{-1} can be found. The flow curves, Figure 3, can be used for the estimation or validation of the volumetric efficiency of the hydraulic motor.

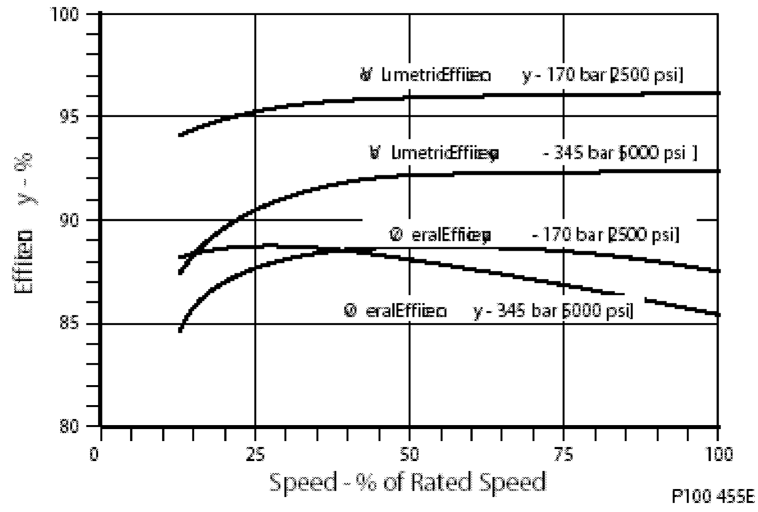


Figure 3: Motor performance as a function of operating speed

The fluid power system of Figure 2 drives the hydraulic motor through a cyclical routine which requires a high speed for a short time and then returns to a low speed according to a periodically changed voltage.

Assuming that the working pressure is constant, the variables of the system are following:

The pressure p_a at the port A of the hydraulic motor, the pressure p_b at the port B of the hydraulic motor, the rotation angle of the motor shaft φ, the angular velocity ω, the flows q_{va} and q_{vb} through the A and B ports of the proportional 4-way valve, the flows q_{ma} and q_{mb} through the ports A and B of the hydraulic motor, the input current to the proportional valve or the corresponding voltage to the amplifier of the proportional valve.

For the components which contain large quantities of oil as pipes and hydraulic actuators it may be assumed that the change in pressure in the dynamic state is proportional to the net inflow of oil, that is:

$$dp / dt = (E / V_o) \,.\, \Sigma Q$$

where:

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

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

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

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

Q_{out} is the outgoing flow from V_o.

In consequence the pressure increase in a pipe element at a junction is proportional to the algebraic sum of the incoming and outgoing flows and inverse proportional to the included oil volume.

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

The description of the dynamic behaviour takes also into account the non-linear character of hydraulic systems as well as the special characteristics of the hydraulic elements used so that the produced model represents the behaviour of the system elements more accurately. For the simulation process the 20-Sim software was used.

The model was validated in comparison to the real process measurements in order to determine experimentally the parameter values that include uncertainty and to define the acceptable limits of deviation between the measurements and the simulation results.

MONITORING PROCESS

The physical hydraulic system was connected with the computer using appropriate electronic and electrical devices, Figure 4. The proportional valve is controlled by a voltage of 0 to ± 10 V via the electronic amplifier of type VT5005.

The amplifier controls the proportional valve and transforms the input voltage U to the input current I. The input voltage U_2 = 0 to ±10 V is converted to a current I_2 = 0 to ±1,8 A. The amplifier VT5005 is connected to a power supply unit of +24 V DC.

The analogue input of 0 to ± 10 V for the amplifier comes from the card PCI-multi I/O. The input to the actual system is the voltage signal U from the control system and the outputs which are fed to the expert system are the angular velocity ω, the pressures p_a, p_b, and the state signals from the devices of the power unit.

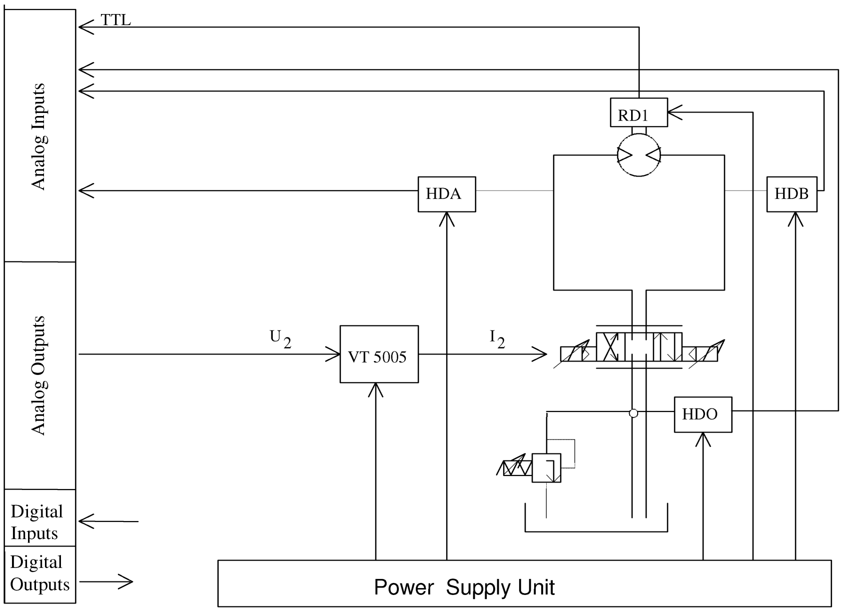


Figure 4: Control diagram of the Valve-Motor system

An effective monitoring unit for on-line condition monitoring requires suitable algorithms and methods to generate information about the condition of the system as well as the ability to communicate this information at any time. It also requires an environment which is capable to evaluate the information provided. The data acquisition system was developed using suitably connected modules of the the DASYLab software.

FAULT DIAGNOSIS PROCESS

Faults are detecting when the limits of a threshold on a residual signal generated from the difference between measured and estimate values are exceeding. In this case the possible deviation from zero is evaluated by a decision system. The role of the decision system is to determine whether the residuals differ significantly from zero and to decide which the faulty component is. Figure 5 presents a diagram of diagnosis process using the model-based approach.

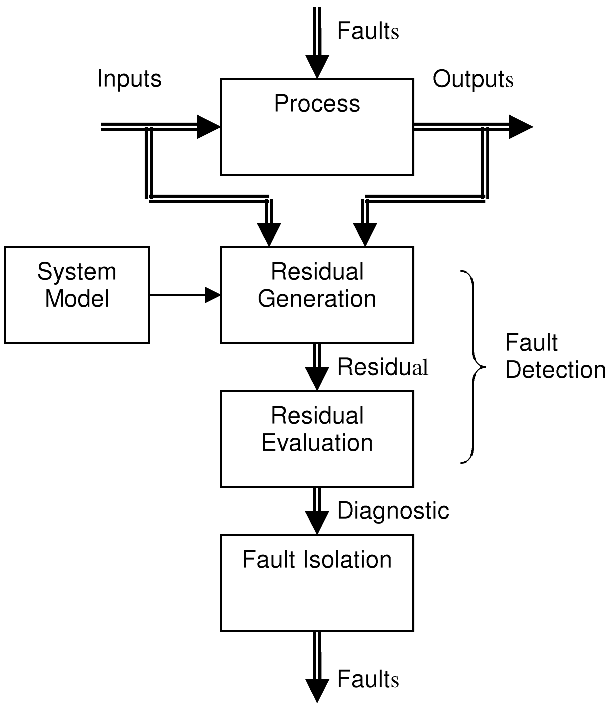


Figure 5: Fault diagnosis using system models

Faults are diagnosed by determining the changes in variables that have significant contribution to the diagnostic process and by relating these variables to specific process equipment faults.

This information is used by the expert system as input together with the text file information coming from the data acquisition process and experiential knowledge to determine more precisely a faulty element after transformation of the comparison results to linguistic variables. Experiential knowledge was complementary used to the scientific knowledge of the mathematical model in order to model more precisely the expert's reasoning activity, to gain the efficiency of heuristics and to respond to the real world requirements for diagnosis of faults.

EXPERIMENTAL RESULTS

The system was tested under various cases of malfunction using simulated faulty conditions.

In this process the effectiveness of the developed model was verified by examining the effects of changes in parameter values, used in the model, on the simulation results. Some parameters, as the friction torque M_r, the moment of inertia J_m and the oil elasticity E were varied. For a variation of $\pm$ 5 %, and $\pm$ 10 % of these parameters the variation of the simulation results was studied.

It was observed that the maximum deviations of the pressure p_a and p_b are approximately 0,5 bar for a variation $\pm$ 5 % of the moment of inertia J_m and 1 bar for a variation of $\pm$ 10 %. These variations can not affect the effectiveness of the fault detection process.

CONCLUSION

Fault detection and diagnosis is an important and challenging issue in many engineering applications and nowadays it continues to be an active area of research.

Faults in fluid power systems are often caused by an incipient leakage. In consequence model-based technologies that are able to detect the effects of leakage in systems and propose corrective actions are beneficial for the production technology. This method enables effective detection of process abnormalities by continuous monitoring of the systems changes and triggers fault diagnostic activities. The research identifies new opportunities to further development of automated procedures for these systems and as consequence a higher degree of reliability in an industrial environment.

The experimentation results show that the developed system is reliable and the method can respond to the requirements of the practice.

REFERENCES

Angeli C. 2008. "On-line Expert Systems for Fault Diagnosis in Technical Processes" *Expert Systems*, Vol. 25, No2. 115-132.

Angeli C. and A. Chatzinikolaou 2005. "Troubleshooting in Hydraulic Systems using knowledge-based Methods". *International Journal of Engineering Simulation*, Vol. 6, Nr.1. 24-29

Frank P.M, X. Ding and T. Marcu 2000. 'Model-based fault diagnosis in technical processes". *Transactions of the Institute of Measurement and Control* 22. 57-101.

Freddi A.; S. Longhi and A. Monteriu 2009. "A model-Based Fault Diagnosis System for Unmanned Aerical Vehicles". In *Proceedings 7th IFAC Safeprocess 09*, Barcelona, Spain

Gentil S.; J. Montmain and C. Combastel 2004. Combinining "FDI and AI approaches within causal-model-based diagnosis". *IEEE Transactions Systems Man and Cybernetics*, Vol. 34, No. 5. 2207-2221

Ghoshal S. and S. Samanta 2009. "Multiple Fault Isolation to a Servo-Valve Controlled Motor Transmission System". In *Proceedings 7th IFAC Safeprocess 09*, Barcelona, Spain.

Isermann R. 2005. "Model-based fault detection and diagnosis status and applications". *Annual Review of Control*, Vol. 29, No. 1. 71-85.

Jelali M. and A. Kroll 2003. "Hydraulic servo systems – Modelling, Indentification & Control – (advances in industrial control)". Springer Verlag, London.

Kashi K., D.; Nissing, D. Kesselgruber and D. Soeffker 2006. "Fault Diagnosis of an Active Suspension Control System". In *Proceedings 6th IFAC Safeprocess 06*, Beijing, China. 498-503

Korbicz, J.,; J. Koscielny; Z. Kowalczuk and Cholewa W. 2004. "*Fault Diagnosis: Models, Artificial Intelligence, Applications*". Springer Verlag, Berlin.

Medjaher K.. and N. Zerhouni 2009. "Residual-Based Failure Prognosis in Dynamic Systems". In *Proceedings 7th IFAC Safeprocess 09*, Barcelona, Spain.

Meuser M. and M. Schmidt 2006. "Standardized Measurements of Hydraulic Directional Control Valves", *o+p*, Vol. 02/06. 76-81.

Muenchhof M., and S. Clever 2009. "Neural-Net Based Fault Diagnosis for Hydraulic Servo Axes", *In Proceedings 7th IFAC Safeprocess 09*, Barcelona, Spain.

Patton R.; P. Frank and Clark R. 2000. "Issues in Fault Diagnosis For Dynamic Systems". Springer-Verlag, New York.

Interfacial Interactions between Multiwall Carbon Nanotubes and Nylon 6 Polymer Molecules in *in-Situ* Polymerized Nanocomposites

Yong K. Kim
Department of Bioengineering
University of Massachusetts
N. Dartmouth, MA 02747 USA
E-mail: ykim@umassd.edu

P. K. Patra
Department of Mechanical Engineering
Department of Biomedical Engineering
University of Bridgeport
Bridgeport, CT 06604 USA

KEYWORDS: MWCNT, Nanocomposites, WAXD

ABSTRACT

Multi-wall carbon nanotube (MWCNT) / nylon 6 nano-composites were prepared using in-situ polymerization technique. Micron sized fibers were then extruded from prepared nanocomposite using single screw extruder. The prepared fibers were characterized for the dispersion and orientation of carbon nanotubes using SEM, non-isothermal crystallization studies using DSC and mechanical properties. Wide angle x-ray diffraction (WAXD) and simple theoretical models of filled polymers were used to approximately determine the effect of the MWNTs on the molecular weight of the prepared nylon 6 polymer.

INTRODUCTION

Carbon materials are found in various forms such as graphite, diamond, carbon fibers, fullerenes and carbon nanotubes (CNTs). The carbon nanotubes have been shown to be easily oriented by mechanical stretch and other mechanisms by many researchers [1, 2]. However the dispersion and adhesion between the carbon nanotubes and matrix have been the subjects of considerable research for a long time. Melt-mixing, ultrasonication, solvent evaporation, chemical treatments and use of surfactants have been extensively used for the uniform dispersion of the carbon nanotubes [3, 4, 5, 6, 7.]. Most of the studies are unsatisfactory as the harsh treatments, higher energies and long exposures of the above treatments are generally required for the uniform dispersion of the carbon nanotubes inside the polymer matrix.

For carbon nanotubes a number of surface treatments have been proposed to improve adhesion at the CNT/matrix interface. Interfacial adhesion is typically improved by attaching various organic functional groups or molecules covalently or non-covalently. These techniques also involve the harsh treatments and strong chemical exposures to the carbon nanotubes [8, 9, 10].
Though these techniques have been shown to be successful in some cases the use of harsh treatments for the dispersion and adhesion can damage CNTs as well as degrade polymer resulting in less or no improvements in the properties of the final CNT-Polymer composites.

To avoid the damage to CNTs or polymer we prepared multi-wall carbon nanotube (MWNT)-nylon 6 nanocomposites using in-situ polymerization technique assisted with few minutes of ultrasonication. The short duration of the ultrasonication treatment was expected to separate carbon nanotubes in a low viscosity of the monomer solution without any damage to CNTs. Interaction of growing polymer chains with carbon nanotubes at nano or angstrom level was expected to render good adhesion properties.

Interfacial interactions between nanoparticles such as clays or carbon nanotubes and polymeric molecules are still far less understood. Polymer type, chemical functionalization of nanoparticles and the method of preparation nanocomposites notably affect these interactions. Strong interactions between hydrogen bonds of polymeric molecules and π-bond network of pristine carbon nanotubes were proposed by Lordi and Yao [11] based on the molecular dynamics studies. In this paper we report indications of such interactions in *in-situ* prepared nanocomposite of nylon 6/MWNT.

EXPERIMENTAL

Neat Nylon 6 and Nanocomposite Fiber Preparation

Purified (HCl treated) multi-wall carbon nanotubes (MWNTs) with average diameter of 30 nm were received from Catalytic Materials. ε-caprolactam, polyoxyethylene (POE) and N-acetylcaprolactam were purchased from Fisher Scientific. Chemicals, ε-caprolactam (40g), polyoxyethylene (POE, 0.88 g), multi wall carbon nanotubes (0.5 % and 1 % on weight of polymer) and N-acetylcaprolactam (20 drops) were taken into reaction flask. The mixture in the flask was then heated slowly. The molten solution was then subjected to 4 minutes of ultrasonication (Cole Parmer Ultrasonic Processor 750 WATTS Model, at 80 % amplitude, 5 sec. ON, 5 sec. OFF) to break the agglomeration of carbon nanotubes. 0.15 gm of NaH was added when the mixture in the tube melted and when temperature reached to 120 °C. Heating was kept continued for 4-6 minutes more until the reaction mixture became much more viscous.

Neat nylon 6 and nanocomposite fibers were prepared using a Brabender single screw extruder (Intelli-torque) and single hole fiber die (diameter = 0.016 inches L/D ratio = 4). Temperatures were set to 250, 230, 230 °C for zone 1, 2, and 3 respectively and screw speed was set to 3 rpm. The extruded fibers were stretched with draw ratio 3 and 4 using Instron fiber clamps at room temperature. Stretched samples were used for further characterization.

Characterization of the Neat Nylon 6 and Nanocomposite Fibers

The prepared nanocomposite and nanocomposite fibers were observed using SEM (JEOL JSM-5610). Samples were characterized for crystallization studies using DSC (Q 1000 TA Instruments) and for mechanical properties using an Instron testing machine. X-ray diffraction studies were done on the identically prepared films of the neat nylon 6 and nanocomposites to understand the effect of carbon nanotubes on the morphology of the nylon 6 crystals. Bruker wide angle X - ray diffraction instrument was used to analyze all the samples.

RESULTS AND DISCUSSION

MWNT Dispersion

The ultrasonic treatment given to the solution of molten ε-caprolactam and multiwall carbon nanotubes found to have significant impact on the interactions between the carbon nanotubes and ε-caprolactam molecules. Just few minutes of ultrasonication produced very homogenous solutions of carbon nanotubes. The solution was stable for a long period of time and remained homogenous over the period of polymerization (2-4 min.) after the addition of initiator, NaH. The ultrasonication creates tremendous amount of energy and ε-caprolactam molecules are possibly forced through the carbon nanotubes bundles and very close to the surface of carbon nanotubes which can assist the setting up the interactions between hydrogen bonds polymeric molecules and π-bond network of pristine carbon nanotubes. The presence of such interactions has been indicated in crystallization, X-ray diffraction and tensile testing studies which will discussed later. The SEM studies conducted on the cross section of the nanocomposite fibers showed well separated carbon nanotubes as shown in Figure 1. The energy created by ultrasonication separates the carbon nanotube bundles which remain separated even after polymerization as polymer chains grow in between well separated carbon nanotubes. The detailed mechanism of dispersion has been discussed in our previous publication [12, 13].

Figure 1: Dispersion of MWNTs: SEM image of cross section of broken fibers

DSC Crystallization Study

Neat nylon 6 and nanocomposites samples were tested for melting and crystallization properties. Though melting points of neat nylon 6 and nanocomposites samples were same, nanocomposite samples showed wider melting peak which indicated the presence of smaller and/or defective crystals. The crystallization behavior of both the samples was observed in the subsequent cooling cycle. The sample was kept isothermal at 240°C for 5 minutes to completely erase the pervious thermal history. In cooling cycle (Figure 2) it was observed that crystallization in nanocomposites starts significantly earlier (about 1 min earlier for -40°C/min cooling ramp) or at higher temperature (about 25°C higher) compared to neat nylon 6 samples. This indicates that carbon nanotubes act as nucleating agents at the early stages of crystallization. However at the later stages of crystallization the rate of crystal growth was significantly hindered. This is apparent from the wider crystallization peaks as shown in Figure 2 and the plot of relative crystallinity (%) vs. time (Figure 3).

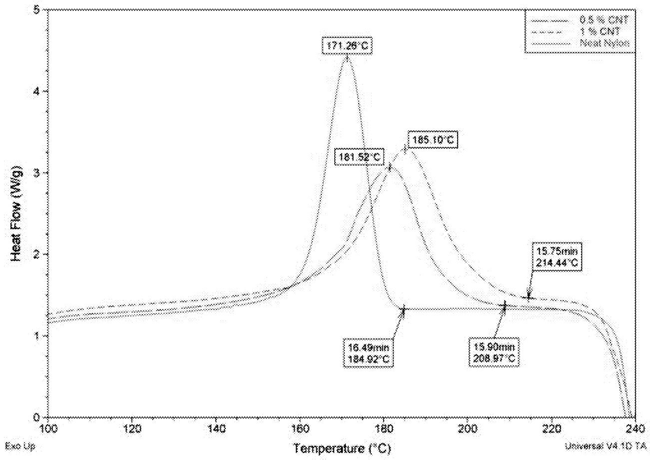


Figure 2: DSC crystallization curves for neat nylon 6 and nanocomposites

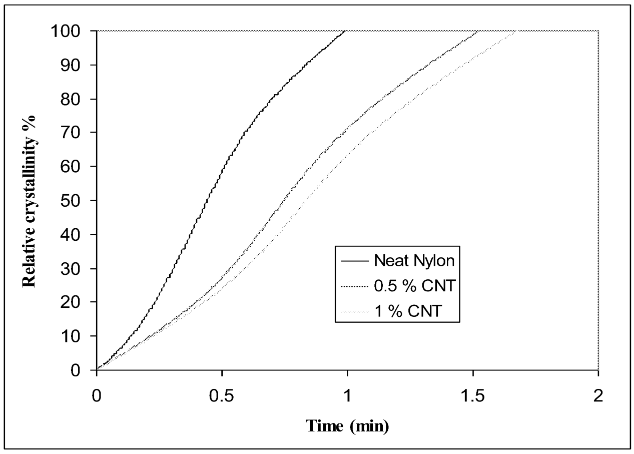


Figure 3: Relative crystallinity with time for neat nylon 6 and nanocomposites

The crystallization half time ($t_{½}$) for nanocomposites samples, 0.76 and 0.83 min. for 0.5 % and 1% MWNTs samples respectively was significantly higher compared to that of neat nylon 6 samples (0.42 min.). Furthermore the Avrami's paramters (growth rate parameter K and nucleation parameter n) were determined from the plot of $ln\,[-\,ln\,(1-X_t)]$ vs. $ln\,t$ (Figure 4), where X_t is the relative crystallinity at time t. The Avarami equation was used to determine values of n and K from the slopes and the interception of the best fit (**equation 1**).

$$ln\,[-ln\,(1-X_t)] = n\,ln\,t + ln\,k \qquad (1)$$

As shown in Table 2 the nanocomposite samples showed lower values of growth rate parameter K which confirms the hindered crystal growth. The values of n around 1.4-1.5 suggest the rod shaped crystal geometry and thermal nucleation type [13].

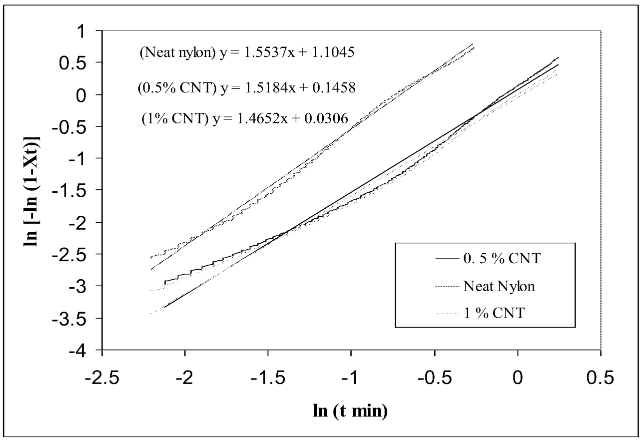


Figure 4: ***ln [- ln (1-X$_t$)]*** **vs.** ***ln t*** **curves for neat nylon 6 and nanocomposites**

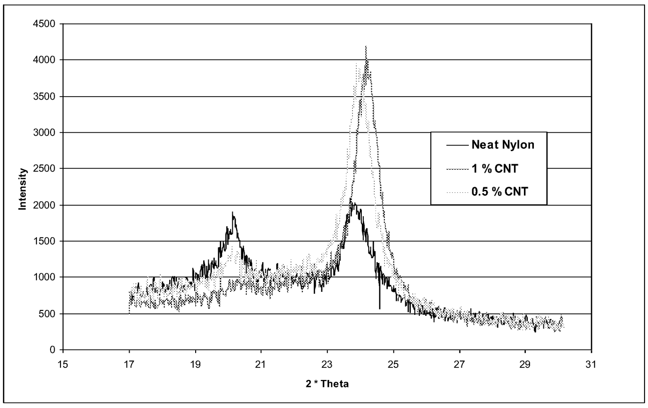


Figure 5: Wide angle X-ray diffraction spectra for neat nylon 6 and nanocomposite samples

Table 1: Avrami parameters for neat nylon 6 and nanocomposite

Sample	t ½ (min)	Avrami parameters	
		K (min^{-1})	*n*
Neat Nylon	0.420	1.054	1.548
0.5 wt. % MWNT	0.760	0.176	1.497
1 wt. % MWNT	0.825	0.039	1.454

WAXD Results

We performed wide angle X-ray diffraction studies on the neat nylon 6 and nanocomposite samples to understand the effect of carbon nanotubes on the crystal structure of nylon 6. Figure 5 shows the Wide angle X-ray diffraction spectra for neat nylon 6 and nanocomposite samples. Nylon 6 alpha crystal structures give two strong and characteristic diffraction signals at spacing 0.44 nm (2θ ≅ 20 deg.) and 0.37 nm (2θ ≅ 24 deg.), respectively. A number of studies suggest that interfacial interactions with nanotubes result in an interfacial region of polymer

With morphology and properties different to the bulk. Kohan [14] showed that these diffraction signals represent a projected inter-chain distance within a hydrogen-bonded sheet and inter-sheet spacing, respectively. However, both 0.5 wt. % and 1 wt. % nanocomposite samples showed reduced intensity for 0.44 nm spacing and huge increase in intensity for 0.37 nm inter-sheet spacing. The decreased intensity of the 20° peaks suggests that (200) planes are parallel to the surface of the film which means preferred orientation of the crystals in the sample. There is random orientation of crystals in the case neat nylon 6 which is evident by the fact that 20 deg. and 24 deg. peaks have similar intensities. As two films were prepared in exact identical ways, the preferred orientation of the nylon 6 crystals should be something to do with the inclusion of carbon nanotubes. It is now well known that carbon nanotubes or nanoparticles can be oriented in matrix materials even with the moderate level of shear forces. Thus, it appears like carbon nanotubes first get oriented along the direction of film due to the shear forces involved during the compression of film and the crystals forms along the curved surface of the carbon nanotubes and hence are preferentially oriented along the film direction.

.

Tang and Xu [15] prepared PPA (polyphenylacetylene)-wrapped MWNTs by polymerizing phenylacetylene monomer in the presence of carbon nanotubes and observed the helical wrapping of PPA around the nanotubes due to the strong interactions between the acidic acetylene hydrogen of the monomers and the Π electrons of the nanotube surface. These hydrogen atoms interact with the fluctuating dipoles of the Π bonds, which are oriented, normal to the surface of the carbon nanotubes. They also mention that since PPA can exist in various forms, *in-situ polymerization* was necessary to wrap the nanotubes, which influences the polymer morphology.

As in this research work nanocomposites were prepared in same way (*in-situ polymerization),* the reduced rate of polymerization and reduced molecular weight can be explained with results of the interactions between the hydrogen atoms of the ε-caprolactam monomer and the Π electrons of the nanotube surface. Reduced crystallization growth and altered crystal morphology observed in X-ray diffraction study further indicates the presence of such interactions between polymer and carbon nanotube.

Furthermore molecular dynamics study conducted by Lordi and Yao [11] showed that such hydrogen bond interactions with Π electrons of the pristine nanotubes can result in very strong interface between polymer and carbon nanotubes provided the polymer molecules are wrapped helically around the carbon nanotube .

Thus improved mechanical strength and modulus can be results of such strong interactions. Strong interface enables the proper stress transfer from nylon 6 matrix to the carbon nanotubes to give reinforcing effects.

Similar to the modeling phase the interaction of more than one user has influence on the execution. The next paragraph elaborates possible conflicts which might occur in a multi-user simulation environment.

Improvement in Mechanical Properties

The mechanical properties of nylon 6/MWNT nanocomposite fibers are shown in Table 2. All the samples showed high variability in properties. Variability was higher in the case of nanocomposite samples. The reasons behind the high variability can be attributed to the absence of the deaeration, mixing of the purging compound and localized MWNTs agglomerates in the case of nanocomposites.

Table 2: Average mechanical properties for nylon 6 and nanocomposite fibers

Properties	Neat Nylon DR-3	Neat Nylon DR-4	NC 0.5 DR-3	NC 0.5 DR-4	NC 1 DR-3	NC 1 DR-4
Modulus (MPa)	1149	1302	1553	1837	1353	1641
SD Modulus (MPa)	96	112	189	258	118	169
Strength (MPa)	205	224	261	327	226	277
SD Strength (MPa)	12	10	29	30	20	34
Breaking strain %	33	24	39	23	38	17
SD Breaking strain %	6	5	7	4	7	3

Note: DR - draw ratio, NC – nanocomposite with 0.5 wt. % MWNTs, NC 1 - nanocomposite with 1 wt. %. MWNTs, SD - Standard dev.

Table 3: % Change in mechanical properties of nanocomposite samples compared to neat nylon 6 samples

		% Change			
		NC 0.5 DR-3	NC 0.5 DR-4	NC 1 DR-3	NC 1 DR-4
Modulus	Neat Nylon DR-3	**+35**		**+18**	
	Neat Nylon DR-4		**+41**		**+26**
Strength	Neat Nylon DR-3	**+27**		**+10**	
	Neat Nylon DR-4		**+46**		**+24**
Strain % at break	Neat Nylon DR-3	**+19**		**+14**	
	Neat Nylon DR-4		**-7**		**-31**

Table 3 shows the % change in mechanical properties due to reinforcement of the carbon nanotubes in neat nylon 6. The maximum increase of 41 % and 46 % in modulus and strength respectively were observed in case of nanocomposite fibers with 0.5 wt. % MWNT stretched 4 times. Less improvement can be observed in case of nanocomposites with 1 wt. % MWNT likely due to the lower molecular weight of the nylon from the interference of CNTs to polymerization. Improvements in modulus and strength of nanocomposites confirmed that there exists a strong interfacial interaction between nylon 6 matrix and embedded carbon nanotubes.

CONCLUSION

Nylon 6/MWNT nanocomposites were prepared by an in-situ polymerization technique. Different studies were done on the nylon 6/MWNT nanocomposites and extruded nanocomposite fibers and the following conclusions were made.

1. SEM analysis of the cross section of the stretched and broken nanocomposite fibers showed that carbon nanotubes were well-separated in the polymer matrix. Thus in-situ polymerization technique was to found to be effective technique to separate carbon nanotubes at low ultrasonication energy and time and without any use of solvent or chemical treatment. Carbon nanotubes also appeared to be nearly oriented along the direction of the fiber axis due to the effect of the extrusion and stretching.

2. Crystallization studies of the nylon6/MWNT nanocomposite fibers showed that the diffusion of nylon 6 molecules was significantly restricted, which resulted in reduced crystal growth rate.

3. In non-isothermal crystallization studies, results showed that, though carbon nanotubes acted as nucleating agents, rate of crystallization growth was reduced significantly. X-ray diffraction study showed the reduced d-spacing for nylon 6 α-crystal geometry due to irregular hydrogen bonds network in the crystals or twisted crystal morphology. Reduced crystallization growth and altered crystal morphology further indicate the interactions between nylon 6 polymer and carbon nanotubes.

4. Tensile testing of neat nylon 6 and nanocomposite fibers showed that reinforcement of nylon 6 fiber by carbon nanotubes increased the modulus and the strength of the fibers significantly (95% confidence level). Compared to the neat nylon fibers, maximum improvements, 41% in modulus and 46% in strength were observed for the 0.5 wt. % MWNT nanocomposite fibers which were stretched four times. But improvements were reduced to the 26% in modulus and 23% in strength for the 1 wt. % MWNT nanocomposite fibers, stretched four times, possibly due to the reduced molecular weight (23%). The results further showed that additional stretching of the fibers from 3 times to 4 times, possibly further oriented the carbon nanotubes and resulted in additional improvements in modulus and strength.

FUTURE WORK PLANNED

We are planning to develop a molecular model of the interaction between the nylon 6 molecules and MWCNT surfaces to understand the observed interfacial phenomena in future.

ACKNOWLEDGEMENT

The authors wish to thank to Professor Steve Warner for his suggestions. We also acknowledge Dr. Terry Baker of Catalytic Materials for providing us MWNT nanofibers. Finally authors are grateful to the National Textile Center (NTC) for funding under US Department of Commerce grant no. 0207400.

REFERENCES

1. Michael Sennett, Elizabeth Welsh, J. B. Wright, Appl. Phys. A, Vol. 76, pp 111, 2003.
2. Jin, L., C. Bower And O. Zhou, Applied Physics Letters, Vol. 73, pp 1197, 1998.
3. Bower, C., R. Rosen, J. Lin, J. Han and O. Zhou, Applied Physics Letters, Vol. 74, pp 3317, 1999.
4. Sandler J, Shaffer MSP, Prasse T, Bauhofer W, Schulte K, Polymer, Vol. 40, pp 5967, 1999.
5. Gong, X., J. Liu, S. Baskaran, R. Voise And J. Young, Chemistry Of Materials, Vol. 12, pp 1049, 2000.
6. Qian, D., E. Dickey, R. Andrews, Applied Physics Letters, Vol. 76, pp 2868-2870, 2000.
7. R. W. Siegel, S. K. Chang, B. J. Ash, J. Stone, P. M. Ajayan, R. W. Doremus, Scripta. Mater., Vol. 44, pp 2061, 2001.
8. http://www.univie.ac.at/spectroscopy/funcars/funcars.html
9. Ajayan, P M., Ebbesen, T. W., Nature, Vol. 362, pp 522, 1993.
10. S C Tsang, Y K Chen, P J F Harris and M L H Green, Nature, Vol. 372, pp 159, 1994.
11. Lordi, V. And N. Yao, *Journal of Materials Research*, Vol. 15, pp 2770, 2000.
12. Kim, Y.K, P. K. Patra, Chapter 14 MWNT-nylon6 nanocomposites from polymerization, Brown and Stevens (ed.), Nanofibers and nanotechnology in textiles, CRC, Boca Raton (2007)
13. Mhetre S., Kim Y., Patra P., Mat. Res. Soc. Symp. Proc., **788**, p. L11.17.1, 2003
14. Kohan M., Nylon Plastics Handbook, Hanser/Gardner Publications Inc., Cincinnati, 1995.
15. Tang, B and H. Xu, *Macromolecules*, Vol.32, pp 2569, 1999.

BIOGRAPHIES

YONG K. KIM is a chancellor professor of materials and textiles since 2003. He was the co-chair of the Department of Materials and Textiles at the University of Massachusetts – Dartmouth from 2002 to 2004. He earned Ph.D. in Fiber and Polymer Science from the North Carolina State University at Raleigh, NC in 1981 before joining the University of Massachusetts-Dartmouth. His research interests are nano/micro composites, biotechnology and digital printing.

PRABIR K. PATRA is an assistant professor of Mechanical Engineering and Biomedical Engineering at the University of Bridgeport, CT since 2009. He was a postdoctoral research associate in professor Kim and Warner's nano composites group at the University of Massachusetts Dartmouth from 2002 to 2007. He then worked for Prof Ajayan's nanomaterials group at the Rice University in Houston TX from 2007 to 2009 before joining the Bridgeport University. He earned his Ph.D in Material Science and Engineering from Indian Institute of Technology at Kharagpur. His expertise is in nanoscale materials and electroactive materials.

TRAINING SIMULATION

A FUZZY PETRI NET APPROACH TO ESTIMATE THE LEARNING OF SKILLS

Brigitte Finel*
Sophie Hennequin*
Nidhal Rezg**

*Ecole Nationale d'Ingénieurs de Metz
1, route d'Ars Laquenexy
F-57078 Metz cedex 3
**UFR MIM - Université Paul Verlaine de Metz
Ile du Saulcy
F-57045 Metz
E-mail: finel@enim.fr, hennequin@enim.fr, rezg@univ-metz.fr

KEYWORDS
Training system, modeling learning of skills, fuzzy Petri nets, schedule design.

ABSTRACT

In this paper, a fuzzy Petri net is proposed to represent the learning of skills helping the schedule design of training systems. The basic idea is that the concepts of motivation and good learning of skills are somehow difficult to model. This modeling is necessary however, to change and to improve the training system but also to implement a good schedule (not too costly and efficient). This work has been initiated by our National Engineering School (ENIM) in order to reduce our total costs and to assure our students the learning of sufficient and efficient skills. This work is in progress, and in this paper, we only present our fuzzy Petri net.

INTRODUCTION

Since its definition (Zadeh, 1965), fuzzy logic has been largely used and developed for the modeling of subjective, imprecise, vague, and uncertain concepts and also for the control of complex systems (Zimmermann, 1990, Yager 1994). However, fuzzy logic could provide sometimes a very complex model (in the sense of a large definition of fuzzy rules) which is timeless computational. Furthermore, the definition and validation of logical properties is not always simple with the theory of fuzzy logic (Vasant, 2004).

On the other hand, Petri nets have shown their ability to represent discrete events systems (Murata, 1994). Unfortunately, Petri nets can not correctly represent the notion of uncertainty even if stochastic Petri nets can integrate randomness. That is why fuzzy Petri nets have been defined (Lipp, 1994) and (Chen, 1992). Since those early works, many different types of fuzzy Petri nets have been proposed in the literature (Aziz, 2010). In this paper, the objective is not to propose a new kind of fuzzy Petri net but only to use a simple structure of fuzzy Petri nets to represent human factor in the learning of a course.

Indeed, a lot of works have been developed to model and simulate timetabling (Burke, 2007 and Qu, 2009 and Petrovic, 2005) or to reduce total costs in training systems, but to the best of our knowledge, no study has been done to represent students who are both customers but also the "product" (Renauld, 2010) on which training systems work.

In this article, a fuzzy Petri net is defined to represent the learning of skills for a given course. The fuzzy Petri nets concept is presented in section 2. In section 3, we present our approach based on a basic fuzzy Petri net. Finally, we give some conclusions and perspectives.

BASIC CONCEPTS

Since its development by C. A. Petri in 1962, Petri nets (PNs) have been largely used to model, to real time control and to supervise discrete events systems (such as manufacturing systems, computer systems, telecommunication, robots, etc.) because of their graphical representation associated with a mathematical description. Petri nets are a powerful and practical tool to manage conflicts, concurrent events, etc. A very good review on Petri nets could be found in (Murata, 1994). Systems, however, are not always easily represented by precise mathematical models. So, the concept of fuzzy Petri nets has been defined to allow the modeling of imprecise data and/or uncertainty (Lipp, 1994). A good classification on different kind of fuzzy Petri nets is given in (Aziz, 2010).

A simple fuzzy Petri net (FPN) is an oriented graph given by $FPN=(P,T,D,I,O,f,\alpha,\beta)$ where:

- $P=\{p_1,p_2,\dots,p_n\}$ is a finite set of places,

- $T=\{t_1,t_2,\dots,t_m\}$ is a finite set of transitions,

- $D=\{d_1,d_2,\dots,p_n\}$ is a finite set of propositions,

with $P\cap T\cap D=\varnothing$ and $|P|=|D|$,

- I and O are the sets of input and output functions such as $I:P\rightarrow T$ is the set of the arcs from places to transitions and $O:T\rightarrow P$ is the set of the arcs from transitions to places,

- $f:T \rightarrow [0,1]$ is an association function, a mapping from transitions to real values between zero and one,

- $\alpha:P \rightarrow [0,1]$ is an association function, a mapping from places to real values between zero and one,

- $\beta:P \rightarrow D$ is an association function, a bijective mapping from places to propositions.

Compared with PN, the principle of a FPN is the following (Figure 1) : a variable (crisp) is modeled by a first place, then a first transition is given to make the fuzzification (i.e. transform the crisp variable into a fuzzy variable which corresponds to a second variable then the modeling of the system is given in an fuzzy environment (the Petri net structure) to evaluate the fuzzy value of the output variable, thereafter a last transition is given to make the defuzzification to transform the fuzzy output into a crisp value given in a last place.

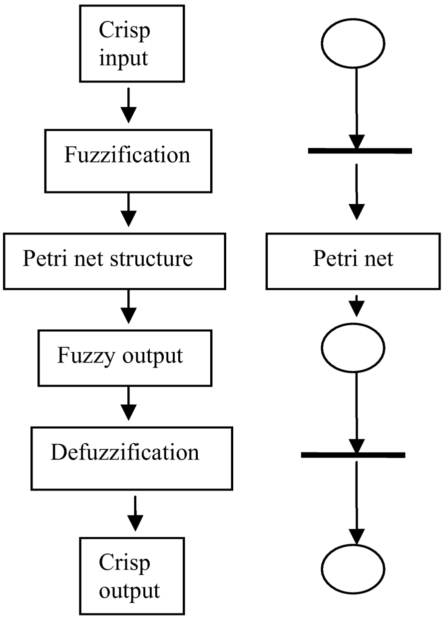


Figure 1 Principle of a FPN

In the following section, we present our fuzzy Petri net for the modeling of skill acquisition and the validation of the learning of a given course.

PROPOSED FUZZY PETRI NET

The basic idea in this paper is to represent by a simple fuzzy Petri net the fact that a course is known by students or not, i.e. that a course skills are acquired or not. In this paper, we limit our study to the design phase of a training system and we suppose that the formalization of the training system is always done (we know what kind of material, resources and prerequisites are necessary for a course). Furthermore, we suppose, in order to facilitate the reading (to simplify our proposed fuzzy Petri net), that the prerequisite of a course is offered by only one other course. We also suppose that all needed resources are available for the considered course (in this paper, we do not focus on timetabling which has already been defined by other colleagues).

Our fuzzy Petri net is composed by two input variables which are: acquisition of the previous course, denoted by Ac_{n-1} and the percentage of motivation, denoted by M. In terms of fuzzy logic, Ac_{n-1} is defined by 5 fuzzy sets, given by triangular functions, which are: *VL, L, M, H, VH* (Figure 2). This first input is not fuzzy but a fuzzification step is necessary to define our FPN. The second input is given by also five fuzzy sets: *VL, L, M, H, VH* (Figure 3) which are defined by Gauss laws. The output of our fuzzy Petri net model is the acquisition of the considered course, denoted by Ac_n.

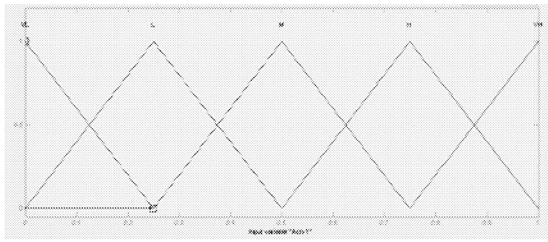


Figure 2 Fuzzy sets for Ac_{n-1}

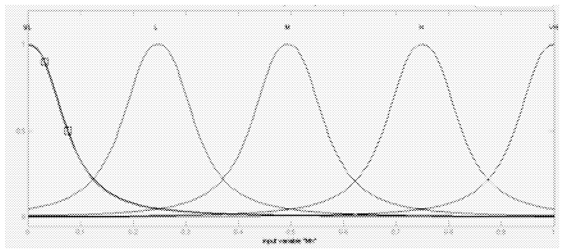


Figure 3 Fuzzy sets for M

This output is also given by 5 fuzzy sets, given by triangular functions, which are: *VL, L, M, H, VH* and are identical to the input Ac_{n-1}.
The 25 rules are given by the fuzzy associative memory (FAM) (Sudiarso, 2002) (see Table 1) and are defined as the following:
R_1: If (Ac_{n-1} is VL) and (Mn is VL) then (Ac_n is VL).

Table 1 FAM

Ac_{n-1}/ M_i	VL	L	M	H	VH
VL	VL	VL	L	L	M
L	VL	L	M	M	H
M	L	L	M	M	H
H	L	L	M	H	VH
VH	L	M	M	H	VH

Our fuzzy Petri net is then given by the following figure (Figure 4):
In a first step, we simply implement our fuzzy Petri net in Matlab to see the values of Ac_n. The obtained results are given in Figure 5.
For our proposed FPN, we have decided that student motivation has a stronger impact on the final results than the acquisition of the precedent course which can be forgotten from one semester to the other.
In a second step, we will add our fuzzy Petri net into the Petri net which has been defined by our colleagues to

represent the timetabling of our National Engineering School.

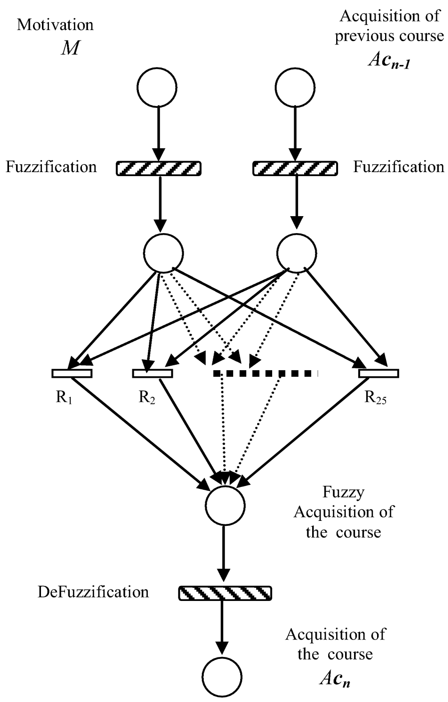


Figure 4 Fuzzy Petri net

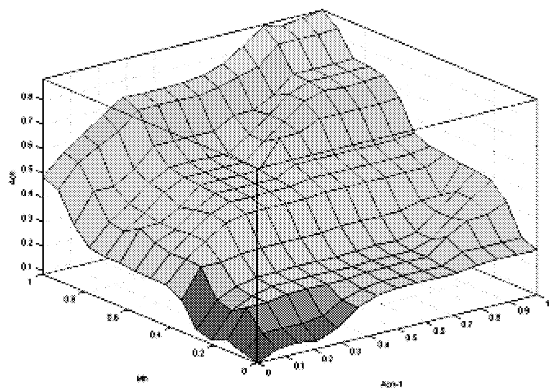


Figure 5 Results for Ac_n

CONCLUSION

We proposed a fuzzy Petri net for the modeling of learning skills for a given course. The advantage of our method is firstly to describe easily human concepts and secondly our proposed fuzzy Petri net is very simple and furthermore consistent (Murata, 1989) and could then be included in the Petri net which has been defined to represent our timetabling.

To improve our fuzzy Petri net, we have to consider various ways: i) we can use fuzzy colored Petri nets to represent more than one prerequisite, ii) we must also consider all the concepts necessary to the acquisition of a course such as the percentage of time spent in the course, the return of knowledge about the course based on evaluation of teachers and the course, etc.

We have to analyze our global Petri net in order to evaluate performances of our training system and also to improve it and then the final objective is to simulate our global Petri net to design a more efficient and cheaper training system.

REFERENCES

Aziz, M. H., L. Erik, M. Parnichkrum, C Saha, (2010), Classification of fuzzy Petri nets, and their applications, *World Academy of Science – Engineering and Technology*, vol. 72, pp. 394-401.

Burke, E. K., B. McCollum, A. Meisels, S. Petrovic, & R. Qu, (2007), A graph-based hyperheuristic for educational timetabling problems. European Journal of Operational Research 176(1),177-192.

Cardoso, J., X. Mao, R. Valette, (2006), State class graph for fuzzy time Petri nets, *Proceedings of the European Simulation and Modelling Conference*, Toulouse, France ISBN 90-77381-30-9, Eurosis-Eti Publication, pp. 350-357.

Chen, S.-M., J.-S. Ke, J.-F. Chang, (1992), Fuzzy reasoning based on fuzzy Petri nets, *International journal of information and management sciences,* vol.3, n.1, pp. 39-51.

Lipp, H.-P. and Günther, R., (1994), A study of fuzzy Petri nets concepts, *proceedings of European Congress on Fuzzy Intelligent Technologies,* Aachen, Germany, pp. 1461-1468.

Murata, T., (1989), Petri nets: properties, analysis and applications, an invited survey paper, *Proceedings of the IEEE,* vol.77, n.4, pp. 541-580.

Qu, R., E. K. Burke, B. Mccollum, L. T. Merlot, S.Y. Lee, (2009), A survey of search methodologies and automated system development for examination timetabling. Journal of Scheduling, 12(1), (pp. 55-89).

Petrovic, S., V. Patel, Y. Yang, (2005). Examination timetabling with fuzzy constraints. In Burke, E. K. and M. Trick, (Ed.), the 5th International Conference on the Practice and Theory of Automated Timetabling (PATAT'05), LNCS, 3616, (pp. 313-333), Springer-Verlag

Renauld, R and Roy, D., (2010), Modélisation di système de pilotage d'un organisme de formation, 8e Conférence Internationale de MOdélisation et SIMulation - MOSIM'10, Hammamet – Tunisie

Sudiarso, A. and Labib, A.W., (2002), A fuzzy logic approach to an integrated Maintenance production scheduling algorithm. *int. j. prod. res*., vol. 40, no. 13, pp. 3121-3138.

Vasant, P., R. Nagarajan, S. Yaacob, (2004), Decision making in industrial production planning using fuzzy linear programming. *IMA Journal of Management Mathematics;* Jan 2004; Volume 15, no 1, pp. 53-65 (ISSN: 1471-6798).

Yager, R. and Filev, D., (1994), Essentials of Fuzzy modeling and control. *John Wiley and Sons.* USA: 1994 (ISBN: 0471017612).

Zadeh, L.A., (1965), Fuzzy Sets, *Information and control*, Volume 8, n.3, pp. 338-353.

Zimmermann, H., (1990), Fuzzy Set, Theory and Its Applications, Second Revised Edition, Kluwer Academic Publishers, ISBN 0-7923-9075-X.

Modelling a Simulated Pharmacy Patient

Summons, P.F.[1], Newby, D.[2], Athauda, R.[1], Park, M.[1]
[1]School of Design Communication and Information Technology
[2]School of Medicine and Public Health
The University of Newcastle

KEYWORDS
Virtual Patient, Assessment, Pharmacy Education, Knowledge Elicitation, Reasoning

ABSTRACT

This paper reports on a pilot study undertaken in a collaboration of 3 Australian universities. The pilot simulated a pharmacy patient as a virtual pharmacy patient (VPP) to study its effects on student learning when it is used as a formative assessment tool for pharmacy students in interviewing and diagnosing a patient. The paper briefly describes the system that was developed and used for student assessment in 2010. It focuses on the issues and associated design strategies encountered during development of the VPP reasoning module.

INTRODUCTION

Three Australian universities (The University of Newcastle, Charles Sturt University, and Monash University) were successful in a collaborative Australian Learning and Teaching Council grant for a AUS$180,000 proposal for a pilot study to develop a virtual pharmacy patient (VPP) system and to evaluate its effects on student learning when used as a formative assessment tool compared to the use of live actors simulating patients (Newby et al, 2011). These universities represent a broad cross-section of the current range of pharmacy programs available at Australian schools as they include major metropolitan, regional and rural areas, and also cover both undergraduate and post-graduate pharmacy programs. The VPP system was used in the assessment of pharmacy students from all participating universities in 2010.

A previous paper (Summons et al, 2009) reported on the initial stages of the project. This paper focuses on the the issues encountered in the development of the VPP reasoning module and their influence on the final system. The next section provides a background to the rationale for developing a virtual patient (VP) and a brief overview of some current virtual patient systems. The general architecture of the VPP is then overviewed, followed by a section discussing the issues encountered during development of the reasoning module, their solution strategies and enhancements to the initial design of the VPP. The final section presents some of the results of VPP implementation and concludes with some possibilities for future developments or research opportunities.

BACKGROUND

Pharmacists practice in a range of clinical settings. However, the majority of pharmacists work in private practice in the community where they play a pivotal role in delivering primary health care for a range of minor illnesses. A significant part of the emphasis of education in undergraduate and graduate Pharmacy programs is devoted to developing the clinical skills required to fulfil this primary health care role. These include communication (e.g. history taking), diagnosing and differentiating minor illnesses from those that require referral to a medical practitioner, choosing the most appropriate treatment for these minor illnesses, and counselling on the correct use of the suggested treatment.

The above skills are especially important in the community setting where the pharmacist is usually reliant on only an oral history to make the diagnosis on which to base their subsequent recommendation for treatment. Pharmacists are not able to physically 'examine' patients as doctors can, and do not have access to other diagnostic tools such as blood tests, X-rays etc to confirm their findings. Fortunately, over 75% of minor illnesses can be identified using an oral history alone. These communication and diagnostic skills are currently generally assessed using written exams, or using Objective Structured Clinical Examinations (OSCEs) with simulated live patients (people trained to respond like patients). The problems of current teaching and assessment methods include:

- tutorials are not conducted using 'real' or 'simulated' live patients because of the large number of students in pharmacy programs;
- there is a lack of adequate numbers of suitably trained 'simulated' patients for use in either tutorials, practice situations or assessments;
- the costs of training and employing real or 'simulated' live patients is prohibitive;
- live patients must be consistent and may respond inappropriately if the student asks questions about an area in which they have not been 'trained'
- feedback on students' interactions with patients on placement may be inconsistent and depend on their supervisor;

The current state of computerised simulation of patients in medicine, dentistry, pharmacy, or nursing personalise the patient by employing images of a patient that are either static pictures of real patients, or dynamic three-dimensional images (either video images of real patients or computer generated images using virtual reality technology and avatars to represent patients).

Some systems have advantages of being scalable in allowing new scenarios to be provided by teachers within a fixed framework, but provide mainly static interaction with the student, allowing only pre-formulated questions or selections of prepared questions (Zary et al, 2006). Other systems encouraged student interaction and realism by providing video clips as responses to student questions (Bergin and Fors, 2003; Farrar, 2002). Video images allow system responses from the virtual patient to convey emotional aspects, such as pain or frustration, in addition to the textual content provided by systems that use static images, however they require a lot of time for the initial filming of every possible answer and they cannot easily change patient features (age, race, gender) without a complete remake of the video. This drawback can be overcome by using avatars (Cavazza and Simo, 2003).

One of the best examples of an avatar based virtual patient operating within a virtual reality world was the DIgital ANimated Avatar (DIANA), created by the University of Florida. DIANA is a female virtual character who played the role of a patient with appendicitis, while a virtual interactive character (VIC - a male virtual character) plays the role of an observing expert (Lok et al, 2006). Students interact with a life-size projection of the DIANA avatar using voice recognition technology.

A similar system has been developed at Keele University in the United Kingdom (Keele University, 2009; PJ Online, 2008). Although no published information exists for the Keele project, it was demonstrated at the Monash Pharmacy Practice Symposium in Prato, Italy. However, evaluation of the DIANA virtual patient, and the demonstration of the Keele virtual patient, identified the following limitations of these types of models:

- students felt the avatar needed more ‘emotion’ and needed to be more ‘expressive’ (Lok, 2006);
- only 60% of student questions were recognised by the virtual patient (Stevens, 2006);
- there was difficulty with the speech recognition technology and students had to train before the assessment to increase the recognition capability of the system. These difficulties *“...brought the students out of the relationship (with the virtual patient sic) and made them cognizant of the product rather than the process”* (Lok, 2006).

The design of the software for the Newcastle Virtual Pharmacy Patient (VPP) system described in this paper was influenced by these previous virtual patient systems. As this was to be a pilot implementation of the virtual patient system it was decided to build a flexible and scalable interface capable of easily generating new patient scenarios within the Pharmacy discipline and which was also easily extensible for re-use in other disciplines.

In 2010, the Newcastle VPP system was used to assess the oral-history skills of pharmacy students at the 3 participating universities in their coverage and style of investigative questions to diagnose three clinical scenarios: a cough; Gastro-oesophageal reflux disease (GORD); and constipation; each with 3 levels of severity: mild, moderate and severe (a total of 9 assessments for each student). Students were randomly assigned a condition with a random severity to assess. The VPP system presented a student with all 3 conditions at randomly selected severities before re-assessing the same conditions at different levels of severity.

VPP OVERALL ARCHITECTURE

The VPP was designed to interact with different user groups: the student being assessed interacts with the VPP to conduct an assessment and obtain feedback; pharmacy lecturers interact with the VPP system for management of students reports and creating/editing assessments; and finally, the system administrator interacts to add/edit virtual patient scenarios and for user (student and academic) registration and management.

The overall architecture for the VPP system is outlined in the Figure 1. The VPP system was developed as client-server based software system. The server side consisted of the database server while the client-side consisted of the VPP Interface module which interacted with the different user groups (Students, Pharmacy academics and Administrators).

All interactions to the server were via a database interface module. The Student Assessment module assigned student’s to a specific assessment scenario and tracked their progress. It included interfaces to the VPP Face Image and Speech modules. The Image and Speech modules were synchronised to represent a patient’s face that spoke to provide feedback to student’s questions.

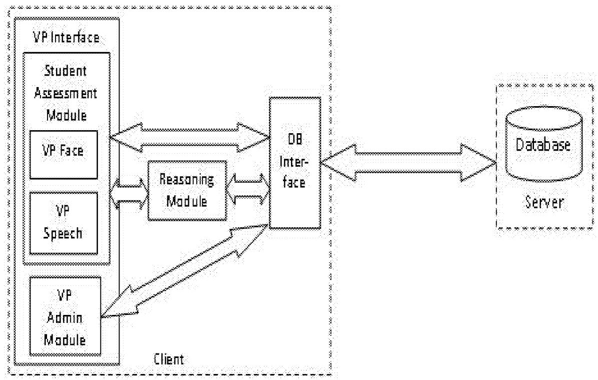


Figure 1 – VPP Architecture Overview

The Reasoning Module had code to recognise the question asked by the student being assessed. These questions were matched to templates held in the server database and appropriate answers were provided to the student via the virtual patient interface (VPP face and speech).

IMPLEMENTATION

The software system of the virtual patient was divided into 3 main development modules:

- Facial Imaging and Speech
- Reasoning (Student Question Recognition and Virtual Patient Answer)
- Student Assessment/Feedback; Lecturer Administration and Reporting

Two facial images, for a male and a female patient (Figure 2) were developed for the pilot assessment of the virtual pharmacy patient (Newby et al, 2011). Different resources (hardware and software) in assessment sites at different universities required different versions of the virtual patient program to be built and tested – some with known limitations, such as low resolution and slow speed. These do not occur if the SQL Server database is installed centrally on a fast server and the reasonably configured client assessment machines are networked to it.

Figure 2 – Facial Images used in the VPP Assessment

For the pilot study a server-based architecture was adopted for the prototype virtual patient. Microsoft SQL Server 2005 was used as a central database to store data (student logins, student questions input during assessment, assessment results, and the data required for the reasoning logic). The virtual patient program was written in Java (using NetBeans IDE) and some data required by the reasoning logic was ported to the client to speed up processing.

VPP REASONING MODULE ISSUES

The assessable areas that the students were supposed to visit during their interrogation of the VPP in order to determine a patient condition were broken down into *Categories*. Some categories consisted of standard areas that might apply and should be questioned across many conditions, including such areas as *Medications Taken*, *Duration* (of condition), *Other Symptoms* and *General Opening Questions.* Other categories were particular to a specific condition, such as the categories of *Normal Bowel Movements* for the constipation condition and *Frequency of Cough* for the cough condition. Categories were populated with questions that were expected to be asked by a student to ensure they had investigated that category. The expected questions are termed *Target Questions* and any variations of them are termed *Alias Questions*. For a specific domain category, there can be many aliases associated with a particular target question however each target question is matched against only one Category/sub-category combination for a specific condition and severity. Each target question for a particular condition and severity is also matched to one virtual patient response (patient answer).

Initially the assessment criteria specified that only one question needed to be asked from a particular category and that there was no required order in which the categories need to be addressed, that is no required sequence to the questioning order of the categories.

Although the assessment criteria only required that one question in a category be asked by a student to indicate coverage of that category, the system is capable of finer reporting detail and so the design further divided the categories, with their associated expected questions, into specific *sub-categories*, for example the category *Duration* (of a condition) was sub-divided into *Start* (of the condition), *Duration* (interval of the condition) and *Existence* (of the condition). This enables finer reporting in the future and also caters for the analysis logic necessary to determine appropriate sequencing and style of questions that are asked by students. In some cases the sub-categories were chosen based on grouping the open-ended and closed-ended expected questions contained in the category, for example the *Frequency of Cough* in the cough condition was subdivided into *Frequency of Cough-Closed* and *Frequency of Cough-Open.* In other cases the sub-categories were associated with specific groups of questions, for example *Other Symptoms-Runny Nose*, *Symptoms-Aches, Other Symptoms-Fever,* and so on, for the cough condition.

Modifications in the original assessment criteria necessitated development of additional logic in the VPP reasoning model to cater for assessment of the appropriateness and sequencing of student questions. In addition to the ability to add to the lexicon and manage the domain knowledge base, the VPP provided an interface for pharmacists (lecturers) to specify simple logic rules for the assessment.

Intra-category question logic allowed lecturers to specify questions that should not be asked following a specific question. For a specific Condition/Severity pair and a specific sub-category, pharmacist lecturers can specify a conditional relationship indicating whether particular Target questions are not appropriate to be asked if a specific question has been asked previously within that category. For example, if a student asks "*Do you have any other symptoms*" and has already asked "*What other symptoms do you have*", then the system provides a response but records this as an inappropriate question sequence and will provide this feedback to the student at the end of the assessment.

Inter-category question logic was also added. It became apparent in beta testing that follow-up questions between categories had to be flagged and indicated. The VPP interface enabled pharmacist lecturers to specify simple logic rules for questions that may be required to be asked in a specific category following a specific question being asked by a student from another category. This allowed a requirement for a follow-up question from a category in response to the VPP's answer of a question from another category. The rules depend on the answer from the virtual patient to the initial student question. For example, if a specific question such as "*Are you on medication?*" is asked by the student and the virtual patient's answer is "*Yes*" then follow-up questions regarding the nature of the medication, or of what symptoms the medication is for, are generally required from the student.

The VPP design was also enhanced to provide the ability to assess the *style* of questions that students asked. Pharmacists supplied the rationale for the style of questioning and the types of questions that were expected for that specific category, such as "Only Open-Ended questions allowed", or "Questioning should start with and Open Ended Question".

Free-Text Recognition and Response

From this paper's introduction section it can be seen that the current state of multi-speaker, wide vocabulary and free-text interpretation is still not perfected. It was felt that the technical difficulty with accurate and consistent speech recognition may detract from the student's learning experience, as they may become frustrated and focus on problems caused by the system's inability to correctly interpret and understand their questions, rather than those caused by the wrong type of questioning. For this reason a text input and speech output design was established. Students used earphones for VPP responses and typed their questions to the VPP. This had the advantage of allowing multiple students to run the system independently in a single room without disturbance from other students speaking commands. A spell-checker ability, that the students could enable or disable, was incorporated into the system to minimize recognition difficulties due to spelling.

One of the main problems in this type of system is recognition of free-text student questions. As the sequence to the interview question/answer sequence was not part of the initial development, this meant that probabilistic reasoning techniques for student question recognition by the virtual patient program were not effective. Also, there was no initial ontology or transcripts of interview question/answers, so there was no training set available and therefore it was difficult to apply any sophisticated artificial intelligence techniques, such as neural networks, to the recognition of sample data. The strategy to overcome this was to choose a simple data-matching reasoning system for the prototype and build into it the ability to "learn" from student supplied questions. The aim was to adopt a pragmatic solution to the recognition problem and develop the systems potential as a tool to generate a lexicon for more complex question recognition in later systems.

Some VP systems, such as Web-SP (Zary et al, 2006), avoided the problems of interpreting student's free-form questions by restricting the allowed questions so that they had to be chosen by the student from a pre-formulated question bank composed by the teachers. This limited the scope of the range of questions to those expected by the system creators and may provide "clues" to a student if used as an assessment.

Adoption of a free-form text style of student questioning was one of the earliest design decisions for the VPP. Systems that use allow free-text student questions and employ lexical analysis to obtain the semantic information, such as DIANA (Lok et al, 2006) and the Arizona Virtual Patient (Farrar, 2002) typically have noticeably slow responses to questions. They perform reasonably well with a small vocabulary but when unlimited free-text questions are allowed they are prone to a reasonably low accuracy as the rate of incorrectly interpreted student questions rises. This is compounded if speech recognition technology is used instead of text entry.

As outlined in (Summons et al, 2009) the reasoning module initially used a pragmatic data matching approach. This was enhanced in the pilot by providing a learning ability for a tutorial version of the VPP and adopting a hill-climbing approach to student question recognition that provided an ability to add to the database lexicon. If a student continually asked questions that the VPP could not recognise, when the VPP finally did match a response, the student was presented with a list of questions that they had asked since the time that the VPP had recognised one of their questions. They were asked to indicate if any of the not recognised questions were similar to the question that the VPP had just recognised. If so, they could be added to the database lexicon as an alias for the question that had just been recognised. Even though this would have helped recognition of students who did not have English as a first language, this feature was not enabled as pharmacy academics wanted to make sure that students were asking grammatical questions that would be understood by the public. However, this was found to have a negative effect on the correct interpretation by the VPP to questions from English-as-a-second-language (ESL) students.

The pilot study concerned itself with the assessment of a student's ability to take an oral history. The main design problem was in correctly interpreting a student's question in order to categorise it against a list of expected questions that were necessary to be asked to investigate specific categories associated with the clinical condition. Once the student's question has been recognised and associated with one of the expected questions, a suitable answer can be easily supplied to the student by the VPP. If the VPP did not understand the student question (it could not be matched against any of the expected questions) then it responds that it did not understand the question and requests that the student re-phrase and then re-enter it. The VPP used 3 different (randomly selected) requests to do this so that the student being assessed would not get bored by the system response.

A virtual patient design decision was adopted so that the VPP would only respond to single-purpose questions, as multiple questions would require sophisticated lexical analysis techniques to determine that there was more than one question to be matched. Another early design decision was that only investigative student questions would be allowed as student question input, thus avoiding potential problems when trying to interpret a suggestion or treatment advice as a question, such as when a pharmacist might suggest to a patient that he/she take a medication as opposed to when they are asking the patient whether they are taking a specific medication. Students are directed, in the initial VPP assessment interface screens, not to suggest treatments or to offer explanations or indicate diagnoses as part of their questioning, but to enter these in separate input areas that allowed modification of diagnosis and recommended treatment by the student at any stage during their assessment.

Each student question is tokenised into individual words after it has been input into the VPP, and then punctuation and irrelevant words are removed. The VPP stores the question in both tokenised form and also in its complete original form for reporting the chronological sequence of student question and VPP answer, but the VPP system logic operates on the tokenised form of the student question.

Matching of the tokenised student question is done initially against a list of tokenised phrases that represent variations in phrasing (aliases) of the expected target question (each variation will only map to one expected question). As the pilot was not intended to produce new reasoning algorithms, initially the reasoning was only based on data matching of the tokenised question against a Target Question (also tokenised) in the VPP.

Academic Pharmacy participants at a workshop in Hobart (2009) trialled the prototype VPP for alpha testing and it was found to have a 50% question recognition rate. They liked how the VPP allowed them to reflect on how they interacted with the virtual patient. Overall, their comments were very positive, such as: *"This is a fabulous student resource" and "A fantastic tool".*

Based on the results of the 2009 alpha testing, extra recognition capability was added to the reasoning module. A secondary filter scheme, using keywords with wildcards to match specific target questions, was implemented as a fallback for the initial data matching to tokenised target questions. If a student question was not recognised initially by the data-matching process (tokenised student question to tokenised target question), then an attempt to match it with a keyword "filter" (consisting of keywords and wildcard characters) was done. If a match with a filter keyword was found then the virtual patient responded to the student "*If you meant to ask* ..." – the virtual patient then provided the target question that the keyword filter was matched to – "*then the answer is*..." – the virtual patient then provided the answer corresponding to the target question for the appropriate condition and severity being assessed. If no match was found to the initial tokenised target question data-matching, or to the subsequent keyword filter, then the virtual patient responded that it did not understand the question and asked the student to re-phrase and re-enter as before. This improved the virtual patient question recognition to be comparable with the DIANA system recognition rate and allowed for simple keyword filters to encompass many aliases. Addition of keyword filters was not incorporated in the teacher interface due to the complexity of finding suitable keywords with filters and assigning them to an appropriate target question. The keyword filters were entered directly into the SQL Server database.

The virtual patient captures student questions that it has not recognised in an assessment. Following assessment sessions the system provides a teacher with the ability to view and analyse these unrecognised questions and, if they are questions that the virtual patient should recognise, add them to the virtual patient lexicon database using the virtual patient teacher interface. If there are multiple valid student questions that the virtual patient has not recognised then a request may be made to the knowledge engineer to incorporate an appropriate category, subcategory, or am keyword filter into the virtual patient.

CONCLUSIONS

Overall results from the assessments at the three collaborating universities indicate that students were generally satisfied with the virtual patient's realism. Students who had worked for less than a year in a pharmacy or not at all tended to indicate more frequently that they felt the virtual patient better prepared them to care for patients and make a diagnosis compared to those that have worked for more than a year (60% vs 45% and 70% vs 56% agreeing /strongly agreeing respectively). In particular, they felt it helped them identify areas of their communication that they could work on (100% vs 56% agreeing/strongly agreeing), and that using the virtual patient would improve their confidence with real patients (90% vs 56% agreeing/strongly agreeing).

Due to time limitations between final testing and the actual assessments only a limited number (192) of keyword filters were added as a secondary data matching recognition scheme in the VPP lexicon database, this improved the recognition rate from 52% in the Hobart alpha test, to an average recognition rate of 52% for international (ESL) students and 62% for domestic students doing the final assessment at Monash, Charles Sturt and Newcastle universities. This achieved one of the design objectives of obtaining a recognition rate that is comparable to other available virtual patient systems. In addition, analysis of the unmatched questions resulting from the assessments over the three participating universities and formulation of additional keyword filters based on those results would improve the recognition rate considerably.

The VPP system is open source and available through the ALTC grant system (Newby et al, 2011). Its pragmatic and simple design has been shown to be capable of being implemented on a variety of software and hardware platforms and with a reasonable level of operation. The tool's pragmatic approach has been seen to be useful as a means of easily generating training data for systems that employ more complex recognition analysis.

REFERENCES

Bergin R, Fors U. 2003. "Interactive Simulated Patient: an Advanced Tool for Student-activated Learning in Medicine and Healthcare" *Computers and Education,* (40:4), pp361-376.

Cavazza, M., Simo, A. 2003. "A Virtual Patient Based on Qualitative Simulation" *Proceedings of IUI'03: International Conference on Intelligent User Interface,* New York: ACM Press, pp19-25.

Farrar, S. 2002. "The Arizona Virtual Patient: Using Question-Answering Technlogy to Enhance Dialogue Processing" *Proceedings of HLT2002 Second International Conference on Human Language Technology Research*, San Francisco, pp222-225.

Keele University: Virtual Patient Demonstration. Accessed online 28th Jun, 2009 from: http://www.keele.ac.uk/schools/pharm/explore/vp.htm

Lok, B., Ferdig, R.E., Raij, A. Johnsen, K., Dickerson, R., Coutts, J., Stevens, A., Lind, S. 2006. "Applying Virtual Reality in Medical Communication Education: Current Findings and Potential Teaching and Learning Benefits of Immersive Virtual Patients," *Virtual Reality* (10), pp185–195.

Marriot, 2007 "Use and evaluation of virtual patients for assessment of clinical pharmacy undergraduates" *Pharmacy Education* (7:4), pp. 341-349.

Newby, D., Jin, J., Summons, P.F., Athauda, R., Park, M., Schneider, J., Kable, S., Marriott, J., Duncan, G., Simpson, M., Xu, R. 2011. "Development of a computer-generated digital patient for teaching and assessment in Pharmacy", Australian Learning and Teaching Council.

PJ Online. 2008 "Avatars Help Keele Students Hone Skills" *The Pharmaceutical Journal* 280(1), p249 [Online Journal] retrieved 28th Jun, 2009 from: http://www.pharmj.com/pdf/articles/pj_20080301_atavars.pdf

Summons, P.F, Newby, D., Athauda, R., Park, M., Shaw, P., Pranata, I., Jin, J.S, Xu, Y.D. 2009. "Design strategy for a scalable virtual pharmacy patient", *ACIS 2009: Proceedings of the 20th Australasian Conference on Information Systems*, Melbourne.

Zary, N., Johnson, G., Boberg, J., Fors, U.G. (2006) "Development, Implementation and Pilot Evaluation of a Web-based Virtual Patient Case Simulation Environment – Web-SP," *BMC Medical Education* (6:10), pp1-17.

ACKNOWLEDGEMENTS

The project work was funded by a Carrick grant. The authors would like to acknowledge the Australian Learning and Teaching Council for grant funding for this work.

Using virtual reality for increasing safety in handling cranes: A presence study.

Iván Peteira
Global R&D Asturias
ArcelorMittal R&D
Ivan.Peteira@arcelormittal.com

Marta Pla-Castells
Dept. of Engineering and Computer Science
University Jaume I - Castellón
Marta.Pla@uv.es

Miguel A. Gamón
LSYM - Robotics Institute
University of Valencia
Miguel.A.Gamon@uv.es

KEYWORDS

Safety, Virtual Reality, 3D, Training, Presence, Immersion, Questionnaire, Steel making facilities, Bridge Cranes

Abstract

Training has been identified as the primary intervention strategy to improve the performance and the accuracy in tasks related to handling industrial cranes. This improvement impacts directly in a safer labour conditions of the crane operators and the rest of workers in the facility where cranes are operating. Training in virtual environments (VE) allows to work in specific conditions and simulating dangerous situations which can not be replicated in the real world. The acquired expertise by the operators using this kind of training, can reduce drastically the incidents during operations. A substantial improvement in the training process could be reached by increasing the quality of the users' experience in the VE. The aim of the presented study is to measure the degree of immersion and *presence* perceived by subjects in a VE which replicates the control of an industrial bridge crane in a steel manufacturer factory. This study has been done by Global ArcelorMittal R&D, in its research centre of Asturias (Spain) using a simulator developed in a previous project. From the results of this study it arises that the users gave an overall positive evaluation of the system and they helped to identify the aspects that can be improved.

INTRODUCTION

Nowadays, safety on the workplace is an important issue to be addressed due to the difficulty to reduce the number of accidents, independently on the type of work, despite social and technical improvements of labour conditions. For this reason, in many European countries, it has become mandatory the adoption of safety measures which vary depending on the nature of the workplace and task performed, but sharing common main directives. Safety measures include structural modifications (buildings and/or machines) and the training of workers to teach them the safest operation modes and practices. Traditionally, the training of the staff is performed on the workplace by means of special courses where theoretical classes are given to provide the fundamental information and, in some cases, practical activities are performed to complete the learning process and evaluate the trainees. In the last years, thanks to the wide development of computer sciences and information technologies, many computer-based tools have been developed aimed to this kind of activities (Vannucci et al. 2009; 2010, Zayas Perez et al. 2007, Lin et al. 2000, Rodenas et al. 2004, Swadling and Dudley 2001, Dong et al. 2009, Wilson et al. 1998, Rouvinen et al. 2005, Garcia-Fernandez et al. 2009, Huang and Gau 2003, Daqaq and Nayfeh 2004)

Virtual reality (VR), described by several researchers (Kalawsky 1993, Durlach and Mavor 1995, Heim 1998, Burdea and Coiffet 2003), is most properly defined as immersive, interactive, multi-sensorial, viewer-centered, three-dimensional (3D) computer generated environments and the combination of technologies required to build them (Cruz-Neira 1993). As this definition suggests, creating a virtual environment (VE) requires immersing humans into a world completely generated by a computer. The user becomes a real participant in the virtual world, interacting with virtual objects. Therefore, the quality of the user's experience is one of the most important considerations in defining the requirements for a VE. As the user's senses and body are involved in the task, it becomes essential to focus on user-centric measures of this quality. Although in the field of human-computer interaction, abstract values such as ease of use, ease of learning, presence and user comfort are significant (Kalawsky 1993), *presence* in VE, which is the subjective experience of being in one place or environment when someone is physically situated in other, (Witmer and Singer 1996) becomes the most important measure.

This concept of experiencing *presence* as a normal awareness or attentional phenomenon is based on the interaction between external stimuli and immersion factors. It is also defined as the mental state in which user feels physically within the computer-generated environment (Slater and Steed 2000). The involvement tendencies depend on focusing one's attention and energy on a coherent set of VE stimuli while immersion tendencies lead to the perception of being a part of the VE stimulus flow. According to Witmer and Singer (Witmer and Singer 1998), both, involvement and immersion are necessary conditions for experiencing *presence*. It is generally held that human performance in VEs is directly proportional to the degree of *presence* induced by the environment, which, in turn, is influenced by the individual's level of immersion in the VE (Witmer and Singer 1998, Stanney et al. 1998). Essentially, a transitive relationship appears where the immersion affects presence, and it directly affects the performance of the system. In this paper we will consider the concept of *presence* as defined in (Witmer and Singer 1998).

Fully immersed observers perceive that they are interacting directly or remotely with the environment. Thus, presence becomes a subjective sensation or mental manifestation that is not easily amenable to objective physiological definition and measurement, with its strength varying both as a function of individual differences, traits, abilities and characteristics of the VE. In general, the more control a person has over the task environment or in interacting with the virtual environment, the greater the experience of presence (Schloerb 1995). The success of using VR as a tool for training and job aiding, therefore, is highly dependent on the degree of presence experienced by the users of the virtual reality environment. So, it is important that we measure the degree of presence of the VR simulator to support training.

If the VR simulator is going to be proposed as a solution for off-line training, it is essential that the environment accurately mimics the real world as perceived by the user. Only then, the effects of training can be expected to transfer from the VR environment to the real world. In order to measure the degree of *presence* of the VR simulator, this study asks for the subjective opinion of humans on the applicability of the system as a support for training of industrial bridge crane in steel manufacturing facilities. This is done by using a *presence* questionnaire (PQ) to ascertain the degree to which individuals experienced *presence* in a VE and the influence of possible contributing factors on the intensity of this experience (Witmer and Singer 1998). This paper describes the VR simulator developed and the study conducted.

DESCRIPTION OF THE VR SYSTEM

Prior to describing the configuration of the simulation system that is the object of this study, the context of the project that led to its development is introduced.

Context

In production systems that link different processes, such as steel making manufacturing process, handling and loading intermediate products is a frequent and critical task. The cranes are used when, in addition to loading or unloading products, it is necessary to move them within the working place, which can be constrained by several obstacles. Therefore, it constitutes a work in which both safety and efficiency are key factors.

Platform cranes (gantry and bridge cranes) are the most prevalent cranes in industrial environments where a convergence of process and environmental factors make the manoeuvres especially dangerous. They are one of the most critical bottlenecks which affect the efficiency in manufacturing rhythms and they are also a focus of potential risks. On the other hand, there are many types of cranes and working scenarios which differ in a wide range of aspects, such as frame, types of beams, types and degrees of freedom of hoist trolley, number of hooks, hook load capacity, speeds and cabin (workplace) or different control panel configuration, just to give some examples. Besides, many other variables must be added to this fact such as the interlocking type and types of loads.

Training of bridge crane operators is a vital and very complex process because of the aforementioned variety of conditions. This training is sought, on the one hand, to obtain an adequate return of investment and, on the other hand, to ensure the safety of workers in the working area according to different rules and standard regulations on the prevention of occupational risks.

To improve the questions above, the R&D centre of ArcelorMittal in Asturias, in collaboration with IRTIC at University of Valencia developed a complete training system for crane operators, in the scope of the project codenamed UCRANE. This system is based on the use of a versatile and easily expandable simulator, both in regard to software, the system of assessment and instructional design, and as regards the scope of supply hardware.

The simulation system used in this study, is a medium-range simulator, installed and located in the research centre of ArcelorMittal in Asturias (see Figure 1).

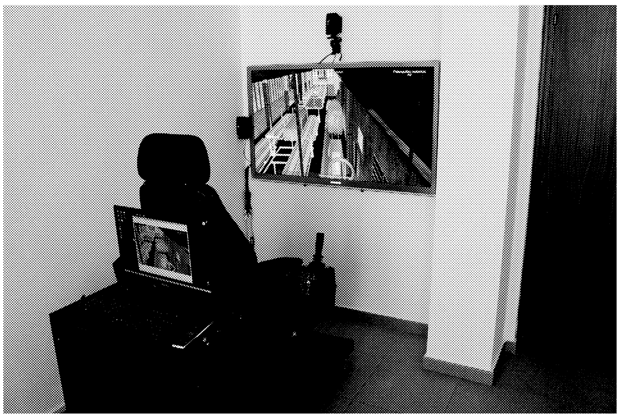


Figure 1: An image of the simulation system

Technological solution

The hardware component which formed the process computer of the simulator are a Pentium Intel® Core^{TM} 2 Quad CPU Q8300 @ 2.5Ghz processors, equipped with 4 GB of main memory and a NVIDIA GeForce GTX 470 for graphics processing with 1280MB of texture memory. The main structure is an industrial seat with two desks that are a replica of a standard crane control panel with two joysticks and various buttons and switches. Under the seat, two electromagnetic bass shakers were added, plugged to a wave amplifier in order to replicate some vibrations according to the movements of the crane. The simulator has got a television where users can view 3D images with stereoscopic vision glasses. In order to improve the vision of the simulation scenario from several points of view, the system has a motion tracker formed by a camera, located in the top of the television which recognize the movements of user's head by means a device located in a cap. Besides, real sounds were added to try to introduce a greater sense of *presence* in users (Storms and Zyda 2000). Figure 1 shows an image of the training system.

Development of the VR environment

The development of the VR environment was based on a detailed analytic methodology. Data on crane operation were collected through observation, expert interviewing and digital data capturing techniques. For some aspects, detailed layouts were requested to the facilities.
The goal of the construction of the virtual environment was to reproduce the appearance of the real working facility. This physical environment is a complex 3D volume with different elements and working tools (see Figure 2).

Figure 2: Image of real environment

The elements of the scenario were built with a high degree of realism using the 3DStudioMax® environment. Most of the models were built with planar polygons for two reasons. First, since the representation of the true complexity of the facility is avoided, fast display rates are maintained (on the order of 25-30 fps) while tracking latencies are minimized. Second, planar polygons (quadrilaterals) are particularly suitable for texture mapping. To provide a realistic appearance of the environment, images of the physical location were used for texture maps.
To generate the environment, no specialized rendering algorithms are invoked because they are provided by the OpenSceneGraph library (OpenScenegraph 2010). Standard direct lighting is used to illuminate the scene. To build the texturize map simple 3D box-like environment and images of the real facility were captured.

Scenarios

The VR system is formed by several scenarios or environments. Every scenario is divided into several training exercises to improve the performance of the users. These exercises will provide a rapid learning curve to the trainees with a moderate effort. This system allows the trainees to start to work with a virtual bridge crane without any previous practical knowledge about its operation. However, psychotechnical tests are performed and theoretical lessons on bridge crane handling are taught to the trainees before they start using the VR system.
The system has four different scenarios. The first one is devoted to learn how a bridge crane works, showing its basic behaviour and operation focusing on its movements, in order to improve the reaction of the users. Other exercises of this scenario are designed to improve the accuracy skills of trainees in low-visibility cases and operator aided manoeuvres. The other three scenarios are based on real conditions in steel plants and, due to the nature and the great variety of bridge cranes, these are simulated depending on the real scenario: primary and secondary elevations, cabin movement together the hoist-trolley versus cabin movement together the bridge, different control panels, etc. These scenarios are focused on a roll workshop, where accuracy and low-visibility tasks are performed; and a slab yard, where the sway of the load is a very important factor due to the kind of clamp that is used and the high weight of the carried load, which can be up to many tons. An image of this scenario can be seen in Figure 3. Finally, a billet yard, where the transportation of loads are made using electromagnetic devices with its consequent problems of energy supply losses and bad coupling between device and load, have also been implemented.
All the aforementioned scenarios comprise different exercises to provide trainees with all the necessary skills before they start working in a real machine.
One important aspect to take into account is the scenarios' set up. This point is performed by the trainer, who defines and configures all the operations that have to be performed in an exercise, together with a set of environmental conditions, possible machine failures and working place events, such as the appearance of crossing

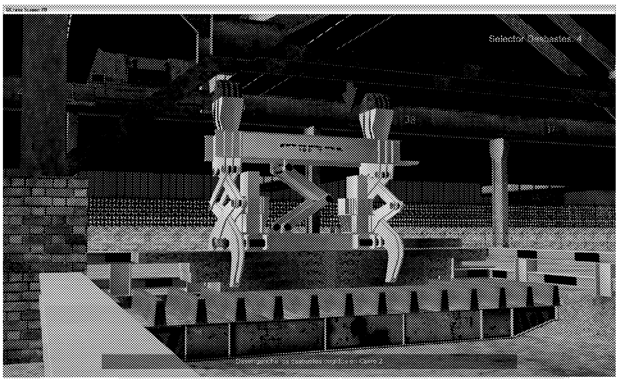


Figure 3: Simulation of the slab yard scenario

Figure 4: Image of virtual environment

workers near the operation zone.

PRESENCE STUDY

Upon this system, an study has been performed in order to evaluate its quality in terms of the degree of *presence* perceived by the users. Next, the experimental design and the tests done with a set of users is described.

Subjects

Twenty-five participants took part in the study (18 males and 7 females). The average age of the participants was 34 years. They were recruited by a request email from the R&D center of ArcelorMittal located in Asturias. All of them were workers and researchers. It is important to remark that participants did not received any compensation for their time. All the participants in the experiment knew the real scenario that is reproduced in the VE, but had no prior experience in handling the crane.

Experimental design

The study was performed using a VR training session consisting on a familiarization scenario for getting acquainted with the system, followed by a normal working scenario. The scenario was the simulation of one of the exercises designed to improve the skills in specific manoeuvres where accuracy is needed. The exercise is formed by two sets of movements. The first one, is oriented to fit a round weight inside a box. The second one, is to pass with the round element, through a kind of a corridor with tight dimensions. The resulting scenario is shown in Figure 4.

The experiment duration was around 35 minutes (a 5 minutes familiarization session, followed by a 20 minutes working session, and 10 minutes to fulfil the PQ).

Procedure

Prior to starting the experiment, subjects were individually told how the working scenario would be. They were taken to the VR room where the simulator was put in operation. Each subject was briefed, also in an individual manner, on the experiment through a description of the objectives of the study and the tasks that they had to do. Subjects were allowed to practice with the VR simulator as part of the software familiarization during a maximum of ten minutes. After the subjects became comfortable, they were asked to begin the working tasks and they notified the researcher when they finished the exercise. After completing all the working tasks, subjects were asked to fill out the PQ v.4.0 (Witmer and Singer 1998). The PQ questions are reproduced next by kind permission of the authors.

1. How much were you able to control events?
2. How responsive was the environment to actions that you initiated (or performed)?
3. How natural did your interactions with the environment seem?
4. How much did the visual aspects of the environment involve you?
5. How much did the auditory aspects of the environment involve you?
6. How natural was the mechanism which controlled movement through the environment?
7. How compelling was your sense of objects moving through space?
8. How much did your experiences in the virtual environment seem consistent with your real world experiences?
9. Were you able to anticipate what would happen next in response to the actions that you performed?
10. How completely were you able to actively survey or search the environment using vision?
11. How well could you identify sounds?
12. How well could you localize sounds?
13. How well could you actively survey or search the virtual environment using touch?
14. How compelling was your sense of moving around inside the virtual environment?
15. How closely were you able to examine objects?

16. How well could you examine objects from multiple viewpoints?
17. How well could you move or manipulate objects in the virtual environment?
18. How involved were you in the virtual environment experience?
19. How much delay did you experience between your actions and expected outcomes?
20. How quickly did you adjust to the virtual environment experience?
21. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?
22. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?
23. How much did the control devices interfere with the performance of assigned tasks or with other activities?
24. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?
25. How completely were your senses engaged in this experience?
26. How easy was it to identify objects through physical interaction; like touching an object, walking over a surface, or bumping into a wall or object?
27. Were there moments during the virtual environment experience when you felt completely focused on the task or environment?
28. How easily did you adjust to the control devices used to interact with the virtual environment?
29. Was the information provided through different senses in the virtual environment (e.g., vision, hearing, touch) consistent?
30. To what extent did you feel completely surrounded by and enveloped by the virtual environment?
31. As you moved through the virtual environment and interacted with it, did you feel like you were inside the virtual environment, affecting or being affected by objects and events in that environment?
32. How much did your experience in the virtual environment seem like you were in a real place, able to directly sense and interact with the environment?
33. In the virtual environment, how strong was your sense of "being there"?

Data collection

Using the questionnaire, the sense of *presence* of the subjects on the working tasks was evaluated as they used the VR environment, focusing on their perceptions of the system. The PQ was used to address their subjective experience in a simulated environment, identifying and measuring the aspects of the virtual environment that engendered a sense of presence. The PQ use a 7-point Likert scale based on the semantic differential principle (Witmer and Singer 1996). Each item is end-anchored by opposing descriptors, but unlike the semantic differential, the scale also includes an anchor at the midpoint. Items marked with an asterisk (*) have to be reverse scored in order to contribute to the subscale and overall totals.

Based on the reliability and the clustering analysis performed by Witmer et al. (Witmer et al. 2005), the four clusters used to collect data for the VR environment were identified as Involvement, Sensor Fidelity, Adaptation/Immersion and Interface Quality. A detailed definition of the aspects measured by each cluster can be found in (Witmer et al. 2005). The questions for the questionnaire are categorized as shown in Table 1. Version 4.0 of the PQ has 33 questions but items

Presence Questionnaire	
Factor	**Questions**
Involvement	1,2,3,4,6,7,8,10,14,17,18,26
Sensor Fidelity	5,11,12,13,15,16
Adaptation/Immersion	9,20,21,24,25,27,28,29
Interface Quality	19*,22*,23*

Table 1: Categorization of the PQ

30 through 33 are recent additions to the questionnaire and they are still being investigated. We will take them into account in the individual question study of the PQ.

RESULTS AND DISCUSSION

The results were analyzed using R (v2.12.1). The Wilcoxon test was used to examine the significance of the deviation from the neutral point 4 on the Likert scale. The results are summarized in Table 2. Every row of the table corresponds to a question. The mean score and its standard deviation are shown in column two. Column three shows the p-values of the Wilcoxon test of that question. Columns four to six show the percentage of responses scored below 4, with score equal to 4 or with score higher to 4. Drawing on these percentages, we can estimate if the question has been accepted or not by users. This result is indicated in the last column as *Accepted (A)*, *Neutral (N)* or *Declined (D)*. If, for a given question, more than a 70% of the responses were higher than the neutral value (4), then it is considered that the question is *Accepted.* If more than a 50% of the responses are below the neutral value, then the question is considered as *Declined.* Otherwise, the response is considered as *Neutral.* This classification, however, is merely informative and used to help read the table.

The results show a significant inclination ($p < 0.05$) of the participants to agree to questions 1, 2, 3, 4, 6, 7, 9, 11, 12, 14, 15, 16, 18, 20, 24, 25, and 29 (mean over 4) and to disagree with questions 22 and 23 (mean less than 4). There was no significant deviation from the neutral value for the responses to questions 5, 8, 10, 17, 21, 26, 28, 30, 31, 32 and 33 in a positive way and to question 19 in a negative way. Question 13 and 23 will be studied separately in next section.

Question	Mean(SD)	W. Test	$\% < 4$	$\% = 4$	$\% > 4$	D/A
1	5.73(0.45)	6.3e-5	0	0	100	A
2	5.89(0.73)	1.0e-4	0	0	100	A
3	5.36(0.95)	8.8e-5	5.26	15.78	78.94	A
4	5.31(0.67)	8.6e-5	0	5.26	97.73	A
5	5.00(1.05)	1.2e-4	5.26	31.57	63.15	N+
6	5.31(0.94)	9.7e-5	5.26	15.78	78.94	A
7	5.36(0.68)	1.0e-4	0	10.52	89.48	A
8	4.52(0.69)	9.4e-5	0	57.89	42.10	N+
9	5.31(0.94)	9.7e-5	5.26	15.78	78.94	A
10	5.21(0.97)	1.1e-4	0	31.57	68.42	N+
11	5.42(0.69)	9.9e-5	0	10.52	89.48	A
12	5.00(0.74)	1.1e-4	0	26.31	73.68	A
13	4.10(1.55)	1.2e-4	26.31	42.10	31.57	N
14	5.21(0.63)	9.4e-5	0	10.52	89.48	A
15	5.26(1.04)	1.2e-4	0	26.31	73.69	A
16	5.42(0.90)	1.0e-4	0	21.05	78.95	A
17	5.00(0.94)	1.1e-4	5.26	26.31	68.42	N+
18	5.68(0.67)	8.6e-5	0	5.26	94.74	A
19	3.47(1.12)	1.2e-4	47.36	31.57	21.05	N-
20	5.05(1.12)	1.1e-4	10.52	10.52	78.94	A
21	4.78(0.71)	1.0e-4	0	36.84	63.15	N+
22	3.21(1.18)	1.2e-4	63.15	26.31	10.52	D
23	2.36(0.76)	1.0e-4	94.73	5.26	0	D
24	4.94(0.84)	1.1e-4	5.26	21.05	73.68	A
25	5.21(0.91)	1.1e-4	0	21.05	78.94	A
26	4.68(1.24)	1.2e-4	21.05	15.78	63.15	N+
27	5.52(0.96)	1.1e-4	5.26	5.26	89.47	A
28	5.05(0.91)	1.1e-4	0	31.57	68.42	N+
29	5.21(0.97)	1.0e-4	10.52	5.26	84.21	A
30	4.73(1.28)	1.2e-4	15.78	15.78	68.42	N+
31	4.78(0.91)	1.1e-4	5.26	36.84	57.89	N+
32	4.42(0.83)	1.1e-4	10.52	47.36	42.10	N+
33	4.31(0.94)	1.1e-4	15.78	36.84	47.36	N+

Table 2: Results of the Wilcoxon and Accept/Decline Tests

Discussion

The analysis of the results allow us to evaluate the degree of presence of the industrial bridge crane simulator. There were no significant responses against the simulator except for question 23. In this question it was asked whether the control devices interfered with performance of assigned tasks and all the user's responses where below the neutral value (see Figure 5-right), indicating that controls are not adequate. Another question that deserves some commentary is Question 13. The users that did the test gave very different responses to this question, fact that is reflected in its high standard deviation (see Figure 5-left). The reason for this wide range of criteria shown by the users can be the object of the question. In Question 13 the user is asked *How well could you actively survey or search the virtual environment using touch.* However, in the VE that is used in the simulator under study the user is not able to directly *touch* any element of the environment, and only the crane's hook interacts with the objects in the scene. The users were told to consider the question was asking about the quality of the reaction of the crane's hook against collisions, but the results indicate that this question should be rewritten for this particular application.

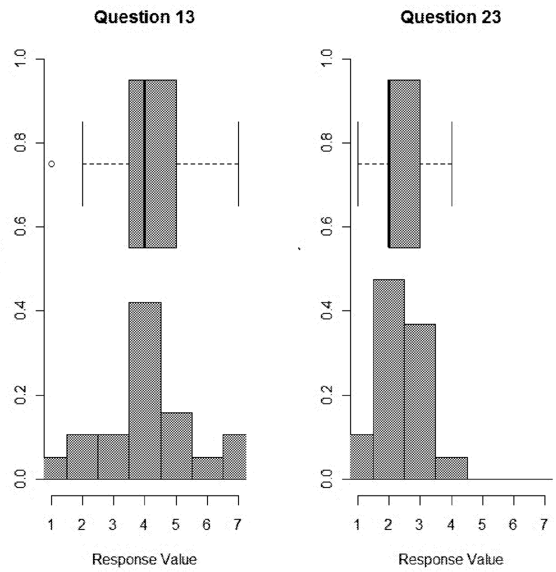


Figure 5: Histogram and boxplot of questions 13 and 23

Analysis by Categories

After discussing the results of the different questions of the questionnaire, next the results are studied for the categorization defined by Witmer et al. (Witmer et al. 2005). Taking into account the classification shown in Table 1, the results of our study for the different categories can be seen in Table 3. In this table, *Mean* column corresponds to the average value of all the questions of the corresponding category. The standard deviation of the data is shown between parenthesis. The K factor is the Kurtosis, which is a measure of whether the data are peaked or flat relative to a normal distribution. That is, *data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak.* A uniform distribution would be the extreme case. Finally, the S factor is Skewness that is a measure of symmetry, or more precisely, the lack of symmetry of the data set.
The results show that for Involvement (I), Sensor Fidelity (SF) and Adaptation/Immersion (A/I) factors, the mean is over the neutral value, i.e, the users feel a high degree of presence in that factors. But for the Interface Quality (IQ) factor the results are clearly less

Results			
Fac	**Mean(SD)**	**K**	**S**
I	5.27 (0.42)	3.13	-1.02
SF	5.03 (0.52)	2.05	-0.40
A/I	5.13 (0.57)	2.16	0.09
IQ	3.01 (0.79)	1.91	-0.33

Table 3: Categorization of the PQ

favourable. The average for this factor is lower than the neutral value (4) and, although the standard deviation is higher than the others, it indicates that the Interface Quality of the application should be improved. This results show that the interface used to move into facility distract the users from doing the working tasks assigned. These results can be observed in Figure 6, which shows the comparison of the histograms and boxplots for the four categories.

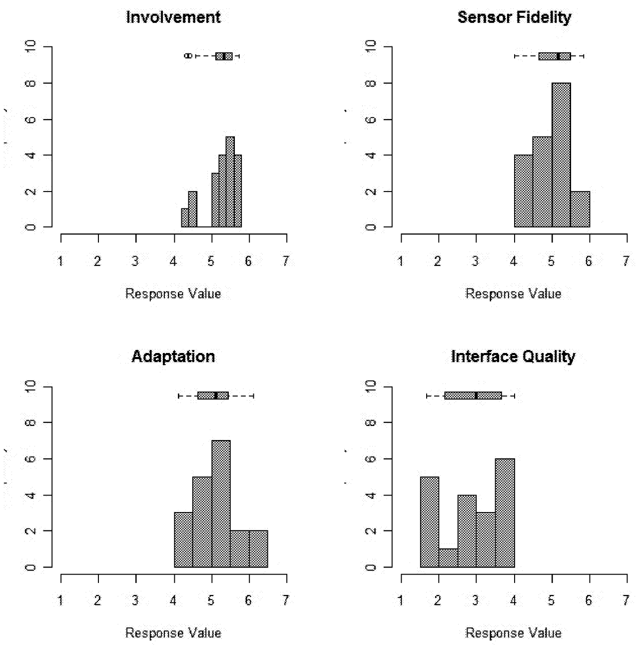


Figure 6: Histogram and boxplot of the four categories

From Table 3 it can be observed that the recolected data has a leptokurtic distribution in the four factors analyzed. That is, it has a high degree of concentration around the mean values of the factors. Negative skewness in three of the four factors indicates a distribution with an asymmetric tail which is *heavier* towards values lower than the mean. These two factors indicate that the majority of responses in each factor are very close to their respective averages and that answers that differ from the average tend to disagreement with the factor analyzed.
Thus, from both individual questions analysis and categorized analysis, the results indicate that the users of the simulator experienced a high degree of *presence* during the test. The users considered, however, that the interface quality is of low quality and, in particular, that it distracts from the tasks that have to be performed.

Despite the fact that the subjects had no prior experience in the use of the crane, it does not affect the responses to the questionaire, as the factors it is measuring are not directly affected by a deep knowledge of the real environment. Anyway, the obtained results have to be confirmed or modified by means of new tests that involve experienced crane users.

CONCLUSION

The presented study evaluated the degree of presence of a VR simulator, used for training of bridge crane operators in a steel manufacturing environment. Four main aspects have been evaluated: usability (Involvement), movement around the virtual environment (Adaptation/Immersion), handling devices (Sensor fidelity) and user interface (Interface Quality). The evaluation has followed the methodology proposed by Witmer and Singer (Witmer and Singer 1996) in order to obtain comparable results. The experimental setup has been done with 25 inexperienced users performing a medium difficulty task.

The results showed the users gave an overall positive evaluation of the system. Among the evaluated aspects, Involvement, Sensor Fidelity and Adaptation/Immersion got an evaluation over the neutral value, while the user Interface Quality is an aspect that the users considered that can be improved. As future work, additional tests will be done with experienced users and with different professional profiles, such as crane operators, technicians or managers, as we expect that different results can be obtained. Also, the results of the study will be used for the improvement of the application, specially in what regards movement and location of the user inside the scene.

ACKNOWLEDGEMENTS

The authors would like to thank all the participants in the study for testing the simulator, giving their time without any compensation other that the minutes of enjoyment. Also, to Dr. Witmer for providing the PQ and the instructions for evaluating the results.

This study was possible due to the existence of the simulator developed during the frame of the projects UCRANE-1 and UCRANE-2 between ArcelorMittal España S.A and the University of Valencia (UVEG), partially granted by the Government of the Principality of Asturias.

References

Burdea G. and Coiffet P., 2003. *Virtual Reality Technology. Presence: Teleoperators and Virtual Environments*, 12, no. 6, 663–664. doi:10.1162/105474603322955950.

Cruz-Neira C., 1993. *Virtual Reality Overview.* In *ACM SIGGRAPH'93 Conference.*

Daqaq M. and Nayfeh A., 2004. *A Virtual Environment for Ship-Mounted Cranes. International Journal of Modelling and Simulation*, 205.

Dong H.; Zou S.; Chen D.; and Xu H., 2009. *Research and Development of Distributed Bridge Crane Training System Based on Virtual Reality. Intelligent Computation Technology and Automation International Conference*, 670–673.

Durlach N.I. and Mavor A.S. (Eds.), 1995. *Virtual Reality: Scientific and Technological Challenges.* The national academies press.

Garcia-Fernandez I.; Pla-Castells M.; Gamon M.A.; and Martinez-Dura R.J., 2009. *UVSIM: A Harbor Cranes Training System. Proceedings of the HMS.*

Heim M., 1998. *Virtual Realism.* Oxford University Press.

Huang J. and Gau C., 2003. *Modelling and designing a low-cost high-fidelity mobile crane simulator. International Journal of Human-Computer Studies*, 58, no. 2, 151–176.

Kalawsky R., 1993. *Tge Science of Virtual Reality and Virtual Environments.* Addison-Wesley.

Lin C.T.; Chung I.F.; Liann S.A.; and Su F.Y., 2000. *Multipurpose virtual-reality-based motion simulator.* In *Intelligent Control and Automation, 2000. Proceedings of the 3rd World Congress on.* vol. 4, 2699 –2704 vol.4. doi:10.1109/WCICA.2000.862545.

OpenScenegraph, 2010. URL `http://www.openscenegraph.org/projects/osg`.

Rodenas J.; Zarza I.; Burgos M.; Felipe A.; and Sanchez-Mayoral M., 2004. *Developing a virtual reality application for training nuclear power plant operators: setting up a database containing dose rates in the refuelling plant. Radiation Protection Dosimetry*, 111, no. 2, 173–180.

Rouvinen A.; Lehtinen T.; and Korkealaakso P., 2005. *Container Gantry Crane Simulator for Operator Training. Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics*, 219, no. 4, 325–336.

Schloerb D., 1995. *A quantitative measure of telepresence. Presence*, 4, no. 1, 64–80.

Slater M. and Steed A., 2000. *A Virtual Presence Counter. Presence: Teleoperators and Virtual Environments*, 9, no. 5, 413–434.

Stanney K.; Mourant R.; and Kennedy R., 1998. *Human factor issues in virtual environments: a review of the literature. Presence*, 7, no. 4, 327–351.

Storms R. and Zyda M., 2000. *Interactions in perceived quality of auditoy-visual displays. Presence: Teleoperators and Virtual Environments*, 9, no. 6, 557–580.

Swadling P. and Dudley J., 2001. *VRLoader: a Virtual Reality Training System for a Mining Application. Proc Simulation Conf and Exhib SimTecT 2001.*

Vannucci M.; Colla V.; Avigni A.; and Avigni A., 2009. *VESPRO: a virtual environment for practicing safety procedures. Proc 1st International Conference on Information Society and Information Technology*, 136–141.

Vannucci M.; Colla V.; Avigni A.; and Avigni A., 2010. *A multimedia interactive 3D virtual environment for emergency simulation, teaching and practice of safety procedures. Proc 7th EUROSIM Congr on Modelling and Simulation EUROSIM 2010.*

Wilson B.H.; Mourant R.R.; Li M.; and Xu W., 1998. *A virtual environment for training overhead crane operators: real-time implementation. IIE Transactions*, 30, no. 7, 589–595.

Witmer B.G.; Jerome C.; and Singer M.J., 2005. *The Factor Structure of the Presence Questionnaire. Presence*, 14, no. 3, 298–312.

Witmer B.G. and Singer M.J., 1996. *Presence Measures for Virtual Environments: Background & Development.* Tech. rep., Draft United States Army Research Institute for Behavioral and Social Sciences.

Witmer B.G. and Singer M.J., 1998. *Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence*, 7, no. 3, 225–240.

Zayas Perez B.; Molina Marin M.; and Islas Perez E., 2007. *Developing a Virtual Environment for Safety Training. Electronics, Robotics and Automotive Mechanics Conference*, 0, 545–550.

MARKET OPTIMIZATION

A WORK ON REAL-TIME DATA DISSEMINATION PROBLEM

Mustafa Mujdat Atanak[1] and Atakan Doğan[2]
Department of Electrical and Electronics Engineering
Anadolu University
Eskisehir, Turkey
E-mail: {mmatanak[1], atdogan[2]}@anadolu.edu.tr

KEYWORDS
Resource management, Computer networks, Model design, Information systems, Real-time.

ABSTRACT

Multicommodity flow problem deals with delivering multiple data items residing on different source machines on a networking system to their destinations in an efficient manner, while respecting the bandwidth and cost limitations that may be imposed on the system components. Real-time Data Grid applications are emerging in many disciplines of science and engineering today. This study defines real-time multisource real time data dissemination problem and divides the problem into two subproblems. First subproblem is to find the routes on which the data transfers occur. Second subproblem is to determine the amount of flow to be committed on each route. Four different approaches related to first subproblem and two approaches related to second subproblem are implemented as part of this study and results are compared.

1. INTRODUCTION

Grid systems provide researchers of various fields with the required hardware and software infrastructure of high performance computation, data storage, and communication resources. The topics of Grid projects vary from high energy physics (EU-DataGrid [1]) to climate modeling (Earth System Grid) to earthquakes (NEESit [2]).

Grid systems execute jobs that has large amount of resource requirements. These jobs may require data sets that are on the order of terabytes. The data may reside on one or more storage element. Delivering these data to the requesting computing element via network elements in an efficient manner is a hard data dissemination problem.

The performance of the Grid system is affected by the performance of the underlying algorithms (such as data dissemination, data replication, data replacement, job scheduling algorithms) as well as the performance of the underlying hardware. Therefore, performance of the algorithms plays a crucial role in performance of the Grid system. This study focuses on data dissemination aspect only. Emerging multimedia based applications require high QoS guarantees such as minimum flow bandwidth or maximum end-to-end delay. In addition, some of these applications impose a hard deadline before which all related data transfers are to be completed.

This study defines a novel problem: real time multi-source data dissemination problem and presents a novel performance measure for this problem. Different algorithms and approaches are employed in the literature for best effort systems. This work presents the comparative performance analysis of these algorithms and approaches for a real time system.

This paper is organized as follows: Section 2 presents a literature review on the subject. Section 3 explains in detail the data dissemination model used in this study. Next section describes the simulation framework and heuristics, which is followed by simulation results and discussions. Last section provides conclusions and future directions of study.

2. LITERATURE REVIEW

This study defines the real-time multi-source data dissemination problem. The problem is, then, divided into two subproblems: 1) assigning routes to requests 2) how to run the traffic.

2.1. Assigning Routes

This subproblem deals with deciding which request is sent via which route(s). Related algorithms from the literature are grouped in 3 main objective functions:

1) Shortest path problem: find the path with shortest distance between two points on a network.
2) Largest flow problem: find the path that yields largest possible flow while respecting network bandwidth limitations.
3) Least cost problem: given predetermined costs of the paths, find the path(s) from one or more sources to one or more destinations with the least total cost.

The solution space of data dissemination problems is extremely large and an *optimal* solution can not be found in a reasonable time. Instead, heuristics are employed to find *good* solutions [3]. For example, heuristics such as Dijkstra [4], Bellman-Ford [5-6], A* [7] are used to solve the shortest distance problem. Similar algorithms are used in todays' state-of-the-art network systems frequently.

In IP-based networks, it is traditional to use protocols that include shortest path algorithms such as RIP (Routing Information Protocol) [8], OSPF (Open Shortest Path First) [9], and IS-IS (Intermediate Systems to Intermediate Systems) [10]. Although these protocols are easy to implement and fast in run-time, they usually are unable to generate efficient schedule tables for networking resources, especially for real-time applications.

Kleinberg's doctoral research study on generating scheduling tables for flows on a networking system is an important work [11]. Kleinberg starts with edge disjoint paths problem. In this problem, for a given directed graph and source-destination pairs {(s1, t1), (s2, t2), · · · , (sk, tk)}, the goal is to find the highest number of pairs between which an acceptable route can be found while respecting the limiting condition. The limiting condition is to use a vertex at most once in the resulting schedule.

2.2. Running Traffic

This subproblem tries to answer the following question: which flow gets how much bandwidth. This study includes two popular approaches from the literature: a) Integrated Service b) TCP.

Running IS Traffic

In order to support real-time data transfers on the Data Grid, the network infrastructure must be capable of assuring end-to-end delay bounds for the data flows. The problem of providing end-to-end delay bounds (or QoS guarantees in general) for applications has been the subject of many studies in the Grid community (e.g., [12], [13]) as well as the network community (e.g., [14]-[16]). In order to support end-to-end guaranteed service in the Internet, the IETF has defined the Integrated Services (Intserv) architecture [14]. Later in [15], the network element (router) behavior required to deliver a guaranteed delay and bandwidth in the Internet were described. Furthermore, Resource Reservation Protocol (RSVP) [16] complements Intserv by enabling the resource reservations on the routers along the path.

Based on [12]-[16] and the related studies, the network is assumed to support end-to-end guaranteed service in which a share of any link bandwidth can be reserved for the transmission of files with deadlines, and then released afterwards. As a result, in DGridSim, when a file with deadline needs to be moved from a source to its destination storage element, the path and bandwidth value which allow the timely delivery of the file over the network will be computed and the computed bandwidth value will be reserved on all links along the path until the end of the transmission by a reservation service. Later, a file transfer service is called upon to deliver the file using the reserved network resources.

Running TCP Traffic

Different models have been proposed to model the throughput of a TCP flow. Chiu states that transfer rate of a TCP connection (or flow) in a bottleneck link inversely proportional to its round trip time (RTT) [17]. The definition of the bottleneck link and the bandwidth sharing mechanism is explained in [18].

Henry Casanova designed a simulator, called SimGrid, with the intent to study scheduling algorithms in heterogeneous platforms [19]. In version 2 (2003), SimGrid starts to use analytical models in the data transfers in favour of the wormhole models used in the previous version [20]. SimGrid by default uses the flow level analytical TCP model MaxMin sharing strategy. With the latest version (v3.3), it also implements Vegas and Reno analytical TCP models. With pseudo-model integration of GTNets, it enables packet level simulation of data Grid networks as well [21]. Performance comparisons with GTNets reveals that while SimGrid runs orders of magnitude faster than packet based simulators, it incurs acceptable inaccuracies from packet based simulators.

Velho and Legrand identified regimes of large inaccuracies in Casanova's work. They presented an improved model that accounts for the slow-start feature of TCP and congestion [22]. In this study, we used the models proposed in [18] and [22] as the underlying bandwidth sharing mechanisms in simulating TCP traffic.

3. REAL-TIME DATA DISSEMINATION MODEL

Definition 1: Network can be specified with a directed graph $G = (V, E)$ such that $|V|=p$ ve $|E|=r$. $V = \{v_1, v_2, ..., v_p\}$ defines the heterogen machines on the system. These machines can be storage elements with limited storage space or routers with more limited or no storage space. $e_{ij} \in E$ is the network element that connects machine i and machine j. Network element e_{ij} can be modelled with a bandwidth c_{ij}, delay d_{ij}, and a cost function ξ_{ij} which can be a function of bandwidth and/or delay.

Assumption 1: Network can be modelled with a directed graph.

*Assumption 2:*Bandwidths are non-negative.

$$\forall (i,j) \in E: \ c_{ij} \geq 0$$

Definition 2: A route $\pi_l \in \Pi$ ($|\Pi|=s$) is the collection of links that connects source and destination machines. A link can not appear more than once in a route.

Assumption 3: Cost functions are additive. A cost of a route is just the sum of the cost links on the route.

$$\text{cost of route } \pi_l, \xi_{\pi_l} = \sum_{(i,j)\in E} \xi_{ij} \colon e_{ij} \in \pi_l$$

Definition 3: A request $d_m \in D$ ($|D|=t$) is modelled with $<t_m, f_m, dl_m, n_m>$ quadruple in which t_m is destination machine, f_m is requested data item, dl_m is deadline, and n_m is the diversity factor which determines the maximum number of splits associated with flow m.

Assumption 4: A data item can reside on more than one source machines. For a request d_m, list of source machine s_m

and file size fs_m are both known. A data item can be sent from more than one source machine via more than one route. In this work, this setting is called Real Time Splittable Data Dissemination Problem (RTS/DDP).

Assumption 5: Link delays are small and they can be neglected.

By means of assumptions 4 and 5, the minimum bandwidth to satisfy the request can be found as

$$h_m = \frac{fs_m}{dl_m - t_c} \qquad t_c\text{: current time}$$

Definition 4: A three dimentional matrix $\Delta_{r\text{-}s\text{-}t}$ with members $\delta_{ijlm} \in \Delta$ can be defined as:

$$\delta_{ijlm} = \begin{cases} 1, \text{link } (i,j) \text{ belongs to route } l \text{ satisfying demand } m \\ 0, \text{otherwise} \end{cases}$$

Definition 5: A diversity factor, n_m, more than one allows the division of flow for request d_m to at most n_m pieces. Each of these pieces is greater than zero. The routes and magnitude of these pieces do not have to be the same. But, such conditions can easily be added to the problem definition.

Definition 6: RTS/DDP is defined as finding two dimentional matrix $X_{s\text{-}t}$ with members $x_{lm} \in X_{s\text{-}t}$: x_{lm} = the magnitude of flow through path π_l for request d_m.

Assumption 6: While solving RTS/DDP, a centralized scheduler that controls all system resources is assumed to be in place. This assumption is consistent with similar studies from the literature.

Assumption 7: The problem has an optimal solution with an upper bound.

Condition 1: The sum of flows using a link should not exceed the capacity of the link.

$$\forall\, e_{ij} \in V\text{:} \quad 0 \le \sum_{l=1}^{|P|} \sum_{m=1}^{|D|} \delta_{ijlm} x_{lm} \le c_{ij}$$

Condition 2: Sum of flows for request d_m is equal to the required bandwidth for the request.

$$\forall\, d_m \in D\text{:} \sum_{l=1}^{|P|} x_{lm} = \begin{cases} h_m, \text{demand } m \text{ is satisfiable} \\ 0, \text{otherwise} \end{cases}$$

For best-effort problems, the aim in most cases is to complete data transfer requests as early as possible. In some cases, the aim may be to minimize a cost function or to maximize link utilization.

Since this work focuses on real-time criteria, the aim differs from best-effort problems. If a schedule to meet the requests' deadline can be found, the request is called *satisfiable*. Percentage of satisfiable request over all requests is called *satisfiability*. In this study, the aim is set as finding a schedule for flow matrix, X, to maximize satisfiability.

To formalize this aim, we define the normalized flow matrix $\bar{X}\text{:}\ \overline{x_{lm}} = {x_{lm}}/{h_m}$. Hence,

$$\forall\, d_m \in D\text{:} \sum_{l=1}^{|\Pi|} \overline{x_{lm}} = \begin{cases} 1, \text{demand } m \text{ is satisfiable} \\ 0, \text{otherwise} \end{cases}$$

In this case, the aim is to maximize the following function.

$$\max \sum_{m=1}^{|D|} \sum_{l=1}^{|\Pi|} \overline{x_{lm}}$$

The rest of the paper uses the problem formulation provided in this section.

4. SIMULATION FRAMEWORK

As shown in Figure 1, the problem is divided into two steps: First step is to find the routes of flows to satisfy data transfer requests. Second step is to to determine how much flow is sent on which route.

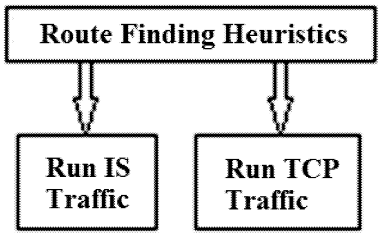


Figure 1: Simulation Framework

4.1. Route Finding Heuristics

In first step, we used four different heuristics defined in literature for best effort systems:

Random
For all the routes from which the file can be sent, send the file via a random route.

Maximum Capacity
For each request, choose the path with maximum capacity. Capacity of a route is defined as the minimum residual bandwidth of links on route.

Minimum Delay
For each request, choose the route with minimum delay.

Minimum Cost
For each request, choose the path with minimum cost function (can be a function of delay and link bandwidths). This study assumes the following cost function:
CostOfLink = Delay + 0.0702/Bandwidth [22] where delay is expressed in seconds and bandwidth is expressed in Mb/s.

4.2. Running Traffic

In this study, we developed algorithms to run two different traffic models: IS (Integrated Service) model and TCP model.

In IS model, it is assumed that the centralized scheduler is able to reserve and use the system resources for the duration of data transfer. The flows are assigned a bandwidth that is only enough to deliver the data items at (not before) the deadline.

Running IS Traffic

NumberOfSatisfiedRequests=0
Shuffle requests
for m=1 **to** *NumberOfRequests*
  Find route π_m of request *m*
  Find lowest residual bandwidth, C_m, on π_m
  Determine h_m: required bandwidth
  if $C_m > h_m$
    Demand d_m is satisfiable
    Commit h_m on links on route
    Find residual network
    Increase *NumberOfSatisfiedRequests*
  else
    Demand d_m is not satisfiable
  end if
end for

TCP model uses bandwidth sharing algorithm proposed by Casanova and Marchal [18]. The model suggests that the bandwidth at the bottleneck links are shared inversely proportional to the round trip times. Velho and Legrand improved the original model to better model the slow start of TCP traffic and congestion in the links [22].

Running TCP Traffic

NumberOfSatisfiedRequests=0
Shuffle requests
Multiply all bandwidths by 0.92 to include congestion effects
Transfers start after $10.4 \sum_{k\mid\mathcal{F}_i \text{ uses } \mathcal{L}_k} L_k$ seconds where L_k is link delay.
while there are *unprocessed* requests
  Arrange bandwidths
  Calculate expected completion times for requests in process
  Increase time to minimum expected completion time
  Calculate completion percentages of requests
  Mark request with minimum expected completion time as *processed*
  Mark the completion time of this request
end while
for m=1 **to** *NumberOfRequests*
  if request completion time < deadline
    Increase *NumberOfSatisfiedRequests*
  end if
end for

Arrange Bandwidths

while there are routes left
  Compute utilization for each link
  Find the bottleneck link
  for each route through bottleck link
    Find lowest residual bandwidth, C_m, on route
    Find the fair share of bandwidth, FS_m, that is assigned inverse proportional to all w_i's on bottleneck. w_i's can be calculated as follows:
$w_i = \sum_{k\mid\mathcal{F}_i \text{ uses } \mathcal{L}_k} \left(L_k + \frac{8775}{B_k}\right)$ where B_k is link bandwidth
    Route is assigned a bandwidth that is the smaller of C_m and FS_m
    Find residual network
  end for
end while

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of the algorithms, a simulation program that can be used to emulate the execution of randomly created data transfer requests on a simulated grid system was developed. The simulator was written in C++ programming language.

With the start of the simulation, a heterogeneous computing system is randomly created based on two parameters, namely the number of *storage machines* p and the number of *links* q. In order to interconnect p links, a network of q links is employed, where the network topology is randomly generated and each machine is randomly connected to a different machine. This simulation model closely mimics a computing system where a set of machines is interconnected by a switched-based network. Note that the network itself can also be heterogeneous where several networking technologies are simultaneously used. During the simulations, all data transfer requests are assumed to come in to the Grid at time zero. For each submitted request, it is associated with a data file available in the Grid. Requested data items are chosen randomly.

Using the simulator developed, a set of simulation studies were conducted. First, a base set of results was established. As shown in Table 1, in base simulations, transmission rates of links are assumed to be uniformly distributed between 0 and 10 Gbits/sec and link delays are assumed to be uniformly distributed between 0 and 5 seconds. In the base set, four algorithms are evaluated in terms of the percentage of the satisfied requests under the following simulation parameters:

Table 1: Simulation parameters in base simulations

	Parameter	Value
Fixed	Number of machines	10
	Number of links	30
	Number of data files	30
	Maximum link delay	0.005 s
Variable	Number of requests	1000
	Maximum deadline	500 s
	Maximum link bandwidth	10 Gbit/s
	Maximum file size	10 Gbits

The results of the base simulation studies are presented in Table 2, where each data shown is the average of satisfiability values of 10 simulation runs. Maximum probable satisfiability value is 100 which indicates that all requests are satisfied. Note that each iteration of the simulation creates a different Grid topology and request set under the given simulation parameters.

According to the base simulation results shown in Table 2, minimum delay and minimum cost perform close to each other. Maximum capacity performs a little worse. Random performs the worst. In terms of the effect of type of the traffic on the performance, IS, in general, performs better than TCP traffics.

As shown in Table 2, AMSEs of IS and TCP are close in maximum capacity, minimum delay and minimum cost (within 1%), and random performs about 20% worse. The results of IS and TCP differs primarily for maximum deadline values that are smaller than 200 seconds. This is the result of slow start of TCP traffic.

Table 1: Results of base simulation

Algorithms	IS	TCP
Random	36,65	35,90
Maximum Capacity	50,27	51,41
Minimum Delay	55,14	53,48
Minimum Cost	54,89	54,34

Later, variable simulation parameters is individually varied to study the impact of the parameter on the performances.

Table 3 shows the average mean square error (AMSE) between results of IS and TCP. Smaller numbers indicate better AMSE performances. As can be seen from Table 3, best AMSE performances are obtained in maximum capacity. Minimum delay and minimum cost performs about 50% worse in terms of AMSE. AMSE of random is about five times larger than the average mean square error of maximum capacity.

Table 3: Average Mean Square Errors (AMSE) between IS and TCP results for different tests

	Random	Max Capacity	Min Delay	Min Cost
TEST1	54,2	11,5	19,0	17,9
TEST2	90,6	77,4	76,0	74,4
TEST3	30,8	5,8	17,1	17,2
TEST4	40,5	8,6	20,5	19,2

Figure 2 shows the effect of varying the number of requests. Number of requests are varied from 100 to 2000. Corresponding satisfiability values are shown. Satisfiability decreases with the increase in number of requests. Random yields the worst performance. Other algorithms perform similar performances.

Figure 3 shows the effect of varying the maximum deadline. Note that a maximum deadline value of 50 indicates that requests are associated with deadlines uniform between 0 and 50 seconds. Maximum deadline is varied from 50 to 1000 seconds during tests. As can be seen from Figure 3, Random performs the worst performance, followed by maximum deadline. Other algorithms yield show similar performances.

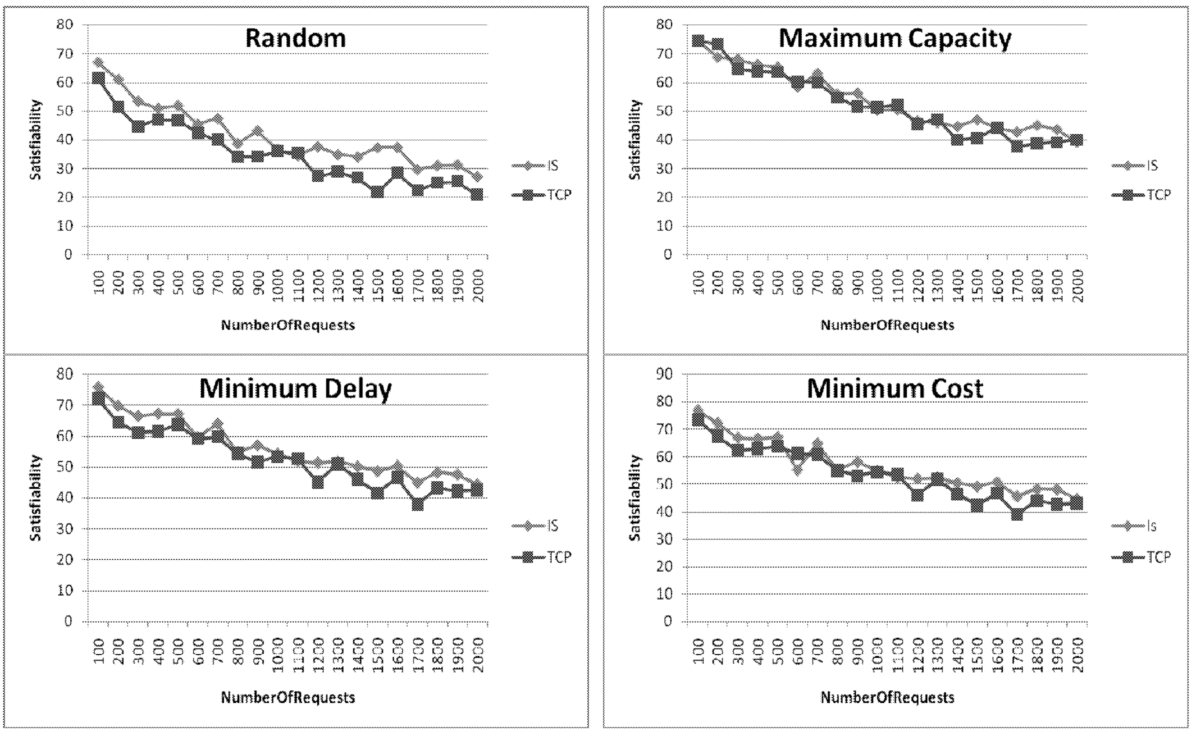


Figure 2: TEST1: Effect of varying number of requests on satisfiability

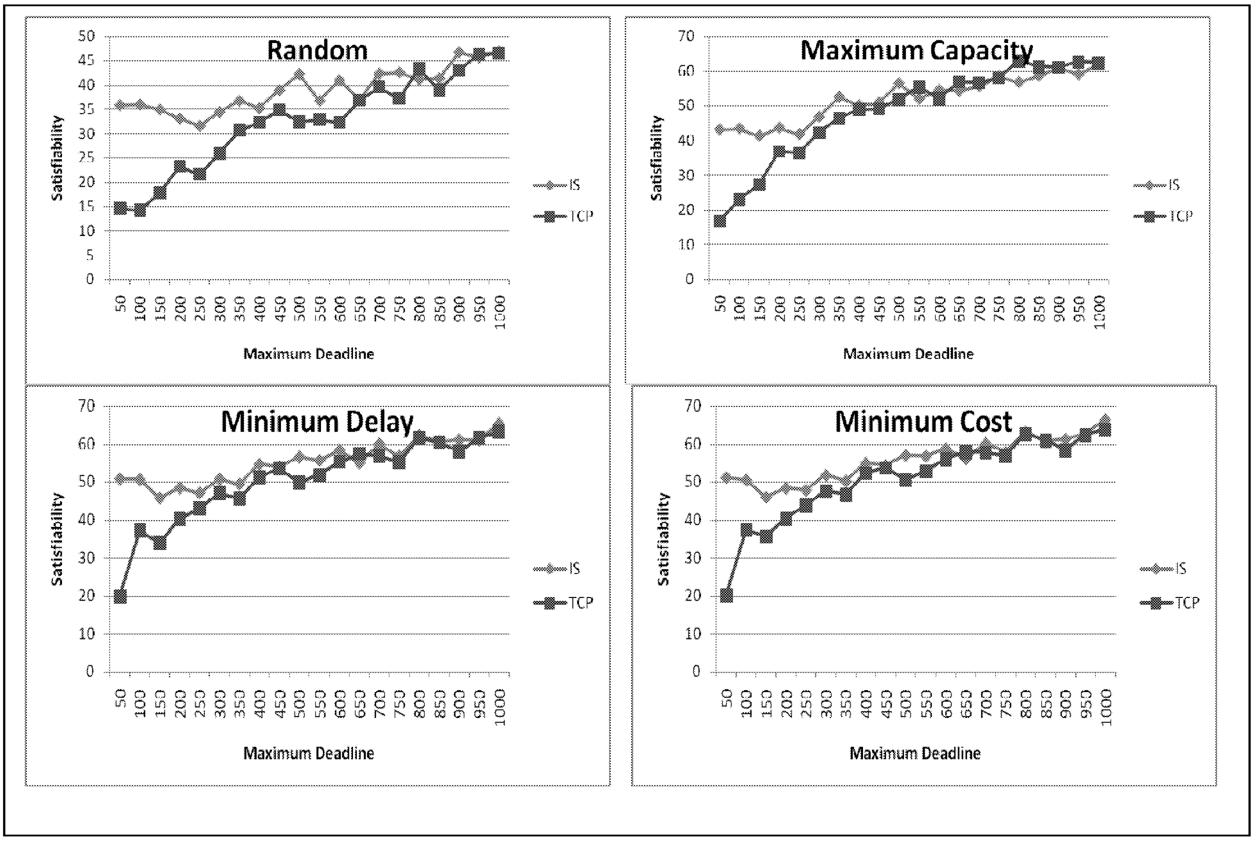


Figure 3: TEST2: Effect of varying maximum deadline on satisfiability

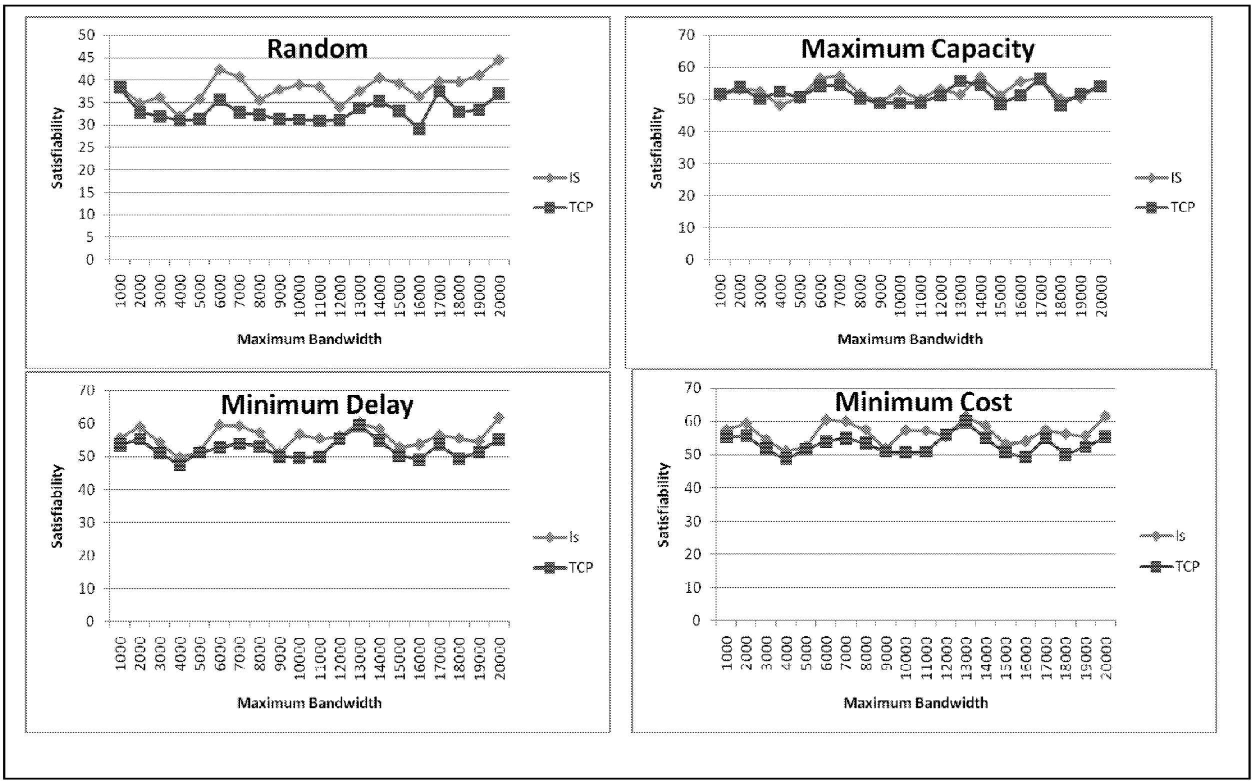


Figure 4: TEST3: Effect of varying maximum bandwidth of links on satisfiability

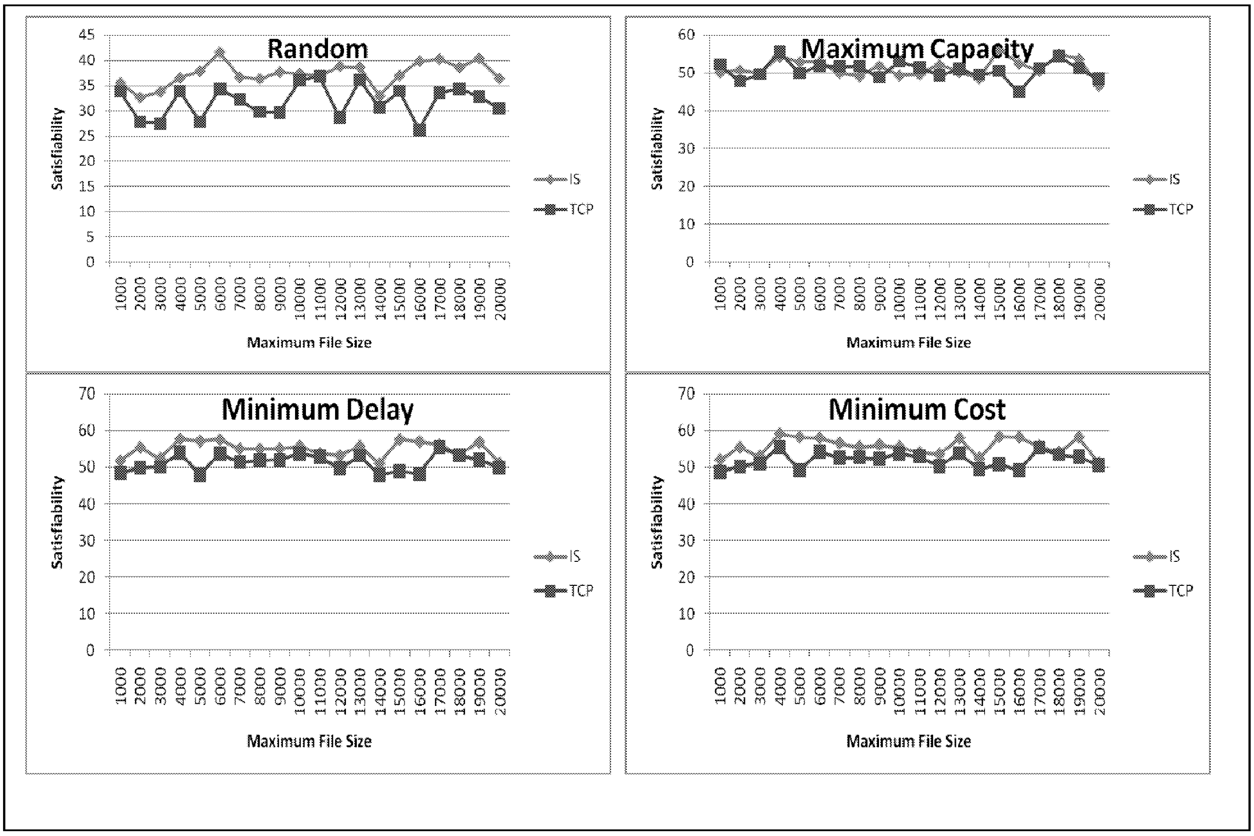


Figure 5: TEST4: Effect of varying file sizes on satisfiability

As can be seen from Figure 3, AMSEs of IS and TCP are close in maximum capacity, minimum delay and minimum cost (within 1%), and random performs about 20% worse. The results of IS and TCP differs primarily for maximum deadline values that are smaller than 200 seconds. This is the result of slow start of TCP traffic.

Figures 4 shows the effect of varying the maximum link bandwidth (between 0 and 20 Gb/s) and Figure 5 shows the effect of varying maximum file size (between 0 and 20 Gb). As seen from the figures, random performs the worst performance, followed by maximum capacity. Other algorithms yield show similar performances.In both cases, in terms of AMSE between IS and TCP, maximum capacity yields best results. AMSE of minimum delay and minimum cost are about 3 times; AMSE of random is about 6 times larger than that of maximum capacity.

6. CONCLUSION

This work presented a real time data dissemination model and introduced a real-time multi-source data dissemination problem. Than divided the problem into two subproblems: 1) Choosing routes for each request 2) Determining the amount of flow to be sent in each route.

From the results presented in the previous section, it is evident that the route finding algorithms and traffic models has significant impact on the real-time Grid performance.

Among four different route finding algorithms, choosing the path with minimum delay, minimum cost function yields similar results. Choosing the path with maximum capacity yields slightly worse results. All these algorithms perform better than randomly choosing the path.Among the different traffic models, Integrated Service shows better performance than TCP based traffic models.

ACKNOWLEDGEMENT

This material is based on work supported by Institute of Scientific and Technological Research of Turkiye (TUBITAK) under Grant No. 108E232.

REFERENCES

[1] The DataGrid Project, http://eu-datagrid.web.cern.ch/eu-datagrid/.

[2] Network for Earthquake Engineering Simulation, http://it.nees.org/.

[3] R. K. Ahuja, T. L. Magnanti, ve J.B. Orlin, "*Network Flows*," Prentice-Hall, Englewood Cliffs, NJ, 1993.

[4] E. W. Dijsktra, "*A Note n Two Problems in Connexion With Graphs*," Numerische Mathematik, 1959.

[5] Bellman, R, "*On a Routing Problem*," Quarterly of Applied Mathematics, 1958, 16, pp 87-90.

[6] Ford, L.R, "*Network Flow Theory,"* Technical Report, P-923, Rand Corp., Santa Monica, CA, 1956.

[7] Hart, P. E., Nilsson, N. J., Raphael, B., "*A Formal Basis for the Heuristic Determination of Minimum Cost Paths,"* IEEE Transactions on Systems Science and Cybernetics SSC4 4 (2): 100–107, 1968.

[8] RFC 1058, *Routing Information Protocol*.

[9] RFC 2328, *OSPF Version 2*.
[10] RFC 1142, *IS-IS protocol specification*.
[11] J. Kleinberg, "*Approximation Algorithms for Disjoint Paths Problems,"* Ph.D Thesis, Dept. Of EECS, MIT, 1996.
[12] I. Foster, C. Kesselman, C. Lee, R. Lindell, K. Nahrstedt, and A. Roy, "*A Distributed Resource Management Architecture That Supports Advance Reservation and Co-allocation*," Int'l Workshop on Quality of Service, pp. 27-36, 1999.
[13] R. J. Al-Ali, K. Amin, G. von Laszewski, O. F. Rana, D. W. Walker, M. Hategan, and N. Zaluzec, "*Analysis and Provision of QoS for Distributed Grid Applications*," Journal of Grid Computing, vol. 2, no. 2, pp. 163-182, 2004.
[14] RFC 1633, *Integrated Services in the Internet Architecture: an Overview*.
[15] RFC 2212, *Specification of Guranteed Quality of Service*.
[16] RFC 2205, *Resource Reservation Protocol (RSVP)*.
[17] D. N. Chiu, "*Some Observations on Fairness of Bandwidth Sharing*," Technical Report, Sun Microsystems, 1999.
[18] H. Casanova and L. Marchal, "*A network model for simulation of grid application*," Technical Report, 2002-40, LIP, 2002.
[19] H. Casanova. "*SimGrid: A toolkit for the simulation of application scheduling*," In IEEE International Symposium on Cluster Computing and the Grid (CCGrid), page 430, Los Alamitos, CA, USA, 2001. IEEE Computer Society.
[20] H. Casanova, A. Legrand, and L. Marchal, *"Scheduling Distributed Applications: the SimGrid Simulation Framework,"* In IEEE International Symposium on Cluster Computing and the Grid (CCGrid), May 2003.
[21] H. Casanova, A. Legrand, and M. Quinson, "*SimGrid: a Generic Framework for Large-Scale Distributed Experiments*," In IEEE10th International Symposium on Computer Modeling and Simulation, Mar 2008.
[22] P. Velho, and A. Legrand, "*Accuracy Study and Improvement of Network Simulation in the SimGrid Framework*," SIMUTools Second International Conference on Simulation Tools and Techniques, 2009.

OPTIMIZATION PROTOCOL OF BIOCHEMICAL NETWORKS FOR EFFECTIVE COLLABORATION BETWEEN INDUSTRIALISTS, BIOLOGISTS AND MODELLERS

Ivars Mozga
Egils Stalidzans
Biosystems Group, Faculty of Information Technology,
Latvia University of Agriculture
Liela iela 2, LV3001, Jelgava, Latvia
ivars.m@gmail.com

KEYWORDS
Biology, Numerical methods, Optimization, Model analysis, Dynamic.

ABSTRACT

Due to the growing number and size of biochemical network computer models an efficient co-operation among experts industrialists, biologists and modellers in industrial biotechnology is becoming a topical issue. In order to speed up the introduction of systems biology achievements into production an organism independent universal collaboration protocol is necessary incorporating a sequence of actions starting with a choice of criterion and model up to industrial tests.
Protocol of cooperation of industrialists, biologists and modellers engineering the performance of a microorganism using systems biology approach to improve the efficiency of biotechnological process is developed. Method is described in form of an algorithm that consists of four consecutive stages starting with determination of the product of interest and optimization criteria and ending by an estimation of industrial process feasibility analysis taking into account the features of the available industrial equipment.
Application of protocol in yeast glycolysis optimization is given as example.
The protocol facilitates efficient resource usage and time savings in biochemical network optimization process due to rational sequence of operations and reduction of optimization duration.

INTRODUCTION

Interdisciplinary scientific fields generally and systems biology (Bruggeman and Westerhoff, 2007; Kitano, 2002) in particular suffers from efficient methods of interdisciplinary collaboration between biologists, modellers, mathematicians, chemists, bioengineers and others solving joint tasks (Aebersold, Hood, and Watts, 2000; Ideker, 2004). Additional problem is the industrial implementation of scientific achievements because of different attitude and expectations of scientists and industrialists about industrially attractive biotechnological process (Otero and Nielsen, 2010). This kind of problem can be addressed by a development of a collaborative optimization protocol.
This paper describes a protocol of cooperation of industrialists, biologists and modellers engineering the performance of a microorganism using systems biology approach to improve the efficiency of biotechnological process. Method is described in form of an algorithm that consists of four consecutive stages starting with determination of the product of interest and optimization criteria and ending by an estimation of industrial process feasibility analysis taking into account the available equipment.

STATE OF THE ART

Currently engineering of intracellular biochemical processes is a growing field in biotechnology aiming to reduce impact of several efficiency related topics: increase of product/substrate ratio, reduction of energy costs, reduction of side products and others. The development of new bioprocesses has also relatively new topics to cope with. Growing number of mathematical models of cellular bioprocesses stimulate increasing use of simulations and optimization of mathematically described process (Li et.al. 2010). Usually dynamic models of intracellular biochemical networks are described in a form of a set of non-linear differential equations that can be optimized using time consuming numerical methods (Hirmajer, Balsa-Canto, and Banga, 2009).
Usually optimization is made by biologists creating a hypothesis of a process improvement involving a small number of system elements because of difficulty to deal with high number of cross-talks and interactions between elements of systems – reactants and enzymes of biomolecular reactions inside the cell. Hypotheses are usually generated by biologists that are specialized in a particular organism. Then hypothesis is tested in a dynamic model. In case of success a biological experiment is following to test the feasibility of a process in a living organism. That is essential because engineering is performed using a small scale models (up to 30 reactions) of the process of interest because full scale dynamic models (thousands of reactions) are not developed yet.
Problem of the above mentioned approach is the high probability that not a complete space of optimal solutions is searched systematically (Schulz, Bakker, and Klipp,

2009) even in 30 reactions models. Thus complex interactions and non-intuitive industrially interesting solutions may be not hypothesized and tested. Additional problem is the limitation on the number of organisms that are known to particular teams of biologists. That increase risk of finding suboptimal solution of a problem because of lack of experience.

Another topic that reduce the intensity of implementation of scientific achievements is that industrialists are usually involved in late stages of optimization process and early rejection of biologically attractive but industrially uninteresting solutions are rejected in a stage when significant amount of resources are already spent.

There is a need for industrially oriented protocol of collaboration of industrialists, biologists and modellers in development of biotechnologically effective engineered organisms exploring all the space of possible solutions with early rejections of industrially uninteresting solutions.

This article describes protocol of optimization of concentrations of reaction speed regulating enzymes. Concentrations of enzymes can be altered by influencing the transcription and translation (Klipp, 2005) intensity of enzymes . Enzyme concentration can be both increased and reduced thus respectively increasing and decreasing reaction flow (speed of the reaction). The protocol can be adapted for a wider range of tasks during optimization of biochemical networks.

OPTIMIZATION PROTOCOL

The protocol consists of four interlinked stages as shown in Figure 1. Each stage is meant to narrow the field of search of the best solutions as early as possible.

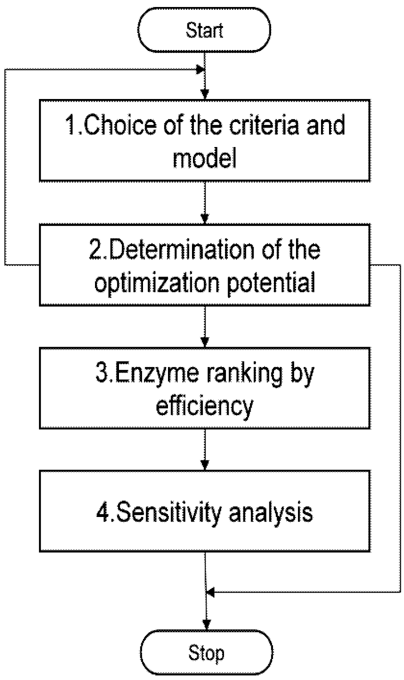


Figure 1 Four stages of optimization protocol

Stage 1: choice of the criteria and model

The first stage – choice of the criteria and model (Figure 2) is collaboration between industrialists, biologists and modellers to set the scope of the optimization work. It has a great influence on costs and duration of the optimization process. Main contributors at this stage are the industrialists. They should clearly express their interests both in mathematical terms and in form of other limitations including legislation, environmental issues and others which may be unknown by biologists and modellers. Still collaboration with biologists and modellers is needed to assure that the criteria's can be calculated from the chosen model. Extension of a model may be needed to link it with the criteria.

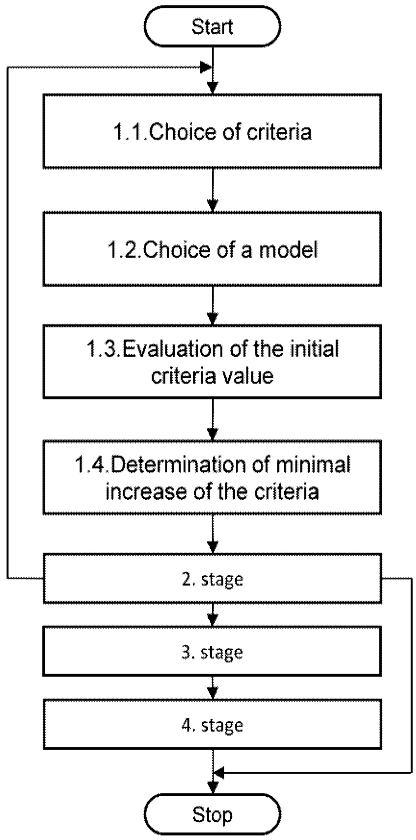


Figure 2 Algorithm for the stage "choice of the criteria and model"

Choice of criteria (1.1.) is a complex task. Too narrow criteria can limit the search in a very narrow area of possible solution space and end without industrially interesting solution at early stage. Too general criteria might cause the opposite drawback – search can be too wide and limitations that were not indicated at early stage can come up at the fourth stage and solutions might be rejected already after spending significant resources to analyze it. A model that includes criteria related processes should be chosen (step **Choice of a model (1.2.)**) from respective databases (Chen et.al. 2010; Oliver and Snoep, 2004) by modellers in collaboration with biologists. The model can be slightly adapted to have a steady state (Klipp, 2005) that is a prerequisite for industrial biotechnological process. The next step is the **evaluation of the initial criteria value (1.3.)** (by modeller) of the original model before optimization. In the next step the **determination of minimal increase of the criteria (1.4.)** should be done by industrialist operating mainly with economical factors. Thus it is defined that criteria values below the minimal increase are not interesting for industrialist and further research is meaningless if minimal increase is not reached.

Stage 2: determination of the optimization potential

The task of the second stage (Figure 3) is the determination of the optimization potential. It is necessary to find out if the chosen criteria and organism can give increase of optimization criteria above the value defined in the first stage. In case of failure the first stage has to be repeated.

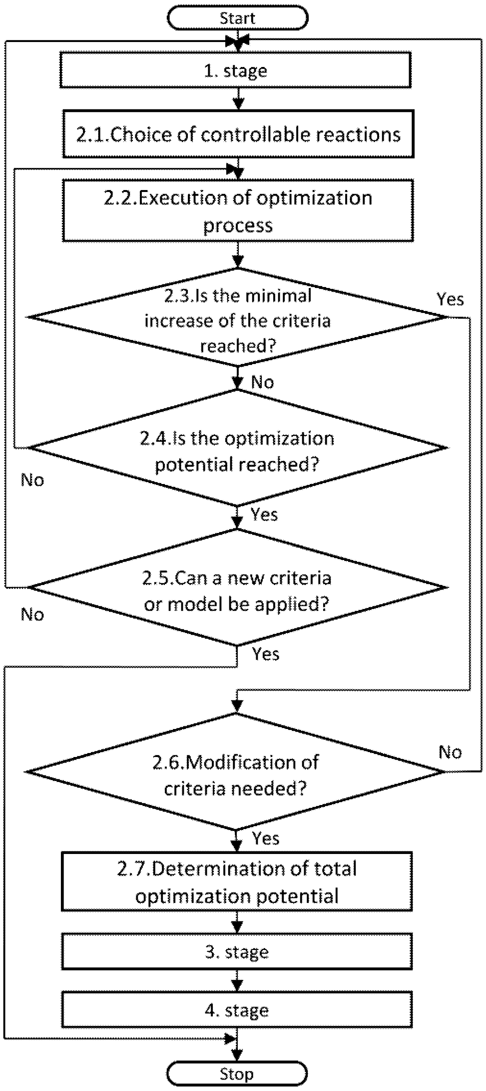


Figure 3 Algorithm for the stage "determination of the optimization potential"

The first step of this stage is **the choice of controllable reactions (2.1.)** in the model that happen by the help of an enzyme and therefore their flow can be adjusted by change of enzyme concentration. Reactions can be selected by biologists and modellers using bases SABIO-RK (Wittig et.al. 2006), BRENDA (Scheer et.al. 2006) or KEGG (Kanehisa et.al. 2010). The next step is the **execution of optimization process (2.2.)** performed by modeller making optimization run where all the reaction related parameters are allowed to change in a wide range, for instance, from -99% of the initial value of the original model up to + 1000%. Modellers should consult with biologists because 1) some reactions may need some biomass related flow and 2) to interpret biologically correctly the reaction parameters of optimal solution. This step can be performed using general numerical optimization software or a systems biology related tools and model formats like COPASI (Hoops et.al. 2006), Potters Wheel (Maiwald and Timer, 2008), SB Toolbox2 (Schmidt and Jirstrand, 2006) and others. In case of a big model with high number of parameters the optimization run can take long time (several days) until optimization potential is clarified. Therefore optimization process should be observed by a modeller asking **"is the minimal increase of the criteria reached?" (2.3.).** In case of a negative answer the question **"is the optimization potential reached? (2.4.)** should be estimated. In other words it means: is the optimization reached maximal value? Positive answer indicates that the optimization has ended and it's potential is not high enough and industrialist should be asked: **"can a new criteria or model be applied?" (2.5.)** to see if adaptation of criteria's can help. Negative answer means that the optimization protocol has not found industrially feasible solution even in modelling level. In case of positive answer to the question 2.3. the criteria has to checked in the step **"Modification of criteria needed?" (2.6.).** The question should be asked to the industrialist as the optimization process and growth of criteria could lead to unrealistic values of parameters. Negative answer to the question 2.6. should be followed by **determination of total optimization potential (2.7.)** done by modeller running the optimization until progress of optimization criteria reaches it's best value for given model. An indication for that is a long stagnation of optimization process.

Stage 3: enzyme ranking by efficiency

The third stage enzyme ranking by their efficiency in increase of the criteria (Figure 4) is performed by the modeller except for the step 3.5. which is done by industrialist. Some of enzymes (via reactions that they control) are strongly contributing in increase of criteria while some of them show very small or no influence on the criteria. Change of concentration of particular enzyme is costly. So it would be of advantage to find out 1) how the values of criteria are growing depending on the number of modified enzymes and 2) which would be the best enzymes to influence in case of modification of set of one, two, three and so on enzymes. For instance in case of one enzyme the most efficient is the enzyme A reaching 40% (increase of 40%) of total optimization potential, in case of two enzymes those are enzymes B and D reaching 65% (increase =65-40=25%) and in case of three enzymes those are A, D and E covering 73% (increase =73-65=8%) of total optimization potential.
The increase of criteria becomes smaller and smaller while the number of enzymes in combinations are increasing. That leads to a situation when adding of more enzymes to be modified becomes too expensive for the small increase of expected efficiency.

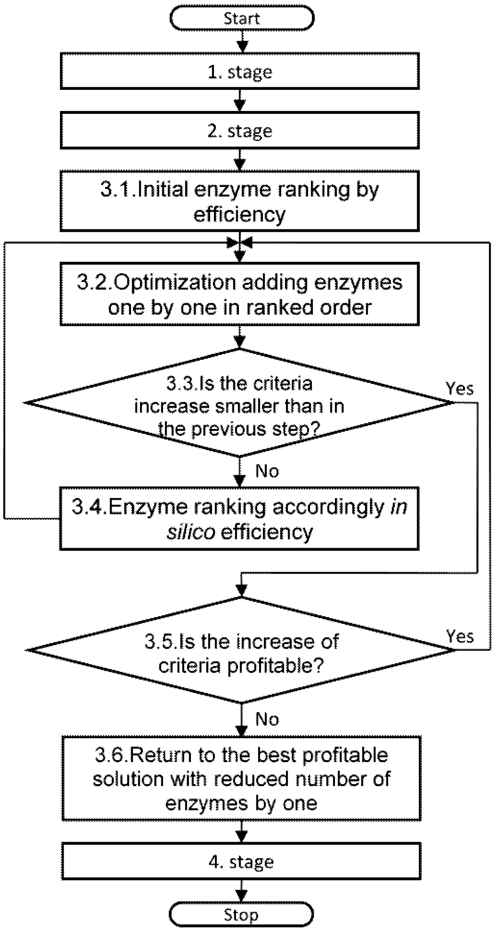


Figure 4 Algorithm for the stage "enzyme ranking by efficiency"

Scanning of all the possible combinations (no order important and no repetition allowed) of enzymes to be modified the number of optimizations K can be calculated by formula:

$$K = \sum_{r=1..j} \frac{n!}{r!(n-r)!} \qquad (1)$$

where r – number of available enzymes to choose from, j - up to which number of enzymes combinations are counted. Thus a model with 15 enzymes that can be influenced in case of up to five enzyme combinations would have 4943 possible combinations to optimize. In case of up to 10 there would be 30826 combinations and in case of 15 there would be 32767 combinations. Target of this stage is to find the best combinations of enzymes for any number of enzymes to be modified in shortest time possible.

The first step is the **initial enzyme ranking by efficiency (3.1.)**. Depending on the model several methods can be used by the modeler to determine the influence of concentration change of a particular enzyme on the other reactions and the value of the criteria. Some of them are Metabolic Control Analysis (Crabtee et.al.1985), (Fell, 2005), (Fell and Sauro, 1985), (Reder,1988), ranking accordingly to the efficiency increase in case of single optimized enzyme or others. It is assumed that the efficiency rank will not be predicted accurately because of nonlinearity of nonlinear differential equations. **Optimization adding enzymes one by one in ranked order (3.2.)** has to be performed by modeller in a cycle with the next step **"Is the criteria increase smaller than in the previous step?" (3.3.)** as long as the answer is negative. Step 4.3. is followed in cycle by **enzyme ranking accordingly in silico (computer simulation) efficiency (3.4.)** changing the efficiency rank of enzymes accordingly the optimization outcome. If the efficiency increase in the step is smaller than for the previous number of enzymes (3.3.) the question **"Is the increase of criteria profitable?" (3.5.)** comes up. Negative answer leads to the end of the cycle and **return to the best profitable solution with reduced number of enzymes by one (3.6.)** that will be analysed in the fourth stage. During the step 3.2. each case of criteria increase and corresponding set of optimization parameter values has to be saved as some of solutions may have weak stability and suboptimal solutions may become the best feasible ones in the fourth stage.

Stage 4: stability analysis of best solutions

During the fourth stage "Sensitivity analysis of best solutions" (Figure 5) the best technically feasible stationary state should be found taking into account the dynamic parameters of both the cellular dynamics and industrial control system. Modellers and industrialists are the main contributors at this stage.

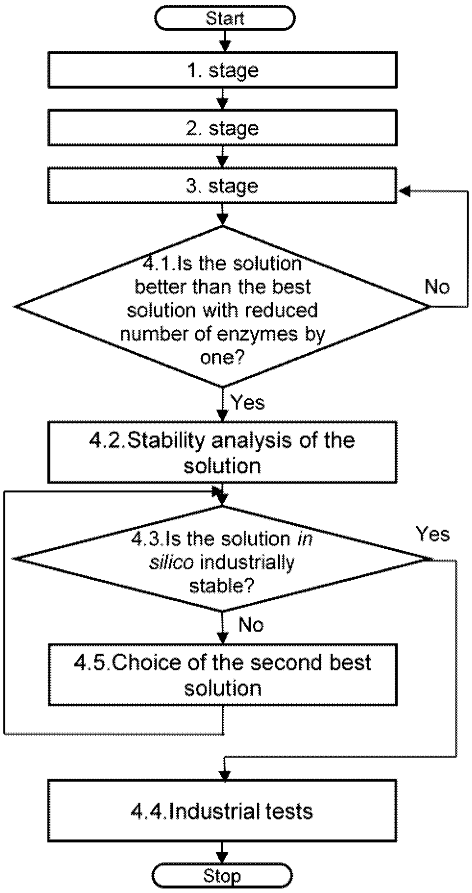


Figure 5 Algorithm for the stage "sensitivity analysis of best solutions"

The stage starts with a question **"is the solution better than the best solution with reduced number of enzymes by one?" (4.1.)**. In case of negative answer the best

solution of combination with less number of enzymes has to be preferred because due to reduced complexity and realization costs. Thus return to the step 3.6. is needed.
Stability analysis of the solution (4.2.) is the next step to be performed. Robustness of bioprocess has to be estimated under known dynamic parameters of cellular biochemical process of interest and the accuracy and dynamic parameters of the available control system of of the bioreactor or other industrial system. Different approaches can be used depending on the available information about the control system. One of methods is the implementation of dynamic model of the control system into the model of the bioprocess and run the stability analysis of simultaneous interaction of both systems.
If the answer to the question **"Is the solution *in silico* industrially stable?" (4.3.)** is positive, the **industrial tests (4.4.)** can be started on level of biological experiments and in case of success may be implemented in the industrial process. In case of negative answer to the question 4.3. the **choice of the second best solution (4.5.)** has to be performed by a modeller and return to the step 4.1 thus forming a cycle.
The optimization progress recorded in the step 3.2. is used in the step 4.5. If the solution that corresponds to the best value of criteria is not stable, the solution with next best criteria value has to be analyzed for stability.

APPLICATION CASE

The protocol has been applied to increase the profitability of yeast glycolysis in production of ethanol from glucose using model published by Hynne and colleagues (Hynne et.al.2001). The application case is related to the production of biofuels. The model contains 24 reactions and 25 reactants (Figure 6). 15 reactions are performed by the help of enzymes with variable concentrations that influence flow of the corresponding reaction. Using the protocol and software tool COPASI combinations of modified enzyme concentrations was found. The optimization criteria was:

$$C = \frac{Ethanolflow}{Glucose\ flow} + 5 * Ethanol\ flow \quad (2)$$

where
Ethanol flow – product produced,
Glucose flow – feedstock consumed.

As a result increase of the criteria from 4,99 (Ethanol flow = 0,804 mmol/min, Glucose flow = 0,832 mmol/min) was increased to 13,3 (Ethanol flow = 2,27 mmol/min, Glucose flow = 1,15 mmol/min).

The fourth stage of the protocol was not performed as it depends on equipment parameters and may be specific from case to case. The increase of ethanol production was reached without ensuring the production of biomass and therefore might perform in biological experiments with lower efficiency than in computer simulations.

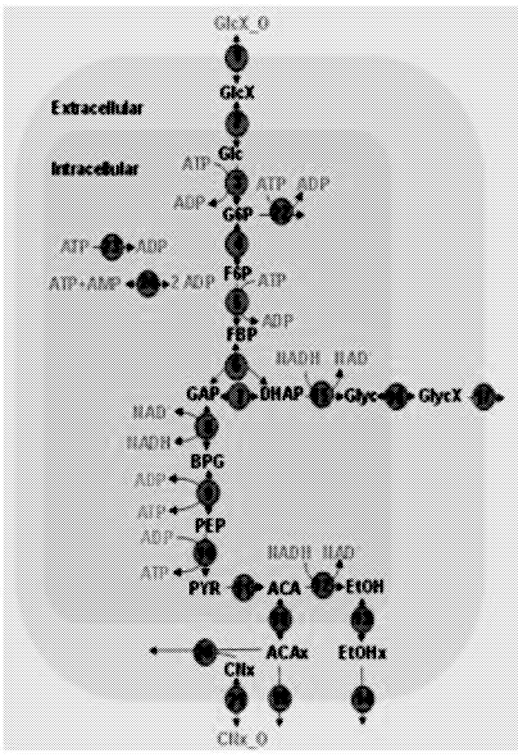


Figure 6 Yeast glycolysis model (Hynne, 2001) Screenshot from JWS online website (http://jjj.biochem.sun.ac.za/database/hynne/index.html)

CONCLUSION

The developed protocol describes cooperation of industrialists, biologists and modellers optimizing the industrial performance efficiency of a microorganism. The protocol consists of four consecutive stages: 1) choice of the criteria and model, 2) determination of the optimization potential enzyme 3) ranking by efficiency and 4) sensitivity analysis of best solutions.
Early rejecting of industrially and/or biologically unfeasible solutions increase the efficiency of optimization.
The protocol propose systematic scan the solution space for the best solutions thus checking also contra intuitive solutions and eliminating subjective approach of particular scientist.
Significant in silico increase of yield and and ethanol production is achieved in case of yeast glycolysis models (Hynne et.al. 2001).

ACKNOWLEDGEMENTS

This work is funded by a project of European Structural Fund Nr. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128 'Latvian Interdisciplinary Interuniversity Scientific Group of Systems Biology' www.sysbio.lv

REFERENCES

Aebersold, R., Hood, L. E., and Watts, J. D. (2000). Equipping scientists for the new biology. Nature biotechnology, 18(4), 359. doi: 10.1038/74325.

Bruggeman, F. J., and Westerhoff, H. V. (2007). The nature of systems biology. Trends in microbiology, 15(1), 45-50. doi: 10.1016/j.tim.2006.11.003.

Crabtree B., Newsholme E.A. (1985) A quantitative approach to metabolic control. Curr Top Cell Regul 25:21-76.

Fell D.A. (2005) Metabolic Control Analysis. In: Alberghina L., Westerhoff H.V. (eds.) Systems Biology, Definitions and Perspectives, Springer-Verlag Berlin Heidelberg, Germany, pp. 69-80.

Fell D.A., Sauro H.M. (1985) Metabolic Control Analysis: Additional relationships between elasticities and control coefficients. Eur J Biochem 148:555-561.

Hirmajer, T., Balsa-Canto, E., and Banga, J. R. (2009). DOTcvpSB, a software toolbox for dynamic optimization in systems biology. BMC bioinformatics, 10, 199. doi: 10.1186/1471-2105-10-199.

Hoops S., Sahle S., Gauges R., Lee C., Pahle J., Simus N., Singhal M., Xu L., Mendes P. and Kummer U. (2006). COPASI — a COmplex PAthway SImulator.Bioinformatics 22, 3067-74.

Hynne F., Danø S., Sørensen P.G. (2001) Full-scale model of glycolysis in Saccharomyces cerevisiae. Biophys Chem, 94(1-2):1, pp.21-63.

Ideker, T. (2004). Systems biology 101--what you need to know. Nature biotechnology, 22(4), 473-5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15085805.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M.; KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 38, D355-D360 (2010).

Kitano, H. (2002). Computational systems biology. Nature, 420(6912), 206-10. doi: 10.1038/nature01254.

Klipp E., Herwig R., Kowald A., Wierling C. and Lehrach H. (2005) Systems Biology in Practice, Concepts, Implementation and Application, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 467 p.

Li C., Donizelli M., Rodriguez N., Dharuri H., Endler L., Chelliah V., Li L., He E., Henry A., Stefan M.I., Snoep J.L., Hucka M., Novère N.L. and Laibe C. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC Systems Biology 2010, 4:92

Maiwald T. and Timmer J. *Dynamical Modeling and Multi-Experiment Fitting with PottersWheel*. Bioinformatics 2008, 24(18):2037-2043

Olivier, B.G. and Snoep, J.L. (2004) Web-based modelling using JWS Online, Bioinformatics, 20:2143-2144

Otero, J. M., and Nielsen, J. (2010). Industrial systems biology. Biotechnology and bioengineering, 105(3), 439-60. doi: 10.1002/bit.22592.

Reder C. (1988) Metabolic control theory: a structural approach. J Theor Biol 135:175-201.

Scheer M., Grote A., Chang A., Schomburg I., Munaretto C., Rother M., Söhngen C., Stelzer M., Thiele J., Schomburg D. BRENDA, the enzyme information system in 2011. Nucleic Acids Res., 39:670-676 (2011)

Schmidt, H. and Jirstrand, M. (2006). Systems Biology Toolbox for MATLAB: a computational platform for research in Systems Biology. Bioinformatics, 22(4), 514–515.

Schulz, M., Bakker, B. M., and Klipp, E. (2009). TIde: a software for the systematic scanning of drug targets in kinetic network models. BMC bioinformatics, 10, 344. doi: 10.1186/1471-2105-10-344.

Wittig U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J. and Rojas, I. SABIO-RK: Integration and Curation of Reaction Kinetics Data. In proceedings of the 3rd International workshop on Data Integration in the Life Sciences 2006 (DILS'06). Hinxton, UK. Lecture Notes in Bioinformatics, 4075: 94-103(2006).

AUTHOR BIOGRAPHY

IVARS MOZGA studied IT in Latvia University of Agriculture. Since 2006 he is a PhD student in Latvia University of Agriculture specializing in computer control of biological systems. Doctoral studies are connected to modeling of biomolecular processes within cellular structures. Researcher in a project of European Structural Fund Nr. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128 'Latvian Interdisciplinary Interuniversity Scientific Group of Systems Biology' www.sysbio.lv

EGILS STALIDZANS studied in Riga Technical University (Latvia), Faculty of Computer Science and Information Technology where defended his PhD in 2005. Research is mainly related to development of mathematical models under information insufficiency. In 2010 he became associated professor and senior researcher at Latvia University of Agriculture, Faculty of Information technology. Since 2006 he leads Biosystems Group. He is interested in application of Systems Biology approach in control of industrial bioprocesses. Project coordinator in a project of European Structural Fund Nr. 2009/0207/1DP/1.1.1.2.0/09/APIA/VIAA/128 'Latvian Interdisciplinary Interuniversity Scientific Group of Systems Biology' www.sysbio.lv

AN EVOLUTIONARY OPTIMISATION FOR ELECTRICITY PRICE RESPONSIVE MANUFACTURING

Tobias Küster   Marco Lützenberger   Daniel Freund   Sahin Albayrak

DAI-Labor Berlin
Ernst-Reuter Platz 7
10587 Berlin, Germany
E-Mail: ⟨firstname.lastname⟩@dai-labor.de

KEYWORDS

Industrial processes, Industrial control, Scheduling, Optimization

ABSTRACT

With an increased amount of energy provision through volatile energy sources such as wind turbines, time dependent pricing of electricity will gain importance and have an impact on overall energy costs, especially for large consumers such as manufacutring plants. Therefore, in future electricity markets, price-based scheduling of manufacturing activities can be used to maintain or increase cost efficiency of manufacturing processes. Software applications taking into account process models and short-time price projections, can serve to compare alternative production plans and identify efficient configurations with respect to all relevant interdependencies. In this paper, we present an approach how to find the most cost-efficient production schedule using an optimisation based on a highly flexible process metamodel.

INTRODUCTION

The potential of simulation based product- and process development has been discovered by the industry long since. To the day, there are many sophisticated and mostly commercial products available for this purpose. While most of these frameworks allow for the simulation of complex production systems and the evaluation of controlling strategies, we were not able to identify any serious implementations, which take energy related criteria into account. Nevertheless, increasing energy costs as well as eco-political restrictions render according considerations inevitable and have the impact to shift the orientation of the producing industry, entirely.

From a business perspective, reducing energy costs will be the main driver for energy aware manufacturing and increased energy efficiency in production. Two different approaches for cost reduction can be determined here: On the one hand, it is possible to decrease the amount of energy consumed by a specific process; on the other hand, it is possible to decrease the costs of a specific amount of energy that is needed for manufacturing processes.To gain cost efficiency through the latter approach - reducing the costs for a given amount of energy - three basic actions can utilised: 1. Time-dependent tariffs can be taken into account. 2. Energy generation capacities operated within the manufacturing plant, such as diesel generators, solar panels, wind turbines, etc. 3. Large energy consumers can participate in short-term market operations on the spot market for e.g. electricity.

By choosing a time slot for energy consuming processes that is superior to another time slot in terms of costs per kWh any of these three actions can serve to decrease actual energy costs. This price-based load shifting also called price response, can be realised without the use of complex software systems for processes with little interdependencies, such as domestic air conditioning. In most manufacturing processes, however, the calculation of shifted process plans must satisfy numerous constraints and take into account interdependencies among single process steps, why an industrial application of this kind needs to facilitate three basic functionalities: 1. Configuration of the overall production process chain or net with all interdependencies and load characteristics. 2. Specification of the production target and other constraints and criteria considered within the optimization. 3. Evaluation of an improved or optimized manufacturing process plan.

In the work at hand, a software system is described that provides these three functionalities and can be used to model and optimise manufacturing processes with respect to a price function. We assume a dynamic pricing system which reflects the market-economic equilibrium of supply and demand. The result to our optimisation is a production schedule, which is optimal with respect to the available energy offer. For our work, we performed a comprehensive survey (see Related Work) in which we examined related work and collected useful approaches for our own concept. Our concept (see Concept) comprises a meta-model in which we define the interdependencies of the examined production chains and the simulation-based optimisation for which we implemented an evolutionary algorithm. We implemented our concept (see Implementation) and performed an exemplary evaluation already (see Evaluation). Although, we identified several improvements (see Discussion), we are satisfied with the capabilities (see Conclusion) of our approach.

RELATED WORK

This work is concerned with the optimisation of production processes towards energy related criteria. Industry has already discovered the potential of optimising production processes, and as a consequence there are many solutions available. A complete survey justifies an entire article of its own, so for this paper, we attend focus on particular works, which influenced our own implementation the most. To provide some structure to this section, we distinguish between industrial, commercial solutions and academic approaches with innovative concepts. The section is wrapped up with a conclusion.

Academic Approaches

Santos et al. (Santos and Dourado 1999) developed an evolutionary algorithm for the optimal assignment of all the production sections, which they also applied in a particular mill in the Kraft pulp and paper industry. The aim of this work was to contribute to the development of an optimal scheduling system for the mass production on energy related criteria. Due to its dimensions, its multi-objective characteristic, and its presence of a high-order constraint, the authors preclude from traditional optimisation techniques and applied an evolutionary optimisation framework with Pareto ranking, and a technique that preserves the feasibility of the solutions. The work of *Santos et al.* addresses similar problems as the ones we are facing. Many aspects (especially energy related optimisation) are highly interesting for our own approach and have to be considered. Nevertheless, we have to extend these considerations by real life requirements such as shift planning or maintenance. Further, we have to extend the mechanism since we not only optimise on the overall consumption, but on dynamic energy costs.
Bernik et al. (Bernik and Bernik 2007) present a mature production scheduling system which applies evolutionary algorithms and a discrete event simulation model. The main advantage of the presented system is to enhance man-machine interaction in production planning, since the computer is able to produce several acceptable schedules using the given data and a set of criteria. The planner then selects the most suitable schedule and modifies it, if necessary. Scheduling is done on two levels. First, the approximate annual resource utilisation is planned, using material stocks and delivery time for a known customer's data. Subsequently, the operative scheduling of a series of jobs without time limits and without constraints related to the identity of the customer (e.g. for any customer) is done. The approach is also used for raw material cutting in order to reduce raw material assortment and related material order complexity. Compared to our approach, the work of *Bernik et al.* puts emphasis on the available resources as well as on the the annual utilisation of small and medium sized companies. Energy relevant considerations are not provided.
Schreiber et al. (Schreiber et al. 2009) developed an approach to determine the lot-size of a given production line. The lot-size is defined as the number of pieces which is processed at the same time at one workplace with one-off (time) and at the same costs investment for its set up. The authors describe an algorithm to determine the lot size so that production cost will be minimal. For this, evolutionary algorithms are used in a simulation model. Energy aspects do not enter the calculation, however, many described principles such as the applied cost calculation for operation, setup, labour, transport or storage for instance are highly interesting and influenced our own implemenentation.

Commercial Software

ADONIS is a free process management tool (Junginger et al. 2000) supporting the collection, modelling, simulation, evaluation and documentation of business- and production processes. In the first place, *ADONIS* is used for the process optimisation but further areas of application such as quality management, organisation management or controlling are covered as well. The software's flexibility is founded on the applied meta-model mechanism. A graphical editor is provided for the design and manipulation of the examined process system. The simulation functionality can then be used to predict the effects of changes in the present process model. In total, four simulation algorithms are provided. *ADONIS* is mainly focused on production aspects and neglects from energy considerations. However, both, the applied optimisation algorithms and the appliance of meta-models and a graphical editing tool constitute valuable input for our own implementation.
The *Siemens Plant Simulation Software* (*PLM*) (Siemens 2010) facilitates the simulation of complex production systems and controlling strategies. *PLM* uses an object oriented representation of the examined factory and further holds for the simulation of business- and logistic- processes. Configuration of frequent scenarios is supported by predefined modules, which enables for quick preliminary results. For the evaluation, a comprehensive set of tools is provided, allowing for instance the autonomous identification of bottlenecks. Next to an automatic optimisation of system parameters such as lot-sizes or sequencing, PLM provides an open system architecture with interfaces to common protocols (ActiveX, CAD, COM, DDE,

HTML, ODBC, Oracle SQL, SDX, Socket, VRML2, XML, to name but a few).
The *PACE* (Eichenauer 2004) framework has been developed for the modelling of complex processes with many dynamic components. Among others, *PACE* can be used to predict and optimise production processes, monetary flow, allocation scenarios and many more. *PACE* models feature a static and a dynamic part and facilitate engineer-quality modelling as well as appropriate system description. Processes can be designed up to an arbitrary level of detail, while the designer decides on the applied precision. Systems can be captured not only with regard to their contents, but also on their spatial layout. *PACE* models are hierarchically arranged and based on so called *High-Level-Petri-Nets*. For the import of topological data many data formats are supported, such as databases or *Excel* tables.

Conclusion

For our work we performed a comprehensive survey on related approaches. For this paper, we only listed particular approaches with innovative concepts, which we adapted for our implementation. To sum up, one can say, that the field of process optimisation is properly covered by comprehensive and mostly commercial products. There are many academic approaches as well, addicted to increase the efficiency of the available commercial opponents. Still, despite the large spectrum, we were not able to discover optimisation on energy related criteria and considerations of dynamic energy tariffs. For this reason, we decided on a respective implementation. For this implementation we applied many features of the introduced related approaches, such as a evolutionary optimisation principle or meta-model based configuration capabilities. The following section emphasises our concept in detail.

CONCEPT

In our approach we use an evolutionary optimisation to rearrange individual processing steps in order to make the best use of times of cheap energy, for instance due to variations in the availability of renewable energies (wind, solar, etc.).
Of course this approach is only feasible if the factory is not used to its full capacity at all times, but only if there is potential for variations. This may be the case if some machines can be used for multiple tasks, only one of which can be carried out at a time, or in case of variable shifts and breaks. Another requirement is the availability of storage area for intermediate products manufactured in advance to make use of cheap energy.
In preparation of the simulation, the first thing to do is to create a model of the factory, the machinery, the resources and the products. We decided on employing a very simple model, being inspired by Petri nets and

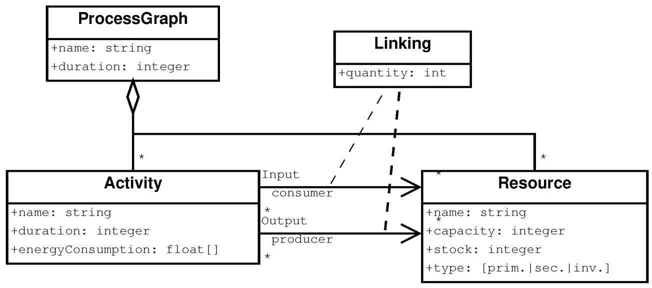


Figure 1: Process Metamodel, slightly simplified.

adding only a bit of domain-specific information on top of that. Basically, all physical entities in the factory are represented by *resources* (products and by-products as well as machinery), and *activities* represent the individual steps in the production as well as supportive processes. We will go into more detail on the model in the next Section.
This model of the factory setup and the individual activities and how they are connected can then be simulated, executing the several activities and consuming and producing resources accordingly. This way, one can simulate how different machines or additional storage capabilities will influence the energy saving potential of the process. Further, the results of the simulation can then be used to re-schedule the execution of the individual activities in the process. In this paper, we propose an automated optimisation using a nature-inspired optimisation algorithm.
The process in the factory is modelled as a bipartite graph of *activities* and *resources*, very similar to a Petri net (Murata 1989): An activity will be "activated", i.e. executable, if both the resources to be consumed by that activity as well as enough capacities for the resources to be produced are available. Other than in a classical Petri net, activities are not executed instantaneously but have a certain duration and there are different types of resources with specific characteristics. Both distinctions account for characteristics of models in the domain under study. In the following paragraphs, all elements are described in detail. Figure 1 shows a slightly simplified diagram of the metamodel used in our approach.

ProcessGraph: The process model as a whole, being made up of activities and resources. The attribute duration stands for a discrete number of steps which the simulation of this process should take.

Activity: A self-contained manufacturing step having the given duration (expressed in multiples of some abstract time step) and energy consumption (one value per time step). Inputs and Outputs are the resources which are consumed and produced by this activity.

Resource: Each resource represents one kind of physical object involved in the production, including raw materials and (intermediate) products, machinery, and even waste heat. Depending on what kind of object it represents the resource's type can be either *primary* or *secondary* (see below). Each resource also has an initial stock, and, most importantly, a capacity, i.e. the number of units the resource can hold. Finally, resources can also cause costs.

Linking: Represents the connection between activities and resources. The quantity specifies the number of units of the respective resource to consume or to produce by the respective activity.

Activities are the elementary steps in the production process and are not to be confused with the physical machines executing these activities. Instead the machines are seen as a special kind of resource. The reason for this is that one activity may be executed by multiple machines (e.g. industrial robots), and at the same time one machine can be involved in multiple activities, automatically imposing constraints on which activities can be executed simultaneously.
Resources are subdivided into three categories:

1. *Primary Resources* typically are resources which are more or less directly integrated into the final product. They include raw materials, pre-fabricated parts provided by external subcontractors, intermediate products and of course the final product.

2. *Secondary Resources* are resources which are consumed or created in the production of an object, but which do not constitute a part of the object itself. Typical examples for such resources may be pressurised air and gasoline for machines, but also for example waste heat.

3. *Inventory Resources* are part of the inventory of the factory, and are as such not consumed or produced, but merely allocated and deallocated. Typical examples are production machinery, transportation and cooling devices.

Based on the kinds of resources being produced and consumed, activities can be differentiated into *primary* and *secondary* activities, as well: Activities which are producing or consuming one or more primary resources, and are thus directly involved in the production, are considered *primary activities*, otherwise we will speak of *secondary activities*.
Electrical energy is not regarded as a resource but treated separately. The reason for this is that while all other resources have to be produced (or disposed of) in some way, electrical energy can be retrieved in (for all practical purposes) unlimited quantities — it just comes out of the power outlet, so to speak. At the same time, unlike the other resources which are assumed to be present at the beginning of the production or produced in the process, electrical energy has a *price*, which moreover can vary over the course of the day.
Further, we see the energy consumption as an attribute of an activity rather than of a machine (i.e. a resource) since the actual energy consumption may depend not only on the machine but on the *activity* in which the machine is involved.

SIMULATION AND OPTIMISATION

The purpose of the optimisation is to find the best possible *production schedule* for a given process model, which is defined by the time (or times) the individual activities of the process graph are executed.
The optimisation process consists of three major aspects:

1. the simulation of a given production schedule,

2. measuring the quality of that schedule, based on the result of the simulation, and

3. finding the schedule with the highest quality.

In the following, we will look at each of these aspects in detail.

Simulation

The simulation of a production schedule keeps track of the resource stocks and the energy consumption in each step of the simulation for the duration of the process. In each step, the simulation checks which activities are to be started according to the schedule, which activities are still running, and which activities are to be ended in the current step, and resources and energy will be consumed or produced accordingly.

- For each activity to be started, the given amounts of primary resources are consumed, and inventory resources are allocated.

- For each activity that is currently running, the given amounts of energy and secondary resources are consumed or produced, respectively.

- For each activity to be ended, the given amounts of primary resources are produced, and inventory resources are deallocated.

Concerning energy consumption and cost, two parameters of the simulation can be adjusted to reflect different settings: First, an *energy price curve* can be provided. In the development prototype, the development of the energy price over a day is approximated using a series of mathematical function, providing for instance a hill-shaped energy price curve. In the near future, we want to extend this feature to take energy price curves from the day-ahead energy market into account. Second, a

base energy level can be specified, being the amount of energy the facility acquires via a flat fee. Energy consumption up to this level has already been paid for, so the *energy price curve* does not apply for that.
Once the simulation is terminated, it holds a record of the amount of consumed energy and the stocks of each resource for each individual step in the execution of the process. These numbers, combined with the resources' capacities and the energy price curve (as acquired from the day-ahead energy market), can be used to determine the *defect*, and finally the *quality*, of that production schedule.

Quality Measurement

The *quality* of a production schedule p is determined by the inverse of its *defect*.

$$quality(p) = \frac{1}{defect(p) + 1}$$

The *defect* of p is the weighted sum of the total energy costs ($p_e \cdot w_e$) and the total over- and undershootings of the several resources' capacities ($p_{r,d} \cdot w_{r,d}$) over all steps of the simulation.

$$defect(p) = p_e \cdot w_e + \sum_{r \in \{p,s,i\}} \sum_{d \in \{l,h\}} p_{r,d} \cdot w_{r,d}$$

Here, different weights ($w_{r,d}$) can (and should) be used for too low and too high resource stocks ($d \in \{l, h\}$) and for the different kinds of resources ($r \in \{p, s, i\}$).
Production schedules, which exceed the maximum or minimum capacities of a resource are not discarded at once, but are given a lower quality rating. For many optimisation algorithms this is necessary in order to overcome a local optimum. For example it is possible that a schedule can be highly improved by swapping two activities using the same machine. During this swap, there may be a few steps in the optimisation progress in which the activities will both occupy the shared machine, but the benefit derived from the new schedule may be big enough to compensate for this temporary defect.

Optimisation

Finding an energy- and cost-efficient arrangement of the several activitities in the process is both a constraint-satisfactory problem and an optimisation problem: on the one hand, there must be no violations of the resources' capacities, and on the other hand the production schedule must be as cost-efficient as possible. In our approach we make use of an Evolution Strategy, as originally introduced by Rechenberg (Rechenberg 1973), which in the following will be introduced briefly.
As the name implies, Evolution Strategy is inspired by natural evolution. Using a $(\mu/\rho + \lambda)$ strategy, an initial "population" of μ individuals (i.e. schedules) is generated. Based on these μ "parents", λ "offspring" are generated by recombining a random selection of ρ parents and slightly altering ("mutating") the result. Finally, the quality of each of the parents and offspring is determined and the μ best individuals are selected to be the parents of the next generation. This process is repeated until a satisfactory production schedule is found.
When applying Evolution Strategy to the problem at hand, three functions have to be implemented next to the quality measurement: How to create the initial population of individuals, how to mutate an individual, and how to recombine individuals.

- The initial population is created by a very simple scheduler, scheduling primary (i.e. production) activities as long as and as early as the primary resources permit. Thus the initial production schedule already contain the maximum possible number of primary activities, but without taking secondary resources or energy costs into account.
- There are several possibilities for mutating an individual, one of which is chosen at random: (a) a random secondary (supportive) activity can be inserted into or removed from the schedule, (b) an activity or a group of activities (primary or secondary) can be moved to another place in the process plan, i.e. being executed earlier or later.
- Due to the many dependencies among the individual activities of a production schedule recombination does not yet work well for this domain and is thus not yet supported.

While Evolution Strategy yields good results most of the time, it is also possible — as with most local search algorithms — that the algorithm gets stuck in a local optimum. To increase the chances of arriving at a solution close to the global optimum, the optimisation should be applied on more than one "population".

IMPLEMENTATION

The approach discussed in this paper is currently being evaluated using a prototypical implementation which can be used to design the manufacturing process to be simulated, to configure and to run the actual optimisation, and to visualise the results.

Metamodel and Model Editor

The process metamodel and a simple graphical editor for creating and configuring process models has been implemented using the *Eclipse Graphical Modeling Framework (GMF)* (see Figure 2).
Following the usual notation for Petri nets, activities are represented by rectangles and resources by circles. Fur-

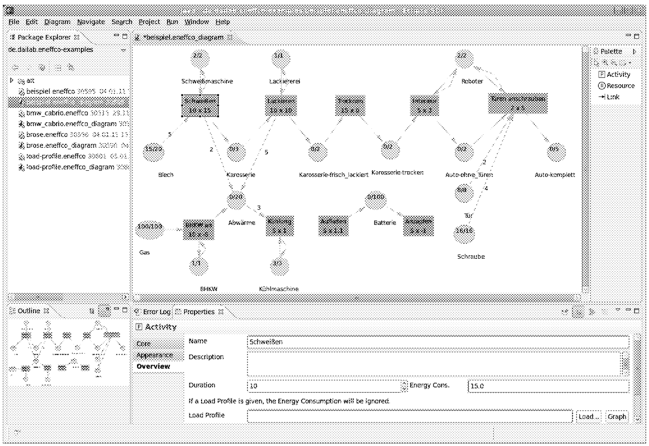


Figure 2: Graphical process editor showing an example process.

ther, we use different line styles and colours for distinguishing the different types of activities and resources (i.e. primary, secondary and inventory). The various attributes can be set directly in the diagram or in the properties view at the bottom of the window. Finally, the XML schema used for saving the diagrams has been optimised for readability.

Optimization Framework

Regarding the optimisation, a generic optimisation framework has been created, which can be used for optimising various domains using various optimisation algorithms. The actual Evolution Strategy algorithm as well as the process model domain have been implemented as plug-ins for this framework (Figure 3).

The system features a large domain-specific area, showing the currently best production schedule as well as a number of controls for selecting e.g. the energy price curve to use, and for visualising the results of the simulation. Further, it provides controls for configuring, starting and stopping the optimisation and showing the results.

Once a process model is loaded and the energy price curve to be used is selected, The optimisation can be started. While the optimisation is running, the upper part of the frame will display the currently best individual, and once the optimisation has come to an end additional charts can be shown, showing the energy consumption and stocks of individual resources over the course of the simulation. Finally, the development of these charts over the course of the whole optimisation can be visualised as a three-dimensional plot using *Gnuplot*.

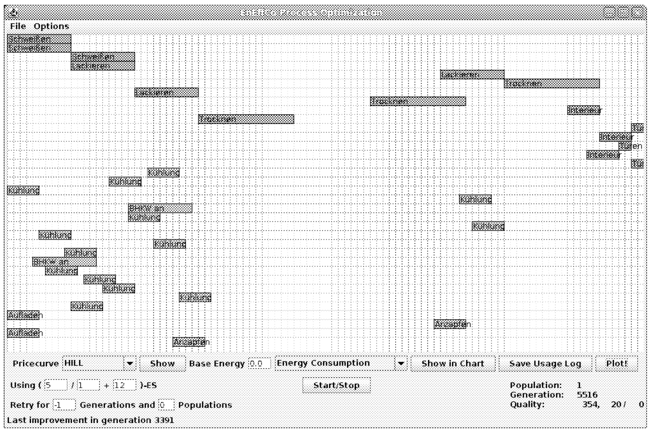


Figure 3: Prototypical user interface for controlling the optimisation and viewing the results.

EXAMPLE

For developing and evaluating the approach a simple, fictional example process was used. In the example a car manufacturing process is pictured, starting with two energy-intensive activities: welding and painting the car body. Once the paint has dried, some interior works are done and finally the doors are attached. For each of the intermediate products there is a separate resource. Further, the process is supplemented with activities and resources for cooling, in-house energy storage and a gas-powered generator. The generator can be used for temporarily lowering the energy consumption, but the price for the gas will add to the production schedule's penalty. The process model can be seen in Figure 2.
While the process surely is very simplified, it demonstrates how different situations can be realised in the process model, for example

- the modelling of the basic production chain,
- one kind of machinery being used for two activities,
- the use of resources associated with a cost, or
- cooling facilities and other supporting processes.

Optimisation Results

The schedule resulting from the optimisation can be seen in Figure 3. Here, the process has been optimised against a hypothetical hill-shaped energy price curve, i.e. with highly-priced energy in the mid of the day and low-priced energy in the morning and evening.
As can be seen, most of the energy-intensive activities (*welding* and *painting*) are taken care of in the morning, with the exception of one instance of the *painting* activity, which has been moved to the afternoon. The high-price period is spend entirely with the *drying* activity, which takes no energy at all. The remaining activities are positioned as late as possible, to get the lowest

possible price for the required energy. Note also, that amount the several instances of the *cooling* activity in the morning there are also two instances of the *charging* activity, charging the aforementioned in-house energy storage, being *discharged* to back up the two *painting* activities highest up the "hill" of the energy price curve. The progression of the optimisation can be seen in Figure 4. As can be seen, the activities are first executed as early as possible (the initial production schedule), and then move quickly to the end of the assumed work day, only to re-distribute again over the course of the day. Finally (in the back of the plot), the activities come to rest so that most of the energy is consumed when the energy price is low.

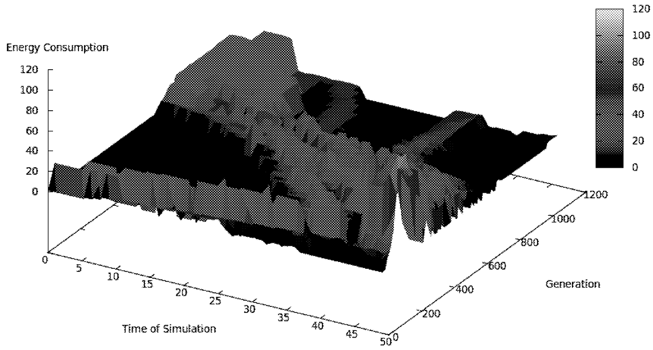


Figure 4: Optimisation of the energy consumption, assuming a hill-shaped energy prive curve

DISCUSSION

First experiments have shown that (a) the metamodel is expressive enough for modelling basic factory settings and at the same time, due to its simplicity, provides easy extensibility, and that (b) the Evolution Strategy based optimisation algorithm works well in the domain of production scheduling, yielding near-optimal results in acceptable time.

Problems with the current approach

Currently the greatest challenge lies in refining the quality measurement function, especially balancing the weightings given to the several factors. While the production schedules are optimised with respect to the energy costs involved, in the course of the optimisation also temporary conflicts in the usage of resources have to be taken into account.
The reason for the early build-up of activities at the end of the example's assumed working day lies in the fact that the initial production schedule does not regard secondary activities, i.e. it does not take cooling into account. Thus, in the first generations there is a massive build-up of waste heat, resulting in a penalty in each remaining step of the simulation. Thus, the first schedules to reduce this penalty are those which move all activities to the end of the simulation, when there are fewer steps left, and thus a lower penalty – even outweighing the penalty resulting from machines being occupied from multiple activities at once. Only when the first schedules introduce instances of the *cooling* activity the situation relaxes and the activities spread out over the whole of the simulated day again.
While the optimisation in the end still finds a near-optimal production schedule, this behaviour results in an unnecessary delay in the optimisation. More importantly, in larger process models such false local optima will be harder to overcome.

Future Work

Most characteristics of manufacturing processes that are relevant for the scope of our work can well be represented through activities and resources, as described before. However, some aspects such as production breaks at lunch time clearly have an impact on the simulation outcome, but can only poorly be modelled with the metamodel at hand. Other aspects which demand for further consideration in the development of a generic metamodel, are energy consumption maxima that may have to be defined for specific points in time. We are currently looking into how this kind of constraints on the production process can best be taken into account.
Apart from elaboration of the metamodel, an additional optimisation method will be introduced, in which parameters of the production process beyond the time of process execution will be varied to explore superior process model configurations. For example the capacity of material storage between two process steps can be increased, to evaluate the impact of increased temporal flexibility on the costs of energy consumed in the manufacturing process.
This will extend the focus of our optimisation from mere scheduling of activities to the actual modification of limiting factors, regarding for example quantities and capacities, for finding manufacturing plant configurations that are cost efficient regarding possible price variations for electrical energy.
However, this will increase necessary computing capacities significantly. We are currently investigating suitable methods of parallelisation for making this feasible.

CONCLUSION

In this paper we described a simulation optimisation, which is based on an evolutionary algorithm and which takes time dependent energy prices into account in order to schedule production processes with regard to their particular consumption. In an optimised production plan, energy-intensive processes are shifted to low-priced time slots, while processes with a low energy usage are scheduled for times of high energy prices. For configuration purposes we developed a process metamodel, which we use for the formal specification of the production processes. Further, we developed a graphical editing tool, which can be used to configure the production processes' detailed characteristics and interdependencies. The configurations constitute the input for our evolutionary optimisation algorithm in which the production process system is optimised towards its energy efficiency. For the pricing model, we adopted a dynamic principle, which reflects the market-economic equilibrium of supply and demand.

For our work, we performed a comprehensive survey. We discovered that there are a lot of similar approaches, some of them at a very mature level. Yet, an optimisation with respect to fluctuating "efficiency intervals" is a novelty so far. To clarify our principle and for evaluation purposes, we defined an easy to comprehend simulation setup, involving a car manufacturing process with some exemplary activities, such as welding, painting, interior works and door attachment. Next to dynamic energy costs we also included alternative in-house energy sources. Using our evolutionary algorithm we performed an optimisation of the described setup and presented the results, which clearly demonstrate the functionality of our implementation. We are currently facing a set of problems, mainly of performance issues.

Yet, convincing results as well as industrial and social demand for according considerations convince us to continue our efforts in the domain of energy-optimisation.

ACKNOWLEDGEMENT

The results presented in this paper were partly established within the project "EnEffCo – Energie-Effizienzcontrolling am Beispiel der Automobilindustrie" (energy efficiency controlling in the automotive industry). This project is funded by the Federal Ministry of Economics and Technology on the basis of a decision by the German Bundestag.

REFERENCES

Bernik I. and Bernik M., 2007. *Multi-criteria scheduling optimization with genetic algorithms.* In *Proceedings of the 8th WSEAS International Conference on Evolutionary Computing.* World Scientific and Engineering Academy and Society (WSEAS), Stevens Point, Wisconsin, USA, 253–258.

Eichenauer B., 2004. *Optimizing Business Processes using Attributed Petri Nets.* In *Proceedings of the 9th Symposium about Simulation as Commercial Decision Help.* 323–338.

Junginger S.; Kühn H.; Strobl R.; and Karagiannis D., 2000. *Ein Geschäftsprozessmanagement-Werkzeug der nächsten Generation – ADONIS: Konzeption und Anwendungen. Wirtschaftsinformatik*, 42, no. 5, 392–401.

Murata T., 1989. *Petri Nets: Properties, Analysis and Applications.* In *Proceedings of the IEEE.* 541–580.

Rechenberg I., 1973. *Evolutionsstrategie : Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Problemata. Frommann-Holzboog, Stuttgart-Bad Cannstatt.

Santos A. and Dourado A., 1999. *Global optimization of energy and production in process industries: a genetic algorithm application. Control Engineering Practice*, 7, no. 4, 549–554.

Schreiber P.; Vazan P.; Tanuska P.; and Moravcik O., 2009. *Production optimization by Using of Genetic Algorithms and Simulation Model.* In B. Katalinic (Ed.), *DAAM International Scientific Book 2009*, DAAM International.

Siemens, 2010. *Plant Simulation — Plant, line and process simulation and optimization.* Project Broschure. URL `http://www.plm.automation.siemens.com/en_us/Images/7541_tcm1023-4957.pdf`. Siemens Product Lifecycle Management Software Inc.

BIOGRAPHY

TOBIAS KÜSTER has obtained his diploma in computer science in 2007 at TU Berlin and is currently working on his doctoral thesis in the field of integrating business process design into agent oriented software engineering.

MARCO LÜTZENBERGER received his diploma in computer science in 2009 from the Technische Universität Berlin and is currently working on his doctoral thesis in the field of multi-agent based traffic simulations.

DANIEL FREUND studied industrial engineering in Information and Communication Technologies with a focus on Project Management at TU Berlin and is now doctoral candidate at DAI Labor. Since 2006 he has been working in various projects focusing on energy efficiency and smart grid value modeling in the residential sector as well as in industrial environments, the distribution grid, in distributed storage networks and in the electric vehicle domain.

PROF. DR.-ING. HABIL. SAHIN ALBAYRAK is the head of the chair Agent Technologies in Business Applications and Telecommunication. He is the founder and head of the DAI-Labor, currently employing about one hundred researchers and support staff.

SIMULATION AND OPTIMIZATION OF AN INTRAOPERATIVE SURGERY WORKFLOW

Hauck Robert
Künzler Urs
Eckerle Jürgen
Bern University of Applied Sciences
Engineering and Information Technology
Computer Perception & Virtual Reality
Quellgasse 21
CH-2502 Biel-Bienne
Switzerland
E-mail: {robert.hauck, urs.kuenzler, juergen.eckerle}@bfh.ch

Hoigné Dominik
Kantonsspital St.Gallen
Klinik für Hand-, Plastische und Wiederherstellungschirurgie
CH-9007 St. Gallen
Switzerland
E-mail: dominik.hoigne@bluewin.ch

KEYWORDS
Virtual reality, Health sciences, Real-time simulation, AI-supported simulation

ABSTRACT

We present a virtual reality based 3D visualization system with the main goal to model, simulate and optimize the workflow of a common surgery in osteosynthesis. The novelty of our research is the simulation and optimization of the workflow of the surgical intervention, as opposed to related projects, which model and optimize the scheduling of services in a clinical environment. Of special interest are the geometry independent workflow description, its simulation, the event-triggered visualization and the 3D character animation in a CAVE environment visualizing a virtual operating room.

INTRODUCTION

HOVISSE (Haptic Osteosynthesis Virtual Intraoperative Surgery Support Environment) is an ongoing medical virtual reality research project which aims to provide a seamless digital support environment for osteosynthesis in trauma care. Its intermediate results have been published in Künzler and al. (2009). One of the primary objectives of this project is to model, simulate and optimize the workflow of a surgical intervention in an operating room. Our virtual reality based 3D visualization system simulates the scenario of a the proximal femoral nail operation. In this operation, an orthopedic implant is used to treat a fracture of a femur bone. It is a very common and standardized intervention in European hospitals.
Our current and future research focuses on the AI component which is used on one hand to process the workflow to be simulated and visualized in real-time and on the other hand to optimize the workflow with respect to the execution time. Especially, a novel multi-agent path finding algorithm, recently developed in our research group (see Eckerle and Roth (2010)), is applied for real-time path planning.
There are several former medical research projects applying AI problem solving methods in which search or planning methods were used to manage and schedule resources, such as clinic personal or hospital beds. In contrast to these projects, our project aims to arrange the medical equipment in the operating room most efficiently. Nowadays, as many as seven to ten people are involved in an medium-sized operation along with an increasing need for medical equipment in the operating room, all acting simultaneously nearby or around the patient. As a result, an optimal placement of the equipment and personnel is critical to guarantee an efficient workflow.
Optimizing the workflow means to minimize the operation time, the main goal in the current system, or to reduce the exposure to X-ray radiation, a goal to be inspected in the future. Until now there is no other system that supports this kind of simulation and optimization of a surgery workflow. As a result, our approach is quite different to former ones.
Our research project includes several tasks: modeling the surgery workflow by a formal language, processing the workflow to be executed by the simulation engine, 3D visualization of the operating room and the medi-

cal equipment needed as well as animating the medical staff while executing an action or moving in the operating room. The workflow is visualized in a 3D CAVE system which allows the visual evaluation of the optimized workflow scenario and an interactive change of the operating room environment. As a result, this project touches several research topics: knowledge representation, man-machine-interaction, 3D character animation and AI problem solving.

RELATED WORK

Medical Systems

A project related to ours is Second Health (Johnston and Whatley 2005). It uses Second Life to create a near-term future vision of what a new generation of hospitals might look like. The system is based on the recommendations of a report on future healthcare delivery in London. The outcomes of this project were four videos showing the ways in which medical processes might be handled in the near future. Since Second Health mainly simulates medical healthcare processes, it bears no direct similarity to our project.

A few medical projects are related to our project: MEDICUS is a knowledge-based planning system which helps to organize the everyday life in a cardiological clinic (Appelrath and Sauer 1998). It supports the management of cardiological interventions and therapeutic activities under limited resources, like the clinical staff, operating rooms and hospital beds, and some other constraints. This system allows to solve the scheduling problem manually by humans or automatically by reactive scheduling algorithms.

Claus (Claus 2005) develops a generic model to describe the flow of treatments in medical departments as well as a suitable software architecture for the implementation of medical information systems. The author introduces the so-called "medical service flow" model for specifying the flow of clinical service processes. This medical service flow model extends the well-known workflow management model, mainly used for business applications. It introduces the concept of services, carried out by service points which represent the care providers. The treatment of a patient is documented in a service flow which is passed on from one service point to the next, along a possible clinical path. The precondition of a service point determines whether a service can be executed next and the postcondition specifies the result of a treatment.

In (Herrler 2007), Herrler examines the simulation and optimization of scheduling processes of clinical treatments in a distributed, dynamic environment. The author developed the Clinical Pathway Modelling Language to model possible clinical paths of services. A specified model can be simulated with a framework based on a multi-agent simulation system, in which a clinical environment contains several autonomously acting units which execute the patient treatments. The author could show that global goals, e.g. minimizing the cost and the execution time, can be optimized exclusively by local scheduling strategies based on an appropriate incentive system.

In summary, the goal of the aforementioned projects is to model, simulate or optimize the scheduling of clinical processes, including an intelligent management of the limited clinical resources by applying operations research or artificial intelligence methods. In contrast to that, in our research the workflow of a single surgical intervention is modeled, simulated and optimized which also includes the geometrical arrangement of medical equipment. Our approach uses a hybrid problem solving process, a scheduling algorithm to solve the logical workflow problem and a path finding algorithm to compute the avatar movement within the virtual 3D operating room.

Multi-agent Path Planning

In our simulation there is a need for a multiple agent path finding algorithm. Silver's WHCA* (Windowed Hierarchical Cooperative A*, see Silver (2005)) is a key algorithm in the group of cooperative path finding algorithms. It executes a sequence of single-agent A* searches, one for each agent. When the current search iteration terminates, a (partial) solution path is reserved in a reservation list. The reservation list maintains all the time-space reservations needed for the move steps along the solution paths computed in previous runs. This way, the other agents are considered as obstacles whose positions cannot be occupied at the given time step. Thus, collisions among the simultaneously moving agents are avoided completely. Because of the high computation cost, the search effort is limited by a maximal search depth d, called search window.

Much effort has been done to reduce the time and space complexity of Silver's original cooperative path finding algorithm, either by using greedy techniques or by using additional techniques like map abstraction and hierarchical search. But until now, all these approaches are limited to or exclusively realized on grid worlds.

We apply a cooperative path finding algorithm recently developed by our research group (Eckerle and Roth 2010). It keeps the simple structure of Silver's WHCA*, but uses a geometric data structure that supports range queries for the reservation list instead of a hash table. As a result, our algorithm can be applied to grid worlds as well as to non-grid worlds without any restriction.

MEDICAL BACKGROUND

Through the use of modern information technology, a lot of progress has been achieved in medicine in recent years. Within the surgical areas the anaesthesia

procedures and the operation techniques were strongly improved. This progress comes along with an increasing specialization of the personnel and diversification of the material. For a medium sized operation, e.g. the osteosynthesis of a broken leg, 7-10 persons from five different occupational groups are directly involved. In addition, technical devices are needed increasingly, for example X-ray-based surgical imaging systems, X-ray-permeable operation tables, navigation systems, etc. (Jacob and Regazzoni 2007). Thus, operations have become extremely personnel, material and cost-intensive.
While the number of involved persons and used devices increases, the room available for the treatment of a patient and the operation field remain constant.The discrepancy of assigned means and limited place becomes a problem. As a consequence changes during the operations are limited and the sterility of the operation field is compromised. Additionally, the use of X-ray-based surgical imaging system leads to a radiation exposure of the operation team. This exposure has to be reduced as far as possible. For medical and financial reasons the operation times are to be kept short.
In this project we analyzed and optimized the workflow in the operating room. As a case study we chose the operation of a broken leg using the implantation of a Proximal Femur Nail (PFNA) (Synthes 2010, Balk and Hahn 2002). This operation is frequent, the implant is common and the operation technique is standardized to a high degree. Despite the high degree of standardization, this operation offers a lot of possibilities for changes in the workflow and thus some potential for improvements (Suhm and Jacob 2003, Matthews and Hoigné 2007). The methods for modeling, analysis and optimization presented here can be applied to many operations. The result of our work intends to contribute to lower costs and improved treatment.

SYSTEM ARCHITECTURE

The software framework we developed is divided into several modules. It is driven by a workflow description, consisting of a set of actions which refer to actors and objects. The main objective of this generic framework was to separate the program logic and the surgery workflow description from the geometrical representation of the referred actors and objects. This allows us to optimize these two entities independently. The system realizes a so-called scaled real-time simulation, the simulation time is proportional to the real time, using a scale factor (see Fujimoto (2000)). After each iteration of the main loop the elapsed time is measured and converted to simulation time with respect to the scale factor. Currently, a scale factor of about 50 is possible, i.e. the simulation can run 50 times faster than in real-time. The software framework has been developed mainly in C++ and OpenGL.
The main framework modules are the simulation engine, the graphic engine, the workflow engine, the path planner and the workflow planner (see figure 1).
The relationship of the modules and the data flow of the system is given as follows:

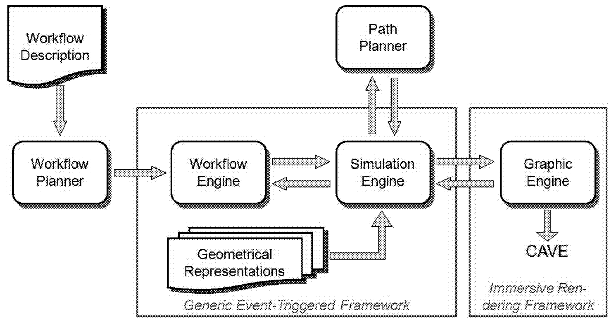


Figure 1: Diagram showing the main modules of our software framework. The arrows depict the coupling of the modules and the direction of data flow.

- The *simulation engine* controls several core components of the surgery workflow simulation such as the event management, the interaction of the user within a CAVE and the animation handling. It applies the discrete event-based simulation technique. Each event corresponds to an action of the surgery workflow, sent by the workflow engine. These events are processed by the animation module of the simulation engine: The simulation advances with respect to the current simulation time after every loop iteration. When a move action has to be processed, a move request is sent to the path planner before execution.

- The *workflow planner* (in a future release to be included) processes the workflow description and generates a workflow which is used as input to the workflow engine. Currently, the workflow is generated by a forward search, later it should be done by a partially-ordered planner.

- The *workflow engine* processes the (partially-ordered) plan and sends repeatedly a group of actions which can be executed simultaneously by the simulation engine. When the simulation acknowledges that an action has been successfully processed, the successor actions are sent to the simulation engine.

- The partially completed *path planning* module, a submodule of the simulation engine, solves the shortest path finding problem for a set of simultaneously executed (move) actions. Additionally, it updates the current location of all the agents before each rendering loop.

- When all the move requests have been processed, all actions to be currently executed by the simulation engine are updated in the current iteration of the animation control loop.

- The *graphic engine* builds a scene, including all scene objects (the equipment of the operating room, instruments, avatars etc.) and the state machine processes the action events. If a primitive action can be applied successfully, the corresponding animation is visualized in the CAVE.

WORKFLOW DESCRIPTION

Medical Field Study

To precisely define the medical operation scenario, we made two 2.5 hour videos from an actual PFNA operation in the Kantonsspital Aarau, Switzerland. Then, based on our videos, the PFNA surgical workflow was completely analyzed and specified by our medical and technical research group in a whole day workshop. As a result, the workflow was documented in a workflow diagram (see figure 2). Each actor is represented by a horizontal stripe and all the actions listed in the temporal order of execution during the surgery of our scenario. The workflow diagram is supplemented by several layout pictures, one is shown in figure 3. One key decision was the level of abstraction to be used: We describe the workflow by high-level actions to reduce complexity and explicitly exclude the manual working of the surgeon conducting the actual surgical intervention, since these actions are highly internalized and intuitively executed. Finally, our workflow description could be reduced to approximately 80 actions.

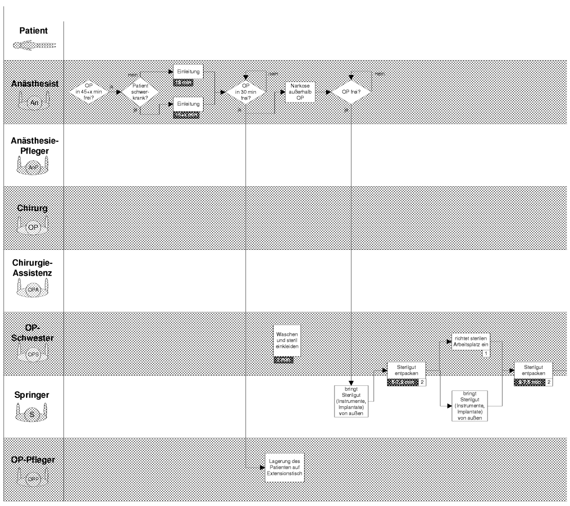


Figure 2: A very small section of the workflow diagram showing the proximal femoral nail surgery workflow scenario. There are seven actors involved (without the patient), each represented by a horizontal stripe. (diagram credits: C. Runde, Fraunhofer IPA)

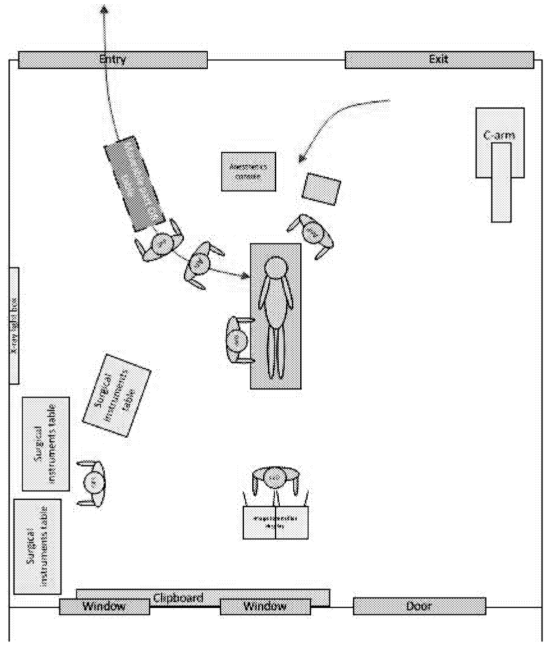


Figure 3: The workflow diagram is supplemented by several layout pictures. Here, the nurse puts away the mobile part of the operation table while the anesthetist pushes the fluoroscope into position. (diagram credits: C. Runde, Fraunhofer IPA)

Problem Representation

First, a key decision was made on what kind of information is to be represented declarative, that is, by a formal language, and which not. As a result, the problem description was separated into two different parts: a workflow description and a geometry description, including the geometrical representations.

Workflow description: The workflow description, which specifies the workflow of the surgery, was specified as a planning problem. This has the main advantage that it allows detecting actions which can possibly be processed simultaneously, although they are usually performed sequentially by humans.

Generally, a planning problem is given by an initial state, by a goal state, and by a set of operations. Each action has a precondition which specifies when the operation is executable and an effect (postcondition) which describes the result of this operation. A solution of the planning problem is given by a set of actions which, when applied with respect to the ordering constraints, transforms the initial state into the goal state (Russell and Norvig 1995). STRIPS, the first planning language developed for the robot Shakey is based on propositional logic (see Russell and Norvig (1995)). The planning language used by our research group is quite similar to STRIPS and is formalized in XML.

We distinguish between two types of primitive actions: Stationary actions that are executed at the actors cur-

rent position, e.g. "the anesthetist makes an infusion", and move actions, consisting of a movement from the actor's current position to a given goal position, for example, "the medical assistant moves the surgery table with the patient". Additionally, there are actions containing a sequence of actions of both types. Only stationary actions have an execution time assigned. The duration of a move action has to be computed by the simulation engine.
For optimization purposes, it has been proved to be advantageous to divide the complete surgery into a sequence of phases. Each phase is characterized by a surgical procedure step, for example "Provisional fixation", and by an initial arrangement of the medical equipment.
Geometry description: This part specifies the 3D geometry of the operating room, the objects and persons involved in the surgical intervention. The initial settings for the 3D scene is defined in XML configuration files. After the initialization, the actual state of the virtual reality scene is maintained by a scene graph, a complex data structure used for maintaining the virtual 3D scene, which contains the position and orientation for all objects and avatars. The scene graph is updated dynamically during the simulation whenever a workflow action is executed and the state of the 3D scene changes.
The connecting link between both descriptions: Our workflow description language purposely does not allow to indicate a position in the virtual operating room by using numeric coordinate values. But we rather use symbolic names in our workflow specification to reference a given location. This allows a complete decoupling of the workflow from the geometrical positions of the equipment in the virtual operating room.

WORKFLOW SIMULATION

The workflow has to be generated from an initial (unoptimized) sequence of actions as it is currently performed in an operating room before it can be simulated and visualized. The workflow generation is based on a complex two-level problem solving process realized by two different modules as explained in the following two chapters.

Workflow Planning

The planning module solves the workflow planning problem and generates a workflow description. This is done by a forward search in such a manner that actions that are pair-wise independent and can be processed simultaneously are recognized and grouped together. The actions of a single group are sent together to the simulation engine and then executed simultaneously. In a later release, we intend to include a workflow planner to optimize and parallelize the currently sequentially executed independent actions.
One of our optimization criterion is the execution time. Optimizing the workflow means to minimize the execution time of the workflow. Our planning problem is familiar with a time planning problem with limited resources which is well-known in literature (e.g. Russell and Norvig (1995)). The actions compete by a set of limited resources which cannot be shared. However, the main difference is that in contrast to the time planning problem, the execution time of a specified move action is not previously known and has to be computed by a shortest path finding algorithm.

Path Planning

For actions that involve a movement, we have a second problem solving step: A move action has to be refined by a sequence of primitive move steps which form a possible path to the goal position. This path is computed by the path planning module in real-time during the execution of the simulation. The path planning module maintains the parallelism of independent move actions.
In the final version, each coarse-granulated (move) action is replaced by a sequence of finer-granulated move steps assigned with a time stamp. However, it is not possible to simply apply A^* or a geometric algorithm because there are typically several actors who act and move simultaneously and collisions could happen. Additionally, a movement in the virtual 3D world has to observe strictly the geometric shape of the objects in the operating room so that it is feasible and has no collisions. Furthermore, the actors compete for limited common resources, for which the resource constraints have to be observed.
Path finding in the virtual 3D room is very time consuming since executing move steps include a lot of expensive operations like collision detection. For that reason, path finding is first done in a simplified, discretized 2D world. The 2D world is given by a projection of the actor's movement onto the ground floor. We apply several techniques well known in the area of computer games. For example, the 2D world is discretized by applying the so-called navigation graph which is a two dimensional regular grid filling of the room (see Buckland (2005)). Since several agents act and move simultaneously, we apply a multi-agent path planning algorithm (see Eckerle and Roth (2010)).
Problems: The corresponding 2D path finding problem simplifies the 3D path finding problem and therefore can lead to collisions in the 3D world. If a path is not executable in the 3D world, that means, the simulation engine has located a collision, the colliding paths should be rejected and an alternative path without any collisions should be generated. However, there is currently no strategy for avoiding collisions.

Workflow Engine

The workflow engine checks after each received notification (i.e. an action has finished) from the simulation

engine whether there are further actions whose preconditions are now fulfilled. These actions can now be fired and the simulation advances. A consistent workflow description guarantees that there is at least one executable action when an action has been successfully executed and is terminated. Thus, the simulation can always make progress.
The simulation engine receives actions to be executed next from the workflow engine. They are added (or when finished, deleted) before the simulation calculates the next step of the actions. All actions currently being processed are maintained in an execution list. The simulation engine requests for each move action of the current execution list a move update from the path planner. If a move action has been finished because the agent has reached its goal position, a notification message is sent to the workflow engine. As a result, the set of successor actions which now can be executed (simultaneously) is sent to the simulation engine.

VISUALIZATION, ANIMATION AND INTERACTIVE CONTROL OF SIMULATION

A primary goal of our simulation is to provide the medical experts a tool to intuitively plan and evaluate the medical procedures and equipment for a specific surgical intervention. Therefore our research project has a strong focus on man-machine interaction aspects. In order to provide the surgeons intuitive and immediate access to the results of the workflow planning simulation, we use a CAVE (Cave Automated Virtual Environment, Cruz-Neira et al. (1993)) system for immersive stereoscopic 3D visualization. A CAVE system consists of multiple projection screens which are mounted in a cubical setup where the actual projection images are adjusted in real time according to the user's tracked position and viewing direction. Our system uses four screens (front, left-side, right-side back-projection screens and a front-projection floor screen) which allows creating an almost real size operating room visualization in which the user's visual horizontal immersion is about 270 degrees. This large screen stereoscopic visualization allows the user an accurate depth and distance perception to immediately identify problems of the workflow optimization which are medically not feasible or potentially risky (e.g. contamination of sterile areas in the operating room). By moving around in this virtual operating room environment the surgeon can view the position of the medical equipment from different perspectives and decide on the suitability of the layout for the tasks to be completed.
Animations of actors are an essential element of realistic simulations (Gillies and Spanlang 2010). For this, the simulation engine contains an animation module that controls all aspects of the 3D animations in the operating room simulation. Actions received from the workflow engine are converted into events that are passed to the simulation engine. For this, each action, which is assigned to either an actor or a medical device, is inserted into an event queue. This queue is processed every time the simulation engine raises an event to update the scene graph and to redraw the frame.

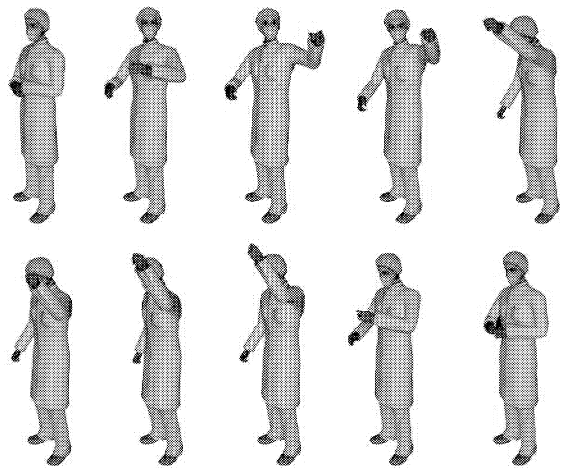


Figure 4: 3D character animation sequence of adjusting an OR lamp.

For each action in this queue, the corresponding 3D animation has to be determined. The animation module can look up the referenced animation in a catalogue, in which every 3D animation is mapped to an object. Using this reference, the animation module plays the 3D animation using the open source 3D character animation library Cal3D (Laurent Desmecht 2006). To synchronize all animations, the 3D character animations use the same calculated time difference between the current and the last iteration as the path finding algorithm uses for calculating the movement of the actors.

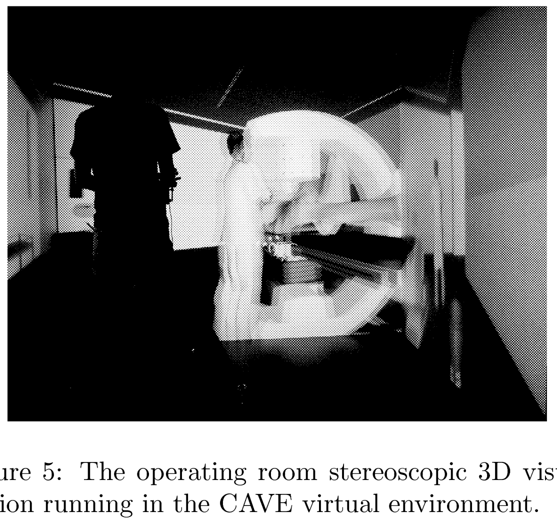


Figure 5: The operating room stereoscopic 3D visualization running in the CAVE virtual environment.

To interactively control simulation parameters and change the initial positions and types of medical equipment, the medical expert can interact with objects of the simulation. For that purpose a haptic device is used to determine the position of the user's hand and to move object in the operating room. A simple button panel widget allows to start, pause, stop and replay distinctive phases of the simulation. Thereby individual workflow phases can be selected and can be iteratively optimized through the repositioning of equipment and the evaluation for medical appropriateness.

Through the use of such an interactive and immersive life size operating room visualization, medical experts are much better supported than in a conventional 2D operating room environment for workflow planning. The change of the environment can be evaluated both by the result computed by an optimization algorithm and by interactive visualization and simulation control.

RESULTS AND FUTURE WORK

We have presented a medical virtual reality based simulation system that uses an innovative approach which is used to model, simulate and optimize a surgical workflow and uses a CAVE for the visualization. The use of 3D character animations supports a virtual surgery simulation by a more realistic visualization. The software framework uses virtual reality and artificial intelligence techniques for the simulation and optimization of the surgery workflow. Initially, the surgical workflow was specified as a planning problem.

The research project is ongoing. There are several possibilities to improve and extend our current system. First, the global optimization of the workflow would require to generate and evaluate all the possible schedules of the operation and to compute the one which is optimal with respect to the selected optimality criterion. This is not yet realized in our system. Second, the workflow could not only be optimized with respect to the duration of the operation, but also to minimize the staff's exposure to the X-ray radiation, or the personnel costs measured by the duration of their stay (work) in the operating room. Third, a final evaluation and verification in a real operating room is still to be done. Finally, our planning language could be extended and our framework generalized such that we are able to apply our system to other problems of simulation and optimization (for example manufacturing problems).

ACKNOWLEDGEMENT

This work has been endorsed and supported by the Hasler Foundation through the funding of the HOVISSE medical virtual reality research project in close collaboration with a Swiss hospital. We acknowledge also the contribution of the remaining HOVISSE WP-C team members, who are not authors of this paper.

AUTHOR BIOGRAPHY

ROBERT HAUCK is a research assistant at the Bern University of Applied Sciences in the research group of Urs Künzler and also a PhD student at the University of Nottingham, UK. His main interest focus on visualization and graphic programming in virtual reality. He is the main software engineer of the HOVISSE project.
URS KÜNZLER is a Professor at the Bern University of Applied Science and leading the research group "Computer Perception and Virtual reality". He is also initiator of the research project HOVISSE.
JÜRGEN ECKERLE is a Professor at the Bern University of Applied Sciences and a member of the research group of Urs Künzler. His main interest focus on simulation, optimization and artificial intelligence, especially Game AI.

REFERENCES

Appelrath H.J. and Sauer J., 1998. *MEDICUS: ein medizinisches Ressourenplanungssystem. KI*, 12, no. 3, 56–60.

Balk R. and Hahn F., 2002. *Die Proximale Femurfraktur. OP-Journal*, 18, no. 2.

Buckland M., 2005. *Programming Game AI by Example.* Wordware Publishing.

Claus M., 2005. *Serviceflow-basierte Modellierung klinischer Abteilungsinformationssysteme am Beispiel der Kardiologie.* Ph.D. thesis, University of Oldenburg, Germany.

Cruz-Neira C.; Sandin D.; and DeFanti T., 1993. *Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. Proceedings of SIGGRAPH'93*, 135–142.

Eckerle J. and Roth M., 2010. *Efficient Multiple-agent Path Planning in Grid and Non-grid Worlds.* In A. Ayesh (Ed.), *GameOn'2010 - 11th International Conference on Intelligent Games and Simulation.* Eurosis, 28–34.

Fujimoto R.M., 2000. *Parallel and Distributed Simulation Systems.* John Wiley and Sons.

Gillies M. and Spanlang B., 2010. *Comparing and Evaluating Real Time Character Engines for Virtual Environments. Presence: Teleoperators and Virtual Environments*, 19, no. 2, 95–117.

Herrler R., 2007. *Agentenbasierte Simulation zur Ablaufoptimierung in Krankenhäusern und anderen dynamischen Umgebungen.* Ph.D. thesis, University of Würzburg, Germany.

Jacob A. and Regazzoni P., 2007. *Medical technology integration: CT, angiography, imaging-capable OR-table, navigation and robotics in a multifunctional sterile suite. Minim Invasive Ther Allied Technology*, 16, no. 4, 205–211. PMID: 17763093.

Johnston C.L. and Whatley D., 2005. *Medicine Meets Virtual Reality 14: Accelerating Change in Healthcare: Next Medical Toolkit*, IOS Press, chap. Pulse!! - A Virtual Learning Space Project. 240–242.

Künzler U. and al., 2009. *HOVISSE - Haptic Osteosynthesis Virtual Intra-operative Surgery Support Environment.* In D. Lalanne and J. Kohlas (Eds.), *Human Machine Interaction: Research Results of the MMI Program.* Springer.

Laurent Desmecht Loic Dachary B.H., 2006. *A skeletal based 3d character animation library.* http://home.gna.org/cal3d/.

Matthews F. and Hoigné D.J., 2007. *Navigating the fluoroscope's C-arm back into position: An accurate and practicable solution to cut radiation and optimize intraoperative workflow. Orthop Trauma.*

Russell S. and Norvig P., 1995. *Artificial Intelligence: A Modern Approach.* Prentice Hall.

Silver D., 2005. *Cooperative Pathfinding.* In *Artificial Intelligence and Interactive Digital Entertainment (AIIDE).* 117–122.

Suhm N. and Jacob L., 2003. *Fluoroscopy based surgical navigation vs. mechanical guidance system for percutaneous interventions: A controlled prospective study exemplified by distal locking of intramedullary nails. Unfallchirurg*, 106, no. 11, 921–928.

Synthes, 2010. *PFNA, Operationstechnik.* Http://www.synthes.com/html/PFNA-Proximal-Femoral-Nail.6869.0.html.

MANUFACTURING OPTIMIZATION

AN INDUSTRIAL CASE STUDY OF WEB-BASED SIMULATION-OPTIMIZATION

Anna Syberfeldt
Ingemar Karlsson
Amos Ng
Virtual Systems Research Center
University of Skövde
P.O. 408, 541 48 Skövde, Sweden
E-mail: anna.syberfeldt@his.se

KEYWORDS
Simulation, optimization, web, industrial case study.

ABSTRACT

This paper presents a web-based simulation-optimization system for improving production schedules in an advanced manufacturing cell at Volvo Aero Corporation in Sweden. The optimization aims at prioritizing components being processed in the cell in a way that minimizes both tardiness and lead times. Results from evaluating the implemented system shows a great improvement potential, but also indicates that further development is necessary before the system can be taken into operation.

INTRODUCTION

Web-based systems have been a strong trend the last few years and a great deal of classical desktop programs is now transformed into web versions. Also the simulation field has been affected by this trend and the term "web-based simulation" is now commonly used in the research community. "Web-based simulation" does not have an exact meaning, but is a broad term including various approaches to integrate the web with the field of simulation. Byrne (2010) states that web-based simulation *"[...] can be defined as the use of resources and technologies offered by the World-Wide-Web (WWW) for interaction with client and server modelling and simulation tools*".

Although web-based simulation is a concept that has been discussed in the research community for 15 years (Fishwick, 1996), the area is still in its infancy (Byrne, 2010). During the last few years, the area has started to grow but there are still only a small number of real applications of web-based simulation (Byrne, 2010). This paper increases this number by presenting a case study of implementing a web-based simulation-optimization system for a real industrial problem. The case study was carried in cooperation with Volvo Aero Corporation, which develops and manufactures high-technology components for aircraft and gas turbine engines. The focus of the case study is on an advanced automated manufacturing cell that simultaneously processes a wide range of different engine components. The machines in the cell, as well as the components being processed, are capital-intensive and the aim of the optimization is therefore to achieve a maximum utilization of the cell.

The case study at Volvo Aero Corporation was carried out as part of a research project called "OPTIMIST" (OPTIMisation using Intelligent Simulation Tools). The overall aim of this multi-discipline (industrial engineering and computer science) research project is to leverage the effectiveness of the Swedish industrial sector by introducing intelligent simulation-based optimization to their daily operations. More information about this research project can be found in Ng et al. (2007).

The next section discusses the underlying reasons for using a web-based approach in the case study by presenting the advantages of web-based simulation-optimization in comparison to classical simulation-optimization systems that are installed on a local computer.

ADVANTAGES OF WEB-BASED SYSTEMS

Compared to classical simulation-optimization systems, several advantages of web-based systems can be identified:

Accessibility
A web-based system is accessible from anywhere with an internet connection, and not only from the specific computer having the simulation-optimization system installed. This also means that a web-based system allows access outside normal business hours.

Scalability
Web-based systems allow for dynamic provisioning of computing resources (Rabinovich and Spatscheck, 2002) and are thereby able to handle an increased number of simulation/optimization requests without performance degradation.

Portability
A web-based system can be run in any web browser on any operating system without requiring recompiling (Suh, 2005). It is not limited only to a traditional computer, but can be run on any device having a web browser (e.g. a mobile phone or an iPad).

Maintenance
Maintenance of web-based systems is easier since they do not have to be installed on each client's computer. Updates are made through a server and reach the clients instantly, which eliminate virtually all on-site maintenance and allows for a frequent update scheme.

Controlled access
Through user logins, a web-based system allows for the configuration of user groups with different privileges based on work tasks. The privileges can be easily changed on the server instead of having the client computer updated.

Licensing
Simulation software licenses are often very expensive, and with a traditional simulation-optimization system it is often required to have one license installed on every computer that runs the system. With a web-based system, the number of licenses can be significantly reduced since the simulation is run on a centralized server.

Altogether, these advantages have motivated the use of a web-based, rather than a traditional local, simulation-optimization system for improving the manufacturing cell at Volvo Aero Corporation, which is further described in the next section.

MANUFACTURING CELL

Volvo Aero develops and manufactures high-technology components for aircraft and gas turbine engines. Today, more than 80 percent of all new commercial aircraft with more than 100 passenger capacity are equipped with engine components from Volvo Aero. Components manufactured at Volvo Aero can be found in military fighter aircraft as well, such as the F/A-18 E/F Super Hornet. As a partner of the European space program, Volvo Aero is also the primary supplier of nozzles and fuel pump turbines for the Vulcain rocket engine. Volvo Aero's facilities are located both in Scandinavia and in the US, and has in total about 3 200 employees. In this work, a factory located at the headquarters of the Volvo Aero Corporation in Sweden is the subject under study.

More precisely, the focus of the study is on an advanced automated manufacturing cell comprising comprises five multi-task machines and five burring stations. The operations that are performed in a machine or at a station vary for different components. Instructions and tools are automatically set up in a machine for the component that arrives, which means that several different components can easily be processed in the cell at the same time.

The machines in the cell, as well as the components being processed, are capital-intensive and it is therefore important to achieve a high utilization of the cell. Finding an efficient processing schedule is, however, non-trivial due to the high complexity of the cell in combination with an unpredictable inflow.

The established method for creating schedules is currently a manual procedure based on trial and error. As there is no guidance on how to change input parameters between iterations, this approach is very time-consuming and requires many iterations and extensive effort by an expert for finding a satisfactory schedule. Furthermore, it does not guarantee that a valid schedule is found, but leaves the validation entirely in the hands of the expert, who is required to consider and carefully control all possible constraints. As there are multiple conflicting objectives to consider when creating a schedule, manual optimization is practically impossible – especially since explicit heuristics for finding a good schedule is missing. Therefore, an automatic simulation-optimization procedure has been identified by the company as necessary for improving the performance of the manufacturing cell. This procedure is presented in the next section.

SIMULATION-OPTIMIZATION

In parallel with the physical build-up of the manufacturing cell, a discrete-event simulation model of it was developed using the SIMUL8 software package. A screenshot of the simulation model can be found in Syberfeldt (2009, p. 73). The aim of building the simulation model was two-folded: 1) to perform what-if analyses aiding the production planning, and 2) to perform simulation-based optimizations. The simulation model has a front-end interface developed in Excel, which for the user facilitates entering input parameters into the model without the need to learn the simulation language. Due to the complexity of the manufacturing process and the various set-up configurations possible in the cell, the simulation model is quite complex. Validity tests, however, indicate that the simulation model represents reality well, and the model is generally accepted among operators working in the manufacturing cell.

When it comes to using the simulation model for optimizations, the main focus is on finding the best prioritizing of engine components being processed in the manufacturing cell. As previously mentioned, several different components are simultaneously processed in the manufacturing cell. In case two or more components arrive simultaneously at a machine or station, a priority number determines which component has precedence. The aim of the optimization is to set the priority numbers of the components in a way that 1) **minimizes tardiness** (overdue components), and 2) simultaneously **minimizes lead times** (the time between a components enters and exit the manufacturing cell). These two optimization objectives are identified by the company as most important to achieve a high utilization of the manufacturing cell.

For optimizing the priorities of components and thereby minimize tardiness of components while at the same time minimizing their lead times, a Hill Climbing algorithms is implemented. Hill climbing is an iterative algorithm that belongs to the family of local search (Russell and Norvig, 2003). The algorithms starts with a random solution to the problem, and attempts to find a better solution by mutating (changing) the solution (Figure 1). If the mutation produces a better solution, this new solution is kept and the procedure is repeated until no further improvements can be made.

```
currentNode = randomStartNode;
 loop until termination criterion fulfilled{
  candidate = Mutate(currentNode)
  if (candidate is better than currentNode)
     currentNode = candidate;
}
return currentNode
```

Figure 1: Pseudo code for a general Hill Climbing algorithm.

Being a local search algorithm, Hill Climbing is good for finding a local optimum but it is not guaranteed to find the global optimum (that is, the best solution out of all possible solutions). Nevertheless, this algorithm has been used in the case study since it is simply to implement and considered to produce as good (or even better) results than other algorithms when the optimization time is limited. The latter is important since a new optimization is performed at every shift change in the cell, allowing the optimization to run for just a few minutes.

In the basic version, Hill Climbing allows only for a single objective to be optimized. In the case study, however, there are two optimization objectives and the basic algorithm has therefore been modified to allow for more multiple objectives. This is done by optimizing one objective at a time without deterioration of the other objective. More precisely, the algorithm starts by optimizing objective 1 (tardiness) since this objective is considered most important by the company. When there have been 15 mutations without finding any better solution with respect to objective 1, the algorithm changes to optimize against objective 2 instead. The optimization of objective 2 is constrained in the sense that a solution improving objective 2 is only accepted if it does not deteriorate objective 1. The whole procedure is described with pseudo code in Figure 2.

```
obj = 1;
obj1Value = undef
numTries = 0;
current = randomSolution;
current.Evaluate()

loop until the user-defined time is up {
 candidate = Mutate(current)
 candidate.Evaluate()

if (obj is 1 AND candidate < current)
   current = candidate;
 else if (obj is 2 AND
          candidate < current AND
          obj1Value is preserved)
    currentNode = candidate;
 else
  numTries = numTries+1

 if(numTries > 15) {
  obj = 2;
  obj1Value = candidate
 }
}
return currentNode
```

Figure 2: Pseudo code for implemented Hill Climbing algorithm.

The Hill Climbing algorithm has been integrated with the aforementioned simulation model to perform simulation-based optimizations. This is done by using the application programming interface (API) provided by the SIMUL8 software. This API makes it possible to control the simulation model programmatically. This API is used to implement an automatic, iterative simulation-optimization process (Figure 3). In this process, the Hill Climbing algorithm generates a set of priority values (input parameters) and feeds them to the simulation model together with a "Run" command. The simulation model performs a simulation based on the provided priorities and computes objective values (tardiness and utilization). The objective values are written to a text file, which is read by the Hill Climbing algorithm. The Hill Climbing algorithm interprets the values and generates a new set of priorities, and then continues the process until the user-defined stopping time is reached. A single simulation-optimization iteration takes about 2 seconds to perform, of which the great majority of the time is spent on the simulation.

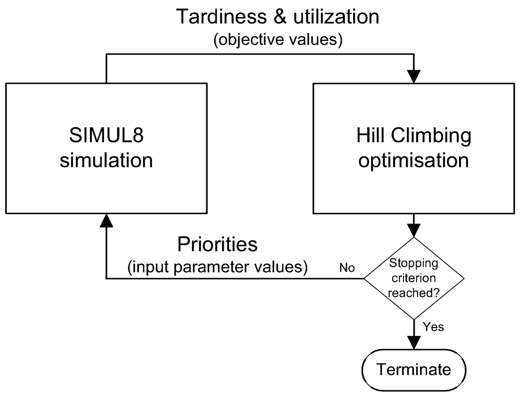


Figure 3: Simulation-Optimization Process.

The simulation-optimization process is enabled using a web-based user interface, which is presented in the next section.

WEB-BASED USER INTERFACE

The user interface is in form of a web page, implemented using HTML, JavaScript and Ajax. When navigating to the web page, the user must first login in. Using a login in-system allows for the configuration of different user groups and the specification of privileges based on position in the company.

When logged in to the system, the first thing shown to the user is an information page describing the web-system and giving instructions about how to perform simulation-optimizations (Figure 4). The intention is that the information page should show not only static information, but also dynamic information about the manufacturing cell related to the current shift. Such information includes shift period, shift leader, operators working in the cell, and known disturbances.

Figure 4: Information and Instructions.

Before starting a simulation-optimization process, the user must first specify the current status and configuration of the manufacturing cell (Figure 5). This data includes shift period, status of components currently under processing, list of entering components, availability of fixtures, availability of operators, and scheduled maintenance. All this data is necessary for the simulation model to be able to mimic the operation of the cell as close to reality as possible.

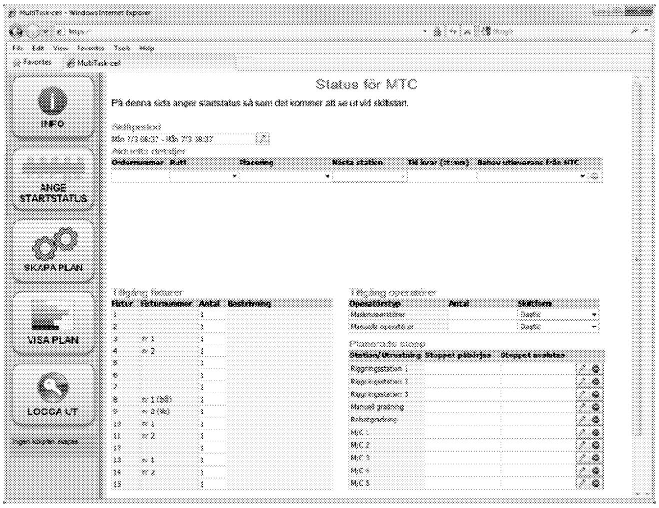


Figure 5: Status and Configuration of the Cell.

When the status and configuration of the cell has been specified, the user can initiate the simulation-optimization process. This is a simple procedure; the user just specifies at which point in time the results should be presented and then presses a start button (Figure 6). When the start button is pressed, a command is send to a web server to initiate a simulation-optimization process. Since the simulation-optimization is run at the server-side, and not on the client-side, the performance of the user's computer is not affected at all by the simulation/optimization. In the current implementation, a single server is used and there can therefore only be one single simulation-optimization process active at the same time. To assure this, the start button is inactivated until the process is completed.

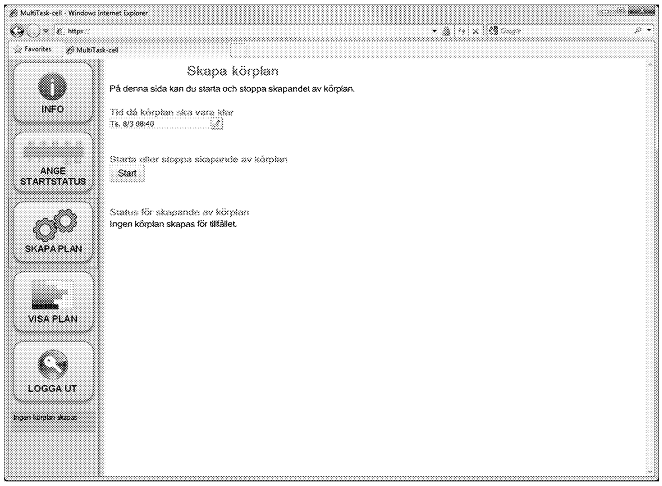


Figure 6: Initiating a Simulation-Optimization Run.

At the specified point in time, the outcome from the simulation-optimization process (that is, a prioritizing of components) is presented to the user. According to requirements specified by the company, the results are presented in form of a graphical production schedule (Figure 7). The graphical schedule specifies which component to process in which machine at each point in time. Along with the graphical schedule are details such as delayed components, periods of operator needs, and initial status presented. Altogether, this information supports the shift leader and the operators in managing the cell in an efficient way.

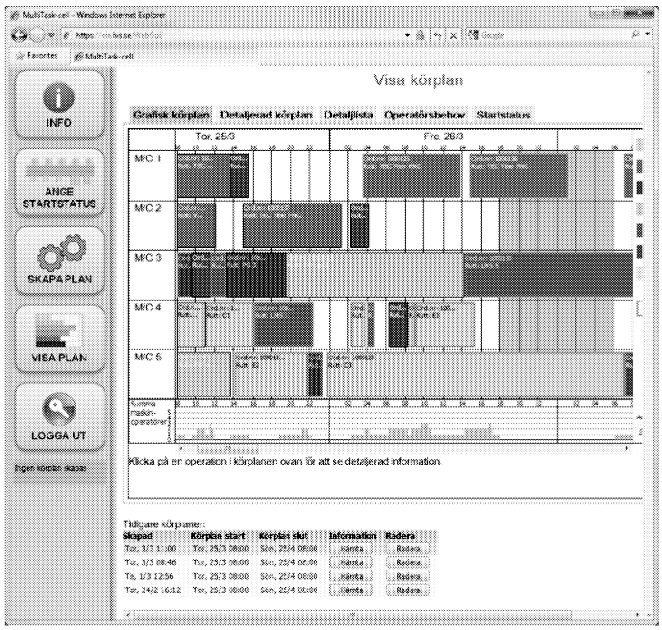


Figure 7: Results in Form of a Graphical Production Schedule.

The next section describes how the web-based simulation-optimization system is evaluated.

EVALUATION

Evaluation of the web-based simulation-optimization system has been undertaken based on two criteria; 1) optimization performance, and 2) company benefits.

With respect to optimization performance, production schedules created by the simulation-optimization are compared to schedules created manually by a domain expert. As previously described, a manual procedure is the current approach for creating schedules. Comparing the schedules with respect to tardiness and lead times (the two optimization objectives) shows a great advantage of the simulation-optimization procedure. The simulation-optimization creates schedules with at least 20% improvements, and not seldom as much as 50% improvements or even more. This is no surprise, considering that the manufacturing cell is highly complex and that manual optimization is virtually impossible. Although it is obvious that the simulation-optimization performs better than a human, it is worth to notice that it is not clear how well the optimized solutions compare with the achievable values of the objectives, as the ideal values are not known. The combinatorial relationships, uncertainty factors, and non-linearities in the manufacturing process means that the problem is too complex to be modeled and solved analytically, and the true optimum therefore remains unknown. It is also important to notice that optimization results have so far only been verified in the simulation model. In future, the optimized production schedules can hopefully be implemented and evaluated also in reality.

With respect to the second evaluation criterion, company benefits, this is evaluated through thorough discussions with company representatives. These representatives include the operative manager of the cell, operators working in the cell, the head of logistics, and logistics engineers. Before the discussions take place, the web-based simulation-optimization system is first presented and demonstrated to the representatives. After that, they are given the opportunity to try out the system themselves. In the subsequent discussions, the representatives were asked about their opinions about the system. All of them rose that they see a great potential in the system when it comes to improving production schedules of the future, especially at high workloads. Also, they see an advantage in that the system heavily reduces the human effort associated with creating production schedules. All representatives agreed that the system is worth trying in real production, but that an additional feature must be implemented before this can take place. This feature is a real-time integration with the manufacturing cell that reduces the need to specify the current status and configuration of the cell before an optimization is started. With such integration, the optimization can be fully automatic and run more frequently. Since the manufacturing cell is relatively new, is has modern, built-in computer systems that make it possible to achieve status information in real-time. A discussion is currently undergoing at the company about how use and integrate these systems with the web-based simulation-optimization system.

CONCLUSIONS AND FUTURE WORK

This paper presents a web-based simulation-optimization system for improving production schedules in an advanced manufacturing cell at Volvo Aero Corporation in Sweden. The optimization aims at prioritizing components being processed in the cell in a way that minimizes both tardiness and lead times. The production schedules being the outcome of the optimization is evaluated by the company and the results look very promising. In comparison with production schedules created manually by domain experts, the optimization is able to find solutions with at least 20% improvements according to validations using the simulation model.

To improve the optimization results even further, future work includes evaluating more sophisticated optimization algorithms. The current Hill Climbing algorithm works well, but is reasonable to expect that improved performance can be achieved with a more advanced algorithm. The algorithm currently under consideration for replacing the Hill Climbing algorithm is a new evolutionary algorithm called Cuckoo Search (Yang and Deb, 2009). Cuckoo Search is inspired by biological mechanisms and mimics the breeding behavior of some cuckoo species that lay their eggs in the nests of other birds. In the algorithm, each egg in a nest represents a solution, and a cuckoo egg represents a new solution. The aim is to use the new and potentially better solutions that are cuckoos, to replace worse solutions in the nests. Initial studies reports that Cuckoo Search is efficient for both engineering optimization problems and for scheduling (Yang and Deb, 2010; Tein and Ramli, 2010), and its original structure make it interesting to take a deeper look at the algorithm.

Besides the potential optimizations possible with the developed web-based system, there is also another important advantage pinpointed by the company. This is the possibility to share results in real-time among stakeholders without additional efforts. With locally installed programs, the results can only be viewed on the specific computer unless they are not printed or e-mailed. This overhead is completely eliminated with a web-based system which allows everyone that has access to the system to view the results simultaneously. This aspect was actually considered the most important advantage of a web-based system according to the company. From the universities perspective, we thought before starting the project that the main advantage of a web-based system would be the obvious benefit of having access to the system from anywhere with an internet connection. But it turned out that this was incorrect, and we learned that possibility to easily share results is an important selling point when creating new projects with other companies in the future.

Besides advantages, it should also be mentioned that the company also raised an important disadvantage with the web-based system, namely security vulnerability. In comparison to a locally installed program, a web-based system is considerable more vulnerable to malicious attacks. This is not only a risk in itself, but there is also a risk that the system is never adopted due to this aspect. Volvo Aero Corporation, similar to many other companies, has a strict IT policy and very high security regulations, meaning that the system might not be accepted if not "fully" secure. For succeeding in introducing web-based systems in industry we therefore emphasize that it is of critical importance to put lots of efforts on security issues.

An additional disadvantage with a web-based system, not mentioned by the company but by the software developer, is limitations in the graphical user interface. The possibilities to create advanced graphical features in the web browser are complicated compared to desktop applications. It might therefore not be possible to achieve a perfect graphical look in a web-based system. However, along with more and more applications being made web-based this will start to change and the problem will most probably be non-existing in the future.

REFERENCES

Byrne, A., C. Heavey, and P.J. Byrne. 2009. "A review of Web-based simulation and supporting tools". In *Simulation Modelling Practice and Theory*, Vol. 18, No. 3, pp. 253-276.

Fishwick, P.A. 1996. "Web-based simulation: some personal observations". In *Proceedings of the 28th Conference on Winter Simulation*, Coronado, California, USA, pp. 772–779.

Ng, A., Grimm, H., Lezama, T., Persson, A., Andersson, M. and Jägstam, M. 2007. Web services for metamodel-assisted parallel simulation optimization, Proceedings of The IAENG International Conference on Internet Computing and Web Services, Hong Kong, pp. 879–885.

Rabinovich, M. and O. Spatscheck, 2002. "Web Caching and Replication". Addison-Wesley, Pearson Education.

Russell, S.J. and P. Norvig. 2003. "Artificial Intelligence: A Modern Approach" (2nd ed.), Upper Saddle River, New Jersey: Prentice Hall, pp. 111–114.

Suh, W. 2005. "Web application development methodologies". In: *Web Engineering: Principles and Techniques*, Idea Group Publishing, London, pp. 76–93.

Syberfeldt, A. 2009. A multi-objective evolutionary approach to simulation-based optimization of real-world problems. PhD Thesis. De Montfort University, UK.

Tein L. H. and R. Ramli. 2010. "Recent advancements of nurse scheduling models and a potential path". In *Proceedings of the 6th IMT-GT Conference on Mathematics, Statistics and its Applications*, pp. 395-409.

Yang, X.-S. and S. Deb. 2009. "Cuckoo search via Lévy flights". In *World Congress on Nature & Biologically Inspired Computing* IEEE Publications, pp. 210-214.

Yang, X.-S. and S. Deb. 2010. "Engineering optimisation by cuckoo search". In *International Journal of Mathematical Modelling and Numerical Optimisation.* Vol. 1, No. 4, pp. 330-343.

AUTHOR BIOGRAPHY

ANNA SYBERFELDT is a senior researcher at the University of Skövde, Sweden. She holds a PhD in Computer Science from the De Montfort University, UK and a Master's degree in Computer Science from the University of Skövde, Sweden. Her research interests include soft computing techniques and simulation-based optimization. Her e-mail address is <anna.syberfeldt@his.se>.

INGEMAR KARLSSON is a senior system developer at the University of Skövde, Sweden. He received his BSc degree in Computer Science from the University of Skövde. His main research interest is simulation-based optimization. His e-mail address is <ingemar.karlsson@his.se>.

AMOS NG is an Associate Professor at the University of Skövde, Sweden. He holds a B.Eng. degree and a M.Phil. degree, both in Manufacturing Engineering from the City University of Hong Kong and a Ph.D. degree in Computing Sciences and Engineering from De Montfort University, Leicester, U.K. He is a member of the IEE and a Chartered Engineer in the U.K. His research interests include production simulation, simulation-based optimisation, multi-objective optimisation, decentralised (agent-based) machine control systems. His e-mail address is <amos.ng@his.se>.

AN APPLIED FRAMEWORK FOR SIMULATION-BASED MULTI-OBJECTIVE OPTIMISATION WITHIN PRODUCTION SYSTEM DEVELOPMENT

Leif Pehrsson[1,2]
Amos H.C. Ng[2]
[1]Volvo Car Corporation, SE-405 31, Gothenburg, Sweden
[2]The Virtual Systems Research Centre, The University of Skövde, SE-541 28, Skövde, Sweden
E-mails: leif.pehrsson@his.se, amos.ng@his.se

KEYWORDS
Simulation, Multi-objective Optimisation, Cost, Sustainability.

ABSTRACT

A method and a framework for the application of Simulation-based Multi-objective Optimisation (SMO) has been developed in order to enhance the prerequisites for decision-making within design and re-configuration of production systems. This kind of decisions often tends to be based on financial information rather than the type of production system parameters to be found in traditional simulation models. Therefore, to combine traditional parameters with new financial and sustainability parameters can be very beneficial for supporting industrial decision-making. The framework has been applied in a number of case studies involving a range of production system issues both within component production and assembly operations. Several types of issues have been explored involving analysis of system behaviour, optimisation of sustainability parameters, in the form of energy consumption aggregated to energy cost, and optimisation of financial parameters in combination with traditional production system metrics. The case studies have adequately proven and verified that the application of SMO, especially including financial functions and objectives, can be very valuable for practical industrial applications.

INTRODUCTION

Decisions concerning production system design and improvement are many times depending on trade-offs between conflicting objectives including traditional productions system parameters such as capacity and lead time in combination with financial parameters such as investment and running cost (Pehrsson 2009). Cost analysis and costing for decision support is central in the effort of improving performance, create value, analysing scenarios and in purpose of applying resources in an effective and efficient way within an enterprise (Professional Accountants in Business Committee, 2009). On the other hand, it is widely accepted that the only general purpose and generally applicable tool for truly complex systems is simulation, particularly discrete-event simulation (DES) (Fu et al. 2000). DES has been considered to be the most promising tool for supporting decision-making in production systems design and analysis. However the traditional simulation objectives within a production system might not be directly transferrable into cost evaluations and financial decision-making. There are examples of merging DES with methods for cost estimation connected to activity based costing (ABC) (Von Beck and Nowak 2000) and the use of DES as a base for cost reduction and performance improvement (Brown and Sturrock 2009). Beside cost , sustainability has been emerged as an important factor for companies worldwide. Just between 2004 and 2005, more than 50% of the companies in the G250 index published corporate sustainability reports, including measurements of their sustainability performance (KPMG International and Amsterdam Graduate Business School 2005). In 2008, the corresponding figure was 80% (KPMG International 2008). The International Federation of Accountants (IFA) has issued an international guidance document for Environmental Management Accounting (EMA) (International Federation of Accountants 2005). According to IFA, some major potential applications for EMA are within the areas of cost effective environmental regulation compliance, strategic positioning from long-turn competitiveness perspective and eco-efficiency initiatives simultaneously reducing cost and environmental impact.

Simulation-based optimisation has empowered DES to be a real optimisation tool by incorporating meta-heuistic search methods, such as genetic algorithms, to the simulation models in seeking optimal or near-optimal solutions. The SMO framework extends the simulation-based optimisation technology by using multi-objective evolutionary optimisation (Deb 2001) to find Pareto-optimal solutions for production systems analysis. An example of an Internet-enabled SBO toolset utilising MOO for integrated concept phase factory design, analysis and optimisation is FACTS Analyser (Ng et al. 2007). The addition of post optimality analysis based on the concept of “innovization” (Deb and Srinivasan 2006) will extend the SMO opportunities into extracting innovative design principles and knowledge through optimisation.

So far, SMO for production system analysis, design and improvement has mainly focused on traditional production system parameters (Ng et al. 2007; Ng et al. 2009). A novel method for optimisation of running cost and investment (Pehrsson et al. 2011) has been developed in order to introduce the possibility to incorporate financial objectives within a framework for production system optimisation and decision support. The integration of sustainability optimisation into the framework is underway and the method and three proof-of-concept case studies are briefly introduced in this paper.

The main purpose of this paper is to explore some of the opportunities for industrial applications opened by such a framework with novel features for analysis of production systems and the creation of decision support for manufacturing management.

SMO FRAMEWORK

A general framework and a method for SMO with cost and sustainability optimisation opportunities have been designed in order to facilitate industrial application within production system development and improvement. The foundation is a simulation model of the production system connected to SMO with meta-heuristic search methods incorporating genetic algorithms and multi-objective optimisation. The addition of functions for running cost estimation and investment integration together with modelling of sustainability parameters such as energy consumption, material consumption and waste opens opportunities to optimise a production system from a multi-disciplinary perspective. In combination with post-optimality analysis, these parameters meet industrial requirements on providing information for financial decision-making, favouring sustainable solutions within design and improvement of production systems. The main principles of the framework can be seen in Figure 1.

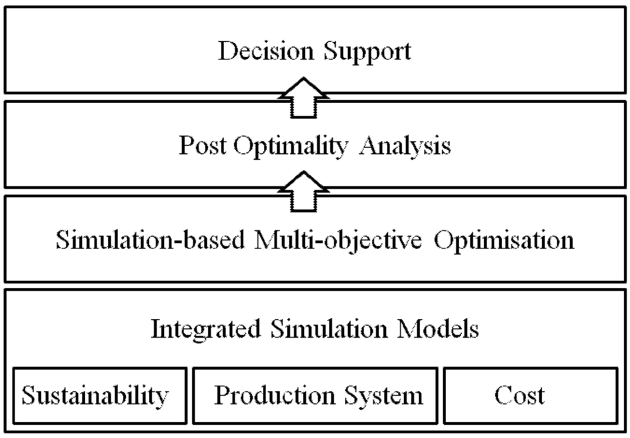


Figure 1: SMO Framework Principle

An essential aspect in providing decision support is presentation of the decision variables and their effect on the objectives. The combination of various analysis tools and presentation techniques can be applied to reveal useful information to the decision maker. Among a number of techniques, some examples are when three-dimensional plotting and colour-coding are used together to present four dimensions, referred to as a Trade-off Plot. Other examples include Parallel Coordinate Plot, Hyper-Radial Visualisation (Liebscher et. al. 2009) and when clustering of data based on data mining is used to highlight areas of interest. An example of three-dimensional plotting of optimised data is found in Figure 2.

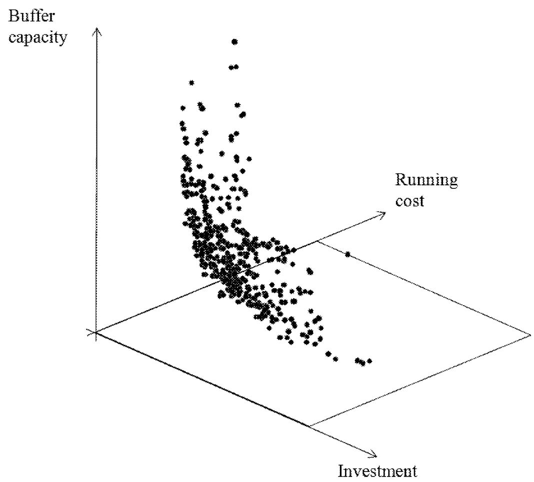


Figure 2: Three Dimensional Plot of Optimised Data

FUNCTIONS FOR ESTIMATION AND SIMULATION OF COST AND SUSTAINABILITY PARAMETERS

The integration of an additional set of parameters into the SMO framework requires a number of new parameter functions and objective functions.

Running Cost Function

One of the essential components of the framework is a model for estimation of running cost within a production system. There are some very detailed production system cost models available such as cost deployment (Yamashina and Kubo 2002) and a general model for manufacturing cost simulation (Jönnson et al. 2007) tending to require in-depth data mapping in the production system.

In this case a fast, accurate method with less data requirements was a necessary prerequisite enabling to meet the decision-making time schedule. One way of overcoming pitfalls due to data complexity is to use a time-based method such as the resource effectiveness model (CMA Canada 1999) as an information source that easily can be translated into the financial domain through a process costing model defined on process time rather than units of production. Time based models also correspond to loss models used in industry (Nord and Pettersson 1997; Pehrsson 2009). A function for transferring the effect of production system improvements into the cost domain is composed, derived from the resource effectiveness model and another similar approach for ABC, referred to as time-driven ABC (Kaplan and Andersson 2007).

In its most basic form the cost model can be written as equation (1):

$$Cr = Ci + \Delta C \tag{1}$$

where Cr = Running cost per year (annual running cost), Ci = Initial total running cost per year, ΔC= Delta cost.

The delta cost is consisting of several components transferring throughput cost consequences, annual cost and cost per unit into one function for simulation model integration. The first part, the delta throughput to cost function, can be written in a few steps summarised in (2) and (3).

$$\Delta Ct = \Delta H \times Ch \quad (2)$$

where, ΔCt = Throughput delta cost, ΔH = Difference in need of production time, Ch = Average cost per hour (additional time) corresponding to “capacity cost rate”

$$\Delta H = Vp\left(\frac{1}{T} - \frac{1}{Ti}\right) \quad (3)$$

where, Vp = Annual production volume, Ti = Initial Throughput, T = Throughput.

A certain delta throughput cost will be valid within a specific interval and several instances of the function might be needed to create a complete model due to constraints in the production setup such as balancing of human resources and different costs on various shifts, shown in (4).

$$\Delta Ct_i = Ch_i \times Vp\left(\frac{1}{T_i} - \frac{1}{Ti_i}\right) \quad (4)$$

valid for $a_i < T < b_i$ and $a_i < T_i < b_i$

The complete throughput delta cost is then as in (5).

$$\Delta Ct = \sum_{i=1}^{m} \Delta Ct_i \quad (5)$$

When combined with the components for annual delta costs and influences on cost per unit, the complete expression including a custom cost component for tailored applications can be written according to (6):

$$Cr = Ci + \sum_{i=1}^{m} \Delta Ct_i + \sum_{j=1}^{n} \Delta Ca_j + \sum_{k=1}^{o} \Delta Cu_k \times Vp + \Delta Cc \quad (6)$$

where ΔCa = delta annual cost, ΔCu = delta cost per produced unit, ΔCc = user definable custom cost component.

Investment Function

The effect of combinations of minor improvements and investments might be difficult to analyse and optimise in combination with a number of dynamic production system parameters. The running cost function integrated in the framework is not sufficient to analyse the complete financial impact of certain improvement combinations without the introduction of investment parameters.
The investment expression in the framework is composed of the three main components processing-time related, up-time related and buffer capacity related investments. With the addition of a custom component for tailored applications the complete investment expression can be written as:

$$I = \sum_{i=1}^{m} Ip_i + \sum_{j=1}^{n} Iu_i + \sum_{k=1}^{o} Ib_k + Ic \quad (7)$$

where, I = total Investment, Ip = processing time related investment, Iu = up-time related investment, Ib = buffer-related investment and Ic= user definable custom component.

Optimal Buffer Allocation Function

During the application of the framework there is often a requirement to optimise buffer allocation and the function used for such optimisations can be expressed as (8):

$$B = \sum_{i=1}^{n} Bc_i \quad (8)$$

where, B = total capacity of all inter-workstation buffers included in the optimisation, Bc = capacity of buffer number i.
This function can be used not only to directly optimize buffer capacity, it may also trigger investments, annual cost and cost per unit.

Sustainability and Energy Consumption Modelling

An initial concept for sustainability modelling has been developed, aligned with EMA and derived from the promising modelling of financial and management accounting aspects. The sustainability modelling will follow the main pattern of the cost modelling in the framework with an initial sustainability performance and delta sustainability reflecting the effect of various options and scenarios to be studied.
In its most basic form, the sustainability model principle is described by (9):

$$Sa = Si + \Delta S \quad (9)$$

Where: Sa = Annual sustainability performance, Si = Initial sustainability performance, ΔS = Delta sustainability performance.

Energy consumption modelled according to the sustainability modelling principle can be expressed as (10):

$$Ea = Ei + \Delta E \quad (10)$$

Where: Ea = Annual energy consumption, Ei = Initial energy consumption, ΔE = Delta energy consumption.

As a starting point for modelling of electrical energy consumption, a first series of measurements has been conducted on equipment within a facility for production of automotive components. Based on the measurements there is an indication that a total of three levels can be used to describe the average consumption patterns over time, of which two is during operation. The three levels are:

- High, based on the average consumption during machining.
- Low, based on the average consumption during idle times between machining cycles.
- Shut-down, based on the consumption when the equipment is turned off or in stand-by mode.

Based on the three-level theory the delta energy consumption expression (11) can be expanded to include these parameters:

$$\Delta E = \Delta E_h + \Delta E_l + \Delta E_s \qquad (11)$$

Where, ΔE_h = Delta energy consumption, high level, ΔE_l = Delta energy consumption, low level or idle and ΔE_s = Delta energy consumption, shut-down or stand-by.

Each consumption object in a simulation model is connected to the complete energy expression using the sum (12):

$$\Delta E = \sum_{i=1}^{m} \Delta E_i \qquad (12)$$

Transferring a sustainability parameter to cost is done using a cost factor and one example is the expression for energy cost (13):

$$Ce = Ea \times Cfe \qquad (13)$$

where, Ce = energy cost, Cfe = energy cost factor.

CASE STUDIES

A number of case studies has been conducted within industry in order to explore various applications and scenarios for the use of the SMO framework within production system design and performance improvement. These include the defining system behaviour, optimisation with cost objectives and optimisation with a combination of cost and sustainability objectives. The presented case studies were conducted using the NSGA-II, genetic algorithm with meta-modelling enhancement. The findings are intended to be utilised in the continued design of a general framework for SMO with integrated decision support capabilities.

Case Study A – Defining System Behaviour with Interacting Pallet Loops

The main intention with this case study is to explore the opportunities to use SMO for production system behaviour analysis. An assembly line is designed with a pallet system loop for transporting the main assembly object. Previously the line has been subject for a case study concerning optimisation of the number of pallets in the main assembly loop combined with innovisation (Ng et al 2009). However, in this extended case study, a secondary pallet system, used for replenishing the line with components to be assembled, is observed in combination with the main system with the aim of explaining the overall system behaviour. The components handled by the secondary pallet loop require classing against another assembly object due to tight tolerances. When the incoming components are loaded onto the secondary pallet system information about component classing is sent upstream in the main process. Material at the receiving upstream position is not allowed to continue in the flow unless there is classing information required to assemble the other classed component. This creates a limitation from a work in process (WIP) perspective in the main flow modelled as a max-WIP loop, dependent on the number of identified components buffered on the secondary pallet system. A simulation model of the line was created and after validation the optimisation objectives were integrated into the model using the SMO framework. The optimisation objectives expected to reveal the system behaviour were (14):

$$\begin{aligned} &nm \text{ (min)} \\ &ns \text{ (min)} \\ &Tp \text{ (max)} \end{aligned} \qquad (14)$$

where, nm = the number of pallets in the main loop, ns = max-WIP due to identified components on the secondary pallet loop and Tp = throughput.

By plotting the three objectives after optimisation, as seen in Figure 3 and Figure 4, some insights about the system behaviour can be revealed.

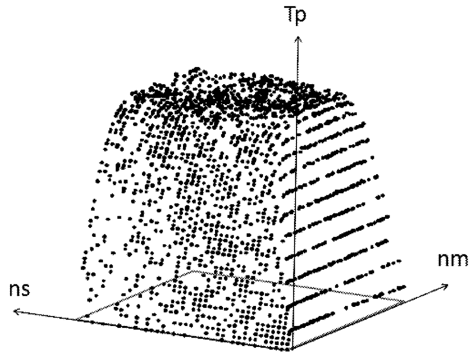


Figure 3: Three dimensional plot of the main decision variables and their effect on throughput.

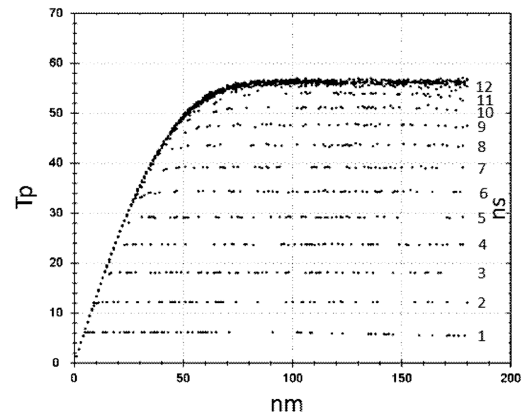


Figure 4: System behaviour plot showing the dependencies between the number of pallets in the main Loop, the Max-WIP loop and throughput

At least 12 identified components with information sent upstream are required to reach maximum system performance from a throughput perspective. The flat top of the curve indicated that there was some other constraint limiting the throughput of the line. A shifting bottleneck analysis (Roser et al.2002) showed some stations significantly restrain the capacity of the line. With targeted improvements in these specific stations the maximum throughput could be increased by 14%. The drawback was that also the requirements on the number of identified components on the secondary pallet loop increased from 12 to at least 17 and with some further improvements to over 20 which was very close to the maximum practical buffer of 24. Keeping the buffer operating at that level would require an extra dedicated operator on the line and it was decided to remove the max-WIP constraint by moving the assembly of the two classed components to a station further downstream in the assembly line. After implementation the relative effect on throughput was observed and it is conforming very well to the figures from the optimisation. The conclusion is that SMO can be a powerful method to extract knowledge about production system behaviour, serving as an important source for the creation of decision support within industry.

Case Study B – Sustainability and Energy Cost in Relation to Maintenance

The main purpose of this case study is to make a first proof of concept for the sustainability components in the SMO framework.

A plant producing components for the automotive industry is planning to expand its production by introducing a completely new line for a new range of products. Due to strategy of the company, sustainability objectives and due to increasing electrical energy cost there is a demand for evaluation of the potential effect from increased maintenance in order to reduce energy consumption. Could SMO with integration of sustainability parameters be of assistance when conducting such evaluations?

A conceptual phase simulation model of the line was built with input and data from existing production facilities and equipment suppliers. Electrical energy consumption patterns for relevant machine types were logged in an existing line and divided into the three categories running, idle and stand by. The energy consumption figures were updated to reflect the behaviour of the new equipment and coolant supply pump stations.

The required time for operation of the line is dependent on the actual capacity or the throughput from the line. An estimation of the required maintenance cost to achieve a certain level of availability was made by the expressions (15):

$$\begin{aligned} &Cm_i = (AvbOP_i - 70)^3 \\ &85 \le AvbOP_i \le 99 \qquad (15) \\ &Cm = \sum\nolimits_{i=1}^{m} Cm_i \end{aligned}$$

where: Cm_i = Delta maintenance cost for equipment i, $AvbOP_i$ = Availability of equipment i and Cm = Delta maintenance cost for the complete system.

The optimisation objectives were (16):

$$\begin{aligned} &Cm \text{ (min)} \\ &Ce \text{ (min)} \qquad (16) \end{aligned}$$

The result from the optimisation plotted in the form of a Pareto front can be seen in Figure 5.

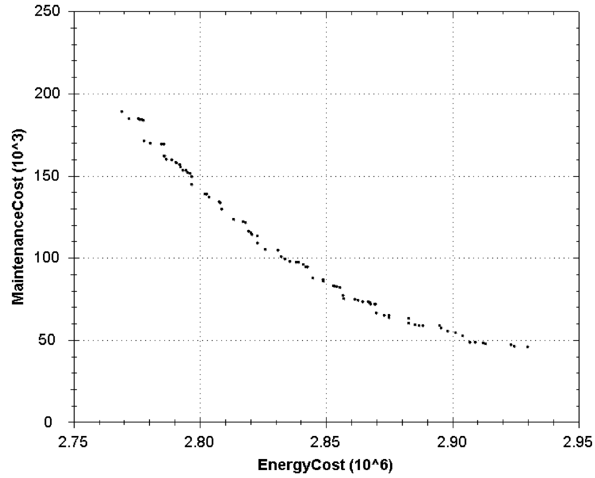


Figure 5: Pareto Front Showing Maintenance Cost vs. Energy Cost

One conclusion is that in this case it seems like energy can be saved to an extent without adding cost to the system. However there is a level above which further savings would add cost.

The main conclusion from this study is that the sustainability modelling has potential for integration in the SMO framework, enhancing the decision support potential of the method.

Case Study C – Cost Optimisation

The cost optimisation components within the framework were tried out through a case study on a production line for components to the automotive industry. An increasing customer demand required the line to be operated on overtime. Some major product changes were planned to be introduced at the same time with a potentially negative effect on capacity. A number of improvements were identified including various investments with potential to reduce the capacity constraints possibly enough to avoid operating the line on overtime. However, there was not enough data and knowledge available to make the decision to invest and reduce the operating time, especially when considering the unknown opportunities due to combinations of several minor improvements.

A simulation model of the original production line was created and validated against real throughput. The model was updated to reflect the planned scenario after introduction of product changes and in order to prepare for optimisation the components for delta throughput cost and investments were integrated in the model using the SMO framework together with a number of identified improvement proposals with attached investments. The model was first run to reflect the planned scenario and the result was a throughput of 28.7pcs/hr, which could not reach the minimum requirement of 30pcs/hr for meeting the customer demand on overtime. To completely reduce the need for overtime, an average throughput of 34.5 would be the minimum requirement. The main target for the line is to reduce the running cost by 20% against the forecast. Again, by applying the shifting bottleneck technique, the major constraints in the line could be identified, supporting the selection of relevant improvement proposals to include in the optimisation. In order to reveal more information about the opportunities to improve the line performance, an optimisation was conducted with the following objectives (17) as defined earlier:

$$\begin{aligned} &Cr\,(\text{min}) \\ &I\,(\text{min}) \\ &B\,(\text{min}) \end{aligned} \tag{17}$$

The optimisation results, considering the conflicting objectives investment against running cost is shown in Figure 6 in the form of a Pareto front. Throughput against buffer capacity is shown in Figure 7.

One conclusion is that by an ideal distribution of buffers the running cost could be reduced from the forecasted M\$4.9 down to M\$4.4 without any additional investments or improvements. Another conclusion is that the stretched throughput target of 34.5pcs/h cannot be reached without re-configured the buffer capacity, considering the improvement opportunities identified so far.

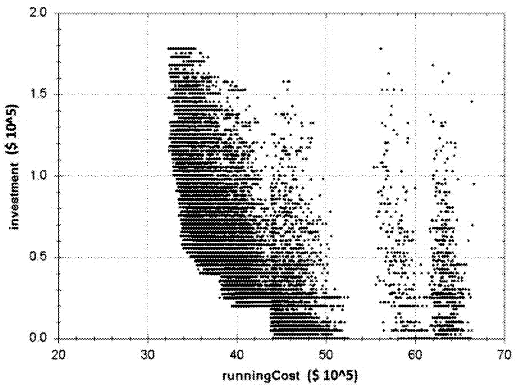


Figure 6: Investment vs. Running Cost

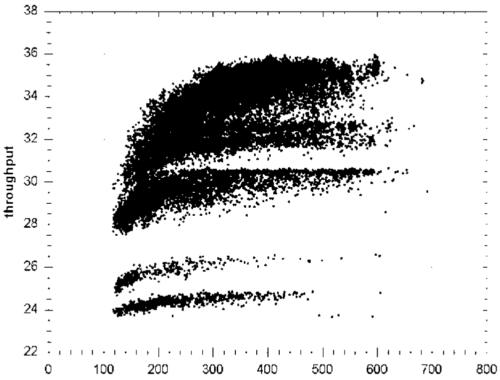


Figure 7: Throughput vs. Buffer Capacity

By considering several minor improvements in combination with buffer capacity re-configuration, a trade-off with \$50 000 investment could be found satisfying the throughput and cost objectives. This would however require extensive re-configuration of buffers not possible in reality. In order to find a feasible solution the search in the data from the optimisation was narrowed down to solutions exceeding 34.5 in throughput with less than 300 in allocated buffer capacity. A search for solutions with as few changes in buffer capacity compared to the existing line was conducted. The best solution of these was then slightly modified by excluding some minor re-configurations in buffer capacity. The result was a solution with \$M0.15 investment reducing the annual running cost by \$M1.29 or 26.3% and increasing throughput 20.9%. Average WIP could be reduced by 12%. despite an increase in the total allocated buffer capacity.

In order to further investigate the opportunities to reveal more information and extract knowledge from the optimisation results post optimality analysis using data mining was applied on the optimised data set. By doing this it was possible to identify the relative importance and relative impact of applying various design options. One major benefit from such an analysis could be to provide an opportunity to select improvement actions in the order of importance in terms of effect on the optimisation objectives. One example of the data mining results can be found in Figure 8.

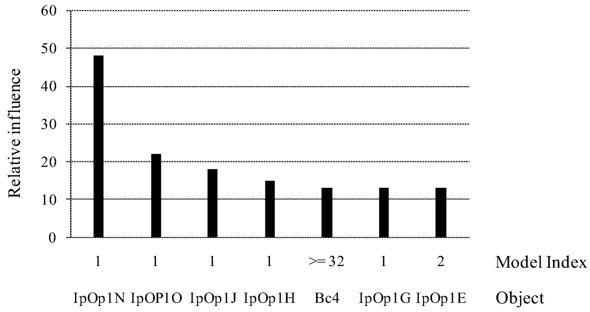


Figure 8: Relative Running Cost Influence

The proposed solution is now implemented and the real improvement performance is conforming very well to the predicted value in investment with the largest +7% deviation above the estimated figure. Part of the validation against investment and throughput can be seen in Figure 9.

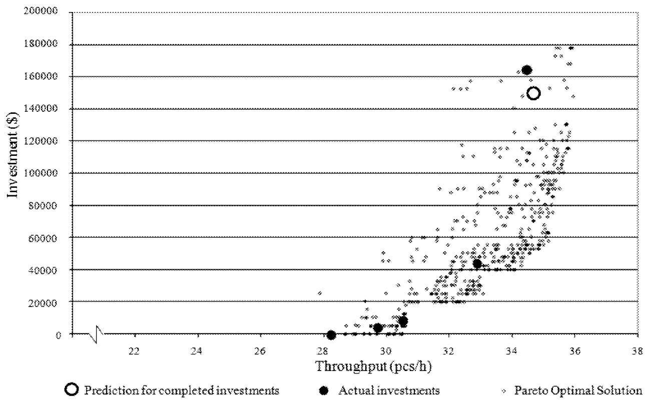


Figure 9: Validation against investment and throughput.

The most important conclusion from this case study is that the capability of the SMO method as a decision support instrument within the industry can be enhanced by incorporating financial parameters and objectives. Performance predicted by the SMO method has proven to be very accurate through the validation against the actual results from the implementation of the recommendations in the case study.

CONCLUSIONS

The potential of applying the SMO framework within various forms of analysis, problem solving and design issues connected to production systems has been explored. The SMO framework opens a wide variety of opportunities for production system analysis, design and improvement. Due to the versatility of the framework, applications within component production as well as within assembly operations have been successful.

Application within the field of analysing the behaviour of a production system in order to create knowledge for decision-making has been proven very beneficial, enabling major steps in throughput performance.

An initial concept for optimisation of sustainability parameters has been applied on optimisation of energy consumption, aggregated to energy cost, with promising results and potential for integration in the SMO framework. Applying SMO, taking into account financial objectives, like investment and running cost, for decision-making support in designing/re-configuring production systems, has been explored. Evaluating several combined minor improvements with the help of multi-objective optimisation has opened the opportunity to identify a set of solutions revealing great financial improvement, which cannot be sought by applying any current industrial procedures.

The most important results from the application of the SMO framework through case studies can be summarised as below:

- Performance predicted by the SMO method has proven to be very accurate through the validation against the actual results after implementation of the recommendations in the case studies.
- Very limited investment was required to reduce the running cost by 26.3% and to increase the throughput by 20.9% by utilising knowledge created by SMO.
- Data mining can be a very useful technique for enhancing the decision support when applied on data from SMO.
- The capability of the SMO method as a decision support instrument within industry can be enhanced by incorporating financial parameters and objectives.
- Sustainability modelling has potential for integration in the SMO framework, enhancing the decision support potential of the method.

The case studies have adequately proven and verified that the application of SMO, especially including financial functions and objectives, can be very valuable for practical industrial applications.

FURTHER RESEARCH

There is a great opportunity to continue the development of a general framework for production system analysis and design with integrated decision support based on SMO. Further research will include the continued development and validation of cost and sustainability functions. Major effort will be put into the exploration of decision support mechanisms in order to enhance the prerequisites to understand and utilise data from production system optimisations in strategically important decision-making within industry.

ACKNOWLEDGEMENTS

This work is partially financed by VINNOVA, Sweden, through the FFI-HSO project. The authors gratefully acknowledge their provision of research funding, and the industrial partner companies, including Volvo Technology Corporation, Volvo Car Corporation and Scania AB in Sweden, for their support in this study.

REFERENCES

Brown, E, J. and Sturrock, D. 2009. “Identifying cost reduction and performance improvement opportunities through simulation.” In *Proceedings of the 2009 Winter Simulation Conference*, WSC 2009, 13-16 December, Austin, Texas, USA.

CMA Canada. 1999. “Strategic Management Series, *Strategic Cost Management, Measuring the Cost of Capacity”*, The Society of Management Accountants of Canada, Mississauga Executive Centre, One Robert Speck Parkway, Suite 1400, Mississauga, ON Canada L4Z 3M3.

Deb, K.. 2001. “*Multi-objective optimization using evolutionary algorithms”*, 3’rd 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.

Fu, M.C., Andradóttir, S., Carson, J.S., Glover, F., Harell, C.R., Ho, Y.-C., Kelly, J.P. and Robinson, S.M. 2000. “Integrating Optimisation and Simulation: Research and Practice”, In *Proceedings of the 2000 Winter Simulation Conference*, 9-12 December, Arlington, VA, IEEE, pp. 610-616.

International Federation of Accountants. 2005- “*International Guidance Document, Environmental Management Accounting”*, International Federation of Accountants, 545 Fifth Avenue, 14th Floor, New York, New York 10017, USA.

Jönsson, M., Andersson, C., Ståhl, J-E. 2007. “*A general economic model for manufacturing cost simulation”*, presented at the 41st CIRP Conference on Manufacturing Systems, Tokyo 2008.

Kaplan, R. S., Andersson, S. R. 2007. “*Time-driven activity-based costing – A simpler and more powerful path to higher profits”*, Harvard Business School Publishing Corporation, ISBN-13:978-1-4221-0171-1.

KPMG International and Amsterdam Graduate Business School. 2005. “*KPMG International Survey of Corporate Responsibility Reporting”*, KPMG International and University of Amsterdam, 250/0605/3.

KPMG International. 2008. "*KPMG International Survey of Corporate Responsibility Reporting 2008"*, KPMG Sustainability Services Partner, KPMG in the Netherlands, Burgemeester Rijnderslaan 20 1185 MC Amstelveen The Netherlands, RRD-105984.

Liebsher, L., Witowski, K., Goel, T. 2009. "*Decision Making in Multi.Objective Optimization for Industrial Applications – Data Mining and Visualisation of Pareto Data*". Presented at the 8th World Congress on Structural and Multidisciplinary Optimization, Lisabon, Portugal, June 1-5, 2009.

Ng. A, Deb, K., Dudas C. 2009. "Simulation-based Innovization for production systems and improvement: An industrial case study." In *Proceedings of The International 3'rd Swedish Production Symposium (SPS'09)*, Göteborg, Sweden, 2-3 December 2009.

Ng, A., Urenda, M., Svensson J., Skoogh, A. And Johansson, B. 2007. "FACTS Analyser: An innovative tool for factory conceptual design using simulation", In *Proceedings of Swedish Production Symposium*, Gothenburg 28-30 August 2007.

Nord, C. and Pettersson, B. 1997. "*Total Productive Maintenance med Erfarenhet från Volvo"*, IVF Industriforskning och utveckling AB, 1997.

Pehrsson, L. 2009. "*Simulation-based optimisation in relation to current production system data and evaluation models"*, Master Thesis in Manufacturing Engineering, University of Skövde, School of Technology and Society, Skövde 2009.

Pehrsson, L., Ng, A.H.C. and Bernedixen, J. 2011. "Multi-objective production system optimisation with investment and running costs." In *Evolutionary Multi-objective Optimization in Product Design and Manufacturing*, L. Wang, A. Ng, K. Deb (eds), Springer (In press).

Professional Accountants in Business Committee. 2009. "*International Good Practice Guidance, Evaluating and Improving Costing in Organisations"*, Professional Accountants in Business Committee, International Federation of Accountants, 545 Fifth Avenue, 14th Floor, New York, New York 10017 USA.

Roser, C., Nakano, M. And Tanaka, M. 2002. "Shifting bottleneck detection", In *Proceedings of the 2002 Winter Simulation Conference*, San Diego, CA, USA. Pp. 1079-1086.

Von Beck, U., Nowak, J, W. 2000. "The merger of discrete event simulation with activity-based costing for cost estimation in manufacturing environments." In *Proceedings of the 2000 Winter Simulation Conference*, WSC 2000, 10-13 December 2000, Wyndham Palace Resort & Spa, Orlando, FL, USA. ACM, 2000.

Yamashina, H. and Kubo, T. 2002. "Manufacturing Cost Deployment", *International Journal of Production Research*, 40(16), pp. 4077-4091.

AUTHOR BIOGRAPHY

LEIF PEHRSSON has a B.Sc. in Mechanical Engineering and a M.Sc. in Automation Engineering from the University of Skövde. He has been working for more than 15 years in various positions within automotive industry. In year 2000, he received the Henry Ford Technology Award for his work on production system development. During the period he held various management positions within manufacturing engineering and he was involved in the development of Volvo Cars Manufacturing System. In 2008, when he was the Senior Manager for production engineering at Volvo Cars Engine, he decided to combine his professional career with Ph.D. studies at the University of Skövde. He is also the chairman of the production engineering network in Skaraborg, West Sweden. In combination with his post-graduate route he is currently holding a position as a Senior Advisor at Volvo Cars Engine.

AMOS H.C. NG is an Associate Professor at the University of Skövde, Sweden. He holds a B.Eng. degree and a M.Phil. degree, both in Manufacturing Engineering from the City University of Hong Kong and a Ph.D. degree in Computing Sciences and Engineering from De Montfort University, Leicester, U.K. He is a member of the IET and a Chartered Engineer in the U.K. His research interests include modelling, simulation and multi-objective optimisation for production systems design and analysis.

OPTIMAL BUFFER ALLOCATION FOR SEMI-SYNCHRONIZED AUTOMOTIVE ASSEMBLY LINES USING SIMULATION-BASED MULTI-OBJECTIVE OPTIMIZATION

Jacob Bernedixen
Amos H.C. Ng

Virtual Systems Research Centre, The University of Skövde, SE-541 28, Skövde, Sweden
E-mails: jacob.bernedixen@his.se, amos.ng@his.se

KEYWORDS
Synchronized Assembly; Optimal Buffer Allocation; Simulation-based Optimization; Multi-objective Evolutionary Algorithms.

ABSTRACT

A practical question in industry in designing or re-designing a production system is: how small can intermediate buffers be to ensure the desired production rate? This topic is usually called optimal buffer allocation as the goal is to allocate the minimum buffer capacities to optimize the performance of the line. This paper presents a case study of using simulation-based evolutionary multi-objective optimization to determine the optimal buffer capacities and positions in the re-configuration of a real-world truck axle assembly line in an automobile manufacturer. The case study has not only revealed the applicability of the methodology in seeking optimal configurations in a truly multi-objective context, it also illustrates how additional important knowledge was gained by analyzing the optimization results in the objective space.

INTRODUCTION

The impact of limited buffer spaces on the performance of production lines or other types of systems, so called optimal buffer allocation (OBA) problems, are studied extensively in the literature (Conway et al. 1988)(Buzacott and Shanthikumar 1993). Generally stated, OBA problems can be classified into either primal or dual (Gershwin 1987). In a primal problem, the objective is to minimize the total buffer space subject to a production rate (throughput) constraint, e.g. (Yamashita and Altiok 1998). Apart from minimizing total buffer capacity, many researchers have approached the primal OBA problem of minimizing WIP levels subject to a throughput constraint (e.g. So 1997). In a dual problem, maximization of the throughput (and/or minimization of WIP) which subject to a total buffer space constraint is desired (Altiok and Stidham 1983)(Tempelmeier 2003).

A common approach in multi-objective studies of optimizing throughput and WIP levels simultaneously is to use a simple cost model as the objective function, in the form of:

$$Maximize\ \left(r \cdot TH - c \cdot WIP\right) \qquad \ldots (1)$$

This approach can be found earlier in Altiok and Stidham (1983) and more recently in Koh and Bulfin (2004). Similarly, instead of using c as the cost per unit WIP inventory, Hillier (2000) uses c as the cost per unit of buffer space so that the objective function comes with the form:

$$Maximize\ \left(r \cdot TH - c \cdot \sum_{i=1}^{n-1} B_i\right) \qquad \ldots (2)$$

Where B_i is the buffer capacity at workstation i, or ws_i, which limits the maximum number of units that can be stored before station i. In other words, this simple cost function is also suitable for studying multi-objective primal problems as B_i is equivalent to the buffer capacity allocated in front of ws_i.

Equation (1) and (2) are called composite objective functions as it combines several individual performance measures into one single objective. In other words, it effectively converts a multi-objective optimization (MOO) problem into a single-objective optimization so that a single optimal solution can be sought using many existing optimization techniques. Nevertheless, the major drawback of applying this kind of cost functions is that the optimal solution obtained by using this procedure is very sensitive to the relative preference vector (r and c in the above equations). As schematically illustrated in Figure 1, with the assumption that $r_1 > r_2 > r_3$ and $c_1 < c_2 < c_3$, it is likely that the preference vector (r_1, c_1) would lead to the finding of only solution A_1, (r_2, c_2) for finding only A_2 and (r_3, c_3) lead to the finding of A_3. At the same time, it can be seen in this example that to find the solution A_3 is more desirable because with certain increase of total buffers invested (B1-B3), the TH gain between A_1 to A_3 is dramatic when compared with the very small increase of TH between A_3 and A_4.

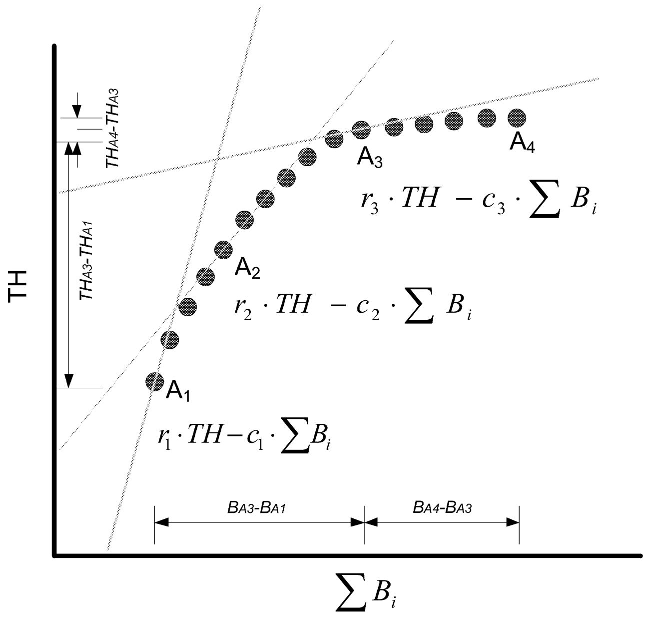


Figure 1: The effect of using different preference vectors.

On one hand, the choice of the preference weights and thus the obtained trade-off solution is highly subjective to the particular decision maker. On the other hand, without the detailed knowledge about the performance of the system in the objective space, it is a very difficult task to select the appropriate preference vector. Therefore, for a production system designer/manager, it would be very useful if all the optimal trade-off solutions between the conflicting objectives, in form of the posterior Pareto frontier (all of the points indicated in Figure 1), can be generated using an efficient MOO algorithm (Deb 2001). With such an approach, the manager would gain more knowledge of the performance of the optimal settings of the system under study before the most suitable configuration among the "best" trade-off solutions is selected.

The aim of this paper is therefore to propose a novel method of using simulation-based evolutionary multi-objective optimization to determine the optimal buffer capacity in production systems design and analysis. With the Pareto-based optimal solutions generated using an efficient MOO algorithm, namely NSGA-II (Deb et al. 2002), to assist the determination of the "optimal" buffer level to a desired throughput, the method is particularly suitable for assembly line configuration and re-configuration. The method is illustrated with a real-world case study performed in a Swedish automobile manufacturer. Since the target line in the case study is in determining the optimal buffer capacity and position in a synchronized assembly line, the concept of modelling this kind of production line is described in the next section. Details of the case study including the simulation model, optimization results and analysis will be covered in the Section "An Industrial Case Study". Finally, we conclude this paper by summarizing the findings.

INDUSTRIAL AUTOMOTIVE ASSEMBLY LINES

Serial Assembly Lines

Assembly lines in industry, particularly automotive manufacturers producing cars or trucks, are usually consisted of multiple workstations arranged in tandem, usually called *serial flow lines* or simply *flow lines*. All of these workstations can be automated, semi-automated, manual or a combination of these modes. The parts visit each workstation along the serial line commonly conveyed by some kind of material handling system like automated guided vehicles (AGV). In a mass production of standardised products, each workstation performs only one assembly *task* or a fixed set of tasks on all workpieces. In other words, the total time that each workpiece spends in a workstation is the cycle time (CT_i) of the ws_i, which is equal to the summation of the task time (T) of the k tasks assigned to the workstation and the transportation time, $t_{i,i+1}$, from ws_i to ws_{i+1}:

$$CT_i = \sum_{j=1}^{k} T_j + t_{i,i+1} \qquad \text{… (3)}$$

Provided that the transportation times (sometimes called transfer time) between any workstations are the same. To assign evenly distributed workload, in terms of equal cycle time, to each workstation (i.e. $CT=CT_1=CT_2=\ldots=CT_n$, for a flow line with n workstations), is the ultimate goal of the research in assembly line balancing (ALB)(Boysen et al. 2008). Therefore, in a perfectly balanced assembly flow line, the production rate (or throughput, TH) of the line is simply the reciprocal of CT, i.e. $TH=1/CT$. In the era of globalisation, while highly customised products become the business norm. Where highly customised products in a production context can be interpreted as; workpieces that will require different tasks in each workstation. Thus the basic goal of ALB to seek perfectly balanced workload (tasks) to all the workstations in the assembly line remains unchanged. But instead of as an one-off assembly line design problem, it demands the continuous re-configuration (re-balancing) of the assembly line to cope with the frequent changes in products and demands.

The ALB problems have been studied extensively in the literature. In general, ALB problems belong to the class of NP-hard combinatorial optimization problems. To find feasible solutions that fulfil the precedence constriants of the assembly tasks is a very challenging optimization problem. The computation time increases exponentially with the increased number of tasks and number of workstations. In the last two decades, it has emerged many ALB optimization algorithms. These (mostly based on branch and bound procedures, heuristics or genetic algorithms) have produced significant contributions in the field of ALB. However, since most of the algorithms developed tend to treat all task times, T_j, as deterministic, their applicability to solve real-world assembly design problems, in which the execution times of the tasks are subjected to certain variability, is therefore highly questioned. In a fully automated assembly line, variability is usually attributed to the disturbances of the workstation due to random breakdowns, tool changes and sequence-dependent setups. These breakdowns and stoppages, although occur stochastically, can be measured over the long run. So if F is the downtime frequency and T_{di} is the average downtime per stop of the workstation i, then the actual CT' of each workstation can be formulated as:

$$CT_i' = CT_i + FT_{di} + t_{i,i+1} \qquad \ldots(4)$$

The variability of manual or semi-automated assembly workstations become more complex. While regular worker breaks can be effectively included in T_{di}, the variations in task execution times due to natural human variability and different skill/experience levels, would require the explicit modelling of CT_i' as probability distribution functions. It is therefore conjectured that stochastic simulation is the only way to effectively estimate the performance of real-world assembly lines. For an existing assembly line, the probability distribution functions can be obtained by performing input data analysis on the historial data gathered from the workstations. For new assembly lines, the variability can only be estimated based on some hypothetical distribution functions (e.g. normal distribution) or previous simulation studies.

Synchronised Assembly Lines

The control of the flow of workpieces in serial assembly lines can be classified into synchronised (all workstations are paced), unpaced synchronous and unpaced asynchronous control. In a unpaced asynchronous flow line, a workpiece is passed from one workstation to another once all the tasks assigned to the workstation is complete. This type of control can be described as *push* control because the workpiece will be attempted to be "pushed" to the downstream workstation without knowing its current status. Since the workstations are not paced, upstream workstation may subject to *blocking* if the immediate downstream workstation is occupied with the earlier workpiece. The difference in the cycle times of an unpaced line may also lead to *starving* when a downstream workstation is forced to be idle after it has completed and passed a workpiece but no new workpiece is coming in from the upstream workstation. More frequent blockings and starvings are usually experienced in the consecutive upstream and downstream workstations respectively around the bottleneck workstation in an unpaced and asynchronised flow line. Since both blocking and starving of individual workstations will degrade the performance of the entire assembly line, adding intermediate buffers to the workstations to absorb their variabilities in order to avoid blocking and starving is believed to be the most common way used in industry.

Unpaced asynchronous flow lines with intermediate buffers are widely used in practice and extensively studied in the literature as mentioned. There is recent trend to use simulation-based optimization to solve OBA problems (Bulgak et al. 1995)(Spinellis and Papadopoulos 2000). The work of Altiparmak et al. (2002)(2007) used Artifical Neural Network based metamodels to enhance the performance of the simulated annealing search procedure. Can et al. (2008) have made a comparative study to explore the effect of different stochastic components of GA to solve an OBA problem. Their work has recognized that OBA problems characteristically exhibit conflicting objectives (high TP can lead to WIP accumulation), but the optimization study concerned only optimizing TP.

Unpaced synchronous assembly lines become a more and more popular way of controlling assembly lines in the automotive industry today. This kind of assembly line is implemented with some production control logic (PCL) that governs the movements of parts in the flow in a paced or synchronized manner. In a paced flow, all the parts in the workstations are processed and progressed forward from one workstation to the next at regular intervals. These regular intervals, which set the pace of the assembly line, is sometimes called the takt time. Therefore, a paced assembly line can be called takt-controlled line. There are many different ways to determine the takt time. In some production environments, the takt time can be solely determined as a management variable to synchronize with the actual customer demand rate. For a perfectly balanced line, takt time can be determined by the estimated cycle time (total task times plus average downtime plus transfer time) of each balanced workstation using equation (4). In a realistic case in which a perfectly balanced workload allocation cannot be achieved, the ws with the highest workload (slowest) would determine the takt time, τ, of the entire system:

$$\tau = \mathrm{Max}\left\{CT_i + FT_{di}\right\} + t \qquad \ldots(5)$$

Where i is the index for the 1 to n stations and a constant transfer time, t, is assumed. In an ideal case of a perfectly balanced line in which the takt time is set to be the ideal cycle time (Equation 3), there is no idle times. In an unbalanced line, workstations which complete all the tasks before the predetermined time to progress forward to the succeeding station would be idle. In other words, paced lines can eliminate the needs of any buffers because the variabilities of the assembly tasks are absorbed in the idle times in non-bottlenecked stations. It is a common performance measure in assembly line studies to minimize the average idle time of the entire assembly system.

There are several advantages associated with paced flow lines, in particular when compared to unpaced asynchronous lines with intermediate buffers:

- Paced lines avoid the needs of any buffers (Reinhart et al. 2008); investment cost and space required is therefore reduced.
- The flow of the production is highly smooth and controllable even though the workloads allocated to the stations are not perfectly balanced.
- Paced assembly lines are promoted in Kaizen assembly (Ortiz 2006) because the bottleneck station, i.e., the one with the least average idle time, can be easily identified. The improvement of the identified bottleneck station can facilitate the reduction of the takt time and thus is an effective way to continuously improve the performance of the assembly line.

The takt control can be applied in part of or the complete production flow. A part of a production flow controlled by a takt PCL is as a group called a takt area. Within a takt area each station has the same amount of time, the takt time, to perform its operations. This takt time is set anew each time material is moved in the takt area. There are some conditions that must be met before material within a takt area can be moved. These conditions are:

1. The takt time must have been reached.
2. All stations in the takt area need to be operational and have completed their processing.
3. The last station within the takt area cannot be blocked.
4. There must be material waiting and ready to be processed before the first station in the takt area (optional).

The fourth condition is optional and will affect the way in which the takt PCL functions. The use of it will not allow stations within the takt area to be empty except in the initial phase, if the production flow was empty from start. The takt time will be exceeded as soon as at least one of the conditions 2, 3, or 4 isn't met. This will cause the takt interval (the average time between two consecutive moves within a takt area) to be longer than the actual takt time. Controlling too large a part of a production flow, i.e. too many stations, will likely degrade the performance of the production flow in an unacceptable way. This is linked to the nature of the takt PCL in which every station in the takt poses a potential risk of delaying the takt by not meeting the takt time. Hence having more stations within a takt area pose a greater risk of exceeding the takt time.

Buffer Allocation for Semi-synchronised Lines

The splitting of too large a takt area into several smaller ones is one way of dealing with the poor performance of too large takt areas. The result of such a splitting of a takt area could be called semi-synchronised lines. A takt area is split using buffers that can somewhat smoothen imbalances between the new takt areas. This of course is a compromise in which the production control is closer to a traditional push system the more takt areas there are. Important questions when splitting a takt area are:

1. Where the takt area should be split?
2. What should the capacities of the splitting buffers be?

These questions should of course be evaluated mainly in terms of the performance increase they yield, since improving the performance of the production system is the goal. At the same time it is desirable to make as few and small changes as possible to the initial production system. We call this problem the Buffer Allocation for Semi-synchronised lines problem, which is a problem well suited for simulation-based multi-objective optimization (SMO).

Simulation Modelling for Takt Control

Implementing a takt PCL, that control a production flow according the four conditions mentioned earlier, in a discrete event simulation software is relatively easy. That is if the production flow and the takt areas don't change. If on the other hand you are dealing with dynamic takt areas the implementation becomes a bit more difficult. One could try different takt area scenarios by building a model for each scenario. However this brings a lot of manual work in terms of building the models as well as analysing and comparing the results from each scenario. A dynamic takt PCL that can split into several takt areas based on the capacities of the buffers along the production flow is desirable. This kind of dynamic takt PCL will enable SMO to help minimize the manual work needed. In addition SMO will also be able to vastly increase the number of scenarios that one would normally cope with. This type of dynamic takt PCL has been implemented and tried in two industrial cases.

AN INDUSTRIAL CASE STUDY

Description of the assembly lines

The studied cases consist of two similar assembly lines. The initial production setup consists of two takt areas, pre-assembly and main-assembly. Simplified graphs of the two production flows can be seen in Figure 2 and Figure 3.

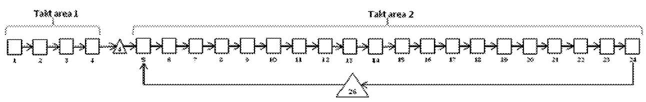


Figure 2: Case 1, Pre-Assembly (Takt area 1) and Main-Assembly (Takt area 2).

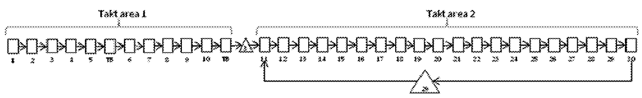


Figure 3: Case 2, Pre-Assembly (Takt area 1) and Main-Assembly (Takt area 2).

The material handling system for the pre-assemblies is either a roller or a chain conveyor system and for the main-assemblies it is autonomous carriers (26 in Case 1 and 29 in Case 2). The takt time is given by the length of a station and the speed of the carrier, 104 seconds in Case 1 and 105 seconds in Case 2. The carriers are controlled by the four takt PCL conditions. All carriers within a takt area are stopped as soon as at least one condition is not met. Condition 2 translates to a carrier reaching the next station without all assembly being completed.

The dynamic takt PCL have enabled the two takt areas in each case to be split into as many takt areas as there are stations within each initial takt area. The extreme with as many takt areas as stations will, as mentioned, represent a pure push system. In these cases the takt splitting buffers are represented by parts of the conveyor system not being assigned to stations. Hence these parts could harbour one or several carriers depending on the length of each part/buffer. This representation of a buffer also implies that the transport time of a buffer is dependent of the length/capacity of the buffer, i.e. larger capacity longer transport time.

The takt splitting buffer capacities have been used as input to the SMO studies. The capacities have been allowed integer values {0,1,2}, where zero indicates that the takt area isn't split at that buffer location. Area restrictions in the factory make it interesting to find the best possible performance with the lowest number of buffers. This is considered in the SMO studies by maximizing the total number of produced parts and at the same minimizing the sum of all buffer capacities.

Each case has historical production data as a reference and also some experience based takt area scenarios. In total there were 8 experience based takt area scenarios that were analyzed manually. These scenarios have been constructed by placing buffers on locations that result in takt areas that on average exceeds the takt time roughly the same amount of time. In this study these scenarios will be compared with the extensive analysis that SMO provides.

Input data analysis

The actual complexity of the two assembly lines have been somewhat simplified in the studied cases. Variants are abstracted into one variant with the takt time as processing time at every station. The dynamic behaviour of the model is instead modelled as failures on each station, i.e. a failure is to be interpreted as a station exceeding the takt time. These takt exceed times have been collected per station during 6 weeks of normal production, and been used to create empirical distributions for the duration of failures. The probability of a failure occurring during the processing of a part, on each station, has been used to achieve realistic intervals between failures. The probability is based on the number of failures that occurred and the total amount of produced parts during the 6 weeks. Results from the failure analysis of the two cases are provided in Table 1. The values in the table are average values considering data from all stations in each of the two cases. The average takt exceed times are almost identical but Case 1 performs a bit worse considering that line being more prone to failures.

Table 1: Average failure statistics for the two assembly lines/cases.

	Average takt exceed time [sec]	Average failure probability [%]
Case 1	39.1	4.7
Case 2	39.7	2.6

Comparing the ideal performance of these lines with their current performance gives an indication of the maximum performance improvement that could be expected. The ideal performance of this type of line that is controlled by a takt PCL is given by the takt time, i.e. without failures the line would produce one part per takt time. Comparing the average takt interval from the main assembly lines of the two cases with the takt time gives the maximum improvement potentials. The gap between current performance and the ideal performance in the two cases are 24.4 % in Case 1 and 16.7 % in Case 2. However these figures only serve as theoretical upper limits on the improvements. Since splitting takt areas reduces the impact of failures, which is not the same as eliminating all failures.

Optimisation results

Results from the experience based scenarios and the optimization, in which 5000 different scenarios were evaluated for each case, are shown in Figure 4 and Figure 5. In both cases the solution in Scenario 3 is identical to the solution found by the optimization when only splitting in one location with a buffer of capacity one. However, if even higher improvements are desireable and it is possible to split in more than one location the optimization finds solutions/locations that are better than the experience based.

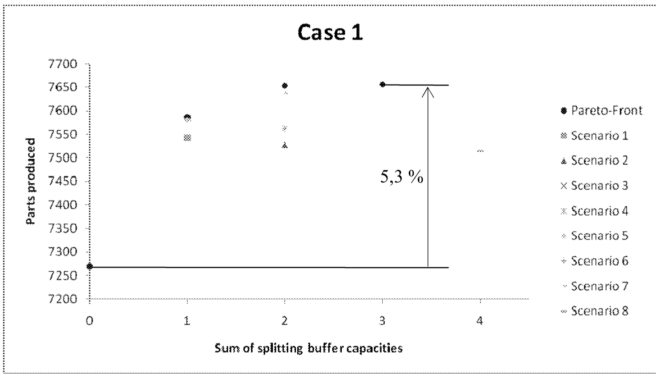


Figure 4: Case 1, Comparison of non-dominated (pareto-front) solutions and experience based scenarios.

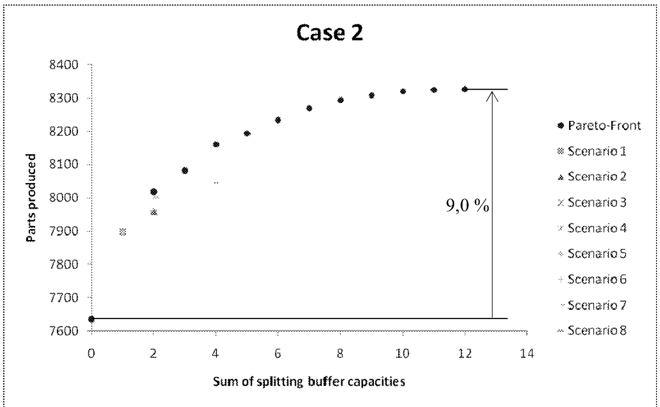


Figure 5: Case 2, Comparison of non-dominated (pareto-front) solutions and experience based scenarios.

Interestingly the SMO performed on Case 2 provides more possible non-dominated solutions to choose from compared to the one performed on Case 1. Also, when only considering the maximization of produced parts, Case 2 with 9 % improvement outperform Case 1 having only a 5.3 % improvement. Overall it seems as semi-synchronised lines have more potential in Case 2, why is that?

Analysis on Takt control

The analysis of all solutions (not only the non-dominated) from the optimizations conducted on the two cases (Figure 6 and Figure 7) can give us the answer to the question risen in the previous section. As expected the results shows a performance improvement (more parts produced) initially when having smaller takt areas (more buffers). However it also shows significant performance degradation if there are too many takt areas. This behaviour is a combination of how the takt PCL functions (mainly condition 4), the capacity dependant transport time of the takt splitting buffers and the limited amount of carriers. A delay caused by an upstream takt area will, if buffers are empty, propagate to downstream takt areas with an additional delay in terms of the transport time of the buffers. Due to this the takt interval will increase in every takt area when moving downstream and degrade the performance of the production. Empty buffers will increase as more splitting buffers are introduced since there is a limited number of carriers. This explains partly why the degradation is more severe and the potential of semi-synchronised lines is less in Case 1, which has got less carriers than Case 2. Part of the difference between the two cases can also be explained by Case 1 performing worse than Case 2, i.e. delays are more common in Case 1.

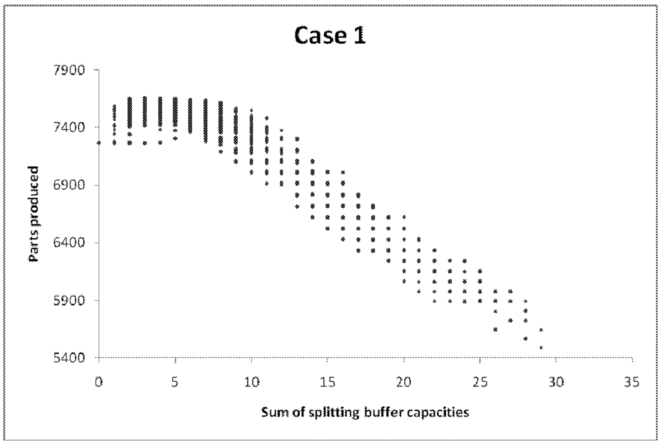


Figure 6: Case 1, Optimization results showing the performance related to different amounts and capacities of splitting buffers.

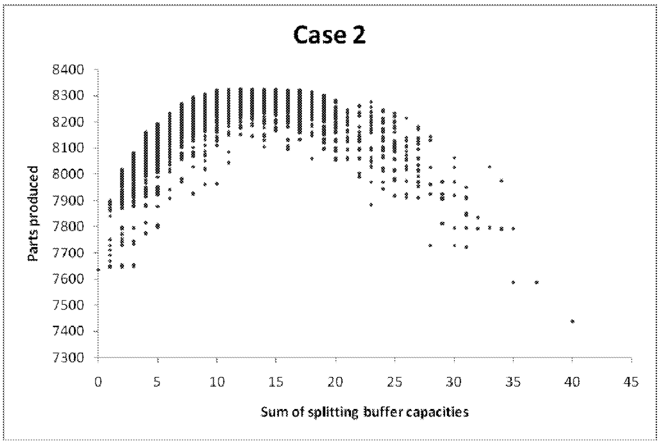


Figure 7: Case 2, Optimization results showing the performance related to different amounts and capacities of splitting buffers.

CONCLUSIONS

The SMO was as expected well suited for the design of semi-synchonised lines. It provided answers like where and how to split too large takt areas in order to improve the performance of the production system. It can also be concluded that the extensive analysis that SMO provides can present solutions that are better than ones based on experience. As in many other cases with SMO additional important knowledge was gained by analyzing the results. Interactions between takt PCL and carrier systems might severly degrade the performance of semi-synchronised lines if there are not enough carriers.

ACKNOWLEDGEMENTS

This work is partially financed by VINNOVA, Sweden, through the FFI-HSO project. The authors gratefully acknowledge their provision of research funding. The authors are also thankful for our industrial partner companies, including Scania AB, Volvo Cars Corporation and Volvo Technology Corporation, for their support in this study.

REFERENCES

Altiok, T. and Stidham, S. J. 1983. "The allocation of interstage buffer capacities in production lines". *IIE Transactions* 15: 292-298.

Altiparmak, F., Dengiz, B. and Bulgak, A.A. 2002. "Optimization of buffer sizes in assembly systems using intelligent techniques". In *Proceedings of 2002 Winter Simulation Conference*, Yücesan, E., Chen, C.-H., Snowdon, J.L. & Charnes, J.M. (eds.), Piscataway, New Jersey, USA: IEEE, 1157-1162.

Altiparmak, F., Dengiz, B. and Bulgak, A.A. 2007 "Buffer allocation and performance modeling in asynchronous assembly system operations: an artificial neural network metamodeling approach". *Applied Soft Computing*, 2007, 7(3), 946-956.

Boysen, N., Fliedner, M., Scholl, A. 2008. "Assembly line balancing: Which model to use when?" *International Journal of Production Economics*, 111, 509-528

Bulgak, A.A., Diwan, P.D. and Inozu, B. 1995. "Buffer size optimization in asynchronous assembly systems using genetic algorithms". *Computers and Industrial Engineering*, 28(2), 309-322.

Buzacott, J.A. and Shanthikumar, J.G. 1993, *"Stochastic models of manufacturing systems Englewood Cliffs"*. Prentice Hall: New Jersey.

Can, B., Beham, A. and Heavey, C. 2008. "A comparative study of genetic algorithm components in simulation-based optimization". In *Proceedings of 2008 Winter Simulation Conference*, Mason, S.J., Hill, R.R., Mönch, L., Rose, O., Jefferson, T. & Fowler, J.W. (eds.), Miami, Florida, USA, 1829-1837.

Conway, R., W. Maxwell 1988. "The role of work-in-process inventory in serial production lines". *Operations Research* 36: 229-241.

Deb, K. 2001. "*Multi-objective optimization using evolutionary algorithms"*, 3'rd ed., Wiley, Wiltshire, UK.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. 2002. "A fast and elitist multi-objective genetic algorithm: NSGA-II". *IEEE Transaction on Evolutionary Computation*, 2002, 6(2), 181-197.

Gershwin, S.B. 1987 "An efficient decomposition method for the approximate evaluation of tandem queues with finite storage space and blocking", *Operation Research*, 35(2), 291-305.

Hillier, M. S. (2000). "Characterizing the optimal allocation of storage space in production line systems with variable processing times". *IIE Transactions* 32: 1-8.

Koh, S.-G. and Bulfin R.L. 2004. "Comparison of DBR with CONWIP in an unbalanced production line with three stations". *International Journal of Production Research* 42(2): 391-404.

Ortiz, C.A. 2006. "Takt time". *Kaizen assembly: designing, constructing, and managing a lean assembly line*. Boca Raton: CRC Press.

Reinhart, G., Werner, J., Lange, F. 2008. "Robot based system for the automation of flow assembly lines". *Production Engineering*, 3(1), 121-126.

So, K.C. 1997. "Optimal buffer allocation strategy for minimizing work-in-process inventory in unpaced production lines". *IIE Transactions*, 29, 81-88.

Spinellis, D.D. and Papadopoulos, C.T. 2000. "A simulated annealing approach for buffer allocation in reliable production lines". *Annals of Operations Research*, 93, 373-384.

Tempelmeier, H. 2003. "Practical considerations in the optimization of flow production systems". *International Journal of Production Research*, 41, No. 1: 149-170.

Yamashita, H. and Altiok, T. 1998. "Buffer capacity allocation for a desired throughput in production lines". *IIE Transactions*, 30: 883-891.

AUTHOR BIOGRAPHY

JACOB BERNEDIXEN is born in Alingsås, Sweden and has a M.Sc. in Industrial Engineering and Management from the Institute of Technology at Linköping University. Since his graduation four years ago he has been working with industrial-based SMO research projects at the University of Skövde.

AMOS H.C. NG is an Associate Professor at the University of Skövde, Sweden. He holds a B.Eng. degree and a M.Phil. degree, both in Manufacturing Engineering from the City University of Hong Kong and a Ph.D. degree in Computing Sciences and Engineering from De Montfort University, Leicester, U.K. He is a member of the IET and a Chartered Engineer in the U.K. His research interests include modelling, simulation and multi-objective optimization for production systems design and analysis.

PRODUCTION OPTIMIZATION

APPLICATION OF MULTIPLE CRITERIA OPTIMIZATION VIA SIMULATION TO INJECTION MOLDING

Maria G. Villarreal-Marroquin
Jose M. Castro
Integrated Systems Engineering Department
The Ohio State University
1971 Neil Ave, Columbus, OH, 43210
United States
E-mail:{villarreal-marroquin.1|castro.38}@osu.edu

Mauricio Cabrera-Rios
Department of Industrial Engineering
University of Puerto Rico at Mayagüez
Mayagüez P.R., 00681
United States
E-mail: mauricio.cabrera1@upr.edu

KEYWORDS
Computer Aided Engineering, Optimization via Simulation, Injection Molding, Continuous Simulation.

ABSTRACT

Injection Molding (IM) is the most important process for mass-producing plastic products. Thus, it is vital to design efficient processes. The difficulty of optimizing the injection molding process is that the performance measures usually show conflicting behavior. Therefore the best solution for one performance measure is usually not the best in some other performance measures. This paper, presents an optimization via simulation method which considers multiple performance measures and is able to find a set of efficient solutions without having to evaluate a large number of simulations. A real injection molding case is used to illustrate and detail the method.

INTRODUCTION

Polymers have been increasingly replacing metallic components in many applications such as automobiles, aircrafts, toys, appliances, office equipment, among other. This is because they are very versatile materials. Nowadays, many consumer products such as computer and automobile components rely on the technology and production from polymer companies. Thus, it is important to design reliable processes to ensure the sustainability of these companies.

In IM, for instance, processing conditions such as melt temperature, mold temperature, pack/hold pressure and duration, and cooling time have to be properly set to ensure the quality of the molded components. Often, these conditions are set by process engineers based on prior experience, resin supplier's recommendations, and/or reference handbooks. These conditions are usually further adjusted by trial and error on the shop floor. This approach is highly dependent on the experience of molding operators and can be costly and time consuming, especially with new resins and/or new applications (Zhou and Turng 2007). However, with recent advances in numerical modeling and computer simulation techniques, a large effort has been made in developing computer simulation tools to help improve and facilitate the modeling of injection molded parts.

The use of simulation for selecting injection molding processing conditions has been the subject of much research in the past (Smith et al. 1998; Gokce et al. 2002; Alam and Kamal 2005). Specialists usually evaluate a limited number of scenarios from which one is finally selected. Nevertheless, this does not guaranty having found a nondominated solution. There is a lot of potential to be exploited in the adequate and efficient selection of optimization techniques for the design of manufacturing processes through computer simulations. Such potential explains the relatively recent and rapidly increasing interest in Optimization via Simulation (OvS) as a field on its own. The objective of an OvS method is to provide a structure to determine the values of the controllable variables (simulation inputs) that optimize an objective function defined as a combination of the simulation outputs (performance measures) (Swisher et al. 2000). Then, an optimization routine uses the calculated values of the objective function and previous evaluations, to select a new set of input values; this is continued until a pre-selected convergence criterion is satisfied (April et al. 2004).

Common OvS methodologies are mainly characterized by the use of a single objective function representing a performance measure (PM). However, in injection molding, we are typically interested on optimizing a set of performance measures rather than a single one. And, as previously mentioned the controllable variables in general, have conflicting effects on the performance measures. Therefore the best solution for one performance measure is usually not the best for some other performance measures. Thus, it is not appropriate to obtain a single solution but rather the set of solutions corresponding to the best compromises (efficient solutions (Pareto Frontier)), from which the engineer can make the best decision depending on his application. This last task falls in the emerging area of Optimization via Simulation with Multiple Performance Measures.

Contributions in the area of OvS have focused on discrete-event simulations and not in continuous or physics-based simulations (like the ones used to model injection molding processes). One reason could be because the computational time required to run physics-based simulation models is usually very long. Consequently, having an iterative algorithm that requires many simulation runs to find the processing conditions to optimize different performance measures simultaneously is not computationally practical. Here a multicriteria optimization via simulation method which integrates design of experiments and metamodeling techniques to reduce the number of simulation runs needed to solve the multicriteria problem is presented. The method is illustrated and detailed using a real IM case.

The paper is organized as follows: first, we introduce the proposed method; then we apply the proposed method to a real injection molding case for illustration purposes. Finally conclusions and future work are presented.

PROPOSE METHOD

The Multicriteria Optimization via Simulation Method proposed here is schematically shown in Figure 1. The method starts with a designed experiment (DOE) from which a simulation run is performed at each design point. Then, the best compromises between all performance measures are obtained using Data Envelopment Analysis (DEA). These solutions are set as the Incumbent Efficient Frontier. At each iteration, a metamodel for each PM is constructed using all the available information. Then, through these metamodels, feasible solutions are generated and a Predicted Efficient Frontier is found. The predicted efficient solutions are then simulated and evaluated against the incumbent efficient solutions to update the incumbent efficient frontier. A series of stopping criteria are evaluated and, if none is met, the new points are added to the existing set of points and a new iteration begins. Otherwise, the method stops and the incumbent efficient frontier is reported.

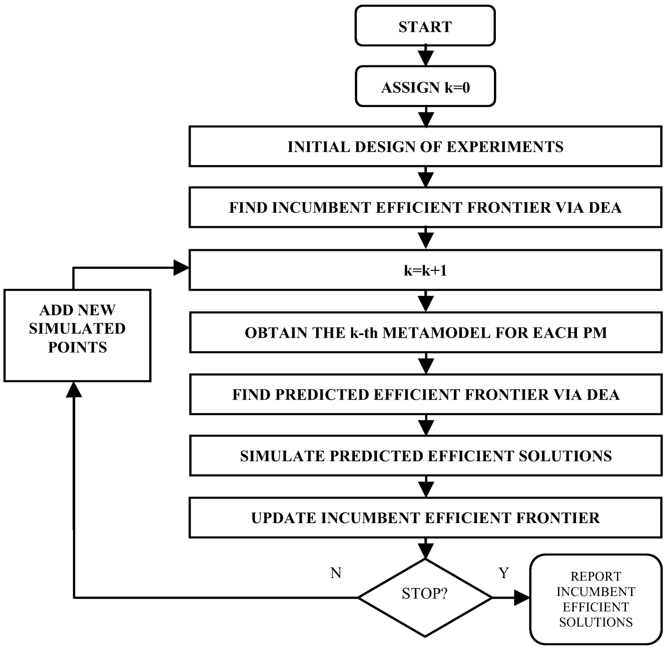


Figure 1: Sketch of Multicriteria Optimization via Simulation proposed method.

Optimization via Simulation in contrast to traditional optimization lacks explicit equations to represent the objective functions (performance measures). Therefore, metamodeling-based optimization has played an important role in OvS. Metamodels are mathematical representations of real phenomena based on a limited number of observations that can take the form of the outputs of a simulation model. Metamodels overcome the need of explicit equations as objective functions, and moreover, they help reduced the computational effort required to evaluate the simulation software. The evaluation time of a metamodel is considered to be much faster than the simulation software. In such cases, new solutions can be estimated inexpensively.

Recent contributions to the area of multicriteria optimization via simulation that used metamodeling are: (Dallino et al. 2009; Zakerifar et al. 2009). Both used Response Surface Methodology (RSM) and Kriging metamodels to represent multiples PMs in inventory problems. Ryu et al. (2009) used RSM techniques to construct quadratic metamodels of PMs in a variety of multiple criteria problems.

The motivations to use design of experiments and metamodels to represent the outputs of the simulation model are the successful results of a the large number of works related to metamodel-based simulation optimization (Andradóttir 1998; Azadivar 1999; Swisher et al. 2000; Fu 2001; Ólafsson and Kim 2002; Fu et al. 2005), and previous work of our research group on the optimization of physical and chemical phenomena in polymer processing (Cabrera-Ríos et al. 2002, 2002a; Castro et al, 2004; Castro et al. 2005, 2007; Villarreal et al. 2008).

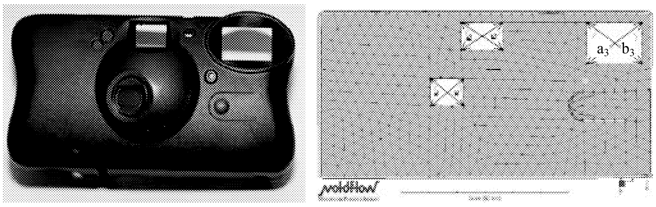
There are several methods to approach multicriteria optimization problems such as simulated annealing and evolutionary algorithms; here Data Envelopment Analysis (Charnes et al. 1993) is used. The idea behind DEA is to use an optimization model to compute a relative efficiency score for each particular solution with respect to the rest of the candidate solutions. The resulting best compromises, identified through their efficiency score, form the envelope of the solution set. These solutions are the so-called efficient solutions, and lie on the Pareto Frontier. The DEA linear programming formulations proposed by Banks, Charnes and Cooper can be found in (Charnes et al. 1993). Both models, the BCC Input Oriented and the BCC Output Oriented, are applied to each of the *n* candidate solutions. A particular solution with an objective function score of 1 (i.e. an efficiency score of 1) using both formulations is considered an efficient solution, and is therefore on the Pareto Frontier.

The main motivations for using DEA to solve the multiple criteria optimization problems are: (1) DEA uses linear programming, which is the simplest optimization problem, (2) DEA can be carried out using simple software, like MS Excel, and (3) once an efficient solution is identified by DEA, one can be sure that it is indeed an efficient solution. This is because the best solution of the linear problem is in fact the optimal solution.

In the next section, we illustrate the proposed method using a real injection molding case.

ILLUSTRATION OF PROPOSED METHOD

An important feature in the manufacturing of plastic components is parts quality. Part quality can be measured in different forms. Here it is measured as a function of part shrinkage. It is desirable that the molded part conserves the designed symmetry and dimensions. This is especially important because the part under study is to be assembled. The part to be analyzed is the front plate disposable camera shown in Figure 2a.



(a) (b)

Figure 2:(a) Real disposable camera (b) Simulation mesh

In our example, it is desired to keep the shape of the biggest rectangular window (upper right corner) after shrinkage. Then, the performance measures of interest are the shrinkages of the diagonals a_3 and b_3 (Δa_3, Δb_3) and the difference between the diagonals shrinkages ($|\Delta a_3 - \Delta b_3|$), see Figure 2b for better reference. The difference between the diagonals a_3 and b_3 has to be close to zero in order to keep the rectangular shape of the window and its shrinkages itself need to be close to zero to keep the original dimensions. Thus, the objective is to minimize Δa_3, Δb_3 and $|\Delta a_3 - \Delta b_3|$. For clarity of illustration, this study only focused on the minimization of Δa_3 and $|\Delta a_3 - \Delta b_3|$.

The camera studied here is injection molded using Poly-Styrene made by Dow Chemical USA (trade name Styron 685D). Packing pressure and melt temperature were held constant at 50 MPa and 200°C respectively. The fill time was kept constant at 1s; and the cooling time was set automatically by the simulation software. An automatic cooling time is the time required to achieve a target average mold temperature (set here at 108 ºC) and the specified percentage of the part that needs to be frozen (set here at 100%). Mold temperature (T_{mold}) and packing time (t_{pack}) were considered as controllable variables and are varied in the ranges of [20,70]°C and [1,10]s respectively. The simulation software MoldFlow™ is used to analyze this process.

The optimization problem for this application is defined as follows:

$$
\begin{array}{ll}
\text{Find} & \mathbf{x} := (T_{mold}, t_{pack}) \text{ to} \\
\text{Minimize} & f_1(\mathbf{x}) := |\Delta a_3| \quad \text{and} \\
\text{Minimize} & f_2(\mathbf{x}) := |\Delta a_3 - \Delta b_3| \\
\text{Subject to} & 20^{o}C \le T_{mold} \le 70^{o}C \\
 & 1s \le t_{pack} \le 10s
\end{array} \qquad \mathbf{(1)}
$$

In optimization problem (1), the values of shrinkage Δa_3 and Δb_3 are estimated using MoldFlow™.

In order to solve problem (1), the described multicriteria optimization via simulation method was applied. Referring to Figure 1 the method is as follows:

Initialization

i. *Initialize iteration counter:* k=0.

ii. *Initial DOE*: The method begins with a Central Composite Design with 9 points (a single center run). The design points and the evaluation of both performance measures are graphically shown in Figures 3 and 4.

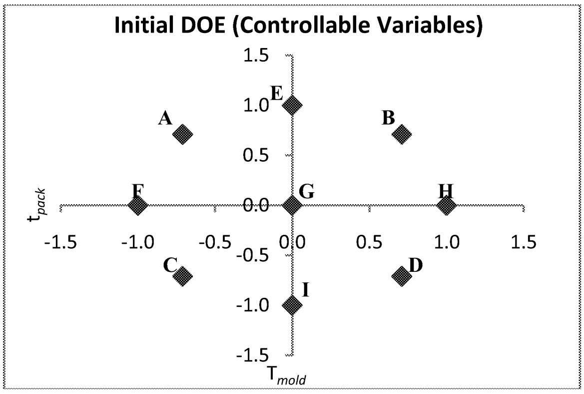


Figure 3: Initial DOE (Coded Values)

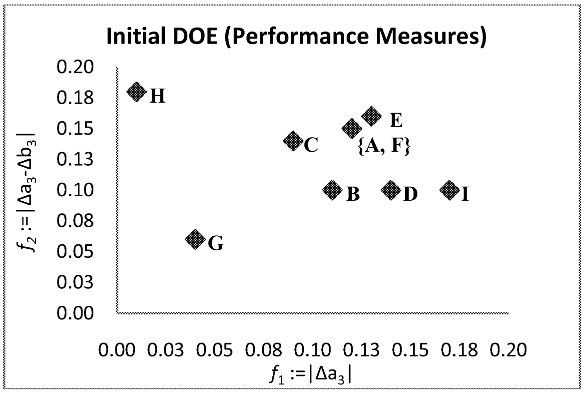


Figure 4: Evaluation of initial DOE

iii. *Select incumbent efficient frontier:* Considering the optimization criteria detailed in formulation (1), the non-dominated solutions are found via DEA. This set is the current incumbent efficient frontier ($\boldsymbol{P}_{k\text{-}best}$), $\boldsymbol{P}_{0\text{-}best}$={H,G}.

Iteration k:= k+1=1

1. *Update iteration counter*: k = 0+1=1

2. *Obtain metamodels*: Using the available points, build the k-th metamodel for each of the M PMs $\widehat{f_m}(\cdot)_k$ (m=1,2,…,*M*), here M=2. The proposed metamodels are linear regression models of the highest possible estimable order. Equations (2) and (3) show the resulted metamodels, where $\widehat{f_1}(\cdot)_k$ estimates $|\Delta a_3|$ and $\widehat{f_2}(\cdot)_k$ $|\Delta a_3 - \Delta b_3|$; x_1 represents T_{mold} and x_2 t_{pack}. The corresponding coefficients of determination (R^2) are 99.52% and 88.88%. Notice that to avoid dimensionality effects the values of $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ were transformed into the scale of [-1, 1].

$$
\widehat{f_1}(\boldsymbol{x})_1 = -0.63 - 0.69x_1 - 0.25x_2 - 0.37{x_1}^2 - 0.36{x_2}^2 + 1.42x_1x_2 + 0.50{x_1}^3 + 1.71{x_2}^3 \qquad \mathbf{(2)}
$$

$$
\widehat{f_2}(\boldsymbol{x})_1 = -0.99 + 0.25x_1 + 0.50x_2 - 0.08{x_1}^2 + 1.53{x_2}^2 + 0.95x_1x_2 - 0.88{x_1}^3 - 1.54{x_2}^3 \qquad \mathbf{(3)}
$$

3. *Optimization*: Using the metamodel of each performance measure, a set of potential feasible solutions is generated using random values of the controllable variables (**x**). Then, DEA is applied to find the efficient frontier of the generated points, called here predicted efficient frontier($\hat{P}_{k-best}$). Figure 5 shows the generated solutions and the predicted efficient frontier ($\hat{P}_{1-best}$).

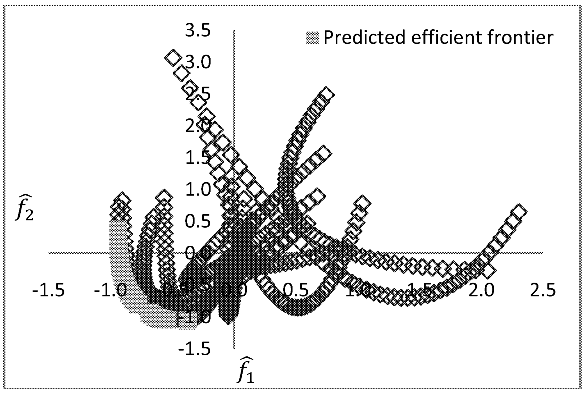


Figure 5: Generated solutions (◇) and efficient frontier (▩).

4. *Simulate the new points*: Estimate, using the simulation software, the values of $f_m(\cdot)$ for each point on the efficient frontier found on 3 ($\hat{P}_{1-best}$). Figure 6 shows the evaluation of these solutions along with the solutions of the initial DOE.

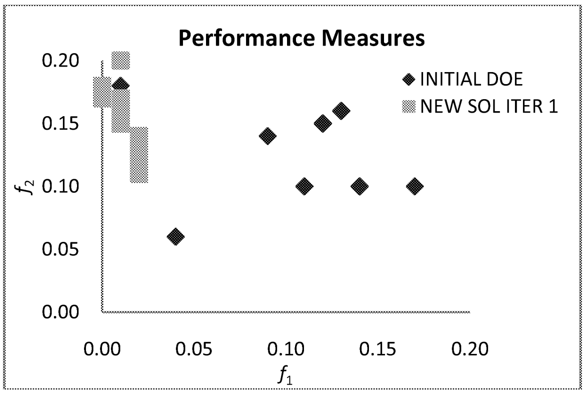


Figure 6: Simulated value of predicted efficient solutions (▩).

5. *Update the incumbent efficient frontier*: Using the actual incumbent efficient solutions ($\boldsymbol{P}_{(k-1)-best}$:=$\boldsymbol{P}_{0-best}$) and the new simulated solutions (from 4), update the incumbent efficient frontier ($\boldsymbol{P}_{k-best}$:=$\boldsymbol{P}_{1-best}$), that is, find the non-dominated solutions among all these. Figure 7 shows the new efficient frontier ($\boldsymbol{P}_{1-best}$) and the last one ($\boldsymbol{P}_{0-best}$).

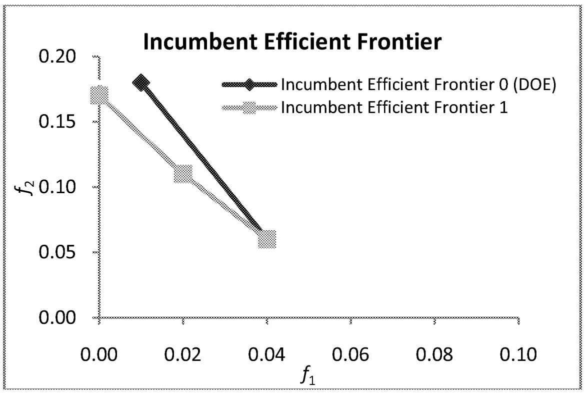


Figure 7: New incumbent efficient frontiers (▩).

6. *Evaluate the stopping criteria*: The following stopping criteria were considered: stop if (1) the incumbent efficient frontier does not change (no new efficient solutions are found) through a determined number of iterations; (2) if the coefficient of determination $R^2 \geq \varepsilon$ for all the metamodels; (3) a maximum number of simulation evaluations have been reached. For illustration purposes we assume that none of the stopping conditions was met.

Because none of the stopping criteria was met, the simulated solutions (squared points of Figure 6) are added to the set of available points to build a new metamodel for each PM, and the main iteration is repeated.

Iteration k = 1+1=2

After constructing the new metamodels, new random solutions are generated. The new metamodels include an additional regression coefficient per added solution to the set of points. Then, the values of the performance measures are estimated using these metamodels ($\hat{f}_1(\boldsymbol{x})_2$ and $\hat{f}_2(\boldsymbol{x})_2$), and via DEA, $\hat{P}_{2-best}$ is obtained. Next, the predicted efficient solutions are simulated and compared with the last incumbent efficient solutions ($\boldsymbol{P}_{1-best}$). Figure 8 shows the change of the incumbent efficient frontier after two iterations. For illustration purposes we stop the algorithm here. The final efficient frontier consists of 6 solutions which are shown on Table 1 and Figure 9.

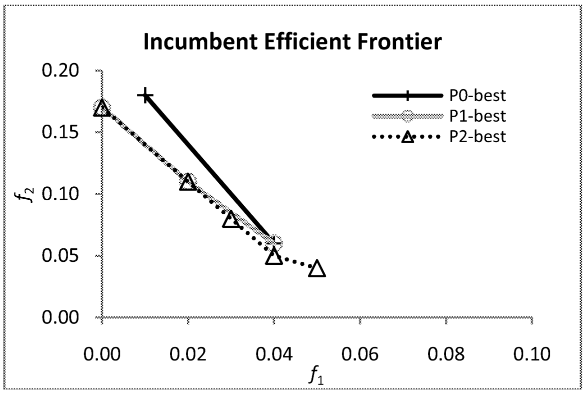


Figure 8: Change of incumbent efficient frontier.

Table 1: Final Efficient Solutions

t_{pack} (s)	T_{mold} (°C)	|Δa3| (mm)	|Δa3-Δb3| (mm)
5	45	0.00	0.17
6	51	0.02	0.11
6	46	0.03	0.08
6	40	0.04	0.05
6	41	0.04	0.05
6	39	0.05	0.04

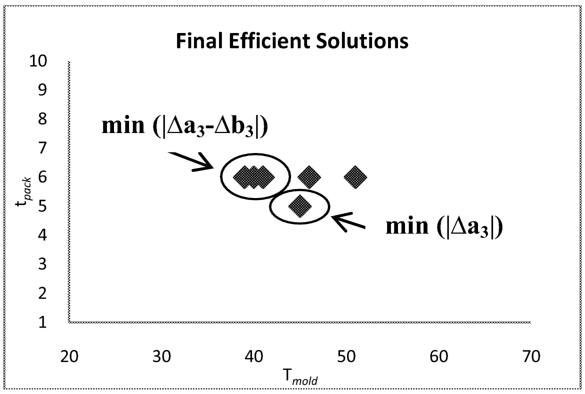


Figure 9: Final efficient solutions.

From the set of efficient solutions the decision maker can select one depending on its particular application.

CONCLUSIONS AND FUTURE WORK

This paper introduced an optimization via simulation method to solve multicriteria injection molding problems. The aim of the method is to help the design of injection molding processes. The method combines design of experiments and metamoding techniques, requiring a small number of simulation evaluations. This makes it attractive for cases where simulations take long time to run. We are in the process of comparing our results to real experiments, as well as to compare the efficiency of the method using other types of design of experiments and metamodeling techniques. Also, we are investigating how to use the efficient frontier to develop Process Windows that can be used by the molding engineers.

REFERENCES

Alam, K. and M.R. Kamal. 2005. "A Robust Optimization of Injection Molding Runner Balancing". *Computer Applications in Chemical Engineering,* Vol. 29, No. 9, 1934- 1944.

Andradóttir S. 1998. "A review of Simulation Optimization techniques". *Proceedings of the 1998 Winter Simulation Conference*, 151-158.

April, J.; M. Better; F. Glover; and J. Kelly. 2004. "New Advances and applications for marrying Simulation and Optimization". *Proceedings of the 2004 Winter Simulation Conference, 80-86.*

Azadivar F. 1999. "Simulation Optimization Methodologies". *Proceedings of the 1999 Winter Simulation Conference*, 93-100.

Cabrera-Ríos M.; J. M. Castro; and C. A. Mount-Campbell. 2002. "Multiple Quality Criteria Optimization In In-Mold Coating (IMC) With A Data Envelopment Analysis Approach". *Journal of Polymer Engineering*, Vol.22, No.5, 305-340.

Cabrera-Ríos M.; K. S. Zuyev; X. Chen; J. M. Castro; and E. J. Straus. 2002a. "Optimizing Injection Gate Location and Cycle Time for the In-Mold Coating (IMC) Process". *Polymer Composites,* Vol.23, No. 5, 723-738.

Castro C. E.; M. Cabrera-Ríos; B. Lilly; and J. M. Castro. 2005. "Simultaneous Optimization of Mold Design and Processing Conditions in Injection Molding". *Journal of Polymer Engineering*, Vol.25, No.6, 459-486.

Castro C. E.; Cabrera-Ríos, M.; Lilly, B.; and J.M. Castro. 2007. "Optimization and Analysis of Variability in Injection Molding". *Journal of Polymer Engineering and Science:* Vol.47, No.4, 400-409.

Castro J. M.; M. Cabrera-Ríos; and C. A. Mount-Campbell. 2004. "Modeling and Simulation in Re-active Polymer Processing". *Modeling and Simulation in Materials Science and Engineering*, Vol.12, S121-S149.

Charnes, A.; W.W. Cooper; A.Y. Lewin; and L.M. Seiford. 1993. *Data Envelopment Analysis: Theory, Methodology, and Applications.* Kluwer Academic Publishers, Boston.

Dellino G. and J.P.C. Kleijnen. 2009. "Robust simulation-Optimization Using Metamodels". *Proceedings of the 2009 Winter Simulation Conference*, 540-550.

Fu, M. C. 2001. "Simulation Optimization". *Proceedings of the 2001 Winter Simulation Conference*, 53-61.

Fu, M. C.; F. Glover; and J. April. 2005. "Simulation Optimization: A review, new developments, and application". *Proceedings of the 2005 Winter Simulation Conference*, 83-95.

Gokce, A.; K-T. Hsiao; and S.G. Advani, S.G. 2002. "Branch and bound search to optimize injection gate locations in liquid composite molding processes". *Composites Part A: Applied Science and Manufacturing*, Vol. 33, No. 9, 1263-1272.

Ólafsson S. and J. Kim. 2002. "Simulation Optimization". *Proceedings of the 2002 Winter Simulation Conference,* 79-84.

Ryu, J. H.; S. Kim; and H. Wan. 2009. "Pareto Front Approximation with Adaptive Weighted Sum Method in Multiobjective Simulation Optimization". *Proceedings of the 2009 Winter Simulation Conference*, 623-633.

Smith, D.E.; D.A. Tortorelli; and C.L. Tucker. 1998. "Analysis and sensitivity analysis for polymer injection and compression molding. *Computer Methods in Applied Mechanics and Engineering,* Vol.167, No. 3-4, 325-344.

Swisher J. R.; P. D. Hyden; S. H. Jacobson; and L. E. Schruben. 2000. "A survey of simulation optimization techniques and procedures". *Proceedings of the 2000 Winter Simulation Conference*, 119-128.

Villarreal, M.G.; M. Mulyana; J.M. Castro; and M. Cabrera-Ríos. 2008. "Simulation optimization applied to injection molding". *Proceedings of the 2008 Winter Simulation Conference*, *1995-2003.*

Zakerifar, M.; W.E. Biles; and G.V. Evens. 2009. "Kriging Metamodeling in Multi-Objective Simulation Optimization". *Proceedings of the 2009 Winter Simulation Conference*, 2115-2122.

Zhou, J. and L.S. Turng. 2007. "Process Optimization of Injection Molding Using an Adaptive Surrogate Model with Gaussian Process Approach". *Polymer Engineering and Science*, Vol.47, No. 5, 684-694.

AUTHOR BIOGRAPHIES

MARÍA G. VILLARREAL-MARROQUÍN is a PhD candidate at The Ohio State University in Industrial & Systems Engineering. She obtained her B.S. in Mathematics (2005) and M.S. in Systems Engineering (2007) from Universidad Autónoma de Nuevo León, México. Her research interests include Simulation Optimization, and applied optimization in manufacturing.

JOSE M CASTRO is currently professor in the Department of Integrated Systems Engineering and Director of the CAPCE at The Ohio State University. After obtaining his PhD in Chemical Engineering from the University of Minnesota in 1980, he was a faculty at the University of the South in Bahia Blanca Argentina for 4 years. He then worked at the Gen Corp corporate technology center for 12 years as chief technologist. After GenCorp he worked for 3 years as Manager of the New Process technology group in Allied Signal Electronic Materials. He joined Ohio State in September 1998. He has published more than 70 peer review journal papers and given numerous invited talks. His area of research is polymer processing and composites manufacturing modeling and optimization.

MAURICIO CABRERA-RIOS obtained his B.S. degree in Industrial & Systems Engineering from Monterrey Institute of Technology (ITESM), in Monterrey, México in 1996. He obtained an M.S. and Ph.D. degrees in the same field from The Ohio State University in 1999 and 2002 respectively in Columbus, Ohio. He was an Associate Professor at the Universidad Autónoma de Nuevo León Graduate, México from 2003 to 2008. Currently, he is an Assistant Professor in the Industrial Engineering Department at University of Puerto Rico-Mayagüez. His research work currently relates to biological data analysis and to applied optimization in manufacturing.

A COMPARISON OF TWO DIFFERENT APPROACHES TO MULTI-CRITERIA OPTIMISATION OF SEMICONDUCTOR FABRICATION

Ingy A. El-Khouly
Khaled S. El-Kilany
Departement of Industrial and Management Engineering
Arab Academy for Science, Technology & Maritime Transport
Alexandria, Egypt
E-mail: ingyelkhouly@aast.edu
E-mail: kkilany@aast.edu

John Geraghty
Paul Young
School of Mechanical and Manufacturing Engineering
Dublin City University, Glasnevin, Dublin 9
Dublin, Ireland
E-mail: john.geraghty@dcu.ie
E-mail: paul.young@dcu.ie

KEYWORDS
Multi-Objective Optimisation, Evolutionary Optimizer, Pareto-Optimal Genetic Algorithm.

ABSTRACT

The process of wafer fabrication is arguably the most technologically complex and capital intensive stage in semiconductor manufacturing. This large-scale discrete-event process is highly re-entrant, and involves hundreds of machines, restrictions, and processing steps. Therefore, production control of wafer fabrication facilities (fabs), specifically scheduling, is one of the most challenging problems that this industry faces. The reason of its high applicability in semiconductor manufacturing is due to the fact that in semiconductor manufacturing the machines used in the product line are extremely expensive and comprise 75% of the total cost of the fabrication facility. Consequently, each wafer revisits the same machines several times to produce different layers. This paper examines the optimisation solution for the operation of a small re-entrant semiconductor fab under two approaches. The first considers employing an evolutionary algorithm to the multi-objective optimisation by weighting each of the objectives in order to obtain a single objective function. This requires some a-priori or external knowledge of the relative importance of the competing objectives and results in a single solution that may be considerably sensitive to the weights. By contrast, the second uses a pareto-optimal genetic algorithm to develop a true multi-objective solution to the same problem. Here no a-priori or external knowledge is required and the decision maker is presented with a set of non-dominated solutions to assist in selection of the most appropriate solution to implement. Both solutions are developed using discrete event simulation models of the Five-Machine Six Step Mini-Fab built in ExtendSim^{TM}.

INTRODUCTION

The semiconductor manufacturing industry is facing increasing worldwide competition. It is a rapidly growing industry, due to the increasing demand of its variant products, which are widely used in most modern devices. Any technology that increases factory output, even by modest amounts, can have significant impact on profitability for manufacturers (Wang et al., 2005). As a consequence, reducing inventories, decreasing cycle time, and improving the utilisation of resources are very important issues in this industry.

Objectives like throughput time and the quantity produced (outs) must be optimized to secure the existence in a global market that is growing. Especially in the front-end of the semiconductor manufacturing process, wafer fabrication; is dominated by cluster-tools and re-entrant flows making production planning and optimisation very difficult (Klemmt et al., 2008). Hence, to improve the manufacturing planning strategies, there is an increasing demand for optimisation using simulation methods to improve the performance of bottleneck machines (Noack et al., 2008).

Effective factory scheduling and dispatching plays a key role in improving equipment utilisation, and as a consequence reducing in throughput time reduction and increasing outs. This production control function decides how wafers should be released into a fab and how they should be dispatched among machines for processing (Hsieh et al., 2007).

This paper examines the operation of a small re-entrant semiconductor fab under two optimisation approaches. The first employs an evolutionary algorithm to the multi-objective optimisation by weighting each of the objectives in order to obtain a single objective function. This requires some a-priori or external knowledge of the relative importance of the competing objectives and results in a single solution that may be considerably sensitive to the selected weights. By contrast, the second uses a pareto-optimal genetic algorithm to develop a true multi-objective solution to the same problem. Here no a-priori or external knowledge is required and the decision maker is presented with a set of non-dominated solutions to assist in selection of the most appropriate solution to implement.

Both solutions are developed using discrete event simulation models of the Five Machine Six Step Mini-Fab[1] built in ExtendSim^{TM}. The Mini-Fab model features six processing steps and five machines arranged in three stations.

The remainder of this paper is organised as follows: in the next section, a brief review of optimisation using simulation is given, followed by a short description of the Mini-Fab model developed for this work. Then, the two approaches

[1] Referred to as Mini-Fab in the remaining of text.

under study are presented. Finally, the conclusions drawn from this work are pointed out.

OPTIMISATION USING SIMULATION

Simulation based optimisation is the process of finding the best values of some decision variables for a system where the performance evaluations is based on the output of a simulation model of this system (Zeng and ZhongzhenYang, 2009). This active field of research and is also increasingly being used in practical simulation applications and being incorporated into simulation software tools (Ólafsson and Kim, 2002). It combines the simulation analysis and the optimal decision making mechanism rather than requiring the user to complete a set of simulation experiments which can take a considerable time. As a result, in the application of these methods the user can often avoid even formulating a difficult problem and thus provide the optimum solution at that instant within a limited time (Gupta and Sivakumar, 2002).

MINI-FAB MODEL DESCRIPTION

The Mini-Fab model features six processing steps and five machines distributed in three stations, each station serves two steps. Station 1 has 2 machines, each machine batches 3 lots at a time, in addition to preventive maintenance for the two machines at this station. Station 2 has 2 machines also that are subject to preventive maintenance and emergency maintenance. Finaly, Station 3 has 1 machine that serves two steps and thus requires setup to switch between different products; in addition, there is a preventive maintenance in this station. Analysis by (El-Khouly et al., 2009) showed that the bottleneck is in station 3. A full description of the Mini-Fab model can be found at http://www.eas.asu.edu/~aar/research/intel/papers/fabspec.html

The Mini-Fab model is one of the most popular systems used by researchers (Ramírez-Hernández et al., 2005, El Adl et al., 1996, Spier and Kempf, 1995, Tsakalis et al., 1997, Vargas-Villamil et al., 2003, Wang et al., 2005, Flores-Godoy et al., 1998), and different simulation models have been developed to evaluate the impact of dispatching rules on a set of predetermined performance measures.

(Flores-Godoy et al., 1998) used Throughput Time (TPT) and Work In Process (WIP); where, (Collins et al., 2003, El Adl et al., 1996) used TPT, WIP and Outs.

Also, other simulation models evaluated the impact of changing lot release policies on the Mini-Fab performance (Li et al., 2008, Wang et al., 2005).

MODEL DEVELOPMENT

A simulation model of the Mini-Fab was developed using the ExtendSim^{TM} v8.0 simulation environment from Imagine That, Inc. It is capable of running with eight different dispatching rules under differing CONWIP lot release levels.

The dispatching rules that are applied to the bottleneck station are: First In First Out (FIFO), Last In First Out (LIFO), Shortest Remaining Processing Time (SRPT), Longest Remaining Processing Time (LRPT), Earliest Due Date (EDD), Critical Ratio (CR), Least Dynamic Slack (LDS), and Shortest Setup Time (SST).

As mentioned earlier, two different approaches to finding the optimal CONWIP level/dispatching rule combination are employed. Both are based on the optimisation using simulation concept; where, the model is run repeatedly with different combinations of CONWIP level and dispatching rules. Optimisation using simulation aims to find the best combination of these two variables that will maximize or minimize specified objectives.

The two approaches presented in this section use a similar search technique that is based on genetic algorithms. However, they differ in the way the objective function is defined. The first technique; referred to as "Evolutionary Algorithm", is included in the ExtendSim simulation environment. The second; referred to as "Pareto-Optimal Genetic Algorithm", is a custom made optimizer developed by (Kernan and Geraghty, 2004) that has also been coded within ExtendSim.

Evolutionary Algorithm

An "optimizer" block is added to the model to control the search for the best set of parameters that maximizes profit or minimizes cost, within specified parameter limits. The "problem" is stated as an objective function or cost equation that which the optimizer evaluates as it changes the parameters by evolution of the current settings. In this manner, the solution space is explored to minimize or maximize the objective to save you going through the tedious process of manually trying different values with each model run.

For such genetic optimizers it is recommended that enough generations are used to achieve optimum/near optimum solution. Yet, optimisation is not foolproof and as any other optimizer can converge to a sub-optimum solution.

The problem is formulated as an integer programming model. The decision variables, objective function, and constraints are defined as follows:

Decision Variables
There are two decision variables, the dispatching rule, and the CONWIP level to be used. The solution space for optimisation is determined by the values of each of these two variables.

Multi-Objective Function
The objective function is to minimize the average Throughput Time (TPT) and maximize the average Outs. Average TPT and average Outs are conflicting in nature, meaning that, by increasing the CONWIP level the Outs performance improves and the TPT performance worsens and vice versa.

Here, a utility function is used to represent the multi-objective nature of this problem where, weights are given to the different objectives; TPT and Outs. The two objectives are measured as a percentage of their best achievable values to overcome the fact that they are of different units of measure; where, the average TPT is measured in minutes and the average Outs is measured in lots per week.

The utility function is given by equation 1, and maximizing the value of the function is realised by decreasing the actual TPT, increasing the actual Outs or both together.

$$\text{Max}\left\{\frac{\frac{TPT'}{TPT}w_1+\frac{Outs}{Outs'}w_2}{w_1+w_2}\right\} \quad (1)$$

Where,

- w_1 and w_2 are the weightings given to TPT and Outs respectively.
- TPT' is the best (minimum) achievable TPT.
- TPT is the actual throughput time.
- $Outs'$ is the best (maximum) achievable Outs.
- $Outs$ is the actual Outs.

It should be noted that TPT' and Outs' represent the ideal values for throughput time and outs without any detractors. The detractors in the Mini-Fab modelled in this work include preventive maintenance, emergency repair, and setup times. Thus, the model has been run excluding all these detractors using the 8 dispatching rules and different CONWIP levels.

The minimum TPT is achieved at CONWIP level 3 and the maximum Outs per week is achieved at CONWIP level 21. The best achievable TPT is 1,055 minutes and the best achievable Outs is 96 lots per week. Then, the multi-objective function is re-written as in Equation 2.

$$\text{Max}\left\{\frac{\frac{1055}{TPT}w_1+\frac{Outs}{96}w_2}{w_1+w_2}\right\} \quad (2)$$

Constraints

The constraints defined are used to limit the solution space where the optimum solution is searched for. These constraints are set for the CONWIP level and for the dispatching rules. The batching process at station 1 results in better performance at CONWIP levels that are in multiples of the batch size (3 lots). This is shown in Figure 1.

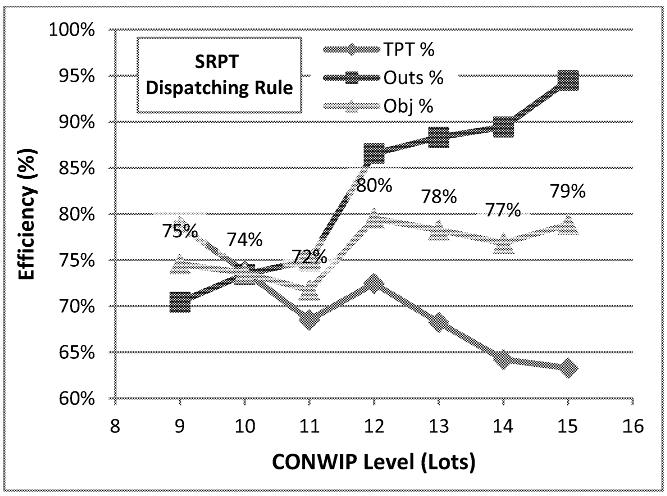


Figure 1: Effect of CONWIP level on the performance.

The value of the objective function is higher at CONWIP levels 9, 12 , and 15. For that reason the constraint of CONWIP level in multiples of 3 has been added.

Therefore, the constraints defined for this model are that the CONWIP level should be in multiples of 3 and should not be smaller than 9 or greater than 21 lot. In addition, dispatching rules start with FIFO and ends with SST (from 1 to 8). Finally, all decision variables are integers.

Optimisation Results

The best result using the "Evolutionary Algorithm" optimizer is a CONWIP of 12 lots combined with the Shortest Remaining Processing Time (SRPT) dispatching rule as shown in Figure 2. This results in an average utility objective function of 79.7%.

This result has been reached after 190 samples and 28 generations; hence, on average 7 samples per generation. Multiple samples are needed for this model due to the stochastic nature of the utility function variables.

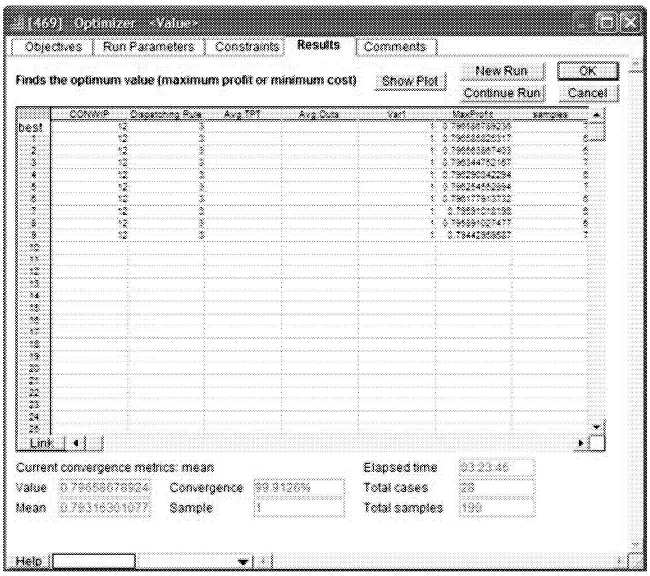


Figure 2: The Results Tab of The Evolutionary Optimizer.

It should be noted that because both response variables were given the same weight, the optimizer selected a CONWIP value that resulted in a greater improvement of TPT than the degradation of the performance of Outs.

If the decision maker is not satisfied with the results and considers TPT or Outs of greater importance; then, the weights assigned to each objective should be changed accordingly and the model should be run all over again. The weightings used will be based on the knowledge of those familiar with the operational performance of the system and the relative importance of one measure over the other.

Pareto-Optimal Genetic Algorithm

A pareto-optimal genetic algorithm is used to develop a multi-objective solution to the same problem discussed previously. Here no a-priori or external knowledge is required and the decision maker is presented with a set of non-dominated solutions to assist in selection of the most appropriate solution to implement.

A "Pareto" block is added to the model as shown in Figure 3, it is a multi-objective Pareto optimal algprithm that uses trial-based-recombination and exo-parental mutation. This block can optimize up to 6 objectives from a decision space of up to 8 variables.

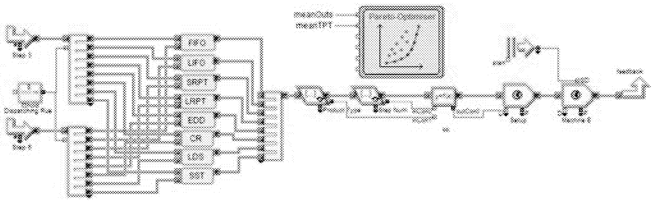


Figure 3: The Model Developed Using The Pareto Block.

Again, the problem is formulated as an integer programming model; where, the decision variables and constraints are as mentioned above. However, the objective function is defined differently as follows:

Multi-Objective Function

The objective function is to minimize the average Throughput Time (TPT) and maximize the average Outs. This is done by connecting the values of the two objective results from the model to the "Pareto" block, and then selecting "Maximise" for objective 1 which represents the average Outs, and "Minimise" for objective 2 which represents the average TPT, in the "Algorithm Setup" tab of the "Pareto" block as shown in Figure 4.

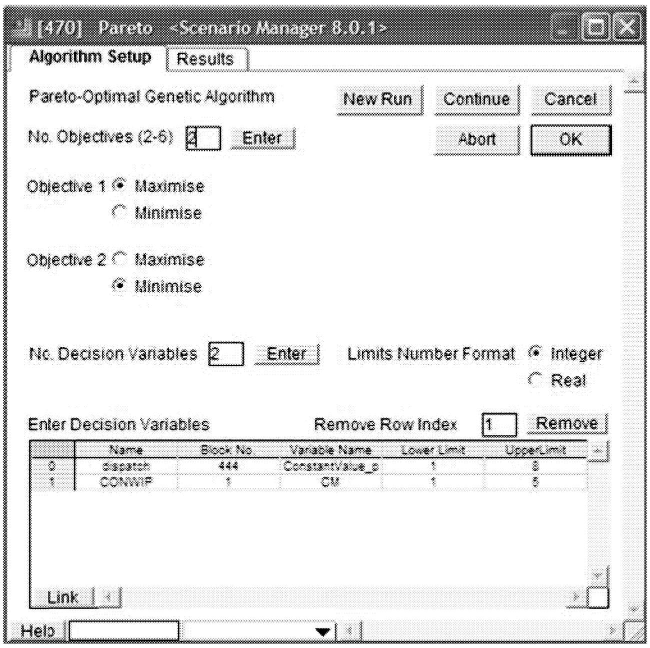


Figure 4: The Algorithm Setup Tab of The Pareto Block.

Optimisation Results

In the Pareto optimal genetic algorithm there is no one best solution, rather a set of solutions is presented together with summary information to assist the user in the selection of the most appropriate.

After running the model with 4 generations for 20 replications, 35% of the space had been searched. The results of this run are reported in Table 1.

Table 1: Results of The Pareto Optimal Genetic Algorithm.

Dispatching Rule	*CONWIP*	*Average Outs Efficiency %*	*Average TPT Efficiency %*	*Average Objective Function %*
EDD	9	70.5%	78.7%	74.6%
SRPT	12	86.6%	72.5%	79.5%
FIFO	18	95.5%	53.3%	74.4%
EDD	21	95.8%	45.8%	70.8%

As presented in Table 1 there are four solutions that the decision maker can choose from. These are presented in Figure 5. In case that both objectives have the same weight, then, the best solution is a combination of the SRPT dispatching rule with CONWIP level of 12 lots, which results in the highest objective function of 79.5%. This is the same solution obtained using the Evolutionary Algorithm.

This technique offers the decision maker a greater flexibility in determining the alternative that best suits his/her requirements. For example, if the Outs is of greater importance than the TPT the decision maker can select the FIFO dispatching rule with CONWIP 18 lots, which results in increasing the Outs efficiency from 86.6% to 95.5%; howerever, TPT efficiency decreased from 72.5% to 53.3%.

On the other hand, if the TPT is of greater importance than the Outs; then, the EDD dispatching rule with CONWIP 9 lots should be selected as this will result in increasing TPT efficiency.

Hence, if the objectives are to be given weights, the user can explore the impact of the different weights on the optimal solution without running the whole analysis again. This results in considerable time saving as the search space is considerably smaller than the original problem.

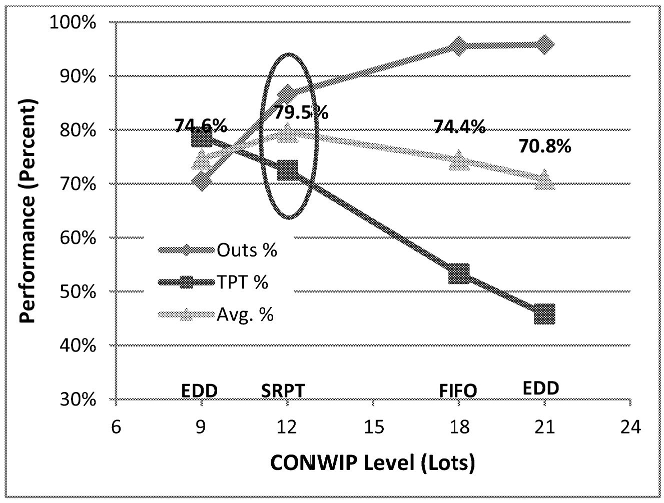


Figure 5: Effect of Changing The WIP Level on The Objectives

COMPARISON OF THE TWO APPROACHES

This work examines the optimisation solution for the operation of a small re-entrant semiconductor fab under two approaches. The first considers employing an evolutionary algorithm to the multi-objective optimisation by weighting each of the objectives in order to obtain a single objective function and the second uses a pareto-optimal genetic algorithm to develop a true multi-objective solution to the same problem. Table 2 summarizes the differences between the two approaches based on four different characteristics.

First, the elapsed time taken to arrive at the solution is considered. In the first approach a convergence of 99.9% is reached after 3 hours, 23 minutes, and 46 seconds, which is a long time relative to the second approach that requires only 1 hour and 20 minutes.

Second for consideration is the form of the output; a single best solution from the first approach while the second approach provides a set of solutions from which the decision maker selects the best.

Third, the formulation of the multi-objective function, in that the first approach requires the use of a utility function to combine the two competing elements into a single objective function, whereas the Patero Optimizer block is designed to support up to 6 different objectives.

Finally, the number of generations used to find the solution gives an indication of how long the search algorithm will take. In the first approach 28 generations are used to find the best solution. In contrast, in the second approach only 4 generations are used.

Table 2: Differences Between The Two Approaches.

Comparison Point	*Evolutionary Algorithm*	*Pareto-Optimal Genetic Algorithm*
Simulation Run Time	Long	Relatively Short
Solution Obtained	A single best solution	A set of solutions, and the decision maker selects one
Multi-objective Function	Only one objective; hence, the need for the utility function	Up to six different objectives
Number of Generations	28	4

CONCLUSION

Optimisation using simulation has advantages and disadvantages. One of the key advantages is that it combines the flexibility of simulation with the intelligence of optimisation without requiring a detailed derivation of a mathematical model. On the other hand, one of the observed disadvantages is that the optimisation strives to find the best possible solution within a given space of the solutions.

The results of the first approach confirmed that combining the SRPT dispatching rule with the CONWIP level of 12 lots, which results in an average utility objective function of 79.4%. This is done under the assumption that both the TPT and the Outs are of the same importance and thus are given the same weight.

The results of the second approach in case that both objectives have the same weight, then, the best solution is a combination of SRPT dispatching rule with CONWIP 12 lots, which results in the highest objective function of 79.5%. This is the same solution obtained using the Evolutionary Algorithm. However, this technique offers the decision maker greater flexibility in determining the alternative that best suits his/her requirements. Therefore, one can select a solution without running the whole model again, hence, it is considered to be time saving.

In conclusion, the two approaches are similar to each other in the search technique applied, the only difference is the objective function, where there is only one objective that can be optimized in the first approach; hence, there is a need for a utility function, while, in the second approach there is up to 6 objectives that can be optimized.

REFERENCES

Collins, D. W., Flores-Godoy, J. J., Tsakalis, K. S. & Hoppensteadt, F. C. (2003) Diffusion bay simulation and its impact on the overall FAB performance: a simplified example. *IEEE International Symposium on Semiconductor Manufacturing.*

El-Khouly, I. A., El-Kilany, K. S. & El-Sayed, A. E. (2009) Modelling and Simulation of Re-Entrant Flow Shop Scheduling: An Application in Semiconductor Manufacturing. *International Conference on Computers and Industrial Engineering.* Troyes-France IEEE.

El Adl, M. K., Rodriguez, A. A. & Tsakalis, K. S. (1996) Hierarchical modeling and control of re-entrant semiconductor manufacturing facilities. *The 35th IEEE on Decision and Control.*

Flores-Godoy, J. J., Yan, W., Collins, D. W., Hoppensteadt, F. & Tsakalis, K. (1998) A Mini-FAB simulation model comparing FIFO and MIVP(R) schedule policies (outer loop), and PID and H∞ machine controllers (inner loop) for semiconductor diffusion bay maintenance. *The 24th Annual Conference of the IEEE Industrial Electronics Society, 1998. IECON '98.*

Gupta, A. K. & Sivakumar, A. I. (2002) Simulation Based Multiobjective Schedule Optimization in Semiconductor Manufacturing. *Winter Simulation Conference.*

Hsieh, B.-W., Chen, C.-H. & Chang, S.-C. (2007) Efficient Simulation-Based Composition of Scheduling Policies by Integrating Ordinal Optimization With Design of Experiment. *IEEE,* 4**,** 553-568.

Kernan, B. & Geraghty, J. (2004) A Multi-Objective Genetic Algorithm for ExtendSim. *First Irish Workshop on Simulation in Manufacturing, Services and Logistics.* Limerick, Ireland.

Klemmt, A., Horn, S., Weigert, G. & Hielscher, T. (2008) Simulations Based and Solver Based Optimization Approaches for Batch Processes in Semiconductor Manufacturing. *Winter Simulation Conference*

Li, S., Xiaohui, Z. & Li, L. (2008) Simulation and analysis of scheduling rules for semiconductor manufacturing line. *IEEE International Conference on Industrial Technology, ICIT 2008.*

Noack, D., Gan, B. P. & Lendermann, P. (2008) An Optimization Framework For Waferfab Performance Enhancement. *Winter Simulation Conference.*

Ólafsson, S. & Kim, J. (2002) Simualtion Optimization. *Winter Simulation Conference.*

Ramírez-Hernández, J. A., Li, H., Fernandez, E., Mclean, C. & Leong, S. (2005) A framework for standard modular simulation in semiconductor wafer fabrication systems. *Winter Simulation Conference.*

Spier, J. & Kempf, K. (1995) Simulation of emergent behavior in manufacturing systems. *Advanced Semiconductor Manufacturing Conference and Workshop, 1995. ASMC 95 Proceedings. IEEE/SEMI 1995.*

Tsakalis, K. S., Flores-Godoy, J. J. & Rodriguez, A. A. (1997) Hierarchical modeling and control for re-entrant semiconductor fabrication lines: a mini-fab benchmark. *The 6th International Conference on Emerging Technologies and Factory Automation. ETFA '97.*

Vargas-Villamil, F. D., Rivera, D. E. & Kempf, K. G. (2003) A Hierarchical Approach to Production Control of Reentrant Semiconductor Manufacturing Lines. *IEEE Transactions on Control Systems Technology,* **11**, 578-587.

Wang, Z., Qiao, F. & Wu, Q. (2005) A New Compound Priority Control Strategy in Semiconductor Wafer Fabrication. *IEEE.*

Zeng, Q. & Zhongzhenyang (2009) Integrating Simulation and Optimization to Schedule Loading Operations in Container Terminals. *Computers and Operations Research*, 1935-1944.

BIOGRAPHY

INGY A. EL-KHOULY joined the Arab Academy for Science, Technology & Maritime Transport (AASTMT) in 2006. Her Master research work (AASTMT, 2009) included Multi-Objective Optimization Using Simulation For Scheduling Re-Entrant Flow Shop In Semiconductor Wafer Fabrication Facilities. She is an Assistant Lecturer; Department of Industrial and Management Engineering. Her e-mail address is <ingyelkhouly@aast.edu>.

KHALED S. EL-KILANY joined the Arab Academy for Science, Technology & Maritime Transport in 1998. His Ph.D. research work (Dublin City University, 2004) included modelling and simulation of automated material handling system of Intel's wafer fabrication facilities. His research interest lies in the analysis and improvement of manufacturing systems performance using simulation. He is an Associate Professor and Chairman; Department of Industrial and Management Engineering. His e-mail address is <kkilany@aast.edu>.

JOHN GERAGHTY joined Dublin City University in 2003 after completing his PhD thesis at the University of Limerick on the performance of several production control strategies, such as Kanban, CONWIP, Hybrid Kanban-CONWIP etc, in the presence of uncertainty in customer demand. Current research interests include Operations Excellence and Supply-Chain Management. His e-mail address is <John.Geraghty@dcu.ie>.

PAUL YOUNG joined Dublin City University in 1998. Following his PhD (Dublin University, 1991) on monitoring the turning process, he worked on vehicle NVH research. Since 1995 he has been active in manufacturing research looking at process monitoring and simulation of the manufacturing system. He is a founder member of the Enterprise Process Research Centre in DCU. His e-mail address is <Paul.Young@dcu.ie>.

INDUSTRIAL PLANT SIMULATION

DESIGN VALIDATION OF AN AUTOMATIC AND FLEXIBLE STEEL FABRICATION FACILITY BY USING DISCRETE EVENT SIMULATION

Antonio Avai
Dario Pietraroia
Technology Transfer System s.r.l.
20131 Milan
Italy
E-mail: avai@ttsnetwork.com
E-mail: pietraroia@ttsnetwork.com

Alessandro Pirovano
Marco Aspesi
Ficep s.p.a.
21045 Gazzada Schianno
Italy
E-mail: alessandro.pirovano@ficep.it
E-mail: marco.aspesi@ficep.it

KEYWORDS

Steel manufacturing, discrete event simulation, design validation, decision support system.

ABSTRACT

This paper explores the use of discrete event simulation as a method to validate the early design of an automatic and flexible steel fabrication facility. Even though simulation has been used as a design and validation tool in various production and business applications, the steel fabrication sector is still almost unexplored. Anyway, the design of a new facility is very challenging because the process is characterized by the uniqueness of steel products and a high mix. Consequently, simulation, accounting for the complexity and dynamic behaviour of actual systems, can bring relevant benefits. This work is based on a real case study and the results, reported in this paper, are very promising and encourage the adoption of simulation as a usual procedure in the design of such facilities.

INTRODUCTION

The steel fabrication process refers to the production of steel pieces, according to the steel engineer's design, carried out by steelworker companies. They transform raw materials into subassemblies used to build up civil or industrial buildings, electric steel towers, bridges, petrochemical field, etc... Structural steel is largely fabricated off-site then erected and assembled on-site.
The complexity of the steel fabrication process is due primarily to the low repetitiveness in production and a wide range of products. There is a large variety of steel pieces, in terms of their geometry and their processing requirements. However, the total production volume of each part is usually small. This characteristic differentiates the steel fabrication process from most other manufacturing processes in which identical products are produced in great quantities (Song and AbouRizk 2003).
The design of a steel fabrication facility is challenging given the multitude and the big differences of steel components and it requires an effective tool to estimate the performances of new systems. Discrete event simulation technology has been widely used to model production lines (Roser et al. 2003) and to assess their overall performances as well as their behaviour (Boër et al. 1993, Huang et al. 2004, Voorhorst et all, 2008). Nevertheless, the application of this technology in the steel fabrication field is very uncommon and its benefits and potentialities are still unexplored.
This paper describes the experience and findings in using discrete event simulation as a tool to support and validate the design of a fully automated and customized steel fabrication facility.
In the remainder of this paper, a short description of the case study goals as well as the facility is presented. The next paragraph investigates some modelling issues. Then, the results about the design validation are described. Finally, conclusions and future work are discussed.

CASE STUDY OBJECTIVES

The case study deals with the early design of a new customized steel fabrication facility for a steel worker company located in Eastern Europe. This company manufactures steel products in power engineering industry and telecommunication, supermarkets, trade centers, etc...
The steel products, worked with this facility, are construction beams that are structural elements, made of steel, used in steel-frame buildings and bridges. They are characterized by profile (the shape of the cross-section), length and material.
The design process was carried out by Ficep. It is a global leader in designing and manufacturing turnkey systems in the processing of flats and profiles in the structural steel industry. It is one of the few manufacturers in the world producing a complete range of leading edge machines for the steel construction.
One of the main challenges during the design phase is to assess the overall performance of a new layout, mainly the throughput. It depends on many factors both quantitative and qualitative such as:

- The layout type: for instance, the distance between different systems and their placement, the number of loading devices in a plant, etc...
- The number and the type of processing systems.
- The equipment of each processing system and their performances.
- The automation level of the handling system.
- The production mix: usually it is very high and each beam has its own cycle time and can be processed on one or more systems.
- Some operational policies, such as scheduling and dispatching rules, scraps management, etc...

Discrete event simulation is particularly useful in predicting the overall performance of complex systems with an inherent stochastic nature such a steel fabrication facility.
The main goal of this case study was to use the simulation technology during the early design of a new customized facility to support the design team in finding out a solution able to meet the customer requirements, principally, in terms of productivity. Another goal was to exploit the animation, provided by a simulation model, to show Ficep's customer the behaviour of the proposed layouts in order to get immediate feedback, suggestions with the intention to speed up the presale phase. The more satisfied the customer is, the higher the chance to win the contract is.

OVEARLL FACILITY DESCRIPTION

The first layout for this use case was composed by 5 fully automatized processing systems as outlined in Figure 1:

1. A shot blasting system: batches of beams go through this system to be cleaned.
2. A flexible Computer Numerical Control (CNC) sawing system: it cuts the raw materials (beams) in shorter components according to steel engineer's design.
3. A flexible and combined CNC drilling and sawing system (see Figure 2).
4. A CNC drilling system equipped with a linear tool changer.
5. A CNC thermal cutting system: it carries out shaped cuts on profile using a robot with 6 degrees of freedom. Technically, such a process is called "*Coping*".

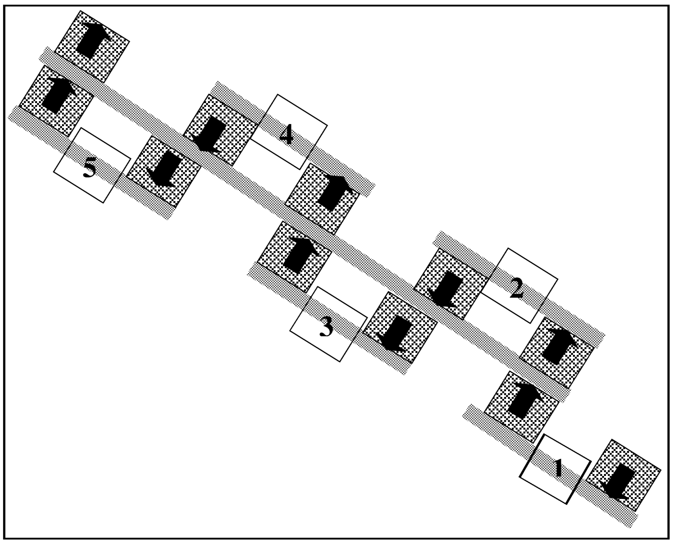


Figure 1: A draft of the layout and its physical flow

Each system, as outlined in Figure 3, is composed by an automatic CNC machine, a roller conveyor passing through the machine and two cross transfer devices: one for feeding the machine with beams and the other for unloading worked parts. All the five systems, above mentioned, are joined together by a central roller conveyor. This is also used to transfer the steel produced components to the unloading device. This facility is really a fully and flexible automated processing system. Beams as well as components can be transferred and processed without any worker intervention. Only loading and unloading operations from this facility are carried out by workers using bridge cranes. In order to save time, beams are packed into batches whose width has to be shorter than the roller conveyor width crossing the automatic shot blasting machine.

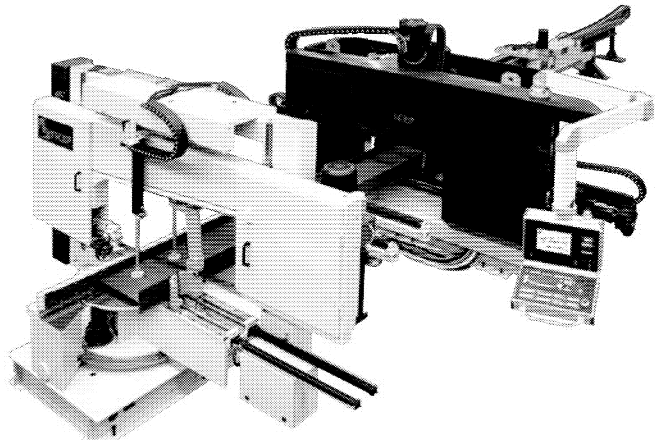


Figure 2: A combined drilling and sawing machine

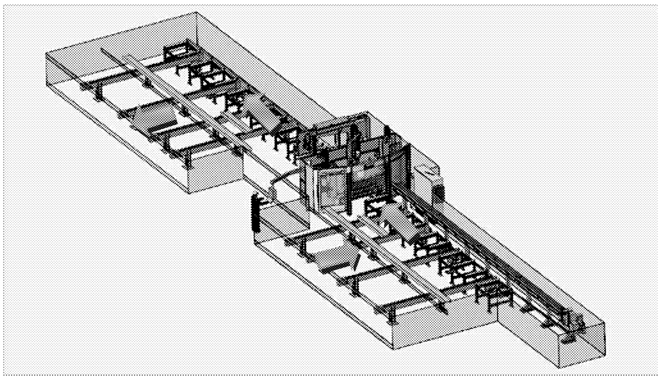


Figure 3: A processing system scheme

KEY MODELLING AND DEVELOPMENT ISSUES

This paragraph describes all the relevant aspects analyzed during the modelling and development phases.
Each simulation model, presented in this paper, was developed in Arena (Kelton et al. 2003), and it is driven by:

- A work order provided by the customer.
- A set of input parameters, selected in cooperation with Ficep, and stored in an Excel file. The goal is to facilitate the sensitivity analysis on some aspects that can be easily changed during the early design phase.

Automatic CNC systems

The modelling and simulation of each automatic CNC system settled on the plant is the core of the model. It is crucial to simulate them with the right level of detail in order to produce correct and reliable estimations.
Because of the production management solution, provided by Ficep, can assess the processing time of each beam in

every CNC system, the modelling phase was focused on the parts movements along with the resources needed to accomplish them, the relative rules and the constraints. In addition, to get more accurate estimations, the system availability was modelled using stochastic distributions fitting the historical data about similar systems.
Another topic aspect to be modelled, mainly for sawing systems, was the scrap and short component management. These parts cannot be automatically handled by the central roller conveyor, but they have to be managed in an opportune way, based on the system configuration. An efficient solution is to equip a sawing system with an automatic unloading shuttering device, placed after the machine (see Figure 4), along with a magnetic gripper able to grasp components of any length. In this way, short parts as well as scraps are automatically conveyed to this device and collected in a special bin that, when it is full, is removed from the plant and replaced by an empty one.

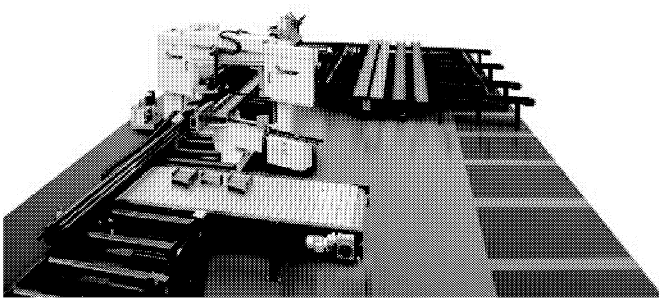


Figure 4: A sawing system equipped with an automatic unloading shuttering device and a magnetic gripper

Cross transfer devices

These devices feed and unfed each working system as well as the whole facility. All the devices settled up in this plant work in a similar way: a single part is being transferred onto the roller conveyor of a CNC machine if free; otherwise it is piled up closer to the previous beam already stored on the device. After a part has completed its movement, the next one starts moving. The simulation of such devices was more complicated mainly because Arena suite has not any feature so suitable for modelling them, taking also into consideration the level of required animation. Each cross transfer device was simulated using a conveyor module and defining several stations, very close between them. Before starting the transfer operation, the model calculates the arriving station based on the beam width and the available room on the transfer device. Then it was modelled the synchronization of the beam movements on each device and the priority between the three following processes:

1. The loading into the transfer device
2. The beam movement on the cross transfer device
3. The unloading operation from the transfer device

In the end, the logic of the transfer devices was validated with the help of Ficep designers.

Routing rules

Generally, the loaded batches can have multiple processing paths, for instances each batch can be completely worked on a combined system or, in alternative, first the raw material is cut on a sawing machine and, then, is drilled on a drilling one. The simulation model has to represent the routing rules, the selection of the processing path for each batch. In the actual system, this decision is taken by the overall PLC after a batch has been shot blasted and before engaging the central roller conveyor. In between, there is a cross transfer device working as a buffer where batches are piled up until the selected system is available.
At this time, the routing rules for the analyzed solution were not still well defined and, moreover, they are influenced by the layout configuration. Anyway, the designers suggested considering the following conditions:

- The workload of each system
- The available room on the relative infeed device

The second condition avoids blocking the central roller conveyor if the infeed device of the selected system is full of parts. The goal of these rules is to keep all the systems as busy as possible.
The routing rules were implemented in the model and they were also improved thanks to the animation as well as the simulation results.
Clearly, if a batch can be worked only on a specific system, it is dispatched directly to its destination if the infeed cross transfer device has enough room; otherwise it is blocked until the device has space.

Animation

An accurate animation is an added value because the model will be shown to the potential customer to get an immediate feedback on the system behaviour. This aspect influenced the modelling phase in order to show all the beams' movements across the whole plant as well as the processing operation such as the sawing one. In the model when a beam is cut, a new component, with the right length, is created and it moves out of the CNC machine while the original beam becomes shorter. All these calculations are made by some procedures written in VBA. Unfortunately, Arena does not allow creating entities' pictures automatically, obliging to design them manually. The animation was refined few times based on the designers' suggestions to make it clearer and more effective.

Report generation

In collaboration with the designers, it was created a customized report in Excel format automatically filled in by the simulation model when every replication is over. This graphical report includes some Key Performance Indicators (KPIs) at the plant level, such as: the overall productivity, the average Work In Process (WIP). It also considers some KPIs at the system level such as: the saturation rate of a machine and its infeed and outfeed cross transfer devices, the average system productivity, the overall processed tons, etc... Moreover, it shows the input parameters used to run the model.

DESIGN VALIDATION AND RESULTS

The design validation process has been carried out in tight cooperation between the design and the simulation teams. First, the Ficep design team draws a layout, then it is simulated by the simulation expert and the results are discussed together to check for problems and feasible improvements. The approved changes are implemented in the simulation model and in the layout drawing, if accepted. The new solution is simulated again until one or two alternative layouts are able to meet the targets in terms of overall performances and geometrical dimensions. The latter has to comply with the physical constraints of the customer's shop floor. Usually the key performance indicator is the average throughput measured as tons per hour (t/h) and not as the amount of components produced due to the low degree of repetitiveness.
In this use case, the target was 5 t/h considering a production mix to build a power station. Because the production mix impacts on the overall performances, the targeted throughput can be accounted only after having defined the final product, the civil or industrial building to be erected.
Each run starts with an empty system and ends when all the worked components are unloaded from the facility.

Simulation model validation

The validation of the model was carried out according to various techniques (Sargent 2009), by simulation and design teams. The main goal of this phase was to check if the model could generate valid productivity assessment under different conditions.
During the model development, the design and process experts from Ficep were asked if the logics implemented in each processing line were corrected. Then, thanks to the animation, the operational behaviour was displayed graphically as the model moved through the time. The design team could check again all the movements of the parts as well as the equipments. Because neither the real system nor all the rules to manage the part movements existed, some logics were implemented according to the know-how of the system engineers from Ficep.
Furthermore, the behaviour of different types of beams was traced through the model to determine if the logic was correct and if the needed accuracy was obtained.
Once the model was completed, some stressing tests considering very simple scenarios as well as the original work order were made to check if the stochastic results about productivity, resources saturation and WIP were reasonable. These tests allowed debugging the models again.
Eventually, several replications considering the same work order were made showing a small amount of variability. It was discovered that twenty replications are enough to get a 10% of relative precision with an α value of 0,05 (Chung 2004).

Initial facility solution

The simulation of the first layout confirmed the potential problem already pointed out by the designers: the outfeed cross transfer of the drilling system engages two different sections of the central roller conveyor and one of these parts is also used for feeding the input device of the same system. This slows down the performances because sometimes a component has to wait before being transferred to the central conveyor. But, most of all, the system can experience a deadlock if appropriate synchronization mechanisms are not implemented.
Small changes were applied to the initial drawing to overcome this potential problem and the assessed average productivity was 4,7 t/h, a little lower than the target. The simulation run lasted 86 hours, about one production week with two shifts per day. Looking at the Figure 5 the system with the highest saturation rate (the blue box) is the thermal cutting one. A system is considered busy not only while it is processing a beam but also during the components handling. Its impact is discussed later in this paragraph.
The saturation rates of the two sawing systems, the first two bars in Figure 5, are very similar, confirming the routing rules implemented in the model work well (in the figure, "N.A." means "system Not Available").
The saturation of the shot blasting line is not represented because all the batches pass through this system without stopping their movement.

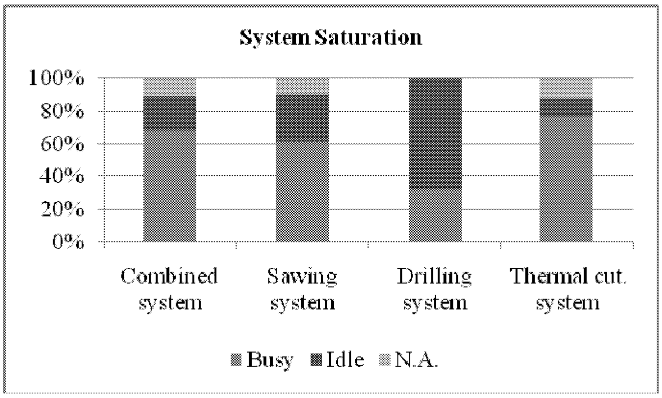


Figure 5: The overall saturation rate of each system

An important issue in the early design phase is the identification of the correct equipment of each processing system and mainly of the sawing one. It depends on many factors and, among them, also on the type of worked products and their mix. The simulation allows to validate this issue and to estimate its impact on the overall performance. For this reason, a hypothetical layout was simulated where all the sawing systems were configured without the equipment (see Figure 4) to manage automatically scraps and short components. Such components are too short and they cannot handle by the central roller conveyor. They have to be removed manually or with the help of a bridge crane if too heavy for a worker.
The average productivity decreased about 18% compared to the initial solution; this percentage is similar to the productivity of a single sawing system. The bottleneck in this hypothetical layout is the combined system as represented in Figure 6.

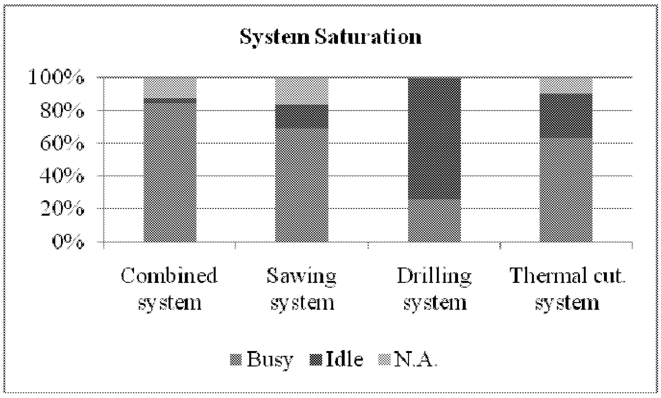


Figure 6: Saturation rate of the hypothetical layout

For this case study, these results confirm it is worthwhile to equip all the sawing systems with such devices as designed in the initial layout.
Coming back to the first solution, the bottleneck is the thermal cutting system. The basic idea to break it was to reduce its workload changing only an operational rule. Usually this system can saw beams and cope components, but the cutting process can be made by the other sawing systems already available in the plant.

Enhanced solution

In this solution, the thermal cutting system makes only coping operations when possible, otherwise it makes all the processing sequences. This solution could increase the traffic in the system, but generally a sawing machine is faster than the thermal cutting system to cut the same beam.
This change was simulated but the result was not completely satisfactory. The throughput is 4,9 t/h and its 95% confidence interval ranges between 4,8 and 4,94 t/h. The throughput is very close to the target. The average WIP is increased, as expected, of about 18%.

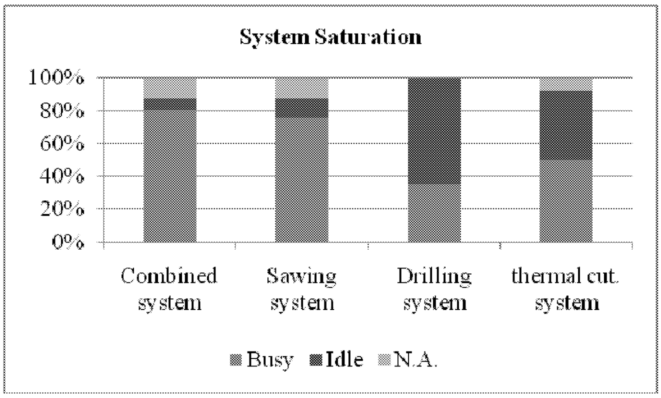


Figure 7: The saturation rate of the enhanced solution

Figure 7 shows the most saturated system is now the combined drilling and cutting one, the first bar in the graph, while the thermal cutting system saturation is decreased.
It was carried out a two model comparison according to the Welch Confidence Interval approach (Chung 2004) to check if the two solutions are statistically different. The confidence interval with an α level of 0,05 is (-0,18; -0,28) and because of this interval does not cover the 0, the solutions are diverse.
Then a sensitivity analysis was made to check if any change of some key input parameters could improve the overall performances reducing the idle period of the most saturated systems. In accordance with the design team, three input parameters were investigated:

A. The performance of the shot blasting system because it feeds the other processing lines.
B. The central roller conveyor speed because it couples all the processing systems.
C. The system availability because it has a direct impact on the productivity.

The 2^k experimental design was selected to reduce the number of alternatives to simulate. Twenty replications were required for each combination to obtain a relative precision of 0,1 at an α value of 0,05. The graph in Figure 8 shows throughout really goes up only when the system availability is increased (factor C+). While considering the factor C-, meaning the availability is unchanged, just the combination A+ B+ C- (the fifth one) allows to reach the target even though its 95% confidence interval is between 4,94 and 5,05 t/h.

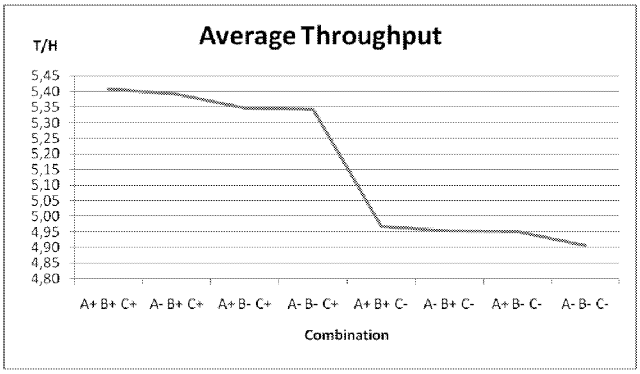


Figure 8: Average throughput vs combinations

Figure 9 represents the average WIP versus the analyzed combinations. This trend allows proofing the model correctness: the lowest WIP is achieved decreasing the shot blasting system performance, the entry point of this plant, and increasing the speed of the central roller conveyor.

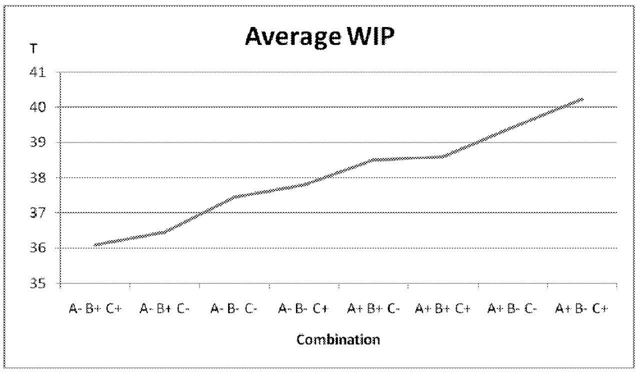


Figure 9: Average WIP vs combinations

Summarizing the results of these experiments, the conveyor speed as well as the shot blasting performance does not impact on the throughput but only on the WIP, while, as expected, the machine availability influences the performances meaningfully. These conclusions are also confirmed by the ANOVA test, performed on the alternative means of the throughput considering a single factor. The results are in the Table 1, where "SS" means "Sum of Square", "MS" is for "Mean of Square", "F" is the result of the test and "P" is the critical value.

Table 1: Anova Analysis

Factor	SS	MS	F	P
Shot blasting performance	0,02	0,02	4,8	4,1
Central conveyor speed	0,02	0,02	4,0	4,1
System availability	1,9	1,9	486	4,1

The F test statistic is much greater than the critical F value (P) only for the last factor, the system availability. Unfortunately, it is challenging to get good and actual estimations about this factor. It also depends on the operational procedures and the maintenance policy of the customer. So it was decided to design a new layout starting from the results and the findings of this simulation model.

New solution

The results of the previous model have mostly pointed out the bottleneck is the sawing capacity and the saturation rate of the drilling machine is very low. For this reason, in the new solution, the drilling system was replaced by a combined one. The cutting capacity is increased while the drilling one is unchanged. The throughput is 5,7 t/h, that is higher than the required one. Its interval of confidence is (5,64; 5,74). The average WIP is decreased about 4%, compared to the previous solution, as well as the average saturation of all the infeed cross transfer devices.
Figure 10 illustrates this new layout is well balanced; the saturation rate of the four systems is very similar and it close to 70%.
Clearly, this solution performs much better than the previous one both in terms of productivity and WIP, but also the investment is bigger. Nevertheless, the cost-benefit analysis is out of the scope of this paper.

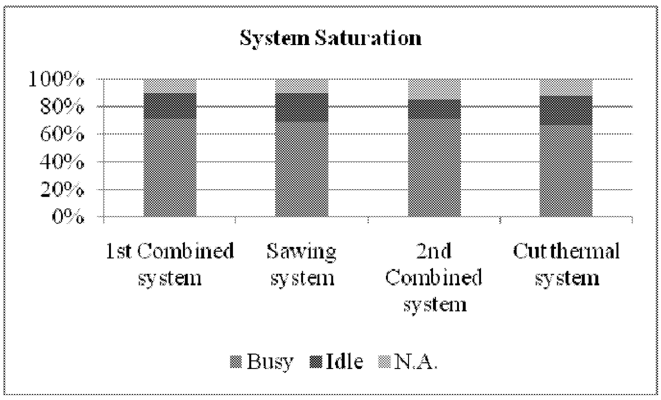


Figure 10: The saturation rate of the new layout

THE LESSON LEARNED

The introduction of the simulation in the design process really brought a relevant benefit. It could solve a pressing and important issue for the design team: the key performances' assessment, in a quantitative way, of a steel processing facility. Furthermore, this technique allowed supporting some designer decisions helping to reduce the number of drawings and to speed up this phase.
At the customer site, the animation provided by the simulation model, helped Ficep salesman to explain the facility behaviour as well as to justify the investment. At the very beginning, the customer was sceptical because it was not so familiar with simulation. After having shown in details all the steps needed to get the simulation results, the bad feeling was overcome. Then, the model acted as a common base to discuss the layout configuration, to test the impact of some changes of the process parameters to the performances, like the width and the speed of the shot blasting line. By using the model, Ficep salesman could assess the productivity of the future system and explain how to achieve this performance, how to load it, how to manage the workload of each processing system and so on. Eventually, the customer was so impressed by the simulation to start reasoning on how to use it in production.
A crucial aspect of achieving suitable results quickly was the tight cooperation between the two teams and their experiences. The simulation team could identify potential problems fast and easily and, sometime, offer some suggestions. The design team could translate them into technical and practical solutions.
This experience and the gained results encourage using the simulation as a usual routine during all the design phases of a new facility.

CONCLUSIONS

This paper explores the use of simulation to support and validate the design of a steel processing facility, mainly to test if the assessed throughput achieves the target. Thanks to the simulation findings and, of course, the designers' know-how, two alternative solutions were found out and were discussed with the Ficep's customer.
Simulation can be a powerful method also during the detailed design of the solution selected by the customer in order to tune up operational parameters and/or to support the design of rules to handle parts across the whole system. Thinking a little forward, the simulation could be useful for the final customer: the steelworker company. The model developed in the design phase, after some refinements, could validate the daily production plan, or it could support the constant improvement process.
Further activities will be focused on investigating the impact on the layout performances and/or configuration of:

- Different production mixes needed to build up similar but not identical final products.
- Different batch loading plans for the whole system of the same work order.

REFERENCES

Boër, C.R., A. Avai., J. El-Charr. And E. Imperio. 1993. "Computer Simulation for the Design and Planning of Flexible Assembly Systems". *Proceedings of International Workshop on Application and Development of Modelling and Simulation of Manufacturing Systems, Beijing*

Chung, C.A, 2004. *Simulation modelling handbook*, Crc Press

Huang, C.N., Y.J. Seong and J.S. Russel, 2004. "Time study on two echelon supply chain". The 12th Annual Conference on Lean Construction (IGLC), 2004

Kelton, W.D., R.P. Sadowski and D.T. Sturrock., 2003. Simulation with Arena, 3rd Edition, WCB/McGraw-Hill, New York, NY.

Roser, C., M. Nakano and M. Tanaka. 2003. "Buffer allocation model based on a single simulation". *Proceedings of the 2003 Winter Simulation Conference*, 1230-1246

Sargent, R.G. 2009. "Verification and validation of simulation models". *Proceedings of the 2009 Winter Simulation Conference, 162-176*

Song, l. and S.M. AbouRizk. 2003. "Building a virtual shop for steel fabrication." *Proceedings of the 2003 Winter Simulation Conference*, 2003, 1510-1517

Voorhorst, F.A., A. Avai and C. R. Boër, 2008. "Optimizing a highly flexible shoe production plant using simulation". *Proceedings of 7th international workshop on modelling & applied simulation,* 233-239

BIOGRAPHIES

Antonio Avai is a partner of T.T.S. (Technology Transfer System s.r.l.), an IT company located in Milan. He has managed research programs with the focus on methodologies and leading information technologies to support all the life cycle of manufacturing processes. He has more than 20 years of experience in discrete event simulation and has authored several papers on this topic.

Dario Pietraroia holds a degree in Industrial Engineering from the Politecnico of Milan. He has been working in TTS for few years where he's responsible to the simulation development team. He has acquired a well know how in simulation applying it on some real industrial cases.

Alessandro Pirovano is currently employed at the Ficep s.p.a. and he is the Technical Manager of the Structural Steel division. He is responsible for the coordination of the design process starting from the idea up to the development of the final solution.

Marco Aspesi received his degree in Mechanical Engineering at the Politecnico of Milan. He is currently employed at the Ficep s.p.a. in the Structural Steel division and he has a long experience in design CNC automatic systems for the steel fabrication industry.

Decision Support System for Modeling and Simulation of Cut-to-size Plants

Robert Schoech
Susanne Schmid
Christian Hillbrand
Industrial Research and Development
V-Research GmbH
Stadtstrasse 33, 6850 Dornbirn, Austria
E-mail: robert.schoech@v-research.at, susanne.schmid@v-research.at, christian.hillbrand@v-research.at

KEYWORDS
Discrete Event Simulation, DEVS, Emulation, Central Hierarchical Control Systems, Decision Support Tool, Cut-to-size Plants, Plant Engineering

ABSTRACT

Planning, implementation and operation of cut-to-size plants are very costly and time-consuming processes. Due to parallel material movements, various possible combinations of the plant building blocks and because each plant is designed as a complete, customized system, the complexity of the plants can be enormous. For these reasons, a decision support system based on discrete event simulation has been developed, which allows the producer of the plants to model, emulate, simulate and animate the plant processes. This allows the system experts to make exact forecasts of the attainable performance. In this paper we describe the development of this decision support system, regarding system architecture, processing of control system orders and task handling in the simulation environment. Finally, based on the findings of the realization and validation of the tool, this paper discusses the opportunities arising from this approach as well as its future potential.

INTRODUCTION

Cut-to-size plants with sorting and stacking solutions are very complex systems. Some of the characteristic problems are parallel material movement, buffer-areas and a large number of simultaneous activities. On cut-to-size plants several kinds of panels can be processed. The main *cut-to-size processes* are *cutting*, *sorting* and *stacking*. These processes are executed on machines like cut-to-size saws, sorting carriages, stacking devices, roller tracks, etc. In combination with their control unit, these machines are specified as self-contained plant units.

The acquisition, planning, implementation and operation of cut-to-size plants impose special challenges. Some of these can be traced back to the fact that a sales process usually takes place before the initiation of planning and implementation, since cut-to-size plants are built according to customer specifications (make-to-order strategy). Plant businesses show a high level of specialization of its product range and services (Pekrul 2006). In case of cut-to-size plants, performance is usually measured by cycle times. Besides that, there exist substantial information and knowledge asymmetries between suppliers and customers (Pekrul 2006). As the latter is not capable of evaluating the whole complexity of a plant, he needs to put a lot of confidence into the provider's projections concerning the performance and benefits of the plant.

An essential sales instrument of cut-to-size plants is the reputation of the provider company. Apart from this, reference plants are used to prove the technical feasibility and efficiency of the plants to the customers (Soellner 2008). This is one of the reasons why the marketing of plants is a very complex organizational process which contains a considerable amount of risk for both parties. The supplier usually faces substantial sunk costs if the customer decides not to purchase the plant upon a rather extensive projection process. In order to reduce supplier-side risks, this paper describes a simulation-based approach to support the process from retail to operation and to make the distribution of cut-to-size plants more efficient. This planning instrument makes it possible for system experts to model, animate, simulate and emulate the cut-to-size processes without a need for any deeper simulation or emulation knowledge.

THEORETICAL PRINCIPLES AND LITERATURE REVIEW

Before providing details about the simulation approach for sales and projection processes, the following section will introduce some basic theoretical principles of simulation and emulation techniques. Based on that we will provide a review of the relevant literature within the field of discrete event simulation for decision support in industrial processes.

Discrete event simulation for supporting industrial process planning activities

As it will be discussed below, simulation techniques are likely to provide a wide range of analytical possibilities for planning and evaluating industrial processes. In the context of this paper, a future manufacturing process as designed in the sales phase can be considered as one specific industrial process.

In the literature, a variety of concepts and applications exist which are concerned with the application of discrete event simulation in Supply Chain Management (SCM). The application scenarios for this decision-supporting method are widespread in this case. They include to a large extent the general optimization tasks which are to be performed by SCM (Archibald et al. 1999, p. 1207). The target variables of simulation studies concentrate mainly on a reduction of cost, the optimization of material and information flows, the reduction of processing and lead times, or a consequent process-related orientation of the company (Banks et al. 2002). A recent study analyzes over 80 articles which describe an application of Supply Chain Simulation either in an industrial pilot project, commercially available software, or a simulation test within a logistics chain (Terzi and Cavalieri 2004). This study shows that only 11 papers are concerned with manufacturing processes of which the vast majority aims at integrating the manufacturing process into the overall supply chain or scheduling of production lots. Only four papers describe scenarios for applying simulation for planning manufacturing layouts: Olhager and Persson describe the successful application of simulation for redesigning manufacturing plants in the electronics industry (Olhager and Persson 2006). The other three articles are related to the simulation software package Supply Chain Builder (SCB) of Simulation Dynamics Inc. (SDI). The first approaches of SDI concentrated primarily on the intra-organizational optimization of value chains. The Plant Builder, for example, is focused on the simulation of internal value chain activities (Siprelle et al. 1999). As an extension, the in-plant distribution as well as the supply chain on the distribution side is included in the simulation model (Phelbs et al. 2000; Phelbs et al. 2001).

Besides that, a number of articles report simulation studies in order to find optimal layouts for flexible manufacturing facilities (Drake et al. 1995; Zhou and Venkatesh 1999; Azadivar and Wang 2000; Aleisa and Lin 2005) or cellular manufacturing layouts (Morris and Tersine 1990; Irizarry et al. 2001). These approaches mostly aim at the (near) optimal solution of a specific problem rather than evaluating a multidimensional scenario like the assessment of feasibility for a future manufacturing plant projection.

By contrast, the starting point for optimizing the sales process is the definition of critical performance indicators (e.g. certain lead or cycle times) which have to be realizable by a projected facility or plant. Subsequently, the main issue is to identify admissible and feasible plant configurations which optimize these performance variables. As this paper will show, discrete event simulation can very efficiently support this task.

Emulation based support for plant systems

Emulation is the virtual reproduction of certain aspects of hardware or software systems (external system) on another system (host system) (Mertens 2006). Therefore emulation can be seen as a special case of simulation, supplemented with the coupling of real functional components. For emulation of cut-to-size plants, job data is generated within a "virtual plant" and processed by the simulation tool.

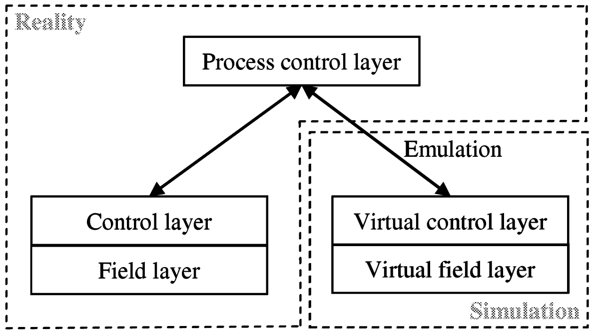


Figure 2 Central, hierarchical control systems

Classical plants are centrally and hierarchically controlled real time systems (see figure 2) (Guenthner and ten Hompel 2010). On a real system the *field layer*, which constitutes the lowest level, represents all mechanical components with their actuators and sensors and controls the material handling. The *control layer* resides above the process level. On this layer, sensor data is processed and control signals for the actuators are generated. This level represents the basis for an automated, unit based plant. Moreover, it coordinates the handover of load data and controls the material flow of the plant units. The *process control layer* is the highest layer in the control pyramid and is usually called *plant server*. The *control layer* and the *process control layer* are connected by means of a communication channel, which passes on the scheduled orders to the *control layer* and receives confirmation when all actions have been processed.

The plant logic of the real time system as described above is mapped to a model within the virtual plant: The *virtual field layer* of the emulated system visualizes all mechanical components and displays the kinematic movement of the material handling in a virtual view. The *virtual control layer* prepares all orders from the *process control layer* and controls each virtual plant unit. Since there is a tight coupling between real time and simulation systems this system can be viewed as an emulation system (Mc Gregor 2002).

GENERAL DESIGN PRINCIPLES

The following chapter provides a general overview of the system architecture for our simulation-based decision support approach. Also, the modeling process and the clear separation of virtual control layer and virtual field layer are discussed. Furthermore, our task handling method for event lists will be described.

System architecture

The application platform of our simulation based decision support system provides simulation libraries with generic modeling functions, which can be used to implement domain-specific modeling environments. A number of domain-specific modeling methods and applications, such as planning of transportation networks or warehousing structures, have already been implemented upon this platform (Dobler et al. 2008; Maerz and Saler 2008). As one part of the described platform, specific methods for cut-to-size plants were implemented. The main intention has been to enable domain experts (i.e. technical sales personnel) to define alternative projected plant configurations within the virtual system and to evaluate them according to the predefined critical performance indicators in order to identify an optimal solution. By using a domain-specific modeling environment with an underlying simulation model built-in, the domain expert can take advantage of emulation techniques without need of simulation expertise. Therefore, the software architecture of the decision support system for cut-to-size-plants is arranged in two levels (see figure 3).

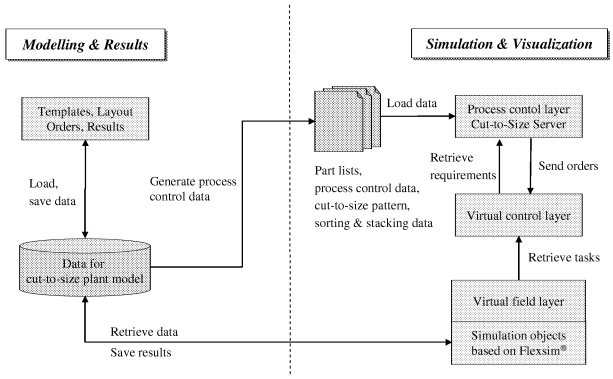


Figure 3 Software architecture
for decision support system

The *Modeling and Results Level* is used for the development of unit-based models (Zeigler and Sarjoughian 2003) in which a plant is designed as a collection of plant units. The modeled plant is persisted and process control data for the cut-to-size server is generated automatically. Depending on this process control data, bills of materials, batch sizes, optimized cut-to-size patterns, sorting and stacking patterns for panels are then computed. The computation results represent the planning data used as a basis to generate the orders for the virtual plant. Furthermore, the simulation run and the visualization of the material flow are triggered at this level. Out of the orders, which are broadcasted by the cut-to-size server, *task sequences* are created (see figure 4). A single task describes an activity that imitates the physical mechanism which is executed on the simulation object.

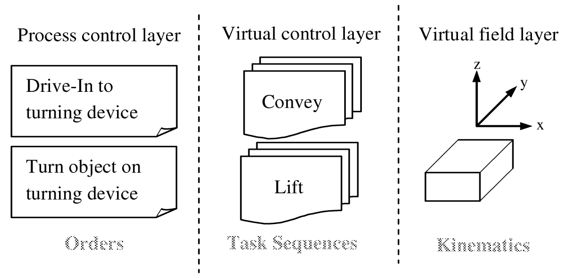


Figure 4 Order handling

The *Simulation and Visualization Level* represents the virtual system, which is controlled. It is based on the simulation engine Flexsim©. Depending on the model data and associated meta-information, the simulation model is created automatically. After the initialization and start of the emulation model, the *virtual field layer* receives the *task sequences*, which are created on the *virtual control layer* and transforms them into discrete events. According to these tasks, kinematic flows are created (see figure 4) and run time information is logged simultaneously.

Modeling process

Our approach to model large-scale plant systems is based on three complementary services (see figure 5).

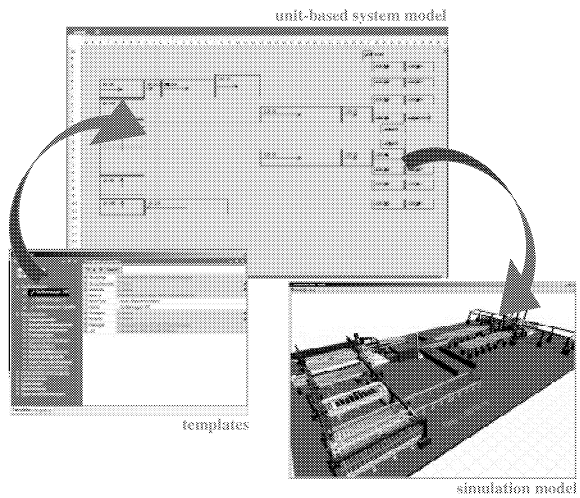


Figure 5 Simulation services

A *plant unit template* represents a self-contained plant unit, which consists of a set of devices. The collection of devices can be viewed with *the template designer*. A device is specified by attributes which describe process- and mechanical information, velocities and meta-information for the instantiation of each class type used in the *virtual field layer* and the *virtual control layer*. They cannot be broken down further.

The *graphical editor* uses the previously described *plant unit templates* and allows the user to instantiate the plant units using drag-and-drop and to couple the material flows using a snap function. Once the model is specified as a *unit-based system model* it needs to be transformed into a simulation model. Therefore, the *unit-based system model* is persisted as a well-formed XML document. The transformation process to build a *simulation model* is automated and consists of three parts: The instantiation of the *virtual control layer* the creation of all simulation objects and the assignment of values to all public properties.

Discrete event task handling

The object-oriented hierarchical simulation model of the plant is based on the functional decomposition approach. The simulation includes the modeled units of the real plant and each unit of a production set is uniquely identifiable and traced during its lifecycle. The model is created according to the modeling process described in the previous section.

After starting the emulation the event list will be served through plant units of the *virtual control layer*. When the *process control layer* sends orders to designated plant units of the *virtual control layer*, all activities are registered as tasks and are stored in the event list in the correct order. The instruction set consists of 22 commands of three types: basic commands, motion commands and item operations.

Basic commands are generally used by the model to handle items. Motion commands describe the kinematic behavior of a plant unit or device. Item operations relate to a panel, respectively parts of a panel.

The implementation of a *blocking rendezvous* pattern (Douglass 2004) for handling the event list makes it possible to have a simulation model in which the behavior is independent of simulation speed. If the *virtual field layer* confirms the execution of a task sequence, the confirmation call is blocked until the *process control layer* has sent new orders to the destination *plant units.* The elapsed time up to the confirmation call is skipped and will not be counted towards simulation time. This is because in real-time, the system which is controlled, has no time delay between order confirmation and receipt of orders. Consequently, the connection between the simulation model and the *process control layer* allows logging of system states and measurement of the performance of the emulated system's cycle time.

Emulation results and analysis

While running the simulation or emulation model, results are written to an Oracle© database which afterwards can be viewed and analyzed by the user through the application platform. On the *virtual field layer* of the simulation system, two kinds of information are logged:

- The information whether a device is *busy*: The status of a device is *busy* if it is moving in some direction (turning, opening, closing, etc.).
- The information whether a plant unit is *occupied*: The status of a plant unit is *occupied* if a flow object is located on it.

With these recorded information several sorts of result analysis can be carried out. These results are shown in tables and charts. The user has the possibility to group the displayed information by devices, plant units or regions of the plant units. Additionally it is possible to compare several simulation or emulation runs. With these options the user is able to analyze throughput rates and cycle times (see figure 6) as well as utilization of plant units and devices of the cut-to-size plants.

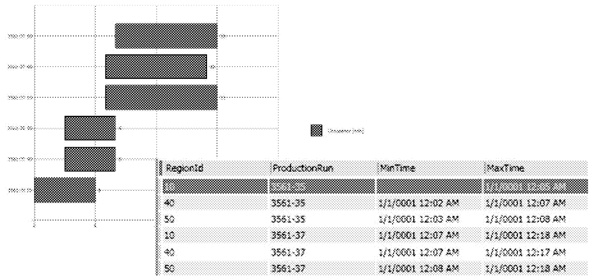


Figure 6 Analysis of cycle time of simulation runs

MODEL VALIDATION

For the verification and validation of emulation models, special aspects have to be taken into account (Rabe et al. 2008):

- an enormous exchange of data based on orders has to be established in an adequate way between the real process control layer and the simulation model
- if the simulation model is particularly developed for the emulation, the model cannot be tested autonomously
- there are interactions with the development of the real control system

For these reasons the verification and validation of the described emulation model is realized on two different levels: the validation of the plant server communication and the validation of the model execution. The applied techniques are described in the following chapter.

Communication with the process control layer

The main challenge with the implementation of an emulation model lies in the establishment of an adequate communication between the process control layer and the virtual control layer. In the realized system, the broadcasted orders from this layer are transformed into task sequences, containing several single tasks that can be executed on the virtual control layer (see figure 4).

The most important requirements are the correct order sequence and the order confirmation at the right time. Depending on the order confirmation, new orders are triggered for processing next task sequences. The implementation of a blocking rendevous pattern (see

also section “Discrete event task handling”) ensure that these rules are adhered to.
To be able to validate the correct order sequence, a *Fixed Value Test* will be applied to verify the deterministic model properties (Rabe et al. 2008). The emulated system’s cycle time, for example, will be measured independent of simulation speed.

Execution of the emulation model

The emulation model cannot be executed autonomously which makes the validation process even more complex and time-consuming. In the verification and validation process of the model execution several validation techniques (Rabe et al. 2008) were applied like described below.
First of all, *animation* of material flows and cut-to-size processes in 3D was used to verify the behavior of the plant. In the current application the visualization of the plant processes plays an important role, due to the fact that the decision support system will be applied as acquisition tool to demonstrate virtual cut-to-size plants to future customer. For this reason, extensive effort was taken to visualize plant units and the contained devices with graphical data files (.vrml) which were generated directly out of the CAD-System of the manufacturer. Nonetheless, when using animation for model validation, it has to be considered, that it cannot guarantee a valid or debugged model (Law and McComas 1991). Still it can be used to demonstrate non valid situations (Law 2007) and especially helps system experts to more easily understand the simulated processes and associate them with the processes of real cut-to-size plants. Based on the animation of the processes, *structured walkthroughs* were carried out with participation of the simulation experts who implemented the model and system experts who have detailed knowledge of cut-to-size processes. The purpose of this procedure was to go through the execution of tasks of each single plant unit and identify mistakes, irregularities and problems within these processes. During these walkthroughts the experts were supported with *operational graphics* in form of displayed transportation speeds and cycle times (e.g. saw cycle, feed cycle, etc.). Related to the presentation of these performance indicators at a certain point of time, is the observation of values in the course of time which can be done within the result analysis of the simulation runs (see figure 6).
Currently, effort ist put in the validation of emulated cut-to-size plants in direct comparison to real plants. The foundations for such *historical data validation* can be fulfilled as cut-to-size runs on real system can be logged. Also the modeling of these real cut-to-size plants into a corresponding emulation model can be realized in an easy way. Still there are some requirements in the preparation of the logged data, which allows a direct comparision of real and virtual plants.

RESULTS AND FINDINGS

The implemented simulation and emulation tool provides the experts with relevant information, models and methods for the acquisition, the planning as well as the operation of cut-to-size plants. The tool facilitates the anticipation of the behavior of real plants on a virtual system and allows plant experts to model, simulate, emulate and animate sequences of cut-to-size plants without being experts in simulation or emulation.
When building up an emulation model, plant units like sawing, sorting or stacking machines are put into a model and linked together according to the modeling process described in the previous chapters. By running the emulation job, data of the virtual machines are generated on the host system, processed in the emulation model and results are written to the database. Subsequently the analysis of cycle times, throughput and utilization of the *plant units* can be realized by the means of tables and charts. The application of the decision support instrument brings various advantages in several phases of the selling and realization of cut-to-size plants.
First of all, by using the decision support tool in the acquisition process vendors of cut-to-size plants can more easily demonstrate the plant concepts, which will help the customer to understand the facts. Especially the animation of material flow helps the user to get a clearer idea of the processes on cut-to-size plants. The possibility of calculating exact cycle times, throughput rates and utilizations of machines improves the accuracy of prediction of performance specifications. Therefore technical expertise, steadiness and reliability can be demonstrated through virtual plants. With this approach costs and time can be saved in the acquisition process. Apart from this, the tool supports the system experts in the plant planning process, in validation of control strategies and in the analysis of bottlenecks in the material flow. The comparison of several simulation scenarios is also possible and allows the system experts to constantly optimize the hardware and software. During implementation of the plant the tool can serve as training instrument to get familiar with the control system and plant processes. While operating the plant it is possible to perform impact analysis of modifications of the virtual as well as the real system. This allows an efficient analysis and resolution of errors without disrupting the normal course of business. With the realization of the described decision support tool a planning instrument could be developed which makes it possible to reduce acquisition and startup times for cut-to-size plants and continuously test and optimize the processes.
Currently the operational use of the implemented simulation-based decision support instrument is being initiated in one company. The concerned specialist for cut-to-size plants was directly involved in the realization process of the simulation and emulation systems.

Additional activities in the further development and extension of the emulation tool are planned in course of this year. Currently the machine units can only be tested separately and not as a combined plant system. The idea is to create a complex system where a combination of reality and simulation (see figure 2) can be achieved. This would mean that real units of the plant could be tested within the whole virtual system. This approach will help the plant experts to be able to find failures on machines or in the material flow much earlier as without the decision support tool.

REFERENCES

Aleisa, E.E. and L. Lin. 2005. "For effectiveness facilities planning: Layout optimization then simulation, or vice versa?". *In Proceedings of the 2005 Winter Simulation Conference.*

Archibald, G.; N. Karabakal; and P. Karlsson. 1999. "Supply chain vs. supply chain: using simulation to compete beyond the four walls". *In Proceedings of the 1999 Winter Simulation Conference*, pp. 1207-1214.

Azadivar, F. and J. Wang. 2000. "Facility layout optimization using simulation and genetic algorithms". *International Journal of Production Research*, 38(17), 4369–4383.

Banks, J.; S. Buckley; S. Jain; P. Lendermann; and M. Manivannan. 2002. "Panel session: opportunities for simulation in supply chain management". *In Proceedings of the 2002 Winter Simulation Conference*, 2, 1652-1658.

Dobler, M.; M. Saler; and L. Maerz. 2008. "Distribution of Concurrent Simulation Runs on a Service-oriented Network Structure". *Paper presented at the ASIM Conference 2008*, Berlin, October 1 – 3.

Douglass B.P. 2004. "Real time UML: advances in the UML for real-time systems". Addison-Wesley.

Drake, G.R.; J.S. Smith; and B.A. Peters. 1995. "Simulation as a planning and scheduling tool for flexible manufacturing systems". *In Proceedings of the 1995 Winter Simulation Conference*, pp.805-812.

Guenthner, W. and M. ten Hompel. 2010. "Internet der Dinge in der Intralogistik". Springer Verlag, Berlin.

Irizarry, M.A.; J.R. Wilson; and J. Trevino. 2001. "A Flexible Simulation Tool for Manufacturing-cell Design". IIE Transactions 33(10), 827-836.

Law, A.M. 2007. „Simulation Modeling and Analysis". McGraw-Hill, Boston.

Law, A.M.; McComas M.G. 1991. "Secrets of successful simulation studies". In: Nelson, B.L.; Kelton W.D.; and Clark G.M. (Hrsg). *Proceedings of the 1991 Winter Simulation Conference.* IEEE, Piscataway, Phoenix (USA). S 21-27

Maerz, L. and M. Saler. 2008: "An Analysis, Planning and Decision Tool for Warehouse Applications". *Paper presented at the ASIM Conference 2008*, Berlin, October 1 – 3.

Mc Gregor I. 2002. "The relationship between simulation and emulation". *In the Proceedings of the 2002 Winter Simulation Conference,* San Diego, California, USA:1683-1688.

Mertens P. 2001. Lexikon der Wirtschaftsinformatik, Springer Verlag, Berlin.

Morris, J.S. and R.J. Tersine. 1990. "A Simulation Analysis of Factors Influencing the Attractiveness of Group Technology Cellular Layouts". *Management Science* 36(12), 1567-1578.

Olhager, J. and F. Persson. 2006. "Simulating production and inventory control systems: a learning approach to operational excellence". *Production Planning and Control*, 17 (2), 113-127.

Pekrul S. 2006. "Bauwirtschaft und Baubetrieb. Strategien und Maßnahmen zur Steigerung der Wettbewerbsfähigkeit deutscher Bauunternehmen". Universitätsbibliothek, Berlin.

Rabe, M.; Spiekermann, S.; andWenzel, S. 2008. „Verifikation und Validierung für die Simulation in Produktion und Logistik". Springer-Verlag Berlin, Heidelberg.

Soellner A. 2008. "Einfuehrung in das Internationale Management. Eine institutionenökonomische Perspektive". Gabler GWV Fachverlage GmbH, Wiesbaden.

Siprelle, A.; D. Parsons; and R. Phelps. 1999. "SDI Industry Pro: simulation for enterprise-wide problem solving". *In the Proceedings of the Winter Simulation Conference*, 241-248.

Terzi, S. and S. Cavalieri. 2004. "Simulation in the supply chain context: a survey". *Computers in Industry*, 53 (1), 3-16.

Zeigler, B. P. and H. S. Sarjoughian. 2003. "Introduction to DEVS Modeling and Simulation with JAVA: Developing Component-based Simulation Models". Arizona Center of Integrative Modeling and Simulation.

Zhou, M.C. and K. Venkatesh. 1999. "Modeling, simulation, and control of flexible manufacturing systems: a Petri net approach". World Scientific Publishing Company, Singapore.

AUTHOR BIOGRAPHIES

V-Research is an Austrian competence center for industrial research and development. In the research field "Process Engineering for Construction, Manufacture and Logistics" the research center focuses on methods to support complex decision making in manufacturing and logistics processes.

Robert Schoech is project manager at V-Research. He develops systems focused on Simulation of industrial processes. His research interests also include track & trace solutions based on GPS, GSM and RFID technology.

Susanne Schmid is research assistant and project manager at V-Research. Her main research areas include the conception, modeling and development of simulation models.

Christian Hillbrand is head of the research area „Technical Logistics" at V-Research and director of the research programme "ProDSS" (Integrated Decision Support Systems for Industrial Processes). He holds a PhD-degree in business information systems from the University of Vienna.

MULTIMODAL TRANSPORT SIMULATION

PREDICTING FUTUR(E-)TRAFFIC

Marco Lützenberger Nils Masuch Benjamin Hirsch Axel Heßler Sahin Albayrak

DAI-Labor Berlin
Ernst-Reuter Platz 7
10587 Berlin, Germany
E-Mail: ⟨firstname.lastname⟩@dai-labor.de

KEYWORDS

Transportation, Urban affairs, Operations research

ABSTRACT

In the coming years, major car manufacturers will mass produce electric vehicles. Despite a quantum leap in EV technologies however, there are many unsolved problems. Short ranges and insufficient charging infrastructures are often named as central issues here. In order to avoid problems and deficiencies it is necessary to simulate the usage of EVs, though contemporary approaches do not fully consider the particular characteristics and limitations. In this work, we emphasise these additional requirements and describe the development of a corresponding simulation framework which is able to predict futur(e-)traffic.

INTRODUCTION

In the coming years, road traffic will undergo a major change. While nowadays electric mobility is applied only in prototypical pilot projects, both industry and government invest large amounts of money in research and development in order to accelerate a breakthrough of this technology [United Press International(2010)]. There are many arguments supporting these efforts, however, a lot of problems have still to be considered.

Compared to conventional driving, the usage of an electric vehicle (EV) comprises several differences. One central issue is charging. While fuel-driven cars provide ranges beyond 600km and a fully developed petrol station network assures untroubled mobility, the EV case is a different one. As it is, only major cities are equipped with charging stations. These "lineups" are patchy at best. Furthermore, the short range of EV models hardly exceeds the 200km barrier, which aggravates the infrastructural deficiency, because charging will be required at least thrice as often as conventional refuelling. In fact, current studies depict the little range and the limited access to charging stations as one of the main show stoppers for the breakthrough of electric mobility. A study [Tremblay(2010)] evaluating the acceptance of electric vehicles has been performed by *Ernst & Young*, recently. The answers to the question: *Which factors would make you most hesitant to choose a Plug-in Hybrid Electric Vehicle (PHEV) or EV as your next vehicle?* are illustrated in Figure 1.

		Access to charging stations	Price	Battery driving range	Reliability/ service ability	Performance and handling
China		69%	57%	73%	64%	57%
Japan		60%	73%	43%	36%	35%
US		75%	74%	75%	57%	49%
Europe	France	74%	63%	81%	26%	46%
	Germany	74%	66%	75%	46%	52%
	Italy	64%	62%	62%	42%	54%
	UK	71%	60%	71%	47%	50%
Average		69%	67%	66%	49%	48%

Highest response rate for each factor Lowest response rate for each factor

Figure 1: Answers to the question: "Which factors would make you most hesitant to choose a PHEV or EV as your next vehicle?", raised within an *Ernst & Young* study [Tremblay(2010)].

The numbers highlight one of the key problems of electric mobility. More than 65% declare short range and insufficient charging station availability as a reason against the purchase of an EV. Considering the required time for a recharging process, the situation becomes even more difficult. Depending on the station and the battery's state of charge, charging currently can last up to several hours, which makes it necessary to schedule charging during the driver's daily routine. Summing up, the mentioned limitations force each EV use to devise plans regarding the scheduled rides, the overall range and the regional charging infrastructure. Contrary to this, the acceptance of electric mobility will be determined by the vehicles' capability to handle these difficulties. In order to detect potential problems, it is necessary to possess instruments which allow for a prediction of the usage of electric vehicles. Regarding the charging infrastructure, a realistic prediction model will allow to evaluate a given setup in several aspects. First, there is the provided capacity. Prediction will allow to determine the maximum number of cars for which the given infrastructure will provide sufficient service. It will become far easier to measure an infrastructure's support for e-mobility. Further, in combination with detailed usage profiles, it will be possible to locate particular regions of deficiencies. Both aspects will help to produce valuable information, e.g. for a determined and well directed urban development of infrastructure for electric mobility.

Additionally, prediction will allow improvements beyond charging infrastructures. Estimates on the practicability and versatility of particular car models can be done as well. Depending on the given car type's attributes (e.g. battery capacity or required charging time), it is possible to classify car types as being suitable, adequate or unsuitable for a given user and usage profile. In reverse, it will become possible to determine a fitting car type to a given usage pattern. A similar relationship may be used for many purposes, such as orientation, purchase decisions, streamlined car development, or to simply inspire confidence in e-mobility.

The development of such a prediction model is complex. Considering the little information available on the use of electric vehicles makes this project even more difficult. Usually, when there is little available data, simulations are considered. The field of traffic simulation has experienced a rapid upswing within the last years. In order to determine in how far existing approaches are able to deal with the described problematic, we did a comprehensive survey. Identifiable approaches as well as the outcome of that survey are part of the following section.

RELATED WORK

There are a large number of traffic simulations available. In order to provide some structure to this section, we will distinguish between the three common categories: macroscopic, mesoscopic and microscopic.

Macroscopic Traffic Simulations

Emme/2 [Inro(2011)] is a commercial software which mainly focuses on simulations of private and public transport, but also comprises demand modelling, matrix balancing procedures, and analysis and automation features. Out of the box, the software comprises over a hundred transport themed maps.

VISUM [PTV America(2011)] is a commercial software simulation for the analysis of transportation planning, travel demand modelling and network data management. It is used in metropolitan, regional, state-wide and national planning applications and also allows for multimodal analysis by providing support for cars, car passengers, trucks, busses, trains, pedestrians and bicyclists. Further, *VISUM* provides a variety of assignment procedures and 4-stage modelling components which include trip-end based as well as activity based approaches.

OmniTRANS [Omnitrans International(2011)] is a commercial, multi-modal and multi-temporal simulation software which is used to analyse the interactions of transport modes in an urban context. The applied algorithms are based on mode chains and allow for the modelling of complex schemes such as park-and-ride, kiss-and-ride and bike-and-ride. Further, it is possible to define the geometric layout and signal plans of intersections. Examinations of multiple time configurations are supported as well.

Mesoscopic Simulations

TransModeller [Caliper Corporation(2011)] is a traffic simulation for wide-area traffic planning, traffic management and emergency evacuation. It is capable of including multi-modal traffic and its main focus is to produce data on the flow of vehicles, the operation of traffic signals and the overall performance of a transportation network.

Rigolli et al. [Rigolli and Brady(2005)] describe a mesoscopic traffic simulator, in which the simulated vehicles are implemented by using a behavioural multi-agent approach to model human drivers. In the simulation, the applied agents generate a qualitative description of the sensed environment, while an included reasoning module is used to generate a set of possible driving options. The simulator features simulations on straight line highways with multiple road lanes and with no entry or exit lanes and comprises different environmental conditions, such as ice or fog.

Cube [Citilabs Inc.(2007)] is a comprehensive, macroscopic transportation analysis system. The software offers a set of modules, each module providing unique functions for unique tasks. While the base module constitutes the system interface, there are modules for forecasting urban, regional, and long distance passenger travel demand, for multimodal micro simulation, for forecasting regional and long-distance commodity flow and truck demand, for estimating and optimising trip tables from traffic counts and other survey data and for forecasting land usages.

Aimsun [Casas et al.(2010)] is an integrated transport modelling software which is used to improve road infrastructure, reduce emissions, cut congestion and design urban environments for vehicles and pedestrians. Next to static and dynamic traffic assignment and mesoscopic simulations, the software supports further features, such as a creation of vehicle actuated signals, an introduction of public transport priority schemes, a definition of traffic management strategies with triggers and actions and many more.

Microscopic Simulations

The *SUMO* [Krajzewicz(2010)] simulation framework is an open-source, microscopic traffic simulation package, which has been developed as testbed for research matters. The software has been designed to handle large road networks and can be easily extended. A set extension modules is already provided (i.e. online interaction, trace export or additional GUIs), while custom implementations are supported as well.

VISSIM [Fellendorf and Vortisch(2010)] is a commercial simulation program for multi-modal traffic flow modelling. With a high level of detail it accurately simulates urban and highway traffic, including pedestrians, cyclists and motorised vehicles. In comparison to most available simulation software, *VISSIM* allows for a modelling of the parking search process.

Paruchuri et al. [Paruchuri et al.(2002)] describe an approach for modelling unorganised traffic in a traffic simulation. The authors describe a microscopic simulation framework in which the vehicles are modelled as autonomous software agents. They use micro and macro goals for the realisation of different driving styles, such as aggressive, normal and cautious. Further psychological traits, such as “confidence” and “rush” are implemented as well.

SCANeR© II simulator [Champion et al.(1999)] is an agent based traffic simulator and has been developed by the French automobile manufacturer *Renault* for ergonomics and advanced engineering studies, for research in road traffic, for human factor studies and for driver training. The software features microscopic simulation of interactively driven entities. Currently, vehicles, trucks, motorcycles, bikes, train, trams or pedestrians are supported. Each simulated object is able to either mimic “unique” or “risky” driving behaviour.

Tang et al. [Tang and Wan(2005)] describe a multi-agent traffic simulation system for urban environments, which is capable of modelling traffic in response to current road network loading at the time of interest. The simulator allows for adjustments to the road network, to the roadway traffic and to the timing of applied traffic signals and illustrates the results in a realtime-3D representation.

Ehlert et al. [Ehlert and Rothkrantz(2001)] describe a microscopic traffic simulator and a model for reactive software agents which are able to control the simulated vehicles. The simulator features simulations within an urban environment with multi-lane roads, intersections, traffic lights and traffic lights controllers. The described model works fine and is capable of realising individual driving styles from slow and careful to fast and aggressive.

Results

The survey emphasised, that each approach has been geared to a set of particular purposes. Some works for instance include pedestrians and support evaluations of multimodal transports. Others apply a far more detailed point of view and support examinations on microscopic level, such as lane change manoeuvres or the effects of different driving styles. An explicit support or extensibility on electric vehicles, however, seems to be missing. On first glance, this endeavour sounds trivial: Integrate some different vehicle- and consumption models and the trick is done. We deem the situation is more complex, because the crucial question about e-mobility is not about the vehicle itself,[1] but to which level a driver is able to deal with the limitations he is entailed with. For this reason, it is necessary to integrate common user patterns as a parameter into the simulation and determine in how far the restrictions of electric vehicles compromise the liberty of a driver.

Beyond that, the limited ranges of electric vehicles imply a strong dependency to the charging infrastructure (CI). An expressive prediction model has to comprehend this condition, yet, we were not able to identify similar features in the examined approaches.

To sum up; in our opinion, the key issue in predicting the future usage of electric vehicles is not solved by extending current works to a very particular vehicle model (many approaches comprise such feature), but by considering the user and the CI as key performance indicators as well. In the following, we introduce an compatible approach.

CONCEPTS

Based on our survey, we have to extend state-of-the-art traffic simulation frameworks in three main categories, namely the electric vehicle, a charging infrastructure and a user profiles. To provide some structure to this section, we will use this classification to explain our approach.

Focusing on Electric Vehicles

Fuel-driven cars provide ranges beyond 600km and a fully developed petrol station network assures untroubled mobility. For this reason, it is usually not necessary to integrate range and consumption issues for traffic simulations. Yet, we already motivated the necessity for this kind of consideration for electric vehicles. For our traffic simulation, we defined a meta model which provides a description for an `ElectricVehicle` type. Extending common vehicle attributes like dimensions, maximum speed or acceleration, our car type addition-

[1] For this, simulation is not required, since many prototypes already exists.

ally references a battery and a consumption table. The assembly is highly modular and allows for easy adaptations of both, which enables for quick updates of the applied characteristics to the latest battery model or consumption measurements. Beyond that, the battery can be considered as an additional simulation parameter and can be used for conclusions on fictitious performances or estimates on critical required capacities for given driving purposes. The applied battery model is comprehensive and supports not only capacity characteristics, but also current conduction values, which can be used to simulate charging and discharging processes in a lifelike fashion. The specific part of the meta model[2] is illustrated in Figure 2.

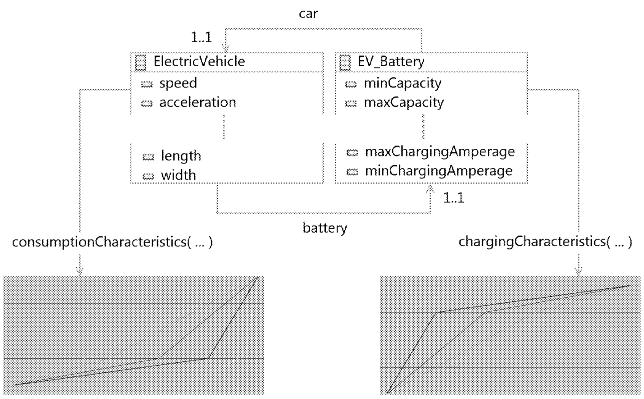


Figure 2: Part of the simulation's meta model with exemplary consumption and charging characteristics.

Based on the presented meta-model and a comprehensive survey, we are able to perform simulations with realistic implementations for the following vehicle types: *Aixam Mega Multitruck*, *Aixam Mega E-City*, *BMW Mini E*, *Compact Power Motors GK1*, *EcoCraft Automotive EcoCarrier*, *Fiat Karabag Micro-Vett 500*, *Fine Mobile Tyike Active*, *Fort Transit BEV*, *German E Cars Stromos*, *Goupil* Goupil G3*, *Mitsubishi i MiEV*, *Volkswagen E-Up*, *Reva Reva i*, *Smart Fortwo Electric Drive*, *Smiles CityEL*, *Tazzari Zero*, *Tesla Motors Tesla Roadster* and *Think Global AS Think City.*

Charging Infrastructure

For the year 2020, the German government predicts a total amount of one million electric vehicles on the German road network [United Press International(2010)]. In order to support such number, comprehensive CI extensions have to be implemented. For a prediction model, it will be highly relevant to include detailed information on the applied charging stations characteristics, their amount and their positions. Since today, these information does not exist, our prediction model had to provide extensive configuration options. We decided to make use of the *Open Streetmap* framework (*OSM*) for this purpose. *OSM* is a collaborative project, which provides free, GPS precision road maps. Next to geographic locations of road networks, *OSM* provides a large amount of semantic information on speed limits, the amount of lanes, traffic lights, traffic signs, car parks and even comprises information on present charging stations. The collected data is validated and gathered by an open community, which ensures an up-to-date condition. For the representation of *OSM* data, the *Extensible Markup Language* (*XML*)[3] is used. The syntax of *XML* simple and it is easy to extend the data for custom purposes. An exemplary *OSM* maps, with the entire set of additional information showing is illustrated in Figure 3.

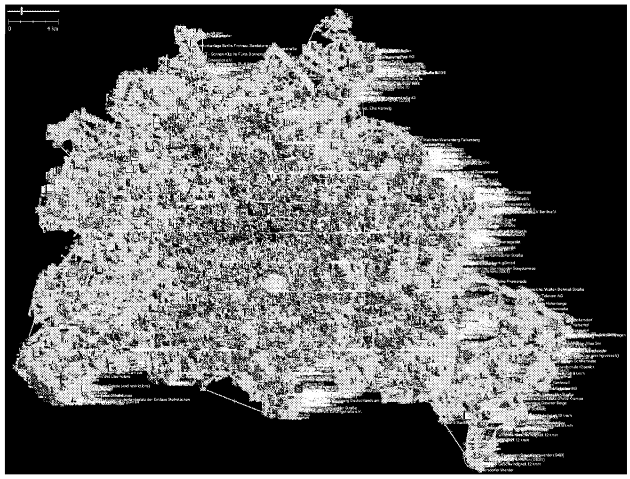


Figure 3: OSM Data of Berlin, visualised by the *Java OpenStreetMap Editor* (*JOSM*).

For our work, we made use of *OSM* and developed a special editing tool, which is able to process the semantic information of the applied maps, illustrate them in a visual manner and supports custom adaptations as well. Using this editor, we are able to use on up-to-date map material but also to extend the given charging infrastructure at will.

The User Profile

In order to predict in how far the limitations of electric mobility affect a user in his daily schedule, we have to find a way to integrate this day schedule as a parameter in our simulation. We applied an appointment based approach in which we assign each vehicle a home position and assemble a day schedule by randomly generating a set of appointments for the user. For the home positions, we extended the underlying *OSM* map by additional information on the population den-

[2] Our meta model is far more extensive, comprising the entire simulation context, including vehicles, roads, parking lots, charging stations, to name but a few.

[3] `http://www.w3.org/TR/2006/REC-xml11-20060816/`

sity [Federal Statistical Office(2010)] and stochastically retrieve home positions by this weighted distribution. We initialise a set of potential target positions by the same principle, though, as there is no obvious correlation between population density and targets, we allow for an explicit marking of frequent target areas within the map. The editing tool we mentioned earlier has been developed, to support this task as well. Figure 4 illustrates the edit tool, showing a target specification with three clusters and different radiuses. Each generated vehicle is now assigned a home location. Further, a set of appointments are generated and associated with a target position. The resulting day schedule is then assigned to the car as individual but realistic driving pattern for the simulation.

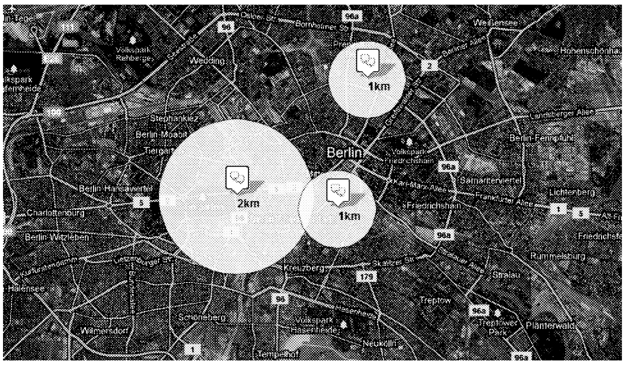


Figure 4: Target specification with three different target clusters for the capitol region Berlin.

For the implementation, we extended the `ElectricVehicle` type with a `Calendar` reference in the meta model. The car's day schedule is stored within this calendar in the form of `Event` types. We distinguish between *Moving Off*-, *Parking*- and *Charging* events, each type featuring a starting time. *Moving Off* events additionally feature a target position. While *Parking*- and *Charging* events occur during a running simulation exclusively, only the *Moving Off* are required for the user profile generation. Regarding the appointment generation process, we have to make sure that the generated schedules are feasible. Given an empty calendar, we start by randomly generating a starting time for a potential appointment. For this purpose, we implemented three different distribution functions (Gaussian distribution with one cluster around 12 p.m., Gaussian distribution with two clusters around 7 a.m. and 7 p.m. and uniform distribution). Next, we retrieve a potential target location from the target position pool and assign it to the appointment. Finally, we randomly determine the appointment's duration. The user can define min and max parameters for that. Once an appointment has been generated, we try to place it within the car's "so far" schedule. Initially this schedule is empty, so the first generation usually succeeds.[4] For each further appointment, routes and cruising times from potentially previous appointments and to potential succeeding appointments are calculated in order to detect conflicts. The user is able to specify a tolerance time which is required between the arrival of the car and the actual starting time of the appointment. The tolerance time is a simulation parameter which guards the vehicles against delays (e.g. caused by traffic jams). If the appointment fits into the schedule, a *Moving off* event is generated and — in conjunction with the appointment — added to the car's calendar. The total number of daily appointments can be specified by the user, again as simulation parameter. So far we have described the three most uncommon extensions to contemporary traffic simulators, which are required for conclusions on the practicability of electric vehicles. In the following section, we will give a detailed description of the simulation process.

THE SIMULATION ROUTINE

The simulation starts with each car at its distinguished home location. Essential for the starting of a car is a *Moving Off* event. These are stored "exact-to-the-minute" within the a car's calendar. The simulation routine processes the day in time intervals of one minute and respectively checks if a *Moving Off* event is scheduled for a car at the current simulation time. In case of an event, the target position is retrieved from the corresponding appointment and a route from the current location to this target is calculated and stored as "currently processed route". According to this route, the car is driven through the map. For this, the current road's maximum speed (retrieved from the *OSM* map), the car's maximum speed (defined by the meta model), a traffic light factor and the current road congestion[5] are used to calculate the effective driving speed. The journey is simulated minute by minute, until the current route has been processed. The car has now arrived at the appointment location. Once the car has arrived, a parking lot is located (again, by consulting the information of the underlying map). A route is calculated and stored within the current route field of the car. In the same manner as described above, the car is driven to the determined parking lot. Upon arrival, the parking lot is checked for vacant places and depending on the availability, the car either initiates the parking (or charging) process or proceeds to the next registered parking lot. At this point, it is important to note that the screening for parking lots is not explicitly considered within

[4]Failure cases are possible! Our simulation simulates the day from 00:00 a.m. till 11:59 p.m., with each car initially at it's home location. A randomly generated appointment on the other side of the city with a staring time of 0:01 a.m. will of course fail. However, our validation routine catches these cases.

[5]We allow to define a specific road congestion for the simulations. This data persistently stored in a description file and loaded at the start of each simulation.

the feasibility validation of the appointment generation process. The level of tolerance can be custom defined as simulation parameters. For this reason, it is possible that a car falls behind schedule. In order to deal with delays, a check is done. A delay detection causes the simulation control either to send the car back to its home location or to forward the car to the next scheduled appointment. This behaviour allows for further interesting examinations. Once a car arrived at a parking lot, the car's battery state of charge is checked and compared with a threshold value (another simulation parameter). Depending on the state of charge, either a *Charging-* or a *Parking* event is generated for the next simulation minute and added to the car's calendar. In the next simulation cycle, the simulation routine switches the car to the according state, in which it will remain until the next *Moving Off* event is occurs. Figure 5 illustrates the rough simulation routine as *BPMN* diagram.

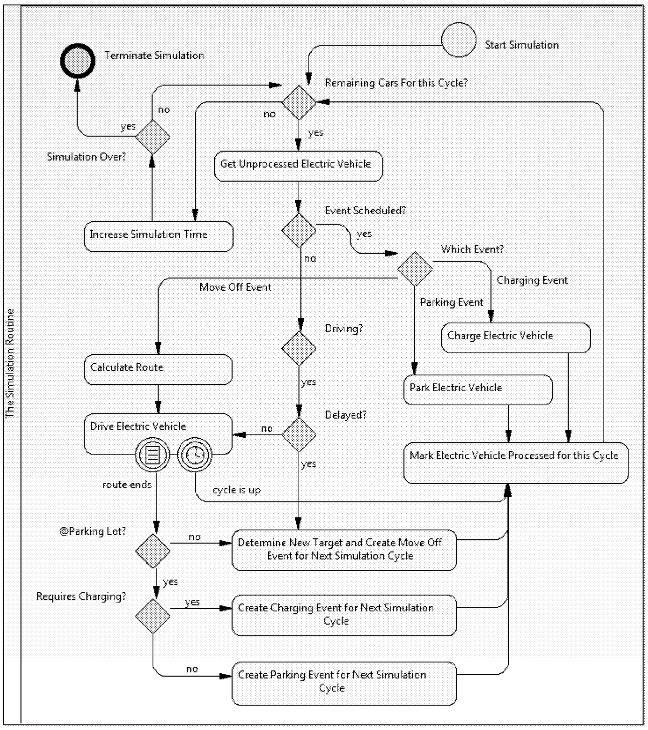


Figure 5: The simulation principle, illustrated as *BPMN* Diagram.

CONCLUSION

In order to draw conclusions on the effects and the practicability of electric vehicles, it is necessary to possess prediction models. In this work, we described an according model in the form of a computer based traffic simulation.

We performed an analysis in order to see how contemporary approaches deal with the problematic and came to the conclusion that state of the art implementations only partially cover the requirements for electric mobility.

We extended common traffic simulator concepts in four main topics. First, we developed a meta model for electric vehicles which – in addition to common car attributes – provides specific information on charging and discharging processes.

Further, we extended our observation focus, to comprehend the user and his daily mobility requirements, in order to focus our examination on the intrinsic problem of electric vehicles to provide unrestricted mobility throughout an entire day. For this examination, we had to generate realistic usage profiles for the simulated cars, hence, we implemented three different distributions and provide the capability to mimic working day-, weekend- and uniformly distributed scenarios. We compared the generated user profiles to the results of a comprehensive mobility study [Federal Ministry of Transport(2010)], which has been conducted by the the *Federal Ministry of Transport, Building and Urban Development*, and achieved results with a margin of error of less than 5%.

Finally, we had to find a way to integrate charging infrastructure information and to provide a possibility to alter them. We realised this feature by including the semantic map framework *OSM* to our simulator, which enabled us to perform simulations on up-to-date map material but also to extend the given charging infrastructure ad libitum.

With the demonstrated approach, we were able to develop a stable simulation prototype, which is currently used in an industrial funded project. The prototype is able to run simulations with a variation of several main parameters (i.e. the applied city or region, number of applied vehicles and models, frequently used target regions, weekday type, etc.) and several secondary parameters (i.e. the average number of daily appointments, the appointment duration, tolerance time for the drives, etc.). Due to a permanent consideration of performance during the implementation, we are able to run simulations with up to 100k vehicles on maps with capital dimensions. Figure 3 illustrates a data set of Berlin, which we used for simulation scenarios. We facilitated performance by using an optimised internal representation of the applied simulation map, which provides fast access to frequently used requests or by implementing an optimised version of the already fast A* algorithm, to name but a few.

The simulation produces plausible and comprehensible results about the limitations, constraints and possibilities of e-mobility. We thus feel confident that we selected the right parameters and that the described approach is able to predict dependencies, influences and handling characteristics of future electric mobility.

Due to the great interest in electric mobility, we will further extend the demonstrated concept and prototype. Regarding the applied meta model, we plan to integrate a representation of onboard-units, which highly influence the range of an EV. We further pursue a more precise realisation of the charging process and its influences on the energy network. An according integration will be able to widen the prediction capability and produce valuable results e.g. for energy providers. Further, as one result to our survey, we identified several approaches which apply psychological aspects concerning the driving styles of the simulated cars. We intend to extend these features to further aspects like human time management and adherence to schedules, which are particularly interesting for our simulation in daytime length. In fact, this work motivated us to develop a comprehensive cognitive model for simulated drivers, which not only comprises short term decisions such as lane change maneuvers, selected velocities, or distances, but also includes strategic behaviour which accounts for the entire day schedule of the driver. We already presented a concept for this objective [Lützenberger et al.(2011)]. The development of such a cognitive model is done as part of the PhD thesis of the first author.

REFERENCES

[Federal Ministry of Transport(2010)] 2010. *Mobilität in Deutschland 2008*. Federal Ministry of Transport, Building and Urban Development, Bonn and Berlin, Germany. URL `http://www.mobilitaet-in-deutschland.de/pdf/MiD2008_Abschlussbericht_I.pdf`.

[Federal Statistical Office(2010)] 2010. *Statistical Yearbook 2010 For the Federal Republic of Germany including International tables*. Federal Statistical Office, Wiesbaden, Germany. URL `http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/SharedContent/Oeffentlich/B3/Publikation/Jahrbuch/StatistischesJahrbuch,property=file.pdf`.

[Barceló(2010)] Barceló J. (Ed.), 2010. *Fundamentals of Traffic Simulation*. Springer.

[Caliper Corporation(2011)] Caliper Corporation, 2011. *TransModeler Traffic Simulation Software*. Caliper Corporation, 1172 Beacon Street, Newton, MA, USA. URL `http://www.caliper.com/PDFs/TransModeler%20Brochure.pdf`.

[Casas et al.(2010)] Casas J.; Ferrer J.L.; García D.; Perarnau J.; and Torday A., 2010. *Fundamentals of Traffic Simulation*, chap. Traffic Simulation with Aimsun. In [Barceló(2010)], 173–232.

[Champion et al.(1999)] Champion A.; Mandiau R.; Kolski C.; Heidet A.; and Kemeny A., 1999. *Traffic generation with the SCANeR© II simulator: towards a multi-agent architecture*. In *Proceedings of the 1st Driving Simulation Conference, Paris, France*. 311–324.

[Citilabs Inc.(2007)] Citilabs Inc., 2007. *Cube 5 — Forecasting Suite*. Citilaby Inc., 1211 Miccosukee Road, Tallahassee, Florida, USA. URL `http://www.citilabs.com/pageloader/broch_cube/pdf/en/cube.pdf`.

[Ehlert and Rothkrantz(2001)] Ehlert P.A.M. and Rothkrantz L.J.M., 2001. *A Reactive Driving Agent for Microscopic Traffic Simulations*. In *Proceedings of the 15th European Simulation Multiconference, Prague, Czech Republic*. 943–949.

[Fellendorf and Vortisch(2010)] Fellendorf M. and Vortisch P., 2010. *Fundamentals of Traffic Sim*, chap. Microscopic Traffic Flow Simulator VISSIM. In [Barceló(2010)], 63–93.

[Inro(2011)] Inro, 2011. *The Emme/2 Website*. URL `http://www.inro.ca/en/products/emme/index.php`.

[Krajzewicz(2010)] Krajzewicz D., 2010. *Fundamentals of Traffic Simulation*, chap. Traffic Simulation with SUMO – Simulation of Urban Mobility. In [Barceló(2010)], 269–294.

[Lützenberger et al.(2011)] Lützenberger M.; Masuch N.; Hirsch B.; Ahrndt S.; Heßler A.; and Albayrak S., 2011. *The BDI Driver in a Service City (Extended Abstract)*. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems, Taipei, Taiwan*. 1257–1258.

[Omnitrans International(2011)] Omnitrans International, 2011. *OmniTRANS 5.1 — A New Dimension Transport Modelling*. Omnitrans International, P.O. Box 161, 7400 Ad Deventer, The Netherlands. URL `http://www.omnitrans-international.com/download/161product-brochure-omnitrans-51`.

[Paruchuri et al.(2002)] Paruchuri P.; Pullalarevu A.R.; and Karlapalem K., 2002. *Multi Agent Simulation of Unorganized Traffic*. In *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems, Bologna, Italy*. 176–183.

[PTV America(2011)] PTV America, 2011. *VISUM — State-of-the-Art Travel Demand Modeling*. PTV America, 9755 SW Barnes Road, Portland, Oregon, USA. URL `http://www.ptvag.com/fileadmin/files_ptvag.com/download/traffic/Broschures_Flyer/VISUM/VISUM_Brochure_e_2009_LowRes.pdf`.

[Rigolli and Brady(2005)] Rigolli M. and Brady M., 2005. *Towards a Behavioural Traffic Monitoring System.* In *Proceedings of the 4^{th} International Joint Conference on Autonomous Agents and Multiagent Systems, Utrecht, The Netherlands.* 449–454.

[Tang and Wan(2005)] Tang W. and Wan T.R., 2005. *Multi-agent Animation Techniques for Traffic Simulation in Urban Environment.* In *Proceedings of the 13^{th} Winter School of Computer Graphics, Plzeň, Czech Republic.* 161–164.

[Tremblay(2010)] Tremblay J.F., 2010. *Gauging interest for plug-in hybrid and electric vehicles in ley markets.* In *Proceedings of the 2^{nd} German Electric Vehicle Congress.*

[United Press International(2010)] United Press International, 2010. *Germany launches electric car initiative.* United Press International. URL `http://www.upi.com/Science_News/Resource-Wars/2010/05/04/Germany-launches-electric-car-initiative/UPI-55231273005809/`.

BIOGRAPHY

MARCO LÜTZENBERGER received his diploma in computer science in 2009 from the Technische Universität Berlin. From 2009 until now, he is part of the Competence Center Agent Core Technologies at the distributed artificial intelligence lab (DAI Labor) of the Technische Universität Berlin, where he is working on his doctoral thesis in the field of multi-agent based traffic simulations.

NILS MASUCH has obtained his diploma in computer science in 2010 at TU Berlin and is currently working on his doctoral thesis in the field of sematic service matching.

BENJAMIN HIRSCH, PhD. received a Master of Science in Artificial Intelligence from the University of Amsterdam in 2001 and a PhD in Computer Science from the University of Liverpool in 2005. From 2005 until 2011 he lead the Competence Center Agent Core Technologies at the distributed artificial intelligence lab (DAI Labor) of the Berlin Institute of Technology, where he did research on the integration of multi-agents, service oriented architectures and business process management. In 2009 he received a OCEB certification as Business Process Management Expert (technical advanced). Dr. Hirsch published numerous articles and is reviewing for many conferences and journals. At EBTIC, Dr. Hirsch works within the iCampus initiative and is leading the iLearning theme.

AXEL HEßLER Dipl.-Inform. Axel Heßler studied computer science at the Technische Universität Berlin. He works as a researcher at the DAI-Labor of TU Berlin with focus on agent technologies. He has worked on a number of projects within the laboratory, including the JIAC agent frameworks. Currently, he is finalising his PhD thesis about agent-oriented methodologies in the context of real-world projects.

PROF. DR.-ING. HABIL. SAHIN ALBAYRAK is the head of the chair Agent Technologies in Business Applications and Telecommunication. He is the founder and head of the DAI-Labor, currently employing about one hundred researchers and support staff.

A THREE THOUSAND VARIABLE NONLINEAR TRANSPORTATION PROBLEM WITH EXCESS DEMAND

William Conley
Austin E. Cofrin School of Business
University of Wisconsin-Green Bay
Green Bay, Wisconsin 5431-7001
U.S.A.
Conleyw@uwgb.edu

KEYWORDS
Cost control, shipping, multi stage Monte Carlo Optimization, transportation, complex systems, management

ABSTRACT

One of the most important discoveries in applied mathematics in general, and business mathematics in particular, in the 20th century, was linear programming and its attendant simplex algorithm. If an objective function and all of its constraints were linear the optimal solution was at a corner point of the feasible solution space. The optimal could be found using the simplex algorithm. Frequently a computer was used to carry out these step by step calculations. Modifications of this approach could be used to solve linear transportation problems that occur in business (as an example, delivering a product in bulk from many locations to many destinations).

Now in the 21st century computer speeds have advanced to the point where one can use them with simulation techniques to approximate the solution to nonlinear transportation problems. Nonlinear ones are like the linear problem just described except that the objective cost function is nonlinear and hence more accurately reflect the world of big business where there are substantial and ever-frequent returns to scale and discounts for shipping in quantity.

Presented here will be a nonlinear such example of 3000 variables that is worked on with the simulation technique multi stage Monte Carlo Optimization.

INTRODUCTION

A book such as (Mizrahi and Sullivan, 1993) presents a fairly standard treatment of linear programming and how to use it, and various versions of it, that can be applied to the classic linear transportation problem of shipping from n locations to m destinations a product in bulk at minimum cost. However, (Conley, 2009) presents a 400 variable nonlinear transportation problem involving the bulk shipping of a product from 20 factories that produced it to 20 warehouses that will temporarily store it for further shipments at a later date. The subsequent 20x20=400 variable nonlinear problem was balanced in that supply equals demand which is generally true in the long run. Presented here is a larger nonlinear example where demand exceeds supply also solved with multi stage Monte Carlo optimization (MSMCO).

THE 3000 VARIABLE PROBLEM

A large manufacturing company produces a product in bulk at 25 factory locations and wants to ship it at approximately minimum cost to 120 of its best customers. There are considerable discounts for shipping it in quantity so the problem is nonlinear and cannot be solved by the classic linear programming simplex techniques (Mizrahi and Sullivan, 1993). However, multi stage Monte Carlo optimization (MSMCO) simulation will be used to at least produce a good approximate solution that is less than the estimated 17,000,000 euro dollar cost if the problem has to be linearized for ease of solution. The problem is further complicated by demand temporarily exceeding supply. Therefore, the company wants to ship its available production to its 120 good customers at minimum cost for now while it redoubles its efforts to increase production to keep its customers happy. Specifically, the factories currently have

113000,	112000,	111000,	110000,	109000
108000,	107000,	106000,	105000,	104000
103000,	102000,	101000,	100000,	99000
98000,	97000,	96000,	95000,	94000
93000,	92000,	91000,	90000,	89000

units available in factories 1 through 25 respectively to ship to its 120 preferred customers who would like to receive 46,600-100j units for j=1,2,3, …. 120 units respectively, where j is customer j.

Therefore, all of the 2,525,000 units available in the factories will be shipped at a total cost of

$$C = \sum_{i=1}^{25} \sum_{j=1}^{120} .17\,(3i+j) * x(i,j) ** (1.0-.00123 * (5i+j))$$

subject to the 25 equations for the units shipped from the factories and the 120 less than or equal constraints for the number of units arriving at the 120 customers. Therefore x(i,j) is the number of units shipped from factory i to customer j (during the shortage where demand exceeds supply) and * is multiply and ** is raise to a power.

The cost equation will be set equal to some value less than 20,000,000 euro dollars in hopes of driving the cost under the linearalized estimate of 17,000,000 euros. Therefore, we have a 26 equation, 120 less than or equal constraint system. We attempt a solution with MSMCO after transforming it to minimize the sum of the absolute values of the differences of the left and right hand sides of the 26 equations subject to the 120 less than or equal constraints. We set up a 50 stage multi stage MSMCO simulation drawing 50,000 sample answers at each of the 50 stages and storing the best answer (minimum) at each of the 50x50,000 = 2,5000,000 function evaluations. This took 31 minutes of computer time on a desktop PC computer.

The complete answer file of how many units to ship from each factory to each customer 25x1200 = 3000 variable values are in the pattern

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_{16}	"	"	"	"
"	"	"	"	"
"	"	"	"	"
$x_{25,116}$	$x_{25,117}$	$x_{25,118}$	$x_{25,119}$	$x_{25,120}$

where i is the factory it leaves and j is the valued customer it arrives at is available by email to the interested reader upon request. The total minimum cost of the best solution (after 31 minutes of runtime) was 13,547,732 euro dollars, which is well under the 17,000,000 euro old linear standard and should save the company money. The 25 factory equations all had errors of less than 1 unit in this solution. The largest error was equal to .03125 of a unit. All 120 less than or equal constraints were also satisfied.

These 120 total amounts rounded down to nearest whole numbers shipped to the j=1, 2, 3. . . . . 120 customers are in order:

22323, 23183, 19118, 18337, 21715
22835, 19934, 22107, 21990, 18183
22287, 20332, 21474, 19504, 19830
20316, 17800, 17711, 18196, 21414
22928, 26910, 19531, 19870, 18878
25193, 21057, 18537, 22185, 21979
18712, 19042, 18757, 20984, 22256
18494, 18330, 17990, 19206, 15699
22288, 25305, 24903, 18356, 27258
21267, 20899, 23480, 18296, 23226
20460, 18563, 15847, 22086, 26045
19027, 23537, 22307, 14995, 26244
21517, 21261, 19016, 22051, 19518
21972, 24452, 18338, 19557, 21620
26668, 21979, 24626, 21428, 24018
24766, 19657, 19716, 19607, 18365
17837, 20374, 20100, 20288, 21794
23382, 22146, 19454, 16896, 21165
21370, 19773, 22410, 17711, 21458
23969, 16515, 19612, 19471, 24694
18129, 20986, 19967, 24356, 19228
24302, 22489, 22602, 15803, 24183
23686, 19191, 19873, 24326, 20457
19944, 20821, 19565, 20372, 34600

This really is a goal programming problem because if the company has a more powerful computer than a desktop PC it could rerun the 26 equation 120 constraint MSMCO simulation and draw more sample answers and continually try to lower the 13,547,732 euro dollar bound. However, too much lowering might cause the 25 shipping equations to have larger errors. Whether this is acceptable or not would of course be a decision that company management would have to make.

However, MSMCO is fairly useful for approximating solutions to most difficult multivariate nonlinear optimization problems (Conley, 2010) in all fields of science and business.

MSMCO SIMULATION

Multi stage Monte Carlo optimization (MSMCO) makes repeated “random” solution attempts, inside the feasible solution region in an ever repositioning and decreasing in size search region always centered about the best answer so far, in hopes of discovering a trail through (in the case of our example here) 3000 dimensional space to the optimal solution (minimum cost) or a useful approximate solution.

The author used this solution technique with pencil and paper calculations in the 1960s to solve small nonlinear optimization problems. However, the problem presented here would take many years to solve with the MSMCO and pencil and paper calculating. We are not in the 1960s anymore and 45 plus years later a 31 minute run on a small computer produced a useful answer to our problem.

Transportation and shipping involve considerable costs for large companies that ship products in bulk. There are almost always discounts for shipping in quantity that make the problems nonlinear and ideal for MSMCO simulation.

A big company could easily have shipping costs of 100,000,000 euros. If MSMCO only saved them one percent on the old linear way of controlling costs, that is still a savings of 1,000,000 euros.

OTHER TRANSPORTATION PROBLEMS

During a temporary shortage of supply a business might agree to ship at least a fixed minimum amount of product to its best customers. Companies might cut their transportation costs by moving closer to their customers. They might send out their delivery trucks or airplanes on the shortest routes (Conley, 1993) and (Lawler, Lenstra, Rinnooykan and Shmoys, 1985) to save fuel. They also could redesign products so they are cheaper to ship, among other management approaches, to reduce large transportation costs.

CONCLUSION

Presented here was a 3000 variable nonlinear transportation problem of delivering a little more than 2.5 million units of a product from 25 factory locations to 120 important customers where demand temporarily exceeds supply. (This happens in the short term sometimes.) The idea is to send out the units of product in amounts $x_{11}, \ldots .x_{25,120}$ for the 25x120 = 3000 shipping routes so that the total shipping cost (represented by the nonlinear cost equation) is a minimum or at least as low as possible. If the cost equation were linear, various versions of linear programming would produce an exact solution to the attendant oversimplified problem.

However, with the nonlinear cost equation approach presented here and solved with the MSMCO simulation technique we get a useful approximately minimum cost solution which includes the incorporated inherent nonlinear discounts for shipping in quantity that really exist in the world of big business. Again, the exact amounts to ship on each of the 3000 shipping routes $x_{11} \ldots x_{25,120}$ are available upon request by email because the file is a little large to present in its entirety in a short paper. However, the answer is summarized and discussed here.

Even though solving large scale nonlinear problems in business and industry is now more possible than ever before with MSMCO and other simulation techniques, it is still an important management decision to choose either the linear simplex approach or the nonlinear simulation approach. They are both useful tools and techniques for the wise manager who wants to reduce large costs, such as the transportation costs featured here.

Additionally, other types of transportation problems were mentioned in this vital field of transportation and shipping in the global industrial world.

REFERENCES

Conley, W. C. 2010. "We can Save Our World if Worlds Collide." In 2010 Proceedings of Industrial Simulation Conference ISC2010 (Budapest, Hungary, June 7-9) EUROSIS-ETI, Belgium, 45-49.

Conley, W. C. 2009. "A Four Hundred Variable Nonlinear Transportation Problem." In 2009 Proceedings of European Simulation and Modeling Conference ESM2009 (Leicester, United Kingdom, October 26-28) EUROSIS-ETI, Belgium, 124-127.

Conley, W. C. 1993. *Multi Stage Monte Carlo Optimization Applied to a Six Hundred Point Traveling Salesman Problem*, International Journal of Systems Science, pp. 609-626.

Lawler, E., Lenstra, J., Rinnooykan, A., Shmoys, D. 1985. *The Traveling Salesman Problem.* John Wiley and Sons, New York

Mizrahi, A. and Sullivan, M. 1993. *Mathematics for Business Life Science and Social Science,* 5th Edition. John Wiley and Sons, New York.

AUTHOR BIOGRAPHY

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

A Model for Mystery Shipping in Logistics

Matthias Klumpp
Christof Kandel
Sascha Bioly
Institute for Logistics and Service
Management (ild)
FOM University of Applied Sciences
Leimkugelstraße 6
D-45141 Essen, Germany
E-Mail: matthias.klumpp@fom-ild.de
www.fom-ild.de

KEYWORDS
Mystery Shopping in Logistics, Mystery Shipping, GPS-based Real-time Shipment Tracking

ABSTRACT

Transparency, security and sustainability in logistics are major trends with increasing efforts in research and business practice. The existing concept of mystery shopping from service quality assessment can be transferred to logistics as a `mystery shipping` concept testing logistics service providers for their reliability (e.g. delivery in time), transparency (validating their track & trace information) and sustainability documentation (evaluating e.g. their CO_2 calculation per shipment). This concepts would require autonomous GPS tracking devices to accompany the shipments used as mystery shipping test setup and an accompanying software system to evaluate e.g. shipment times and CO_2 emissions per shipment.

1. INTRODUCTION

Modern logistics in global supply chains as well as regional and local transport chains is influenced by several trends: Surveys show demand for more transparency and security in supply chains or cooperation with innovative logistics service providers (see DB Schenker Laboratories 2009). In addition, sustainability is a major trend in logistics and relies in many cases on the measurement and reduction of CO_2 emissions. Therefore carriers have to prove a Product Carbon Footprint (PCF) for their transports per shipment (see Aronsson et al. 2008; Sundarakani et al. 2010). Furthermore the importance of track and trace solutions for cargo increases according to the development of satellite based traceability services in logistics (see He et al. 2009; Carlino et al. 2009). The calculation of CO_2 emissions for each shipment is a challenge because of the complexity of logistics services. A lack of traceability on the part of the customer according to load variation or travelled routes makes it nearly impossible to allocate real CO_2 emissions to each item, especially in groupage freight transports (see Lohre, Herrschlein 2010). Nowadays PCF are calculated by the use of linear distances or with theoretical distances on the basis of road maps. But that is not even close to reality: Detours, by-passes of traffic jams or dynamic routing are not included. A possibility to close this gap in CO_2 calculation between theoretical calculated emissions and actual emitted value is the use of a GPS based track and trace system. Because of the documentation of the travelled routes it is possible to reconstruct the real covered distance to calculate travelled-distance-based CO_2 emissions for each shipment. This increased accuracy in CO_2 calculation could be used in order to evaluate for example the carbon footprint specifications of logistics service providers. This could be drafted as a quality control model according to the mystery shopping concept as described below.

2. THEORY OF MYSTERY SHOPPING

The concept of mystery shopping has been an integral part of quality management and quality improvement concepts especially in service industries for quite some time (see for example Erstad 1998; Wilson 1998; Hudson et al. 2001; Wilson 2001; Cook et al. 2002; Beck, Li 2003; Moriarty, McLeod, Dowell, 2003; Norris 2004; Calvert 2005; Van der Wielea, Hesselinka, Van Iwaarden 2005; Gosselt et al. 2007). The following basic *elements* are defining mystery shopping concepts:

- (i) The fact and also specific details like time and place of mystery shopping processes are hidden from the service provider to be evaluated, especially operational personnel in day-to-day business processes.
- (ii) Sufficient and reliable data has to be collected in a systematic manner during the mystery shopping process in order to allow for comparability and benchmarking options.
- (iii) Evaluation mistakes or deception have to be excluded with the highest possible probability (e.g. by loyalty statements, external providers).

This concept can be transferred to logistics as described in the following chapter in order to arrive at similar fruitful benchmarking an quality improvement options as usually addressed in existing mystery shopping concepts in services as e.g. retailing, banking and other service industries.

3. MODEL FOR MYSTERY SHIPPING

The mystery shipping model in logistics can be adapted by referring to the three mentioned basic elements and adjusting them to logistics services as follows:

- (i) The mystery shipping has to be invisible to the logistics service company. A tracking device for service measurement has therefore to be hidden in existing industrial shipments of roughly the size of europallets or skeleton transport boxes.

- (ii) The tracking and information system used for mystery shipping has to allow for at least 30 independently calculated and stored tracking devices in order to provide a representative picture of overall service and information quality of the tested service provider as only or several single shipments may pose an exception. Furthermore the tracking data has to be very precise in order to guarantee a fair evaluation of the logistics service provider.
- (iii) As in other service industries the mystery shipping could be conducted by external consultants and specialists – and also maybe for trial runs by scientific institutions. These have to saveguard the mystery shipping personnel (e.g. sending companies) and processes from any leaks and deception. Furthermore a sufficient storage and documentation of the tracking data is necessary to allow for credibility and therefore service improvements by the logistics service provider.

4. TESTING CONCEPT WITH GPS TRACKING

As outlined before, the use of hidden GPS tracking devices with an autonomous power supply for the length of the transport process are necessary for implementing a mystery shipping concept in logistics. Since 2011 the ild Institute for Logistics and Service Management of FOM University of Applied Sciences, Essen/Germany, owns a GPS based track and trace system by AIS Advanced InfoData Systems GmbH, Ulm. To use the system and to analyse the results, a laboratory was installed at the institute under the name `GPS.LAB`. The specifications of the system are the following:

- The system contains in the basic setup 40 independent GPS modules to track individual shipments in a logistics transport situation.
- Furthermore these modules have high-capacity rechargeable batteries which make it possible to track a shipment over a period of up to 48 hours depending on the communication rhythm interval (GPRS mobile phone SIM card for data transmission with the mobile phone network).
- This data transmission interval can be defined by different parameters in the system, e.g. by a time interval (e.g. every ten minutes), a transport distance (e.g. every three kilometers) or after a change of direction of a defined angle (e.g. more than 45 degrees).
- The modules are equipped with a high GPS receiving and broadcast performance for receiving satellite position signals even out of boxes and containers (steel hull) in order to allow for transfer and packing of shipments in one vehicle (e.g. truck, also train).
- The data communication ('live tracking') is ensured as described by telecommunication via GPRS (mobile phone network).
- The modules are storing and also communicating a multitude of information about position, altitude, speed and the duration of stops regarding a single shipment in logistics.

By reason of these specifications the tracking data now reaches an accuracy which could not yet be realised so far. Using the software `map & guide` by PTV Planung Transport Verkehr AG, Karlsruhe, the GPS performance data can be mapped as a basis for further analysis and calculations. In fact, with the help of such a system logistic processes would be analysed well-founded as the data is verified by real transport and not by simulation (figure 1).

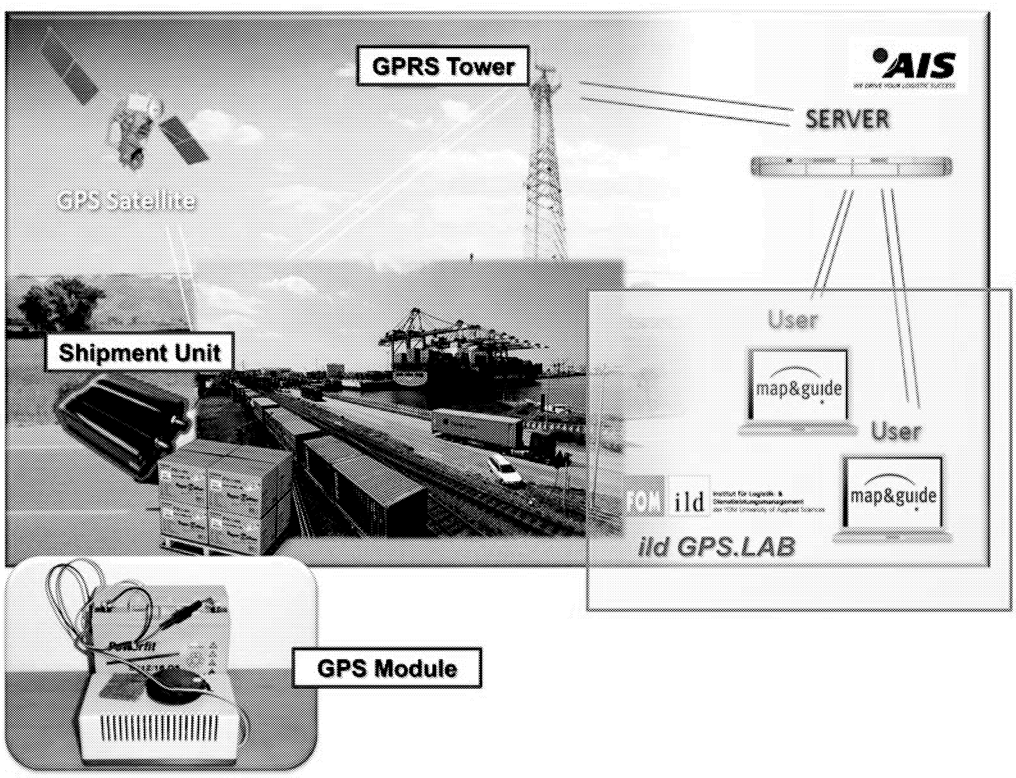


Figure 1: Testing Setup for Mystery Shipping in the ild GPS.LAB

Thererfore due to battery power the GPS.LAB makes it possible to track goods down to the level of pallets, cases, cartons or ideally items during the whole transport to retrace the logistic process in detail and to analyse the performance of supply chains. This will be tested in first trials, when the tracking modules will be tagged at several shipments to make sure that the whole GPS based track and trace system runs with a very high service level and to check the performance of the system for identifying possible applications which are mentioned in the following paragraph.

The first trial results runs for the new GPS.LAB will be presented to demonstrate the benefit of using a real-time track and trace system in production alliances and supply chains according to the following examples in order to evaluate logistics service providers against their promises:

- The consequence of goods tracking is a route of transport tracing, so that a tour can be illustrated on a roadmap with software support. Because of this presentation it is easier to understand complex transport chains, for example to divide a groupage freight tour in forerun, mainhaul and post-carriage distribution. Furthermore a specific hub and spoke structure of a logistics network or the position of delivery on the last-mile can be considered in the context of a PCF calculation. This will be the main using context of the system.
- Moreover the real-time tracing of the shipments makes it possible to estimate the arrival of goods in a production network with several locations, like a concurrent enterprising alliance, so that production planners get a better support, for example to synchronise the delivery of material and preliminary products with the production scheduling. With the information about position and speed, conclusions can be drawn according to the time of arrival at the production. In addition the storage of historic routes provides extra information and supports production planners in future planning challenges.
- A conceivable application for the introduced GPS system in logistics practice is due to the context of 'Supply Chain Event Management' (SCEM). Existing concepts basically focus - among other things - on real-time information (see Klumpp et al. 2010). Thereby the task of such systems mostly lies in realizing data along a supply chain as a prerequisite for (semi-)automated management systems. Supply Chain Management (SCM) needs a smooth information flow for an efficient functionality within a delivery chain (see Nissen 2002) and SCEM has to ensure a permanent monitoring of material and goods flow along the entire chain to make coordinated management action possible (see Beckmann 2003).

Due to the fact that GPS at present is mostly used in logistics for fleet management, the tracking functionality is restricted to a single vehicle. A shipment-based version would enforce the range of functions/features of real-time event management. Regarding the desire of rising the percentage of multimodal transports and the thereby caused carrier exchanges, a real-time determination of the place of residence of goods is possible by using shipment-based GPS-systems with many feasible advantages like e.g. true geofencing. So a GPS-based track and trace system as described above is required for a practical implementation of the drafted mystery shipping model in logistics.

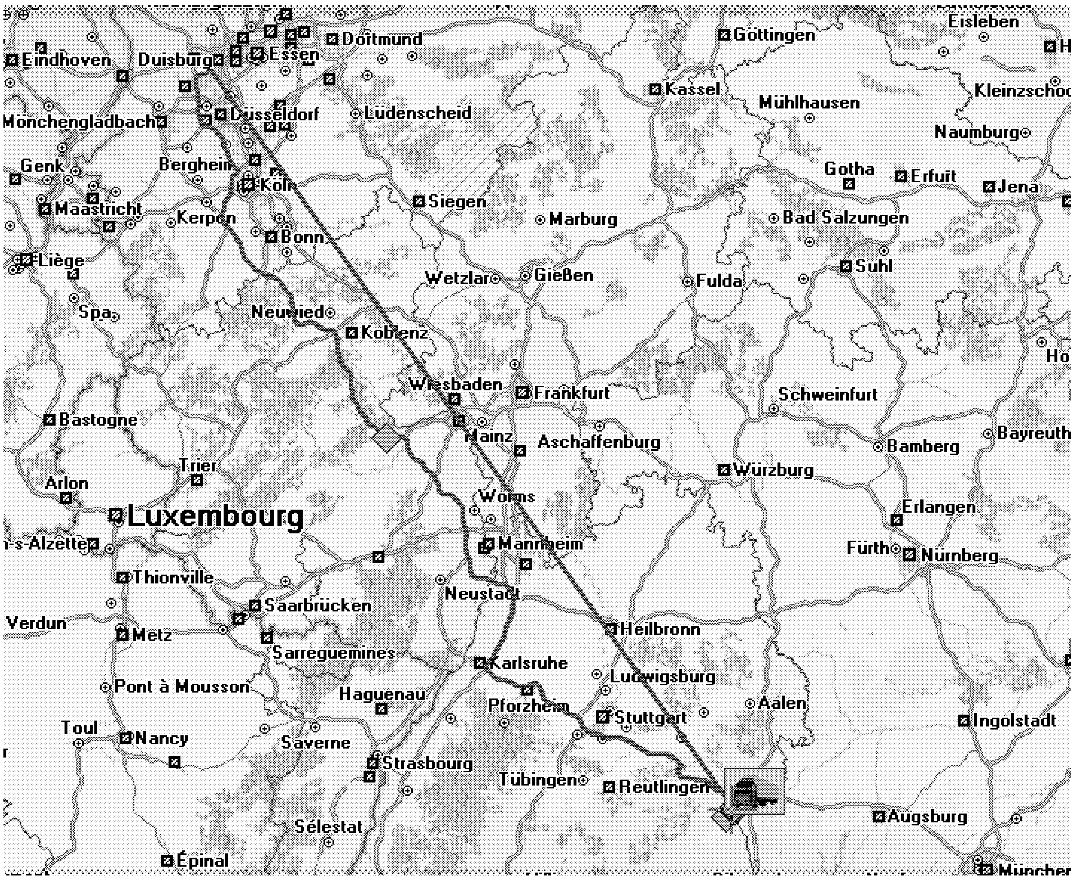


Figure 2: Mystery Shipping Test Run Duisburg-Ulm

5. TESTING RESULTS

For testing purposes a simulation mystery shipping from Duisburg to Ulm in Germany was scheduled and implemented with the help of the German logistics service provider DB SCHENKER as depicted in the above positioned figure 2.
Actually DB SCHENKER did not know about the test run (on an operational level by process-related personnel).

6. EVALUATION OPTIONS

Naturally, as outlined above, many different aspects of a service as e.g. a transport service can be tested and evaluated by GPS-based `mystery shipping`. But several of these generally possible objectives can be highlighted as specific data from the GPS measurement seems to implicate their use:

- Check of start, transport and delivery times (comparison to track & trace data provided or other forms of information from the LSP).
- Check of transport routing (especially in cases for example with exact kilometer-based invoicing as often the case with special, time-critical and courier transports).
- Check of delivery and geographic status in case of transport hindrances.
- Check of eco-performance measures such as GHG emission figures by an exact route-based approach as developed below in table 1. As this first test example shows, an evaluated difference of CO_2 emissions of 33,78% has been shown for the shipment from Duisburg to Ulm.

Pos	km total	km diff.	vehicle	EC empty [l Diesel]	EC loaded [l Diesel]	utility max [pallet]	utility actual [pallet]	CO_2e-emisions [kg]	*Info*
1	525,0	525,0	40t (Euro 4)	20,9	31	36	36	14,349	*DB DU --> DB UL*
2	530,5	5,5	7,5t (Euro 4)	15,5	19,4	15	15	0,226	*DB UL --> stop 1*
3	533,1	2,6	7,5t (Euro 4)	15,5	19,4	15	14	0,113	*stop 1 --> stop 2*
4	534,4	1,3	7,5t (Euro 4)	15,5	19,4	15	13	0,060	*stop 2 --> stop 3*
5	536,4	2,0	7,5t (Euro 4)	15,5	19,4	15	12	0,098	*stop 3 --> AIS*
								14,846	***Shipment based CO_2e emissions***
Air distance	*406*	*406*	*40t (Euro 4)*	*20,9*	*31*	*36*	*36*	***11,097***	***Air distance based shipment CO_2e emissions***
			Difference					**3,749**	***33,78% deviation***

Table 1: GHG Emission Calculation on GPS Tracking Basis for Eco-Evaluation

7. CONCLUSION AND OUTLOOK

The general idea of mystery shipping in logistics holds many promises of evaluating service quality, transparency and also sustainability parameters communicated today by logistics service providers to their customers. Further research has to establish the practical feasibility of this quality assessment concept in logistics with the described facilities of the ild GPS.LAB. This is planned for test implementations with several logistics companies in Germany during 2011 and 2012. These further test results will be re-evaluated and integrated into the described mystery shipping model in order to allow for a sustained and high-quality improvement process for a successful implementation and use in the logistics industry in the future.
Besides this streamlining of the suggested model a lot of supplementary functions are generally possible and can briefly be outlined (see also Wang, Potter 2008; Stopka 2009):

- An auditing approach by a (new) third party institution especially regarding the CO_2 documentation of logistics service providers (e.g. as the German ‘Stiftung Warentest’ or other public auditing institutions).
- Furthermore real-time tracking data could contribute to dynamic scheduling and quality assurance in production and production networks, especially in the new evolving global production supply chains (see e.g. Brewer, Sloan, Landers 1999; Meers, Hennes, Nyhuis 2010; Wannenwetsch 2010).
- Also such audits and even rankings of logistics service providers are feasible on the basis of the described real-time tracking – showing hubs and transport networks and maybe even enabling social audits by the way of evaluating transit spaces, countries and regions (e.g. if one LSP has established an international hub in a specific low-cost-country location, see Hillbrand, Schoech 2007).
- And even further in the future even short-term production networks could decide about their specific set-up in the light of real-time GPS-tracking information of shipments (see Kärkkäinen, Ala-Risku, Främling 2004).

REFERENCES

Aronsson, H., Huge, Brodin, M., Kohn, C. (2008). Logistics Structures – Drivers of Environmental Impact. In: Stentoft, Arlbjørn, J., Halldórsen, Á., Jahre, M. et. Al. (Eds.): Northern Lights in Logistics & Supply Chain Management, Narayana Press, Gylling, 2008, 183-200.

Beck, J., Li, M. (2003). Mystery shopping in lodging properties as a measurement of service quality. Journal of Quality Assurance in Hospitality & Tourism, 4(1/2), 1-21.

Beckmann, H. (2004). Supply Chain Management. Springer, Berlin, 2004.

Brewer, A., Sloan, N., Landers, T. (1999). Intelligent Tracking in Manufacturing. Journal of Intelligent Manufacturing 1999, Issue 10, 245-250.

Calvert, P. (2005). It's a mystery: Mystery shopping in New Zealand's public libraries. Library Review, 54(1), 24-35.

Carlino, M., Confessore, G., Liotta, G. (2009). Potentials of satellite-based traceability services in logistics. In: Proceedings of the European Navigation Conference – Global Navigation Satellite Systems 2009, Naples, Italy, 1-6.

Cook, L. S., Bowen, D. E., Chase, R. B., Dasu, S., Stewart, D. M., Tansik, D. A. (2002). Human issues in service design, Journal of Operations Management, 20(2), 159-174.

DB Schenker Laboratories (2009). Ergebnisse der Messebefragung „transport logistic", Berlin, 2009.

Erstad, M. (1998). Mystery shopping programmes and human resource management. International Journal of Contemporary Hospitality Management, 10(1), 34-38.

Gosselt, J. F., van Hoof, J. J., de Jong, M. D. T., Prinsen, S. (2007). Mystery Shopping and Alcohol Sales: Do Supermarkets and Liquor Stores Sell Alcohol to Underage Customers?, Journal of Adolescent Health, 41(3), 302-308.

He, W., Tan, E.L., Lee, E.W., Li, T.Y. (2009). A solution for integrated track and trace in supply chain based on RFID & GPS. In: Proceedings of the Emerging Technologies & Factory Automation 2009, Mallorca, Spain, 1-6.

Hillbrand, C., Schoech, R. (2007). Shipment Localization Kit: An Automated Approach for Tracking and Tracing General Cargo. In: Proceedings of the Sixth International Conference on the Management of Mobile Business 2007. Toronto, Ontario, Canada, 46-52.

Hudson, S., Snaith, T., Miller, G. A., Hudson, P. (2001). Distribution Channels in the Travel Industry: Using Mystery Shoppers to Understand the Influence of Travel Agency Recommendations. Journal of Travel Research, 40(2), 148-154.

Kärkkäinen, M., Ala-Risku, T., Främling, K. (2004). Efficient tracking for short-term multi-company networks. International Journal of Physical Distribution & Logistics Management, 34(7), 545-564.

Klumpp, M. et al. (2010). Green Supply Chain Event Management. In: Lencse, G., Muka, L. (Eds.): 8th Industrial Simulation Conference 2010 (ISC´2010 Budapest, Hungary), EUROSIS-ETI, Belgium, 195-200.

Lohre, D., Herschlein, S. (2010). Grüne Logistik. Studie zu Begriffsverständnis, Bedeutung und Verbreitung „Grüner Logistik" in der Speditions- und Logistikbranche. Institut für Nachhaltigkeit in Verkehr und Logistik, Hochschule Heilbronn, Bonn.

Meers, S., Hennes, G., Nyhuis, P. (2010). Logistische Herausforderungen in Produktionsnetzen. In: Zeitschrift für wirtschaftlichen Fabrikbetrieb 2010, Issue 11, 949-952.

Moriarty, H., McLeod, D., Dowell, A. (2003). Mystery shopping in health service evaluation. The British Journal of General Practice, 53(497), 942-946.

Nissen (2002). Supply Chain Event Management. In: Sinz, E.J., Mertens, P. (Eds.): Wirtschaftsinformatik 44(5), Vieweg, Wiesbaden, 477-480.

Norris, P. (2004). Reasons why mystery shopping is a useful and justifiable research method. Pharmaceutical Journal, 272(7303), 746-747.

Stopka, U. (2009). Herausforderungen und Potenziale von Mobilfunk-, Ortungs- und Navigationsdiensten in Güterverkehr und Logistik. In: Wissenschaftliche Zeitung der Technischen Universität Dresden 2009, 1-2, 81-89.

Sundarakani, B., de Souza, R., Goh, M. et al. (2010). A Sustainable Green Supply Chain for Globally Integrated Network. In: Wang, L., Koh, S.C, (Eds.): Enterprise Networks and Logistics for Agile Manufacturing, Springer, London, 191-206.

Van der Wielea, T., Hesselinka, M., Van Iwaarden, J. (2005). Mystery shopping: A tool to develop insight into customer service provision. Total Quality Management & Business Excellence, 16(4), 529-541.

Wang, Y., Potter, A. (2008). The application of real time tracking technologies in freight transport. In: Proceedings of the Third International IEEE Conference on Signal-Image Technologies and Internet-Based System 2007. Shanghai, 298-305.

Wannenwetsch, H. (2010). Integrierte Materialwirtschaft und Logistik. Beschaffung, Logistik, Materialwirtschaft und Produktion. Springer, Berlin, Heidelberg, 2010.

Wilson, A. M. (1998). The role of mystery shopping in the measurement of service performance. Managing Service Quality, 8(6), 414-420.

Wilson, A. M. (2001). Mystery shopping: Using deception to measure service performance. Psychology and Marketing, 18, 721-734.

Simulation of logistic decisions within a freight transportation model

Tabitha Maes
Katrien Ramaekers
An Caris
Gerrit K. Janssens
Tom Bellemans

Transportation Research Institute, Hasselt University
Wetenschapspark 5 - Bus 6
B-3590 Diepenbeek, Belgium
E-mail:{tabitha.maes;katrien.ramaekers;an.caris;gerrit.janssens;tom.bellemans}@uhasselt.be

ABSTRACT

This paper presents a conceptual framework of a freight transportation model. The focus of the framework lies on simulating the logistic decisions of firms. In this paper, a first application is discussed on a small test sample of ten zones. Inside is gained into the working of the model and into potential problems when implementing the framework on a country-size scale.

INTRODUCTION

In a growing globalised context and consumption economy, freight transport is of crucial importance. Activities of firms are expanding, even across borders. This causes changes in the logistics activities of firms as they become more dynamic. Public and private decision makers need to take these trends into consideration with regard to their decisions and a better projection of freight traffic flows becomes necessary. Being able to understand the drivers of freight flows makes it possible to forecast freight flows in the future and to calculate the impact of different policies on freight traffic. It will put policymakers in the position to get a better insight in the way the transport of goods comes about. Still, freight demand modelling is lacking behind on the efforts made in passenger transport models. Most freight demand models are not considering logistic aspects, such as inventory policy, shipment size and mode choice, which are connected to the creation of freight flows (Tavasszy et al. (1998) and Liedtke (2009)).

Recent trends in freight modelling are moving to agent-based models, which are part of the group of activity-based models and focus on each freight agent separately (Liedtke and Schepperle, 2004). They are better able to model individual operational decisions and interactions concerning logistics and transports. Furthermore a disaggregated approach is applied, by looking at trips and decisions on a microscopic scale and no longer to aggregate flows between different zones. This enables the understanding and representation of roles that each actor plays in the freight transportation system, as also the interactions between actors. Further, it is possible to incorporate changes in actors and their interactions over time. These elements are of fundamental importance in the development of more behavioural models for the freight system (Roorda et al., 2010). The disaggregated approach of these models, together with the representation of the different actors, enables better modelling possibilities for logistics decisions.

In this paper, a freight transportation model framework is presented and applied on Flanders, the northern region of Belgium. With this framework a first attempt is made to predict future freight flows and the effect of certain policies. Also, the integration of logistic decisions made by the different actors involved in the creation of freight flows can be simulated. The paper is organized as follows. In the next section a framework for freight transportation modelling is presented. The logistic decisions module of this framework is worked out further in section 3 and applied to the region Flanders. Finally, conclusions are drawn and options for future work are suggested.

A NEW FREIGHT TRANSPORTATION MODEL

As basis for the microscopic, agent-based freight transportation model for Flanders, four building blocks are used. These represent the four main steps in the freight transportation model, i.e.: production and attraction, distribution, the logistic transport decisions and the assignment to the network. The four building blocks used are closely parallel to the structure of the traditional four-step transportation model (de Jong et al., 2004).

The implementation of these steps may differentiate

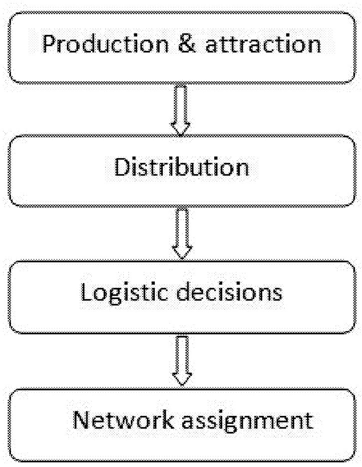


Figure 1: Steps in the freight transportation framework

substantially between different freight transportation models (Pendyala et al., 2000). In the next paragraphs, the four steps of our framework are shortly described.

Production and attraction: This step allows creating a clear link with the economy. Firms may be generated, whereby location, economic activity, size and other attributes are assigned to a firm. Also the creation of multi-establishments firms may be considered, to arrive at a better representation of reality. In these multi-establishments firms there are close interactions and co-operations between the establishments. Furthermore, firms' annual demand and supply are simulated.

Distribution: The supply and demand of the different firms is matched with each other. This involves the choice of a supplier for each firm, as well as the quantity to be purchased. The result of this step is the annual commodity flow between pairs of firms and is represented by Production/Consumption (PC) matrices.

Logistic decisions: In this step the order quantity, frequency and transport mode are chosen. Also whether or not to outsource and the use of carriers and forwarders are modelled. More detail may be found in the next section, where the decisions are worked out on an example region.

Network assignment: The scheduling and routing of individual shipments onto the network is modelled. Also empty trips are accounted for, this is often overlooked in freight transportation models. The impact of various constraints such as equipment and link capacities has to be looked into. Furthermore, different techniques to assign flows to the network may be considered.

The key objective of the framework is to have a model that includes the simulation of logistic decision making. This framework needs to be able to give a more realistic representation of freight flows in Flanders than existing models. The agents considered for this model are the firms who are sending and receiving the goods, respectively called sender and receiver. Next, there are the carriers who actually undertake the transport. Finally, forwarders are modelled, who may be responsible for the entire organization and execution of the transport and which may cooperate with carriers.

As stated by Holguin-Veras et al. (2011) it is important in freight transport to make a clear distinction between the generation of demand and the generation of traffic. The generation of freight demand is determined by the economics of production and consumption. Freight trips, on the other hand, are the output of logistic decisions. The greatest gap in many existing models is in the modelling of logistic decisions. Most frequently a rate is used to determine the link between freight demand and freight traffic flows. To improve this link, the focus of this research will be on the logistic decisions module.

CONCEPTUAL LOGISTIC FRAMEWORK - EXAMPLE

In this section, only step three "logistic decisions" of the framework in figure 1 will be further elaborated. This will start after the distribution step and will take the PC flows between the firms as given. Also the network assignment will be left out of this paper. As for now, the PC flows are on a zonal level, because the production and attraction and also the distribution step are not yet in place. Instead a disaggregation step is inserted, to create firm-to-firm flows. This is demonstrated in the following subsection for the region of Flanders.

In figure 2 the different steps of the logistic decision module are shown. Starting from PC flows the sender/receiver relations become clear. One of these two agents will be responsible for arranging the transport of the goods. For this, he can rely on a carrier or forwarder, or execute the transport himself. Also decisions about shipment size and transport chains have to be handled. In a first pre-processing step the transport chain possibilities are built, after which the sender or receiver calculates a shipment size and chooses the appropriate transport chain. Due to possible interactions with a carrier (price negotiations) or a forwarder (complete planning of the transport), consolidation options can be integrated and long term contracts may be achieved.

In the rest of this section, the different steps of the

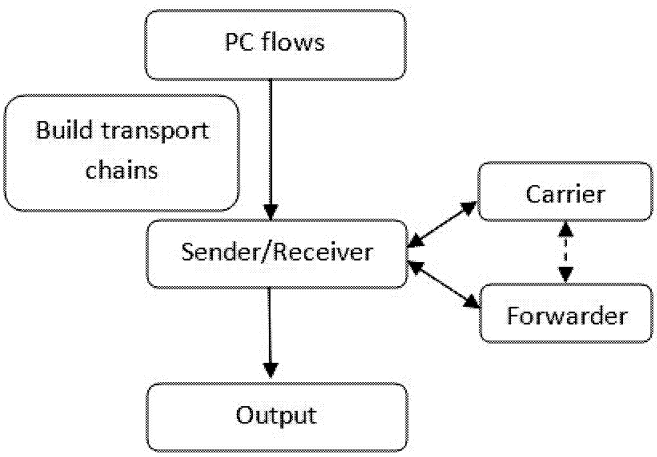


Figure 2: Steps in the logistic decisions module

logistic decisions module are explained. However, the interaction with carriers and forwarders is not yet incorporated. The goal is to include these agents, so that transport rate negotiations between sender/receiver and carriers can be simulated. Hence, a firm can choose between the service of a carrier or use, if present, his own transport fleet. Furthermore, an option will be inserted to rely on forwarders for the organization and execution of the transport orders. This implies that a forwarder will be held responsible, not solely for carrying out the transport, but also for the choice of an appropriate transport chain and optimal shipment size.

The region Flanders in Belgium

To illustrate the conceptual logistic decisions model, it is applied to the region of Flanders, the northern part of Belgium. The 308 communities in Flanders are used as zones in the model. For import and export one zone is created with one production firm and one consumption firm. However, to keep calculation efforts within reach a first example is applied on only ten communities selected from the 308 communities in Flanders. Selecting only ten zones allows the testing of the logistic decisions module, while limiting the calculation efforts.

To start our model, information is needed on the PC flows between the different zones. These flows are divided between nine matrices corresponding to the nine different NSTR (category 7 and 8 are combined) commodity categories used. The NSTR classification is a standard goods classification for transport statistics composed of ten categories, which is often used in Europe. Furthermore, data on the network structure of the different transport modes and the location of the terminals is needed.

In the example the outsource options are currently not applied, so interactions between carrier and forwarder are left out of the example. Future research will integrate these actors. For now, the main goal is to get inside into potential problems within the framework, before implementing it on a larger scale. For that reason only a small test sample is used.

Disaggregation step

To integrate logistic decisions in a freight transportation model, the logistic decisions are best modeled on a microscopic scale. This allows the representation of the different agents and their decisions. Therefore, in a first step the data of the PC flows, which are on an aggregated zonal level, need to be disaggregated to firm-to-firm flows. To do this, the steps of the Aggregate-Disaggregate-Aggregate model (ADA-model) of Ben-Akiva and de Jong (2008) are used. The ADA-model is an activity-based model, which has its main decision protocol at a disaggregated level instead of an aggregated level.

Between two zones only a fraction from all the potential relationships, between senders and receivers of a certain commodity, is actually realized. To calculate the actual number of firm-to-firm (F2F) relations of commodity k between two zones (r and s), the following formula is used:

$$F2F_k\ relations = F_k \cdot P_k \cdot C_k$$
$$With,\ F_k = \frac{N_k}{\sum_s C_k}$$

Where, F_k is the fraction of actually realised links between a senders and receivers of two zones, P_k is the number of producers of commodity k in a zone r, C_k is the number of consumers of commodity k in a zone s and N_k is the number of receivers per sender for commodity k.

Applied to Flanders
Suppose a yearly zonal flow between Bruges and Genk of 60 tonnes of agricultural products. Assume 30 agricultural producers in Bruges and 5 agricultural receivers in Genk. This leads to a total of 150 potential relationships, which will not all be realised. Considering 20 receivers per sender, the actual number of relations can be calculated.

$$F_k = \frac{20}{(308 * 5)} = 1,3\%$$

$$\text{F2F}_k\text{relations} = 1,3\% \cdot 30 \cdot 5 = 1,95$$

This leads to two firm-to-firm relations in the agricultural sector between Bruges and Genk. These two relations will be selected at random from the 150 potential relationships. Then the 60 tonnes of yearly flow are proportionally divided between the two realised links, according to the size of the firms involved.

Data needed in this step, for each commodity category are:

- The number of senders in a zone
- The number of receivers in a zone
- The number or receivers per sender
- Indication of the size of firms (either measured in number of employers, or turnover, or . . .)

As output of this step flows are given at a firm-to-firm level for each of the nine commodity categories and for each combination of two zones. For our small example of ten zones, this results in 900 matrices.

Pre-processing step: Build transport chains

Within the logistical decisions module is a separate step for the creation of the different transport chains. It is run separately, before the logistic decision making process. This pre-processing step is based on Ben-Akiva and de Jong (2008).

The module makes a Total Logistic Cost (TLC) function for each of the transport chains. The TLC function exists of an ordering cost, inventory cost, capital cost of the goods in transport and in inventory and transport cost. The first five cost components depend on the commodity category. The transport cost is carrier specific and is split into several components:

- Variable link-based cost composed of a distance and time-based cost
- Transhipment cost between the different transport modes
- Loading and unloading cost which is vehicle/vessel specific
- Commodity specific cost (ex.: cargo fees)

The TLC of commodity k transported between firm m in zone r to firm n in zone s, with a shipment size q using transport chain l is:

$$G_{rskmnql} = O_{kq} + T_{rskql} + Y_{rskl} + I_{kq} + K_{kq}$$

The optimal transfer points within each predefined transport chain (possible combinations between the transport modes) are determined, based on the TLC function. The module will test the different possibilities of transfer points in each predefined transport chain and choose the transfer point that has the minimum TLC. This procedure will be executed for each combination of zones and commodity types. Afterwards these transport chains will be used by the companies depending on the zone in which they are located, the commodity goods they are transporting and the availability of a transport mode. In this step, an average transport rate is used, as well as the most common vehicle/vessel type, for a commodity category. The transport rate can be later adjusted by simulating price negotiations with carriers, to arrive at the real transport cost. Because these price negotiations depend on the barging power of the firm, they are not included in the pre-processing step. The results of the pre-processing step are used as an initial starting point for all the firms.

Applied to Flanders
The transport modes used for Flanders are road (light road and heavy road), rail, inland waterways and sea (30 TEU, 100 TEU and 200 TEU). Transport by air is only considered for export and import. The load capacity of each of these transport modes is known for the different commodity categories. In this limited example only one vehicle/vessel type per transport mode is considered.

With these transport modes, transport chains are created. In total 30 chains combinations are possible. These chains may be divided over ten categories, represented in table 1. In the first category (road), heavy road consolidated is considered as a third option in addition to light road and heavy road (unconsolidated). This leads to six possible transport chain combinations, with the three road transport modes. For Flanders, the last category is only used for import and export.

Table 1: Transport chains categories

Category	**Number of transport chains**
Road	Six combinations
Direct rail	One direct link
Road-Rail	Four combinations
Road - Inland waterways	Four combinations
Direct sea	One direct link
Road - Sea	Four combinations
Sea - Inland waterways	Two combinations
Road - Sea - Rail	Two combinations
Sea - Inland waterways - road	Two combinations
Road - Air	Four combinations

The chains that do not start or end with road transport are only possible for zones where a rail/inland waterway terminal is located.

In the pre-processing step transport chains are built between the different zones taken into account average shipment sizes for each commodity category. Later these shipment sizes are adjusted to the specifications of the yearly transport orders between firms, for each specific firm-to-firm relation. Not all transport chain options are allowed for every combination of zones and commodity category. Between some zones certain options will not be available. Two adjacent zones both without a sea or inland waterway connection, will only be able to include transport chains of category one, two, three and ten.

In the rest of this section the TLC calculations are shown. As an example, this is worked out for the transport of agricultural goods between Bruges and Genk, and only for the heavy road - rail - heavy road transport chain. An average shipment size of 1,48 tonnes is used to determine the transfer points. The yearly demand is 30 tonnes for each of two realised links. From all the possible transfer points the terminals of Meerhout and Ostend leads to the lowest TLC, so these transfer points are used in this example.

Table 2: Symbols

Symbol	Description	Value
o	Order cost	55 €
Q	Yearly demand (in ton)	30 tonnes
q	Shipment size (in ton)	1.48 tonnes
f	Frequency	21 shipments
D_{ph}	Distance pre haulage	34.13 km
D_{mh}	Distance main haulage	190.12 km
D_{eh}	Distance end haulage	27.6 km
TC_{hr}	Transport cost heavy road	1 €/km
TC_r	Transport cost rail	15 €/km
Cap_r	Capacity rail	1200 tonnes
L_{hr}	Cost to (un)load heavy road	2 €/ton
L_r	Cost to (un)load rail	0.4 €/ton
TT	Travel time (in hours)	26.02 hours
d	Interest rate (per year)	4%
v	Value of the goods	442 €/ton
w	Warehouse cost (per year)	20%

First the frequency of the shipments is calculated and rounded up to the next integer number. This leads to a frequency of $Q/q = 30/1.48 \approx 21$ shipments.

The order costs is $o * f$ or $55 * 21 = 1155$€.

To calculate the transport cost, the following formula is used:

$$[D_{ph} * TC_{hr} + D_{mh} * TC_r * [q/(0.75 * Cap_r)]+ \\ D_{eh} * TC_{hr} + q * (4 * L_{hr} + 2 * L_r)] * f$$

The assumption is made that for rail (IWW, sea) transport the transport cost are proportional to the actual capacity used. Starting with a 75% fill level, this can be later on adjusted when considering different consolidation options. This assumption is only valid if the rest of the train/vessel is occupied by other firms and the cost can be shared. The total transport cost is found in table 3.

The next cost component is the capital cost of goods in transit. To determine the time that goods are in transit a sum is made of the travel time (heavy road and rail transport) and the waiting time for a vehicle or train to be available. Further, the interest rate and value of the goods is needed. This gives the following formula:

$$\frac{TT * d * v * Q}{365 * 24}$$

This cost component has only a limited influence on the TLC. Especially, for small shipment sizes with low value density, this cost is negligible.

For the inventory cost, the cost to store half of the shipment size is calculated based on a percentage of the goods value:

$$\frac{q}{2} * w * v$$

Finally, the capital cost of the inventory is calculated. This based on the same interest rate as for the capital cost of goods in transit.

$$\frac{q}{2} * d * v$$

A summary of all the costs is given in table 3. This procedure has to be repeated for all the possible transport chains and their transfer points. Only the transfer points with the lowest TLC are maintained for the next step.

Table 3: TLC: heavy road - rail - heavy road

Order cost	1155 €
Transport cost	1610,34 €
Capital cost of goods in transit	1,57 €
Inventory cost	65,42 €
Capital cost of inventory	13,08 €
TLC	**2845,41 €**

The data necessary in this step are, besides the cost components presented earlier:

- The average value of the goods in the different commodity categories.
- The average shipment size of each commodity category

The result of this step is a list of 30 transport chains between each combination of the ten zones and for each of the nine commodity categories. This leads to 12.150 possible transport chains built, depending on the availability of a rail/inland waterway terminal for each zone. For each of these transport chains the optimal route and transfer points are set, based on an average shipment size. One problem encountered with this method, is that it is dependent on the initial shipment size used. At the moment, for each transport chain only the transfer points that leads to the lowest TLC are maintained. Starting with a small shipment size sometimes leads to favouring short main haul distances, when consolidation is not possible. If the shipment size is changed in a later step, the chosen transfer points may no longer be optimal.

Determining an optimal shipment size and choosing a transport chain

PC flows at a firm-to-firm level represent the shipments between senders and receivers. For each shipment, given as a yearly flow, sender and receiver have to decide who is responsible for the transport. The agent that will be responsible for the transport has to decide on the transport chain used and the appropriate shipment size.

For each of the available transport chains created in the 'build transport chain' step, a shipment size is determined. This is done by minimizing the TLC function for each chain. As for now, the interaction between firms and carriers or forwarders is not considered. This implies that the transport rate is the same for each firm. In a later addition to the model the interaction between carriers will be integrated. This allows representing price negotiations and different transport rates depending on the offer bids of the carriers.

After the determination of a transport chain and shipment size for each PC flow, consolidation options may be investigated. Based on common legs in the transport chains of different firms, economies of scale may be achieved by consolidating shipments.

Applied to Flanders
For our example of transport of agricultural commodities from Bruges to Genk the selected firms have to determine who is going to be responsible for the transport. Mostly, this will be the sender, or the firm that owns a personal transport fleet. Based on the list of the possibly 30 calculated transport chains, from the preprocessing step, between Bruges and Genk and for the agricultural commodity category, calculations are made.

For each of these transport chains the optimal shipment size is determined based on their yearly flow and by minimizing TLC. Afterwards consolidation options within the different transport modes, for each leg of the transport chain are considered.

In the heavy road - rail - heavy road transport chain between Bruges and Genk with the rail link between Meerhout and Ostend, the optimal shipment size is determined by minimizing the TLC. This is done for each firm-to-firm link. Suppose that one of the two links has a yearly demand of 30 tonnes. Then a shipment size of 10 tonnes gives the lowest TLC, 1165,91 €. So if in the final decision it is opted to use the heavy road - rail - heavy road transport chain, the yearly demand should be divided into shipment sizes of 10 tonnes, with a frequency of three shipments per year.

Output of the logistic module

The agent responsible for the transport will determine the transport chains and the respective optimal shipment sizes. After he has calculated the options on each chain and has considered consolidation with other firms, he has a list of possible transport solutions. Based on the minimum TLC of the transport chains, the firm will choose the optimal transport solution. The output of the logistics module is on an Origin - Destination (OD) level, containing the following information:

- The shipment size
- The transport chain chosen
- Which firm (sender or receiver) is responsible for the transport
- Total logistic cost of the transport
- Whether or not consolidation has taken place

This information can be used for the final step of the freight transportation framework, to assign the freight flows onto the network.

CONCLUSIONS AND FUTURE WORK

In this paper the logistical decisions module of a new freight transportation framework is presented and applied for a small selection of communities in the region of Flanders, Belgium. One of the advantages of this freight transportation framework in regard to other existing models is the integration of logistic decisions on a firm level.

Although the framework is still in a conceptual phase some points for attention can be formulated. Difficulties are to be expected when enlarging the scale of the model, due to an exponential growth in calculations and processing time. This needs to be addressed before the framework can be implemented on a country-size scale. Another difficulty is that the model requires a lot of data. Ways to manage and gather this data need to be looked into.

Future work is needed to evaluate different options to determine an optimal shipment size and transport chain choice. One option is to see if discrete/continuous models, as described in Cavalcante and Roorda (2010) may be applied to multimodal transport chains. This will allow the joint analyse of the mode choice decision with the shipment size, shipment frequency. Another possibility is evaluating different utility maximization techniques, such as recently done by Samimi et al. (2011). Adaptations to the calculation method of the TLC, have to be considered, for example using variable transport costs instead of a constant transport rate. Also the consolidation options have to be further worked out. Finally, the integration of carriers and forwarders needs to be considered. When these aspects are in place, the example can be expanded to all 308 communities in Flanders, to see how the model will behave in a real network.

REFERENCES

Ben-Akiva M. and de Jong G., 2008. *The Aggregate-Disaggregate-Aggregate (ADA) Freight Model System.* In M. Ben-Akiva; H. Meersman; and E. Van de Voorde (Eds.), *Recent Developments in Transport Modelling: Lessons for the Freight Sector*, Emerald, UK. First ed.

Cavalcante R. and Roorda M., 2010. *A disaggregate urban shipment size/vehicle-type choice model.* In *TRB 2010 Annual Meeting CD-ROM.*

de Jong G.; Gunn H.; and Walker W., 2004. *National and international freight tranpsort models: an overview and ideas for further development. Transport Reviews*, 24, no. 1, 103–124.

Holguin-Veras J.; Jaller M.; Destro L.; Ban X.; Lawson C.; and Levinson H., 2011. *Freight generation, freight trip generation, and the perils of using constant trip rates.* 90th annual meeting of the Transportation Research Board, Washington DC.

Liedtke G., 2009. *Principles of micro-behavior commodity transport modeling. Transportation Research Part E*, 45, 795–809.

Liedtke G. and Schepperle H., 2004. *Segmentation of the transportation market with regard to activity-based freight transport modelling. International journal of logistics: Research and applications*, 7, no. 3, 199–218.

Pendyala R.; Shankar V.; and McCullough R., 2000. *Freight travel demand modeling. Synthesis of approaches and development of a framework. Transportation Research Record*, 1725, 9–16.

Roorda M.; Cavalcante R.; McCabe S.; and Kwan H., 2010. *A conceptual framework for agent-based modelling of logistics services. Transportation Research Part E*, 46, 18–31.

Samimi A.; Kawamura K.; and Mohammadian A., 2011. *A disaggregate analysis of rail-truck mode choice behaviors for freight shipments.* 90th annual meeting of the Transportation Research Board, Washington DC.

Tavasszy L.; Smeenk B.; and Ruijgrok C., 1998. *A DSS for modelling logistic chains in freight transport policy analysis. International Transactions in Operational Research*, 5, no. 6, 447–459.

MULTI-CRITERIA EVALUATION OF MULTI-MODAL TRAFFIC SYSTEMS BY DISCRETE-EVENT SIMULATION

Christian Hillbrand and Susanne Schmid
V-Research GmbH
Industrial Research and Development
Stadtstrasse 33, 6850 Dornbirn, Austria
E-mail: christian.hillbrand@v-research.at, susanne.schmid@v-research.at

KEYWORDS
Logistics, simulation, evaluation, green logistics, multi-modal transport.

ABSTRACT

For an efficient utilization of multi-modal transport networks, usually high shipment volumes with very little variability are necessary. Only a few shippers are able to provide such constant flows of goods. In all cases not meeting these ideal conditions, a considerable need for co-ordination between multiple shippers and/or carriers along the multi-modal transport chain arises. From an analytic point of view, the basic conditions for this issue can be formulated as a rather complex optimization model. This paper proposes an approach to model cross-company logistics concepts and a novel evaluation framework based on discrete-event simulation techniques which enable decision-makers in companies to jointly plan consolidated multi-modal traffic systems. By doing so, they are able to optimize for multiple criteria such as transport and handling costs as well as flexibility or environmental factors (CO_2, NO_x, …). A real-life case study describing a pilot region in Romania concludes this paper and demonstrates the impressive potential of the proposed approach.

INTRODUCTION

Progressive globalization and rising cost pressure in the automotive industry are increasingly forcing car manufacturers and component suppliers to set up new manufacturing regions in or move existing locations to low-wage region (Kinkel et al 2004; Sihn et al. 2006). It is for this reason that Central and Eastern Europe (known as CEE) became a popular target for relocation. These new production sites often just assume the supplier and customer structures of the parent plants.
However, benefits tend to fall short of predicted cost advantages, due to rising wage costs (in particular in industrial regions). The automotive industry in particular is thus focusing on the costs of transport logistics arising from intense transportation between Eastern and Western Europe.
The trend towards relocation has shown that the exchange of goods leads to new demands and challenges for transportation and logistics (Feldmann et al. 1996). Business networking strategies and especially cross-company co-operation is one of the key factors to improve in production issues as well as in logistics and hence to survive in competitive markets (Kinkel et al. 2004; Wiendahl & Lutz 2002).
Based on this situation, a new simulation and evaluation model, which supports the development and evaluation of new logistics concepts, was developed. It is used for the validation and evaluation of cross-company logistics models. Due to the new, holistic evaluation approach potentials for optimization in the areas emissions, costs and logistics competitiveness are targeted on developing new sustainable and energy-efficient logistics models.

MODELING MULTI-MODAL TRANSPORT CHAINS

The currently applied logistics processes, especially for the specific needs of individual enterprises in automotive industry do not appear optimal from a holistic point of view. Deficits might emerge from direct transport running far beyond capacity, use of small transport carriers, less-than container load (LCL) with long running times or multiple handling steps as well as bad transportation tariffs due to small quantities. High stocks and capital tied up are results of this inefficiency. Since many companies have a similar source-target-behavior the potential of cross-company bundling to optimize transport efficiency is high.
There are various approaches for cross-company logistics models that conform to the general network model of logistics. These models represent networks transporting rights, goods, finance and information where spatial, quantitative, information-related and temporal differences as well as company boundaries are crossed (Vahrenkamp 2007). Parameters defining the structure of a logistics network are paramount (Rösler 2003):

- Number, locations and functions of source points (= loading locations, making goods available),
- Number, locations and functions of target points (= unloading locations, points of reception),
- Number, locations, functions of connections or nodes between sources and targets.

The network nodes are called transshipment terminals. This implies that only transshipment but not storage in general (no inventory) is foreseen at these locations. Transshipment terminals serve as consolidation terminals where the flows of goods are collected and/or as break-bulk terminals where the flows are in turn distributed (Rösler 2003). The **basic structure** of transportation links can be represented either as **direct connection** ("point-to-point" transport) in its simplest form (single-stage, uninterrupted transport chain) or as a **multi-stage system** with preliminary leg, main leg and subsequent leg with transshipment terminals where the network nodes serve as consolidation terminals where the flows of goods are collected and/or as break-bulk terminals where the flows are in turn distributed.

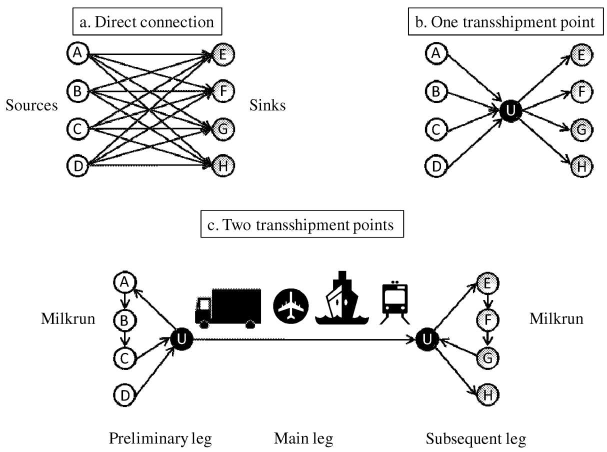


Figure 1: Illustration of consolidation by means of transshipment points (Brauer 1982).

The mixture of logistics systems made up from the given basic structures is decided in the logistics network structure. The processes are designed when the logistics capacities are superimposed on this. The logistics capacity can be subdivided into transport capacity, warehousing capacity and information capacity. In addition to the basic structure of the systems, the speed of traffic flowing between the individual points in the system must be taken into account (Pfohl 2004). The network strategy is also based on geo-economic considerations such as the long-term development of customer demand or the development of the required delivery time.

Summing up, the criteria logistics costs, supply service, adaptability, susceptibility to interference, transparency and time for planning and establishment of the system are important in the moment of developing and evaluation logistic models (Pfohl 2004).

As described in the initial situation, optimization of transports for individual businesses does not appear ideal; therefore companies can align with partners to a logistics cooperation and bundle transport volumes. Bundling, also referred to as consolidation, happens when transport volumes are combined to form larger transport batches in order to allow more efficient and more frequent shipping by concentrating large flows onto relatively few links between terminals, thus lowering transport unit costs and the unit costs of incoming or outgoing goods at their starting or target points. The starting points for the scenarios for transport bundling are the individual parameters of the logistics network structure. The following forms may thus be used:

- Source-point bundling often following the principle of the "milk run" (the shipments intended for a particular destination are collected from several places of shipment, from neighboring places of shipment or from a shipment region and processed together)
- Target-point bundling, where shipments from one place of shipment intended for several destinations or for a delivery region are processed jointly and transported together and
- Transport bundling, where shipments are collected and delivered in one tour

Further forms of bundling can be inventory bundling or temporal bundling, and vehicle bundling and transshipment point or transit terminal bundling as forms of spatial bundling (fig. 1). The number of transports between sources and targets can be reduced by the setup of transshipment points from m x n to m + n, m and n being the number of source and target points (Campbell 1990; Simchi-Levi et al. 2000).

Bundled transport over the long run between 2 transshipment points can raise high potentials due to low transport costs and efficient use of transport capacities (Trip & Bontekoning 2002). Logistic performance is improved by the raised frequency of transports. Overall every bundling type must meet the requirements of savings through consolidation of synergy effects to cover higher transport costs, operation costs of handling points or longer distances of time frames in comparison with direct relations.

The goal to reduce logistic costs while keeping logistic quality at the same level or raising the quality (delivery times, adherence to delivery schedule) is the main focus when designing the transport network. An iterative method is needed to evaluate the impacts of modifications in logistic models regarding ecology, economy or logistic competitiveness.

Transport bundling or cross-company logistic networks are originally based on the idea of good distribution in urban centers. The different approaches can be summed up with the term city logistics (Taniguchi et al. 2001). Other known developments of transport bundling of different suppliers are area contract freight forwarders, bundling and delivering goods for one plant conjointly. Collaborative approaches and the logistic models in this case are mainly based on the following premises:

- Identification of route sections where transport volumes can be handled with efficient transport carriers
- Availability of adequate partner for transport bundling on route sections (legs)
- Possibility of individual businesses to efficient usage of carriers
- Distance from source to target of possible nodes considering impacts of variance from ideal path

- Prioritization from transport volumes given limited capacities of one carrier in the main run as a result of different impacts on target categories
- Possibility to change transport frequency

SIMULATION MODEL FOR CONSOLIDATED MULTI-MODAL TRAFFIC SYSTEMS

The control variables required to achieve the objectives (minimize emissions, reduce costs of logistics and increase logistics competitiveness) are the following:

- Traffic avoidance: organize transport more efficiently (improve vehicle utilization, cut down on transport capacity)
- Bundling of goods flows: consolidate in order to optimize the substitution relationship between transport and inventory costs
- Switching freight transport to other means of transport: inter-modal transport

In the long run, it is also necessary to validate the results of the model's conceptual design. This should be performed in accordance with the main target dimensions - emissions, costs, and competitiveness.

It is in particular the intermodality aimed for in the models that plays an important role in the evaluation of this target dimension. In this point, only a selection of the most harmful emissions - CO_2, NO_X, and amounts of particulates - is analyzed; they are mainly accounted for in the dominant means of transport - road haulage. The emission levels are mainly dependent on the journey, i.e., distance covered by the predefined journey profile and on the allocated transport resource. Diesel or electrical power consumption also plays a decisive role in the output of emissions.

The cost calculation model is somewhat more extensive and can be subdivided into three different categories (transport costs, handling (i.e. transshipment) costs, and inventory costs). When it comes to transport costs, it is important that the model is based on the actual costs incurred, i.e., the overhead costs, road charges, customs clearance, and wage costs, and not on the transport tariffs charged by shipping companies. The road charges are particularly difficult to determine due to differing systems in the individual countries, and they play a considerable role in determining the route.

$$C_{Total} = C_{Transport} + C_{Handling} + C_{Inventory} \quad (1)$$

$$C_{Transport} = C_{Tsp.Truck} + C_{Tsp.Train} + C_{Tsp.Plane} + C_{Tsp.Ship} \quad (2)$$

$$C_{Handling} = \sum (Q_{Chg.\ of\ Resource} \cdot C_{Transshimpment\ (Chg.\ of\ Resource)}) \quad (3)$$

$$C_{Inventory} = C_{Capital} \cdot C_{Warehousing} \quad (4)$$

The third criterion in evaluating optimization models is logistic competitiveness, which is made up of the ability to deliver (a measure of the extent to which the company can guarantee the logistics service requested by the customer - short delivery times compared to the competition are especially important for a high ability to deliver) and delivery reliability (delivery reliability rates the service provision of the logistics process - it indicates the proportion of the complete and punctual deliveries compared to all delivery orders) (Vahrenkamp 2007).

Against this background, it becomes apparent that the cooperative planning of multimodal transport is affected by a multiplicity of factors. Numerous interdependencies between these parameters and the goal criteria such as costs, emissions, or flexibility exist. The evaluation model thus has to solve the conflict of goals like the trade-off between costs and guaranteeing competitiveness by applying stocks.

Beyond that, the identified parameters are in reality very often afflicted with uncertainty. These affects frequently influence the quality of material planning decisions and transport planning considerably.

Owing to dynamic interactions and taking stochastic phenomena into account, a static estimation of the behavior is difficult or almost impossible. Simulation has satisfactorily demonstrated its ability to illustrate and evaluate systems with dynamic behavior.

In order to provide a simulation model for logistics operators, the system-specific characteristics in a suitable simulation environment have to be illustrated. To achieve flexibility in planning regarding changes of logistics models or changes of the basic conditions, it is useful to provide a generic and easily adaptable model. For this purpose, the models of intermodal transport were illustrated within an application platform for simulation. This simulation environment permits the illustration of various technical models in a modeling environment adaptable to the planning domain. Furthermore, the environment permits an automatic derivation of simulation models and a return of the results to an economic or ecological level. In this section, the fundamental methods of the application platform simulation and the development of a simulation model for the planning of multimodal transport are described.

Depending on which simulation tool is chosen from the set of tools with different universal validities and application references, the simulation expert can access more or less preconfigured building blocks. Special simulators contain solutions matched to the specific area of the domain of application, thus simplifying handling. The higher the degree of universality, the more varied the possibility of creating and depositing own functional building blocks becomes; however, the necessity of a simulation-specified training increases as well. The application environment developed by V-Research deals with simulation models in production and logistics.

To execute a conventional simulation study, a substantial amount of time and money has to be invested, which still prevents small and medium-sized businesses to apply those. To prevent a planner, who is usually well trained in his/her respective systems domain, from dealing with the simulation expertise itself, it is helpful to develop an instrument that lets him answer upcoming questions without specific simulation know-how in an accessible amount of time. Thus, the planner is able to define planning tasks, generate models, analyze results, and optimize those results by comparing different scenarios.

The core of the platform represent application related and technically oriented components that base on the utilization of open industry standards (.Net Framework 2008) and reliable, well-tested simulation software. The concept embraces the idea that each simulation study needs certain key functions and procedures; contains customer and project specific characteristics and requirements; should be expandable concerning detailing, functionality, and system boundaries; and lets other users (e.g. customers) enhance the model.

The separation between task specific and resource specific components is reflected in the structure of the application environment. Task specific components summarize all aspects concerning process logic of an application that are illustratable by a (production and/or logistic) system. Examples would be the definition of manufacturing technologies used for processing orders, process outcomes, product structure etc.; logistics sequencing strategies (push, pull, KANBAN, …); organizational classifications (employees assigned to department and resources, shift schedules, …); as well as an architecture for administering simulation runs, result data, and shift schedule data.

These components signify the applied level. The technical level comprises the structure of a production and logistics system. The resources (machines, means of transportation, tools, …) are systematically described in predefined components. In order to create a complete model, these project specific resources are substantiated and incorporated into a basic model. The entire architecture thus differentiates between technical components, application oriented components, and the actual business application. The business application represents the simulation study, which can be made up of a number of simulation experiments. The component model is well suited for customer specific simulation applications within a short period of time.

By describing the behavior and the possibility of providing process cycle time fluctuations using distribution functions, so-called confidence intervals can be determined through a number of replications (simulation runs with independent random variables) which allow a prediction to be made as to the bandwidths where the target dimensions are likely to be with a given level of probability.

The basis for the simulation model is formed by individual logistics building classes (factory, transshipment centre, etc) that can be combined with one another to represent any desired logistics concept (point-to-point transportation, consolidation terminals, milk run). These building classes are created in the simulation environment Flexsim®. Structural, procedural, and resource-related data are required in order to model with the simulation system. By modeling with building classes, it is possible to describe both the structure and the behavior of individual resources independently, as well as their interaction with other building classes. This inherent knowledge contained in the individual building classes is used and extended by configuring the building classes to form an overall model.

The central technical construct in the domain transport planning is the route, which represents a given start-destination-relation, for example between a loader in Eastern Europe and a Western European production facility. This route is either being served by a direct relation or by intermodal transport, which is usually conducted through certain hubs. In any case, the chosen domain model displays the relations from start to destination between Eastern and Western production locations. The actual routes taken were tracked via GPS and then contributed into the system's data. The information includes not only kilometers driven for the running time calculation, but also other data such as altitude profiles. These profiles further refine the simulation model leading to consumption and emission data of the transport resources to be implemented in the model.

The system load of the simulation model is determined by the output of products on the loader's side, which is being processed in form of a tabulated schedule. A profile of each originating plant is defined therein, stating which products are being put out in which amount at what time. The running time of the simulation is thus supplemented with the corresponding number of product objects that later undergo the defined transport process (stated below).

Transport resources are active parts when it comes to bridging regional distances within the domain model. In the case of the afterwards presented application various types of trucks and trains are displayed in the model. Transport resources are matched with certain routes and characterized by a number of attributes. Besides loading and unloading times for the transport resource, triggers have to be defined when a transport resource leaves a location (e.g. every Monday 8 a.m.). Furthermore, additional data can be added to the model, such as cost, consumption, emissions, or any local- or time-specific data. Therefore, the model is formed in a way that specifies consumption data for different types of trucks and diesel locomotives. In case of electrified railways, the proportionate emissions for the necessary power are included as ecologic target factors in the simulation for each country.

In the domain model, a plant can serve as a starting point or as a destination, or even as a hub to add products to or to discharge items from the transport. Product objects are created in the course of simulation, beginning at the starting plant with a time-amount-profile, and they are dissolved after reaching the destination.

The handling point describes a special location along the route, with incoming transports being unloaded and leaving transports prepared for the next trigger with their inherent loading and idle times (see transport resource).
In contrast to the handling point, no change of transport resource and no transshipping take place at a processing point. This simulation block simply represents a location where the transport is halted for a certain handling time, calculated as a stochastic variable. This method enables the display and calculation of customs clearance at national borders. The handling itself is not directly calculated due to model simplification, and it is thus represented by idle time.
Beside of these simulation blocks, the transport process is an integral part of the simulation model. The later presented application includes a process that starts off with the production of products at a starting point and ends at the destination point, where the products are destroyed. On closer inspection, the transport starts, when a specified amount of products is available at a certain location and the starting trigger for a certain means of transportation is activated. Consecutively, the transport process is being planned according to the route; the products are loaded on the transport resource and then transported to the next stop on the route. If the next stop is a plant or handling point, the products are unloaded and, if needed, prepared for the next shipment. In case of the next stop being a handling point, the idle time is applied and the transport continues. Upon arrival and unloading of the transport, the respective products are destroyed in the simulation environment and the simulation run is ended.
Throughout this process, the running time and its contributing factors of each process entity, such as time needed for loading procedures, handling activities, driving times, and emissions depending on the road profile (distance, incline, street conditions), are logged. After the simulation, this logged data supplements the results analysis. A historiography of the granulated simulation data in a database enables detailed and flexible analyses and a deduction of key performance indicators. The (logistic) results drawn from the simulation can subsequently be applied to calculations with data sources concerning emission behavior, for cost analyses, and for comparisons whether logistic targets have been met. The model structure simplifies changes made to the parameters (e.g. varied transport cycles, use of various traffic resources and carriers etc.), and a comparison of the effects on the three target areas economy, ecology, and competitiveness is enabled. Based on the opposed target dimensions, a decision can only be made through comparing different scenarios and considering the pros and cons of all compromises.

OPTIMIZING MULTIMODAL TRANSPORT CHAINS: THE CASE OF A ROMANIAN AUTOMOTIVE CLUSTER

As the preceding sections of this paper showed, simulation approaches can provide helpful techniques for supporting various decisions in co-operative multimodal logistics chains. The simulation model used in this approach can be used to evaluate arbitrary scenarios. Hence, a variety of decision-related issues can be answered like

- the optimal location of transshipment points,
- the preferred modes of transport for preliminary, main and subsequent legs,
- or optimal routing options for main legs.

In order to validate the proposed simulation approach and illustrate the full potential of this technique in supporting co-operative transport decisions this chapter focuses on a case study describing possible optimizations for an automotive cluster in the region Timis in western Romania. Within these region, a considerable number of automotive suppliers are planning the exchange of goods with production sites mainly located in Germany, northern Italy or Spain on a local basis (i.e. without coordinating the transports with other suppliers). As a consequence, the analysis of the initial situation showed that all companies used direct road transport as their only means of transport. Another result of this initial study was that these transport capacities were only partly utilized which in turn shows a first potential for transport coordination.
The identification of new co-operative logistics models dealing with the issues of the described situation should result in a reduction of transport costs as well as air pollution while maintaining a certain level of flexibility for the consignors. Standard logistics concepts like direct transport and part load concepts like milk run or groupage traffic were used as building blocks for alternative logistics scenarios. The use of a block train to handle the main leg is the common characteristic for all scenarios. The requirement to maximize the utilization of this main leg leads to the necessity of coordination of transport demands among the consignors in the region. The design of candidate scenarios for optimized coordinated logistics models is an iterative process requiring expert knowledge and the support of information technology. This approach has to consider the following topics:

- Identification of route sections which can be handled by a more efficient means of transport.
- Existence of potential partners for transport bundling within a region.
- Possibilities to optimally utilize transport carriers by joint planning activities.
- Prioritization of loads, given a carrier's capacity restrictions on the main leg.
- Methodology for handling transport backlogs infringing these capacity restrictions.
- Possibilities to alter transport frequencies in order to scale for different transport demands in the region.

As a result of these considerations, four alternative scenarios for co-operative multimodal logistics models were proposed:

1. Scenario 1 envisages the installation of a block train starting from Arad (Romania) with destinations in Stuttgart, Frankfurt and Wolfsburg (Germany) with two rotations per week (see fig. 2). Pre-carriage traffic for collecting goods from consignors as well as onward-carriage for distributing goods from the respective handling points in Germany to the consignees are designed as road-based direct transport. The shipments are assigned to the main leg in a "first-come-first-served" manner. Excess transport loads as well as loads for other destinations in Italy, Spain and Poland are handed over to direct carriers.
2. Scenario 2 is based on the first scenario but accounts for further bundling potential for direct traffic as well as for preliminary as well as subsequent legs. Here, the milk run concept is applied whenever an efficiency criterion yields this decision.
3. Scenario 3 is based on a shortened block train concept between Arad and Frankfurt twice a week. Pre-carriage as well as onward-carriage are designed as direct loads. This scenario accounts for a reduction of the rail-side complexity emerging from the first two scenarios.
4. Scenario 4 also builds upon the shortened block train concept but further bundles the remaining road traffic similar to scenario 2.

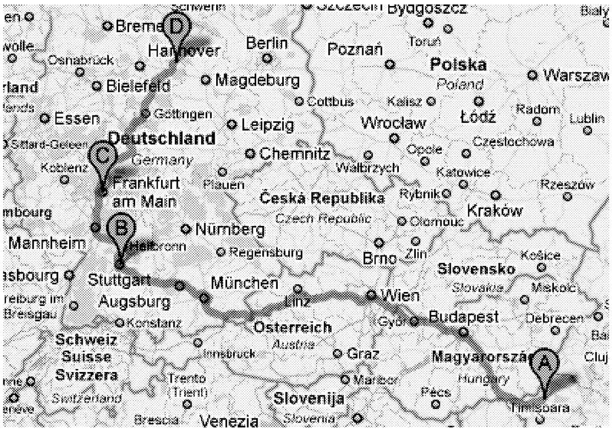


Figure 2: Main Leg in scenarios 1 and 2

As it is obvious from the above considerations, the holistic evaluation of the proposed scenarios represents a complex decision problem. In the proposed approach, the simulation model accounts for various types of transport- and handling-related costs, environmental factors associated with road traffic as well as with rail transport in the respective countries (usage of diesel locomotives, mixture of power sources, etc.) and logistics performance measures. The simulation study analyzed the shipment volumes and frequencies of seven automotive suppliers in the Timis region within a timeframe of ten subsequent weeks.

As a result, the proposed simulation approach shows the potential for optimization within the three distinct perspectives of economic as well as environmental and logistics factors.

The descriptive statistics of used transport concepts already outline the implications for these three dimensions: The share of tonnage handled by direct traffic decreases from 95.4% in the initial model to 35.1% in scenarios 1 and 2 and even further to 23.1% in scenarios 3 and 4. Against that, the shared intermodal transport concepts in the first two scenarios accounts for 62.8% and 66.3% in scenarios 1/2 and 3/4, respectively. This shift of transport paradigms results in 59% of tonne kilometers being transported by rail (compared to 0% in the initial scenario).

As it is obvious from the reduction of road-based traffic, there would be a tremendous ecological impact which reduces CO_2 emissions by nearly 40% and NO_x even by 50% (see fig. 3). This effect is still influenced by the high share of caloric production of traction power in CEE countries. However, this only weakly affects the environmental results without altering them.

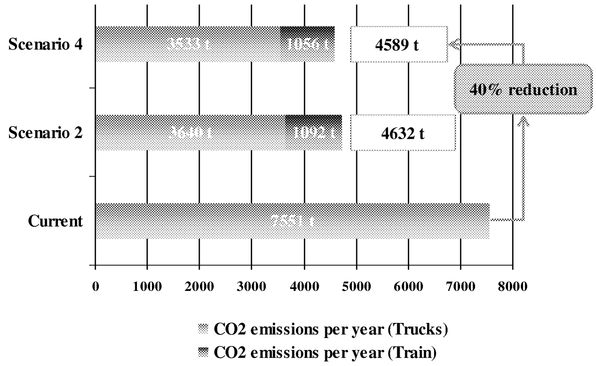


Figure 3: CO_2 emissions per year

Regarding the economic impact of the proposed logistics models one can observe a slight reduction of transportation costs for scenarios 1 and 2 (- 1.2% and - 3.6%, respectively). When reducing the complexity of rail transport in scenarios 3 and 4 the full potential of co-operative transport bundling can be tapped. With the shortened block train concept, significant cost reductions of -16.5% and -14.2% in scenarios 3 and 4 become possible. As a sensitivity analysis shows, this result of the latter two scenarios becomes robust under limited variations of alternative transportation costs as well as shipment volumes. The development of fuel prices as the main driver for road-side transportation tariffs give rise to the expectation that this relation will change in favor of rail-based transportation in the future.

Finally, the logistics perspective can be analyzed by the simulation approach: Due to increased handling requirements and the periodical manner of rail transport the throughput times for a single shipment rise by at least 21% (scenario 1). In a worst case scenario (scenario 2) the combined transport will take 38% longer than the direct load equivalent. However, these figures are based on the assumption that the consignees and consignors of the shipments do not adapt their manufacturing processes to the framework conditions of the new logistics models.

CONCLUSIONS AND OUTLOOK

This article described a simulation-based approach to evaluate different logistics models based on coordinated intermodal concepts. The novel approach of separating the domain model of logistics concepts from the calculation of Key Performance Indicators (KPIs) enables a holistic quantification of manifold impacts arising from a concept shift between road and rail transportation within an economic, an ecological as well as a logistics dimension.

The case study of a Romanian automotive cluster illustrated the analytic complexity of the problem as well as decision-supporting possibilities which are offered by the simulation technique. By using a domain-specific modeling approach the simulation framework was able to quantify the extensive potential which is offered by coordinated planning of transport chains.

Whilst the four candidate scenarios based on multimodal transport concepts yield tremendous optimization potentials with respect to the ecologic perspective and considerable possibilities to reduce transportation costs, they result in higher throughput times. The latter result can be explained based on the assumption that consignors and consignees do not adapt their manufacturing and sourcing processes to the possibilities of the multimodal transport chain. However, for fully tapping the potential of multimodal transport means also with respect to the generation of logistics advantages, it is necessary to adjust the planning processes for manufacturing plants on either side of the transport chain to the forwarding process. This could be achieved for example by explicitly integrating cost saving effects of multimodal transportation into manufacturing lot planning techniques. This in turn could result in a better timing of manufacturing output with respect to periodically scheduled block trains or other multimodal transportation concepts.

REFERENCES

Brauer, K.M.; Krieger, W. 1982. *Betriebswirtschaftliche Logistik.* Berlin.

Campbell, J. 1990. “Freight consolidation and routing with transportation economies of scale.” *Transportation Research Part B* 24, No. 5, 345-361.

Feldmann, K., Rottbauer, H., Roth, N. 1996. “Relevance of Assembly in Global Manufacturing.” *Annals of the CIRP* 45, No. 2, 545-552.

Kinkel, S., Lay, G., Maloca, S. 2004. *Produktionsverlagerungen ins Ausland und Rückverlagerungen.* Report on the BMF research project no. 8/95, Fraunhofer-ISI, Karlsruhe.

Pfohl, H.-C. 2004. *Logistikmanagement: Konzeption und Funktionen, 2., überarbeitete und erweiterte Ausgabe.* Springer-Verlag, Berlin.

Rösler, O.M. 2003. *Gestaltung von kooperativen Logistiknetzwerken: Bewertung unter ökonomischen und ökologischen Aspekten.* Deutscher Universitäts-Verlag, Wiesbaden.

Sihn, W., Palm, D., Matyas, K., Kuhlang, P. 2006. *Automotive Region Eastern Europe –AREE: Chancen und Potentiale des "Detroit des Ostens" für Automobilzulieferer*, Fraunhofer Austria, Vienna.

Sihn, W., Palm, D., Schmitz, K, Leitner, R. 2009. *Automotive Region Central and Eastern Europe – Produktionsstrukturen von Automobilherstellern und ihrer Zulieferer.* Fraunhofer Austria, Vienna.

Simchi-Levi, D., Kaminsky, Ph., Simchi-Levy, E. 2000. *Designing and Managing the Supply Chain – Concepts, Strategies and Case Studies.* Irwin McGraw-Hill, Boston.

Taniguchi, E., Thompson, R.G., Yamada, T., Van Duin, R. 2001. *City Logistics: Network Modelling and Intelligent Transport Systems.* Pergamon, Oxford.

Trip, J., Bontekoning, Y. 2002. “Integration of small freight flows in the intermodal transport system.” *Journal of Transport Geography* 10, 221–229.

Wiendahl, H.-P., Lutz, S. 2002. “Production in Networks.” *Annals of the CIRP* 51, No. 2, 573-586.

Vahrenkamp, R. 2007. *Logistik: Management und Strategien, 6.Auflage.* Oldenbourg Wissenschaftsverlag, München.

AUTHOR BIOGRAPHIES

V-Research is an Austrian competence center for industrial research and development. In the research field “Process Engineering for Construction, Manufacture and Logistics” the research center focuses on methods to support complex decision making in manufacturing and logistics processes.

Christian Hillbrand is head of the research area „Technical Logistics” at V-Research and director of the research programme “ProDSS” (Integrated Decision Support Systems for Industrial Processes). He holds a PhD-degree in business information systems from the University of Vienna.

Susanne Schmid is research assistant and project manager at V-Research. Her main research areas include the conception, modeling and development of simulation models.

CONTAINER TERMINAL SIMULATION

IMPROVEMENT OF KEY PROCESSES IN A CONTAINER TERMINAL: AN APPLICATION OF DISCRETE SIMULATION

ILEANA PÉREZ VERGARA
Email: *ileper@ind.cujae.edu.cu*

ROSARIO GARZA RÍOS
Email: rosariog@ind.cujae.edu.cu

Instituto Superior Politécnico "José Antonio Echeverría"
Ciudad de La Habana, Cuba

HENDRIK VAN LANDEGHEM
Industrial Management Department
Ghent University, Belgium.
Email: hendrik.vanlandeghem@ugent.be

J. CARLOS VALENZUELA
Departamento de Matemáticas
Universidad de Cádiz, Spain.
Email: jcarlos.valenzuela@uca.es

Keywords
Discrete simulation, logistics.

Abstract
To alleviate congestion at the entrance of a Maritime Container Terminal simulation models were used to explore alternative ways of operation. By introducing parallel operations, and rearranging weighing machines, we could obtain higher service levels and less flow time for the trucks. Results were obtained using Arena and proved valid in practice.

Introduction.

A maritime container terminal (MCT) is a complex system in which several processes of either transport interchange or logistic management (maritime and overland traffic) of containers are developed (Marìn et al, 2004). One of the most important aspects to bear in mind when analysing the improvement of the performance of a MCT is to identify all the processes that are developed and all the relations between these processes. Although, in general, there are two key processes: loading/unloading of ships and the reception/overland delivery of goods, some other processes (as the storage and the internal transport of containers), which are considered as supporting processes, also take place (see Figure 1).

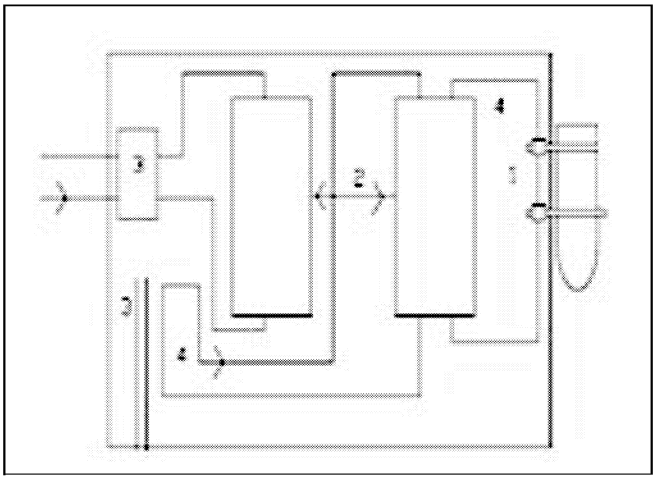


Figure 1. Subsystems in a maritime container terminal

1=Load/unload 2=Storage 3= Reception/Delivery 4= Transport

Both the large number of external agents, who may take part in the maritime and overland traffic and the huge amount of generated data, results in a complex relationship context for the MCT. This context becomes even more complicated because of the wide range of types of containers, ships and terrestrial transport vehicles that interact in the system.

These types of systems are classified as discrete dynamic systems due to their behaviour and to the changes that are generated in the processes (Cimino et al, 2010). Therefore, we could model it by using discrete simulation as a tool for analysing the system.

Situating the Maritime Container Terminal

An MCT is considered an intermediate level in the supply chain and it is associated to the import/export process of goods: raw material, materials or products to sale. The breaches of the commitments (basically focused in terms of production/delivery time) that acquire the MCT with the providers and the clients have very negative repercussions in the delivery times. Therefore it harms the costs and the satisfaction of the intermediate and final clients, which are not able to count on products in the settled time. Due to this fact, it is very relevant to looking for solutions that manage to adapt the production capacities and/or services in this type of facilities with the objective of verify the quality indicators that the managers take as strategies for the supplied service.

The maritime containers terminal is a system integrated by several sub-processes, physically connected and sharing information with terrestrial and maritime transport networks (Longo, 2010) as shown in Fig. 1. Each process is related to some task,

1. Container's load-unload process: It constitutes the maritime interface.
2. Internal connection subsystem
3. Container's storage process: which occupies most of the terminal's surface and whose organisation is strongly related not only to the traffic that both previous systems demand but also to the choice of handling systems that this subsytem is going to manage.

4. Terrestrial reception and delivery: it is formed by terrestrial gates for trucks and trains and the relations that are established with the facilities/installations disposed to facilitate the management of the high volume of traffic and information that in this zone is acquired, as well as the necessary spaces to carry out these operations.

	Activity (point) 1	Duration (min)	Resources (maq/op)
0	Arrival	LOGN(0.843, 0.545)	………………
1	Access Inspection	UNIF(2 , 3)	1 Operative
2	Inspection 1	UNIF(2 , 3)	3 Operatives
3	Confection of the entry turn	GAMM(1.21, 1.71)	3 Operatives
4	Gate system	UNIF(2 , 4)	1 Operative
5	Weighing machine	5	2 weighings
6	Load of full cont.	LOGN(15.2, 18.7)	3 (2carrier,1transtrainer)
7	Load of empty cont.	UNIF(2 , 3)	2 frontal vació
8	Exit inspection	BETA(0.573, 4.53)	1 Operative
9	Exit inform	BETA(0.898, 2.46)	1 Operative
	Punto 2		
1	Entrance turn	UNIF(2 , 3)	1 Operative
2	Load of full cont.	LOGN(3.06 , 1.68)	2 carrier
3	Exit inform	UNIF(2 , 4)	1 Operative

Reception and delivery process

The main aim of the reception and delivery process is to facilitate the receipt/delivery of goods by terrestrial transport in an efficient and rapid way, but compatible, in conditions of operational security, with the high number of documentary exchange and, in fact, of information which is needed.

A high waiting time was detected for trucks in the container's reception/delivery process in the MCT. This problem prompted the research, in order to find a solution to this problem.

A diagnosis was realized and the main problems were determined. Among these we can highlight:

1. The parking capacity in the MCT, in the Revision Area is not sufficient
2. The Gate system, which gives landside access to the MCT (in which trucks are received and the containers and the documentation are inspected), does not allow to guarantee the constant flow of trucks requesting service.

Simulation model

Bearing in mind that the civil and technological infrastructure of the Terminal was designed for a lower level of operation that the current one and that several restrictions existed blocking its extension, it was decided to study the system and to analyse possible solutions by means of a simulation study. This simulation considered the logical interactions between activities and operative decisions of the process. This was implemented in the simulation model developed with Arena 7.01 (Kelton & Sadowski, 2004). The trucks were defined as entities, according to their activities, differentiated by the attribute 't':

T1: trucks that operate only with full containers
T2: trucks that operate only with empty containers
T3: trucks that realize double operations

In Table 1 we can find a summary of the sequence and duration distributions (obtained using the Arena Input Analyzer) of the activities and the resources associated to each working station in the current situation, used as input to the simulation model.

Table 1. Times and resources associated to each activity in the process

The variables that were measured are:

- Total number of vehicles served by operations
- Average waiting time in gate: the time from arrival to the moment the first operation on the vehicle starts
- Average time in system: between arrival and exit of vehicles at the MCT
- % utilization of the gate

An experimental design was realized and 5 replications were run until steady state.

The model permitted us to determine that the average number of trucks waiting in the road access was 24.38, with maximum of 25 and with an occupation level of the gate of 100%. This clearly indicated congestion at the gate.

Alternative solution

The obtained results motivated a proposal to realize simultaneous operations. These operations will take place in an exterior area with respect to the Terminal, which a maximum capacity of 25 trucks, because extension of the internal area of revision was not possible due to infrastructure restrictions (Figure 2).

The activities that are proposed to be simultaneously realized are the following: Inspection at road access, Inspection 1 and the Execution of the entrance turn. Moreover, each of the entrance gates will have a weighing machine that will permit us to weigh the trucks at the moment of access to the MCT.

In internal services, no changes should be proposed because it was shown that the available manipulation teams had sufficient capacity and they could be used, depending on the demand, to provide load/unload services of containers to the trucks. That is to say, the new model will use the same amount of teams available nowadays.

The exit gates will have a weighing machine to weigh trucks depending on the operations that it realizes. Exit gates without weighing machine are defined for those trucks that are uncharged. At the exit gates the inspection operations and the exit consignment will be done simultaneously. The new model needed an estimation of the times for the simultaneous activities at entrance and exit gates, so several

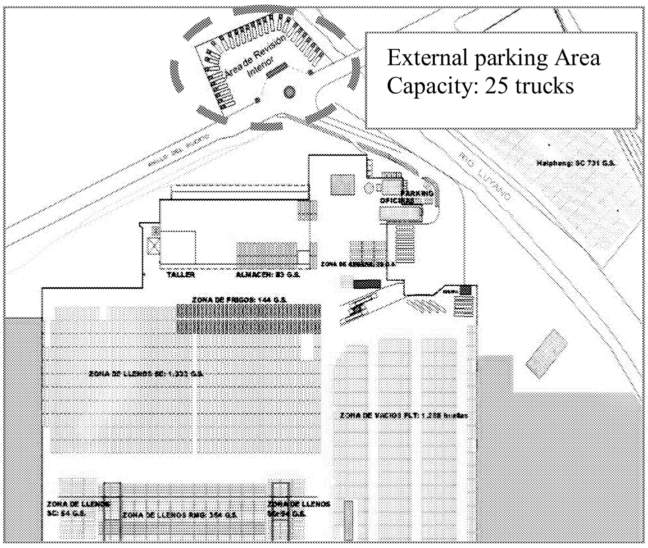


Figure 2. Proposed solution

tests were realized in collaboration with the MCT's expert (see Table 2)

Table 2. Estimated times for operations in entrance and exit gates.

Service time	Entrance	Exit
Entrance/exit full/empty trucks (requires weighing)	UNIF(2 , 8)	UNIF(5 , 8)
Entrance/exit full/empty trucks (requires weighing)	UNIF(5 , 8)	UNIF(2 , 8)
Trucks without weighing	UNIF(2 , 8)	UNIF(2 , 8)

This proposal required several modifications in the distribution of the Terminal, as it is depicted in Fig. 3.

A new simulation model was designed that would permit us to simulate the behaviour of the system under the new conditions and to determine the amount of gates needed to guarantee an appropriate flow of trucks in the Terminal.

The obtained results for the proposed model permitted us to determine that 5 gates will be used in total. Three of them should have a weighing machine, to guarantee a better flow of service. A comparative analysis was done of the obtained results in current conditions and under the proposed solution (see Table 3).

The results show that with the new solution it is possible to get significant increments of the service level obtained in the MCT: the average number of trucks handled rose to 53, with a decrease of an average of 57 minutes of average waiting time and a significant decrease of time in the system of trucks.

Table 3. Simulation results NEW situation.

Variables		Average		Impact
		Now	Proposal	
Total of operarions (trucks)	Trucks loaded	63	145	+82
	Trucks unloaded	174	128	-46
	Truck double op.	59	76	+17
	Total Operations	296	349	+ 53
Average waiting time in gate. (min).	Waiting time at Entrance gate.	59,15	2.15	- 57
Average time in the system. (min)	Time for trucks load	97	45	-52
	Time for trucks unload	79	42	-37
	Time for trucks double op.	102	68	-34
% use of gates	Use of gates	100	92	-8

These results helped the Operations Manager of the Terminal to increase the operation indicators, service levels and the productivity when the needed resources were calculated. Besides, both models were used as "cases to study" for the operatives related to the activity.

Conclusions

The use of a simulation model permitted us to:

- Point out the source of the detected problems in the diagnosis of the Terminal
- Design the necessary resources to guarantee service levels for the given demand
- Evaluate the management of the organization
- Consider operative and strategic decisions based on indicators
- Use the designed models as "case studies" in the training process of operators and supervisors.

References.

Marín M. F.; García G. I.; Ortiz O. A.; Arreche B. L. M.., 2004, Gantry crane operations to transfer containers between trains: a simulation study of a Spanish terminal. Transportation Planning and Technology, Vol 27, 4 pp: 261-284.

Cimino A, Longo F. and Mirabelli Giovanni A, 2010, General Simulation Framework for Supply Chain Modeling: State of the Art and Case Study, International Journal of Computer Science Issues, Vol. 7 (2), No 3.

Law A., Kelton D. Simulation Modeling and Analysis. Tercera Edición. McGraw Hill. 2004.

Longo F., 2010, Design and integration of the containers inspection activities in the container terminal operations International Journal of Production Economics, Vol. 125 (2), pp. 272-283.

Layout Planning of Less Than Truckload Terminals

Uwe Clausen, Ina Goedicke, Larissa Mest and Jan Kaffka
Institute of Transport Logistics
Department of Mechanical Engineering
Technische Universität Dortmund
D 44227 Dortmund, Deutschland
E-mail: mest@vsl.mb.uni-dortmund.de

KEYWORDS
Logistics, Traffic simulation, layout planning, forwarding agency, freight traffic, LTL

ABSTRACT

Strategic planning and operational decisions affect the efficiency of the business of a forwarding agency. Especially the layout planning has high impact on handling costs. The detailed simulation of internal and external material flows in logistics nodes of road haulage enables the testing of various layout scenarios for forwarding agencies and the identification of the “best” solution. Simulation studies can be conducted efficiently concerning the comparison of different layout concepts by using a toolbox for generating.

INTRODUCTION

Forwarding agencies dealing with LTL (less than truckload) offer transportation services for piece goods, which include the collecting of the advised consignments at the consignor and the transport of the shipments in appropriate time and quality to the indicated consignee. The various piece goods are in general handled as groupage freight by the shippers. I. e. the transport of several small consignments, which are first collected in the local range of a forwarder at different consignors (forerun), then consolidated and transported by line haul as collective consignment to another forwarder, and finally distributed to the consignees by the receipt carrier. In order to provide these transports, the shipments collected at local traffic routes in one region first have to be sorted at a terminal for the specific long-distant relations.
Pieced goods shipments are in most cases heterogeneous, hence the application ranges of mechanized sorting and unloading technologies are reduced and the handling is personnel and cost-intensive. An efficient "good" dispatching and/or planning of resources therefore has large effects on the costs of a terminal, because an improvment of the terminal by the change of technology is often not possible. The investment and operating cost in the terminal, accounting for approx. 35 % of total costs of a LTL terminal, are determined by the layout and the arrangement decisions of the operating forwarding agency.

System Description

The Chair of Transportation Systems and Logistics, Technische Universität Dortmund, has developed a transportation traffic logistics simulation environment in order to analyse the interaction and the effects of strategic planning, e.g. layout decisions, on operational procedures. Based on the simulation software Enterprise Dynamics, the tool ED Transport permits the illustration and simulation of material and information flows in logistic nodes (e.g. terminals, distribution centres) of road haulage. Due to the fact that there is hardly any rationalisation potential in using different technical items, e.g. handling technology, the optimization of strategies and operations is the most important task to increase the economical performance of these agencies.
The modelling of logistic nodes is done on a microscopic scale by using several atoms, which enables to simulate each individual consignment and thereby the connected internal and external processes such as vehicle movements and transports with forklifts. The simulation-borders of the system are at the gatekeeper of the terminal, therefore all processes on the yard area of a forwarding agency are included in the model (Neumann, 2007).
The structure of a LTL terminal can be described by two subsystems: the yard and the terminal building itself. These subsystems are linked with each other by doors. The basic task in the operation of a LTL terminal consists of routing vehicles in the yard as well as unloading, buffering and loading shipments in the terminal. The main focus of this paper is on the terminal processes.

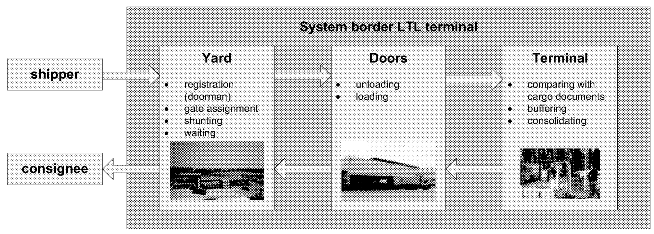


Figure 1: Structure of a LTL terminal

Process Chain In LTL Terminals

Vehicles for unloading arrive at the gatekeeper and are assigned to an unloading door of the terminal. After unloading a shipment to an unloading buffer (buffer which serves only for temporary storage of unloaded consignments) by the freight carrier or the personnel of the terminal (depending on the used strategies), the consignments are removed from the unloading buffer by elevating trucks (usually motorized) or fork lifts. The consignments are normally brought to the loading buffer for a specific destination (the information can be found on the attached barcode or delivery notes). Partly, the goods are transported to universal buffer areas, but this does not enable the systematic pre-assembling for dispatching and thus avoids fast loading of long-distance vehicles. For loading vehicles, a loading dock near the relation-specific buffer is allocated in order to minimize the internal material flows.

LAYOUT PLANNING

The term layout planning indicates strategic decisions, which have to be made within the scope of new building or expansion or restructuring of LTL terminals respectively. This comprises the determination of shape and size of the terminal, the number of doors and their positions at the wall. Additionally, the function of the doors (unloading or loading, short-distance traffic, long-distance-traffic or mixed) has to be determined as well as which relation is assigned to which gate. This is done on a tactical / mid-term planning.

For an existing terminal, operational questions concerning size and spatial arrangement of internal areas (for buffering and pre-assembling of unloaded shipments) as well as the position of travel paths have to be answered in the process of layout planning.

Layout decisions determine e.g. the path length between doors with strong relations and therefore also have an effect on handling and personal costs. So far, these decisions are made manually or based on the experience of the dispatcher, which are feasible and yield to the required performance (throughput). Systematic optimization of both strategic planning and operational processes rarely takes place. For this reason a large potential to reduce costs can be assumed.

Optimizing Concerning Handling Costs

Due to the low level of automatisation, handling personnel is the main operating resource inside the terminal. The amount of the workers needed every day depends directly on the distances travelled to handle the shipments inside the terminal, e.g. by fork lift or on foot. On the different planning levels the operators have several possibilities lowering the distances.

On the strategic level the layout of a terminal causes the basic distances between the doors. On the tactical level the assignment of trailers to load doors lowers the transport performance inside the terminal. Trailers for destinations with a high amount of shipments should be assigned to doors near the unloading zone. In addition, also the assignment of destination to the different buffer areas has to be done in a same way, because the destination buffer always should be located near the destination loading door. On the operational level the allocation of trucks to unloading doors and the strategy for unloading the trucks are important for the performance.

Scientific authors have proposed solutions to layout problems for freight terminals (Peck, 1983; Tsui and Chang, 1990, 1992), but in every case, freight flows from unloading doors to destinations were assumed to be independent of the layout. Under such conditions, the supervisor's policy is equivalent to a First Come First Serve policy (Gue, 1999).

To optimize the assignment of trailers to load doors on the tactical level, Chmielewski developed a helpful doorassigner, using a branch and bound algorithm (Clausen, 2009). But a systematic optimization of both strategic planning and operational processes rarely takes place, especially concerning the varying amount of shipments and the inclusion of layout restrictions. For this reason a large potential to reduce costs by using simulation can be assumed, which enables to consider the different levels of layout planning and their interactions in one model.

This paper focuses mainly on the tactical level:

- position of unloading and loading doors
- assignment of trailer to load doors
- decisions concerning size and spatial arrangement of internal areas (e.g. buffering) and of travel paths

For the improvement of such terminal-layouts, complex process-chains have to be modelled, taking varying amounts of shipments, congestion and the path capacity for forklifts under consideration. A 'good' dispatching and/or planning of resources have large effects on the costs of a terminal, because an optimization of the terminal by the change of technology is often not possible. Assignment decisions determine e.g. the path length between unloading and loading doors and therefore have an effect on handling and personal costs.

MODELLING

As described above the layout and organisation of a LTL terminal is essential for its performance. Therefore, it is demanded by terminal operators to model the complex characteristics and dependencies within a terminal when analyses are done with the help of simulation. Three levels of information have to be evaluated and integrated into the simulation model:

1. Geometric dimensions and positions of the physical layout components (e.g. doors, buffer areas) form the basis.
2. The second step has to define further characteristics of each layout component. As an example it is necessary to determine whether a door is used for loading or unloading.
3. Finally, all information to implement the process chains within the terminal has to be provided (esp. concerning the assignment of destinations to layout components).

It is an additional challenge that the different scenarios evaluated within a simulation project often require several simulation models, for example due to repositioning of doors or buffer areas. Due to the large number of components (typically about 300-600) and their complex dependencies it is huge effort required to create a scaled model manually. Since the LTL-business is subject to largely dynamic market conditions, simulation studies cannot be made within long-term time horizons. Bottlenecks have to be identified and strategies for enhanced performance have to be determined quickly based on current shipment amounts and characteristics.

To fulfil these requirements of the terminal operators the simulation expert is demanded to model complex systems within short time periods. The first aim is to enable simple development of different layout scenarios without taking care of all complex requirements of a simulation model. Furthermore, it has to be possible to transfer the information of such a basic layout plan into the simulation software for further analysis. As a consequence, a special toolbox has been developed in order to support the simulation expert in his tasks. It has been an additional aim that the toolbox enables also planning experts and even the customer himself to transfer their ideas and know-how into data for simulation scenarios. This approach reduces data handling between different project participants and integrates the whole project team into the study which has been identified as an essential requirement to save time in simulation studies and increase the reliability of project results (Gocev and Rabe, 2010). The modules of the toolbox are implemented in established software tools and based on each other in correlation with the three levels of information described above. Depending on the specific scenario different efforts have to be made to create the simulation model. Only the third level is changed if new process chains within the existing layout are evaluated. The third and the second level are renewed in case of a change concerning functions of different layout components. All three levels have to be modified if edificial changes of the terminal building have to be considered.

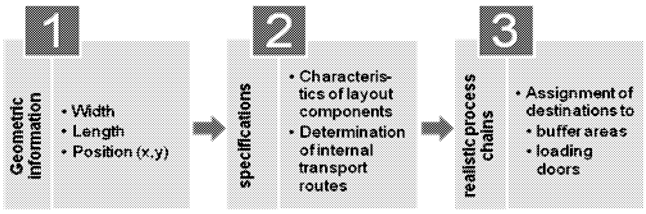


Figure 2: Modules of the toolbox

Level of Information

The first level of information as mentioned above contains geometric dimensions and positions of the layout components. In a graphical user interface, as for example Microsoft Visio, the essential data of construction drawings and manual drawings (e.g. of internal areas) can be transferred into one functional scaled drawing considering all relevant components. Based on an as-is layout different arrangements of buffer areas or internal transport routes can be created with little efforts by planning experts or terminal operators. The scaled visualization of the system supports the development of improved operating strategies as well as the communication with the customer. The toolbox transfers the graphical information into tables that can be easily understood by the simulation software. Each component has a distinct identifier and is represented in a row of the table with its geometric parameters.
The second level of information contains all characteristics of the layout components besides the geometry. For example, it has to be determined which doors are used for loading, unloading or both activities. Furthermore, doors can be grouped into zones with a specific function and the internal areas have to be linked to the routes of internal transport. Other characteristics can be integrated individually due to the aim of the simulation study. Of course, it is possible to adjust those settings later on in the simulation model manually but due to the large number of components and the complexity of their dependencies this contradicts the overall aim of the toolbox. The required characteristics are therefore integrated into the table of geometric information on each layout component in a standardized form. Since every component is represented by a row in the table and each characteristic by a column the concept is extendable if additional features have to be considered in a simulation study. The table combines simulation know-how with the as-is information or further optimisation ideas of logistics experts.

When the first two levels of information are documented for a simulation scenario they are linked with the simulation software. Based on the standardized tables the basic simulation model of the terminal is created by using the atoms of the simulation suite representing the layout components of a LTL-terminal. If further system aspects have to be modelled, e.g. a special sorting area, they can be integrated into the basic layout in the next step individually.
The third level of information provides the required data to implement realistic process chains. As described above the performance of a LTL-terminal largely depends on the internal transportation distances that have to be covered. Especially destinations with large shipment amounts determine the work load for transferring shipments from unloading doors to buffer areas and from buffer areas to loading doors. Thus, it is the main task to model the assignment of destinations to loading doors and buffer areas realistically. The assignments are not fixed for each layout component in the basic model to minimize the efforts for modelling different simulation scenarios. With the help of tables the correlation between a specific loading door or buffer area and its destination is defined for a simulation run. Changes in the configuration of used doors and areas can be simulated without building a new layout in the simulation environment. Additionally, the approach enables the assignment of several destinations to one loading door which is important to model the line-haul departures as well as the departures for local delivery. The required tables are prepared externally from the simulation software and are based on the distinct identifiers of the layout objects in the graphical layout of the first modelling step. In that way, also planning experts involved into the project or even the terminal operator as the customer can define an assignment that should be analysed within the simulation study.
The completed tables are copied into a structure within the simulation environment already arranged when creating the basic layout. This also applies for the input data defining the incoming vehicles of the terminal. Afterwards, the simulation model is ready for the experiments.
The modelling process as described above meets the requirements of simulation experts working for LTL-terminal operators. It aims at fast and decentralised creation of simulation scenarios and integrates the Know-How of other experts and the customer into the modelling process.

EXPERIMENTS AND RESULTS

Different LTL terminal have been simulated within research projects in the last years to identify the individual “best” solution for reducing handling costs. Two of them – with different focuses concerning layout planning – are described in the following paragraph. Both terminals are forwarding agencies in Germany.

Determining The Position Of Unloading Doors

In this terminal, one of the tasks during the simulation project was to identify the best position for a fixed number of unloading doors. The scope for the first terminal is as follows

- terminal of 6300 m² with 240 unloading doors and 36 loading doors

- 4300 Handling units (e.g. pallets) per day for inbound and outbound groupage, unequally distributed on 23 national and 33 regional destinations

Two different layout concepts and the resulting spatial arrangement of the internal areas can be found in Figures 2a and 2b. In both cases, destinations with a high amount of shipments were positioned near both unloading zones to minimize the travel distance and travel time.

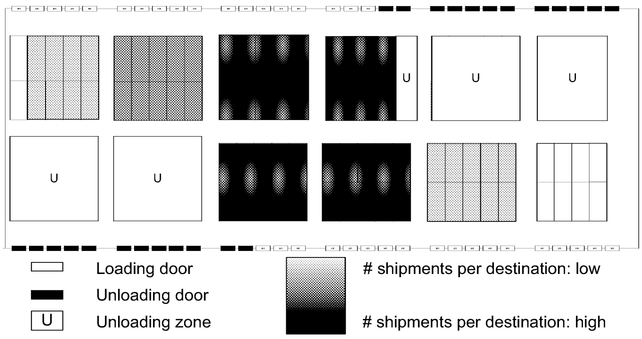


Figure 2a: Layout with unloading doors in the corners

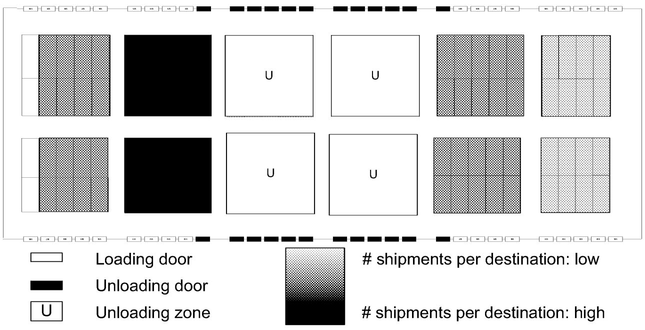


Figure 2b: Layout with unloading doors centered and opposite to each other

For both layout concepts, simulation experiments have been run to determine the full and empty runs of forklifts and the related time. In both cases, the available fork lift with the shortest distance to the specific unloading zone has been chosen to transport the shipment to its buffer area.

Table 1: Results of different layout concepts concerning the position of unloading doors

	Corner	Centered
Avg. time full run (sec)	28.56	25.48
Avg. time empty run (sec)	22.58	17.47
Total time for internal transport per shipment (sec)	51.15	42.95
Percentage Full run	55.8	59.3
Percentage Empty run	44.2	40.7

Some results of the simulation are shown in Table 1. The total time for internal transport per shipment (including full and empty run) can be reduced significantly by 16 % by positioning the unloading doors at the center. In addition, the percentage of empty runs is reduced significantly. Even though the travel time per handling unit is only reduced by 8.2 seconds, the reduction per day, with an average shipment volume of 5000 shipments per day, is about 11.39 hours. This time reduction can lead to a better service level, because the shipments are placed on the loading buffers earlier and therefore can be loaded sooner, or might result in a partly reduction of staff.

Arrangement Of Travel Paths and Door-Destination Assignment

The scope of the second model is a terminal of 11,700 m² with 11 unloading doors and 110 loading doors (8 of them not used for groupage handling). 4,500 shipments per day are handled for inbound and outbound groupage, unequally distributed on 36 national and 80 regional destinations.

During the simulation study, different layout concepts have been evaluated. One focus was on the rearrangement of internal areas and travel paths, keeping the position and number of unloading doors in the corner of the building (terminal part 2). Lower handling costs therefore only can be obtained by shorter internal transportation distances due to better position of internal areas (concerning the main haul destinations) and a new position of travel paths in terminal parts 3 and 4. In terminal part 4, only local destinations and export destinations are handled. Thus, there was no new door-destination assignment needed. (Figure 3a and 3b)

In the current layout, the travel path for fork lift is at the sides of the hall, which leads to crossing material flows of fork lift transports and loading the vehicles. By separating the travel path from the aisles for loading, the loading of vehicles can be done faster due to the elimination of crossing material flows. In addition, a new assignment of trailers to load doors has been done to reduce the internal transportation distance. Based on the number of shipments for each destination and several constraints of positioning, a new door-destination table was implemented in a modified simulation model concerning the travel paths.

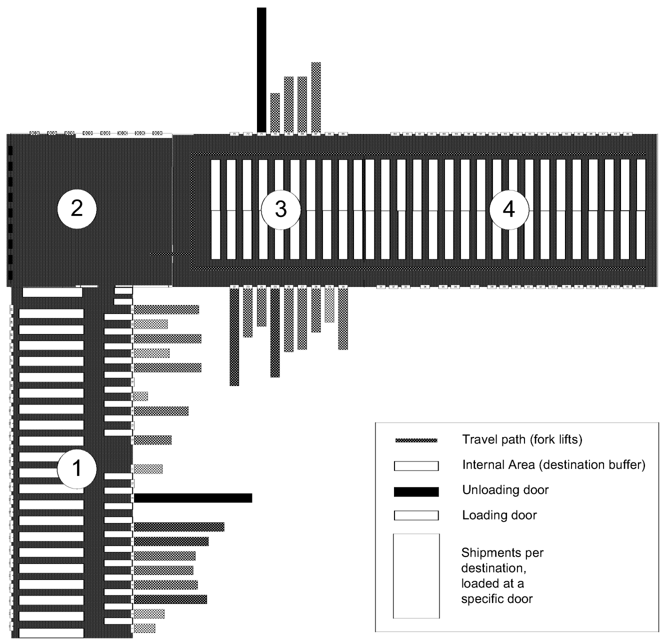


Figure 3a: Layout with travel path near the loading doors and historical arrangement of internal areas

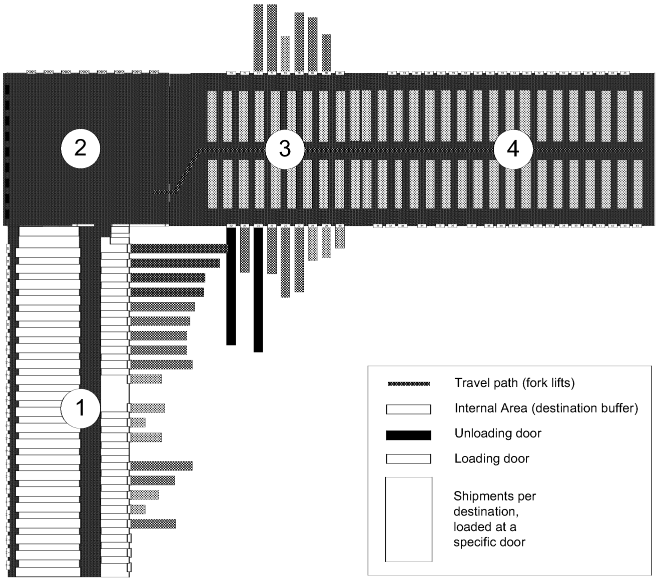


Figure 3b: Layout with centered travel path and new door-destination assignment

Due to the new layout concept and the new door-destination assignment, the transport time of the fork lifts could be reduced by 2.2 %. This leads to a reduction of the time at the doors for the long haul vehicle by 25 minutes, which is a reduction of 7.2%. As a result, the long haul vehicles are altogether able to leave 13.5 hours earlier each day and therefore the cut-off-times will be kept significantly better.
For the local vehicles, the reduction of loading time by 9.1% is mostly due to the faster loading according to the separation of aisle and fork lift travel path, which results in an earlier leave of 8.4 hours per day.

Table 2: Results of present and improved layout concept

	Present Concept	**Improved Concept**
Transport time forklifts per day (hr)	179.0	175.0
Time per local vehicle for loading at a door (incl. waiting for shipments) (mins)	77	70
Time per long haul vehicle for loading at a door (incl. waiting for shipments) (mins)	347	322

CONCLUSION

This paper showed that improving LTL terminals with simulation brings a benefit for terminal operator. Especially the tasks of layout planning can be simulated to identify aspects for reducing handling costs. It was outlined by case studies that the position of unloading and loading doors influences the internal transport time per shipment as well as decisions concerning size and spatial arrangement of internal areas (e.g. buffering areas) and of travel paths and door-destination assignment of trailer do. All of these tactical level decisions allow enhancing the performance of LTL terminals.
As the aspects of layout planning are strongly depending on the strategic decisions made but are also influencing them on the other hand, the need of an integrated solving tool is quite obvious. The developed toolbox for generating layouts of LTL terminals allows for that purpose the efficient creation of different simulation models for scenario analysis. By using this tool in combination with simulation as a method for optimizing LTL terminals, it is possible to evaluate different scenarios within a simulation project within short-term time horizons.

REFERENCES

Clausen, U.; Chmielewski, A.; Naujoks, B.; Janas, M., 2009: “Optimizing the Door Assignment in LTL-Terminals”. In: Transportation Science (Journal of the Institute for Operations Research and the Management Sciences), informs, 43 (2), 198-210

Gue, K.R. 1999. “The Effect of Trailer Scheduling on the Layout of Freight Terminals”. In: Transportation Science (Journal of the Institute for Operations Research and the Management Sciences), informs, 33, 419-428

Gocev, P.; Rabe, M. 2010. “Graphical Layout Planning Module within a Semantic Web Framework for Modelling and Simulation,” In: Integration Aspects of Simulation: Equipment, Organization and Personnel, Karlsruhe 2010.

Neumann, L. 2007. “Velos – a traffic simulation tool at the interface of internal and external logistics”, Proceedings of the 21st European Conference on Modelling and Simulation, 537-541.

Peck, K. E. 1983. “Operational Analysis of Freight Terminals Handling Less Than Container Load Shipments”, Ph.D. thesis, University of Illinois at Urbana-Champaign.

Tsui, L.-Y. and Chang, C.-H. 1990. “A Microcomputer Based Decision Support Tool for Assigning Dock Doors in Freight Yards,” Comput. Indust. Eng. 19, 309–312

Tsui, L.-Y. and Chang, C.-H. 1992. “Optimal Solution to a Dock Door Assignment Problem,” Comput. Indust. Eng. 23, 283–286.

IN-HOUSE TRANSPORT LOGISTICS

AGV CONTROL - THE KEY TO SUCCESSFUL TRANSPORT SYSTEMS WITH AUTOMATED GUIDED VEHICLES IN THE PRODUCTION ENVIRONMENT

Sander M. Slagboom
Hans P.M. Veeke
R. van der Stappen
G.Lodewijks
Faculty of Mechanical Engineering, Maritime and Materials Engineering
Delft University of Technology
Mekelweg 2
2628 CD Delft, THE NETHERLANDS

Keywords: AGV, Production environment

ABSTRACT

In the production environment of Vector Aandrijftechniek BV, a new AGV system is required to increase flexibility and efficiency of internal transport of parts and products. In this paper we consider which way of control is required in such an environment: centralized or decentralized. Four control architectures are designed according a modular concept and one of these architectures is chosen to be simulated in order to show the feasibility of the concept, which has different functions that each can be located centralized or decentralized. The functions considered are planning, dispatching, traffic control, order execution and routing and route execution and positioning. In a simulation model the routing strategy has been applied both centralized and decentralized to show the differences between both ways of implementation.
The results show that there are only very small differences between both strategies and about 3 AGVs can handle the required production of 70000 drives a year (including peak production). The conclusion is that the modular oriented centralized concept works properly . It is recommended to control the AGV as simple as possible. In case of the routing strategy the choice will depend on future requirements and financial inducements.

Keywords : AGV, Production environment, centralized and decentralized control.

1 INTRODUCTION PROJECT AGV CONTROL

1.1 Eurodrive project

Vector Aandrijftechniek BV is a company which produces electromechanical drives and is a subsidiary of SEW-Eurodrive. For the future, Vector has the vision to become a “model Eurodrive” and plans to move to a new building. With this move, the strategy will be to fit an optimized (called lean) production process. A project has been initialized for the development of this process. Within this project different topics will be addressed such as an electronic data medium and a painting robot. One of the subjects of the reorganization considers the internal logistics. Production islands are configured for an optimized assembly process but because there are lots of storages throughout the building a lot of movement are needed to fill up these storages. Vector searches a production mechanism where redundant transport processes and storages are minimized, match within the needs of the process and processes are automated to lower the costs and improve the controllability.

1.2 AGV Project

At the moment current transports from warehouse to production islands are done manually, an assembler picks the parts of a motor and a gearbox and walks to the island to assemble the whole drive.
The transport between production islands and the test station is done by conveyor belt. In the desired situation the location flexibility of the island must be higher and the efficiency needs to be improved. The technique, which Vector would like to implement to acquire these improvements, is an AGV system.

In figure 1 the flow of materials through the new building (concept) is shown. In the center the warehouse, the islands around it, the oil fill and test station are positioned, at the upper left the painting, drying and packaging stations are situated in sequence. The production in 2008 is about 57000 drives but the design will be dimensioned to accommodate an amount of 70000 drives a year on average (with a maximum of 100000 drives).

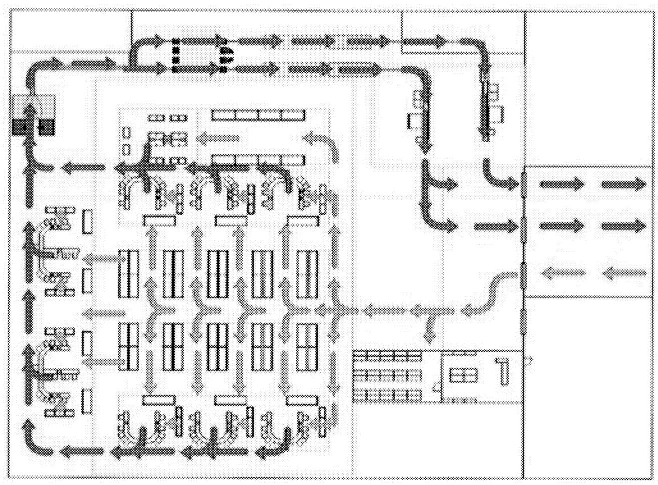


Figure 1. Concept layout of new production site

The idea is to use the AGV's for the transport from the warehouse to the oil fill and test stations, with the islands in between. The parts are picked in the warehouse and transported to the islands. An assembled drive will be transported from the islands to the oil fill and test station.
The average number of transport orders, will be 75 per hour in the desired situation (assuming 1875 hours a year).

1.3. Goal of this research

A previous paper (van Leeuwen et al, 2010) discussed the type of AGV's and layout that would be needed to achieve a predefined production volume. In this paper we consider which way of control is required in such an environment: centralized or decentralized.
The AGV's should have a supervising control system, which manages all the AGV's, or each AGV should have the intelligence to operate by itself. Those kinds of control also can be described as centralized (supervising management) and decentralized (autonomous intelligence). Vector does not know which way of control can satisfy the requirements of the desired AGV system. Therefore the research question is:

"Which way of control suits the new AGV system best?"

Looking at literature about the topic AGV Control, we can find a number of subjects that are related to this main question. These subjects represent questions that should be answered to complete the design. They concern:

- Kind of control; what are the differences between centralized and decentralized control?
- Dispatching; how are the AGV's dispatched or scheduled?
- Positioning; how does an AGV know its position?
- Routing; how can an AGV reach its destination?
- Safety (collision); how to avoid collisions with the environment and moving objects?
- Deadlocks; How to avoid them?
- Parking; what should an AGV do, when no jobs are available?
- Battery management; where and how does the AGV get its energy?

The main goal of this project can be described as follows:
Design a (de)centralized control system for the required AGV-system that can be implemented in the production facility of Vector.

The outline of this paper is as follows. In the next paragraph, a short review of literature is given, from which four concepts of AGV control will be introduced and compared. The third paragraph will explain details of the simulation model that was constructed to experiment with (de)centralized control. The results are shown in paragraph 4 and finally conclusions will be drawn and recommendations will be done.

2 AGV CONTROL

2.1 Literature review

Papers about the subject of AGV control are widely available. Some of them are broadly oriented in vehicle based material transport (Le Anh, 2005) and sometimes the focus is more on the use of agents as an approach of decentralized control (Lindeijer, 2003; Weyns, 2006). Also a lot of papers are focused on one single subject like routing (Duinkerken et al, 2006), planning (Nishi et al, 2005), deadlocks (Yoo et al, 2005) and positioning (Bosien et al, 2008). Besides research papers on the development of new algorithms, other papers focus on the implementation of AGV Control in real operating production environments (Weyns et al, 2005; Mes et al, 2007). Most of these implementations seem to work technically well in a laboratory environment but are not (yet) implemented in practical systems.

Overall we can see that decentralized control functions try to simplify the control system by using intelligent algorithms in order to improve flexibility. Unfortunately the predictability of the system as a whole becomes low and therefore complex problems may occur. The centralized control had the disadvantage that intelligent algorithms cause longer computational times, and make the control more complex.

2.2 Architectural design concepts

The AGV control system will get orders from the process controller which depends on the information of SAP. An order will be released if the parts are in the warehouse and a buffer at the islands is empty, but also when a drive is finished. To execute the transport job with an AGV the control has to maintain different functions to translate the transport order into movements of the AGV.

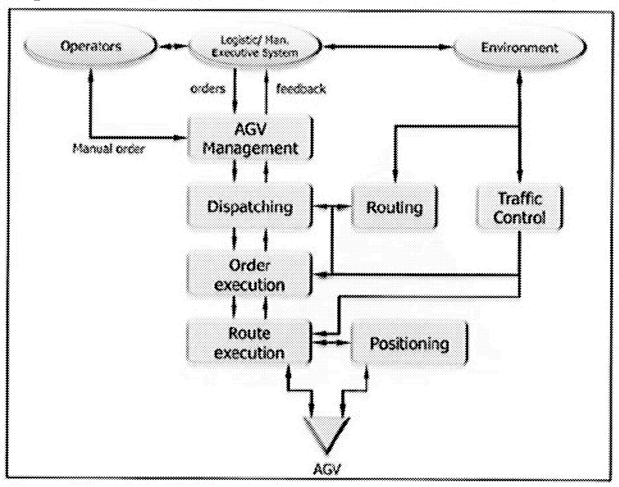


Figure 2. Model of AGV control

These functions, determined from literature study, can be found in figure 2 and they are:
- AGV management; receives the orders from process control, manages the AGV(s) and knows the status of each AGV, creates jobs for maintenance or battery

charging and monitors AGV's during operation. It will also be responsible for providing information to the user of the system or for manual control by an operator.

- Dispatching; receives the information of available AGV's and transport orders from AGV management. It assigns a transport order to an AGV (or vice versa, depending on the algorithm used).
- Order execution; receives the assigned transport order from dispatching and will translate it into a job, consisting of the loading and delivery point and the route.
- Routing; gets a request from order execution and calculates a route from a starting point and the destination(s) of an AGV.
- Traffic control; it gets the information about the actual routes and the positions of the AGV's within a certain time window. It can order AGV's to stop or redefine routes (depending of the algorithm used).
- Route execution; receives the route from order execution and translates it into jobs for drive control. These jobs could be something like move, stop, charge and load or unload.
- Positioning; measures the actual position of an AGV and compares it with the required route position. With this position a route can be checked and calculated or adjusted.

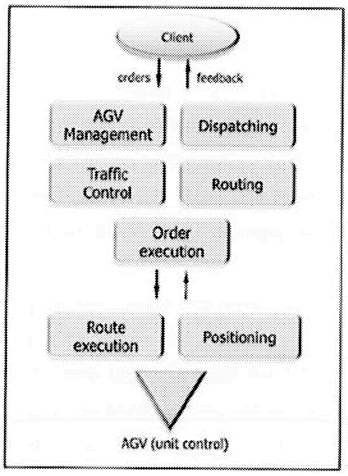


Figure 3. Modular centralized concept

From these seven functions four concepts are designed: a centralized concept, a decentralized concept, a centralized oriented concept and a decentralized oriented modular concept. These concepts are compared with each other. The conclusion is that the functions of AGV management, dispatching and traffic control can best be centralized because of the efficiency and simplicity (especially with a low number of AGV's).
The route execution and positioning are expected to perform better near drive control and should be decentralized. Order execution and routing are expected to perform slightly better in a centralized way. Routing needs in general information that is only available centrally. The selected concept is a centralized oriented modular concept and is shown in figure 3. This concept will be used in the simulation model.

3 SIMULATION MODEL EURODRIVE

3.1 Model description

The goal of the simulation model is to show the detailed working of a designed concept and to find out which differences occur between a centralized and decentralized routing strategy. The model uses Delphi in combination with the simulation package TOMAS (Veeke, Ottjes, 2002).

The layout used in the simulation is shown in figure 4, and more in detail in fig. 5. The AGV is represented by a blue square and drives in a layout of segments and transfer points. The location of these segments and transfer points can be defined by the user, and is in this case the designed production facility of Eurodrive. Tablets, one with the motor parts and one with the gearbox parts, will be loaded at the warehouse and delivered at the dedicated locations at the islands.

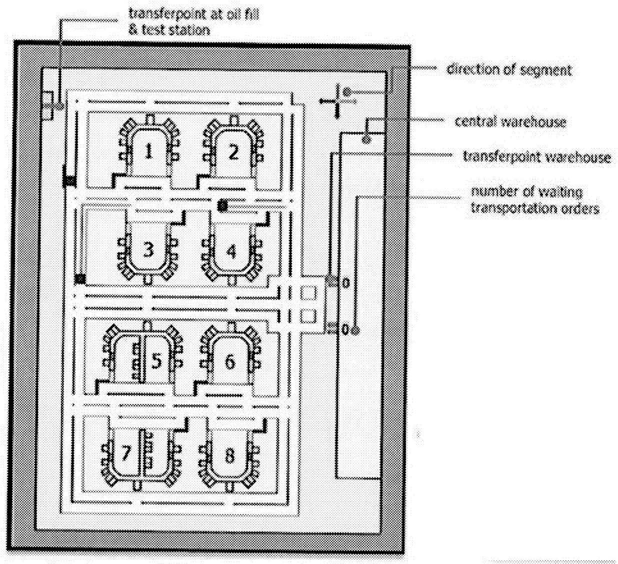


Figure 4. Simulated layout

Some assumptions have been made:

- An AGV is represented by a square and can only move in the horizontal or vertical direction. This assumption will disable the effect of swerving. In reality this effect not be neglected, but the design of the layout can be made in such a way that the effect has only little influence (wide lanes and intersections).
- The speed of the AGV is constant, both in straight parts and in angular parts. The angles are rectangular and the extra distance needed is analogous to lowering the speed about 20% for a real angle path.
- An AGV can have two possible transport orders. From the warehouse to an island or the dedicated parking place and from an island to the oil fill and test station and the parking place near the station. The orders available at the warehouse will be handled first.

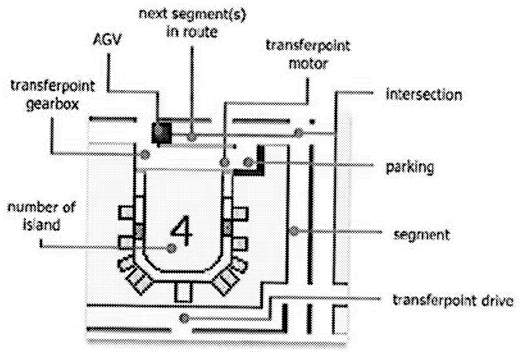


Figure 5. AGV driving near an island

The working principle of the simulation is described below in summary.
At the start transportation orders are created, which normally come from a client system, but in the model they come from a tablet generator. Two tablets are generated according a predefined inter-arrival time at a random transfer point at the warehouse. Central control will notice the new order and search for an available AGV. If available, a route will be defined in detail. The AGV will then execute its route and the traffic controller will intervene when necessary. When an AGV is ready with the transport order it will move to a parking place nearby, and not block the transfer point for other AGV's.

Besides the generator of orders at the warehouse, islands will generate orders when a drive has been assembled. After delivery of tablets to an island, the island will take the tablets into process.
After a defined period of time, the drive will be assembled and a new transportation order is created. These orders are assembled in the same way as from warehouse to an island, but only one tablet with a drive will be transported.

3.2 Input parameters

Figure 6 shows the interface of the model. The following input parameters are available in the setup of the simulation model:

- Routing strategy
- Runtime
- Average number of tablets/hour and standard deviation
- (Un)load time
- Number of AGV's available
- Maximum speed and acceleration of an AGV

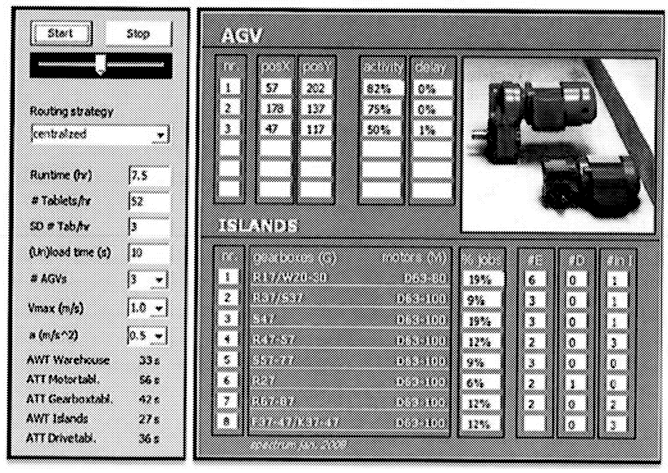


Figure 6. Interface of the simulation program

In table 1 the input settings for each experiment can be found. These settings are based on the values of the graph in figure 7. This graph represents all the numbers of drives delivered per day (blue points) when the current production is scaled to the desired production (100% = 70000 drives). The blue line shows the effect of taking the average of two consecutive days. It happens when the amount of ordered drives is too high, the assembly can be done the day before delivery. Besides the number of deliveries, two lines are shown, the red line represents the desired peak (140%) and the orange line represents the high peak (200%). The top peak is not visible in the graph, but these settings are used to simulate the effects of a simulation with a higher number of AGV's.

			Speed (acc)	1.0 m/s (0.5 m/s^2)		2.0 m/s (1.0 m/s^2)	
	# tablets/year	# tab/hour (SD)	(un)loadtime	10 s	5 s	10 s	5 s
current	(55693)	30 (5)		2, 3	2, 3	2, 3	1, 2
desired	70000	37 (5)		2, 3	2, 3	2, 3	2, 3
desired peak	100000	53 (3)		3, 4	3, 4	2, 3	2, 3
high peak	140000	75 (0)		4, 5	4, 5	3, 4	3, 4
top peak	200000	107 (0)		6, 7	5, 6	4, 5	3, 4

Table 1. Input settings for simulation

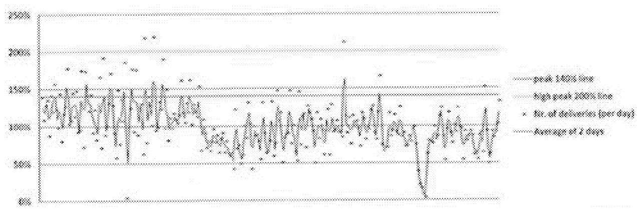


Figure 7. Number of drives delivered per day

From table 2 the input data for the generator and the islands are calculated. Here we see the actual number of drives produced in a year. With this information the share of each island, the ratio between different assembly times and the absolute assemble times are known. The number of FTE per island can also be calculated (production uses a maximum efficiency of 70%).

		Normal motor				SS or repaired motor						
island	gearbox type	time (s)	number	%	FTE*	time (s)	Number	%	FTE*	Total FTE	Total nr.	%
1	R37/S37	972	2691	30%	0,39	515	6133	70%	0,47	0,86	8824	16
2	S47	958	3451	42%	0,49	503	4715	58%	0,35	0,84	8166	15
3	R47-57	1120	2195	46%	0,36	585	2588	54%	0,22	0,59	4783	9
4	W20-30/R17	1253	11303	100%	2,10	514	0	0%	0,00	2,10	11303	20
5	R27	870	2568	45%	0,33	538	3143	55%	0,25	0,58	5708	10
6	K37-47/F37-47	1425	2562	51%	0,54	566	2501	49%	0,21	0,75	5063	9
7	S57-77	1122	3521	52%	0,59	583	3187	48%	0,28	0,86	6708	12
8	R67-87	1223	2773	54%	0,50	625	2365	46%	0,22	0,72	5138	9
Total			31064		5,3		24629		2,0	7,3	55693	100

Table 2. Quantities and times of assembly islands

4 RESULTS OF SIMULATION

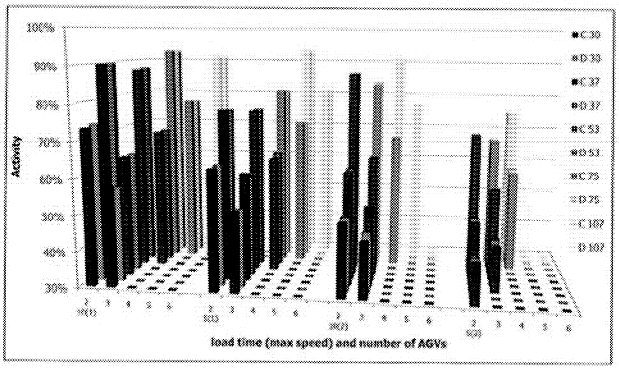


Figure 8. Activity of the AGV's

Figure 8 shows the activity of AGV's. The activity is the relative time an AGV is busy with a transport order including delays. The color shows the amount of tablets per hour (which stands for the amount of drives to assemble). It shows that with the desired peak production (53 p/hr, red) the use of three AGV's is sufficient in most cases. There is almost no difference between centralized and decentralized routing.

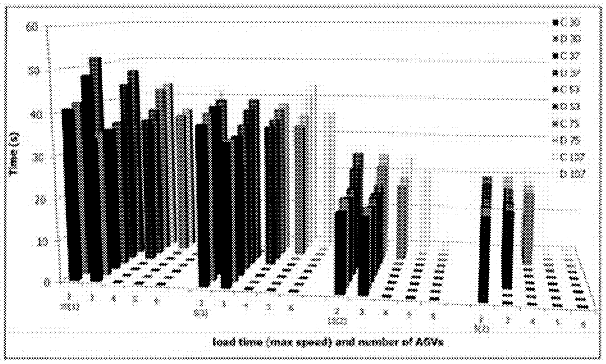


Figure 9. Average waiting time at warehouse

In figure 9 the average waiting time of tablets at the warehouse is shown. These results are dependent on the speed of the AGV. With a moderate activity one may expect an average waiting time of 35 sec. with an AGV- speed of 1m/sec and 18 sec. with 2 m/sec. When the activity rises above 75%, the average waiting time rises a few seconds higher.

Figure 10 shows the average waiting time of tablets at the islands. Also here the waiting time depends on the speed of the AGV, but the waiting time will not increase when the activity is high. An AGV picks up warehouse tablets first, so if the activity is too high, the island tablets will not be picked up anymore and the number of tablets with high waiting times is therefore limited.

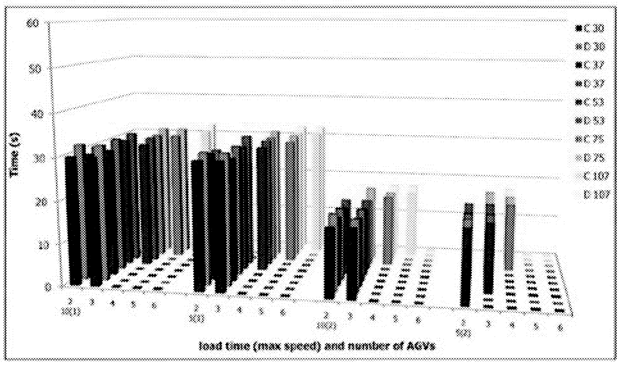


Figure 10. Average waiting time at islands

Also when we look at the centralized and decentralized routing strategy in the graphs of the average waiting times no big differences can be observed.

Finally the number of AGV's needed can be determined with the use of the graph in figure 11.This graph shows for each case(speed and loading time) the number of required AGV's. 100% represents the desired production of 70000 drives per year. The white dots in the graph represent the current number of drives delivered per day (with an average of 57000 per year). When the same distribution is expected for the desired production, 3 days a year a peak of 600 drives per day can be expected (similar to 150000 per year). With a loading time of 10 seconds and a maximum speed of 2 m/sec, 4 AGV's are required to handle the flow of transport orders.

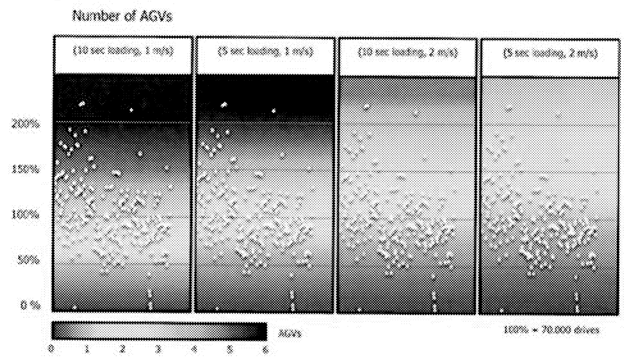


Figure 11. Number of AGVs

5 CONCLUSIONS

The design of the centralized oriented modular concept is working properly. It can be used to operate as AGV control for the desired AGV system with the industrial environment of the model Eurodrive.

No significant differences between a centralized and decentralized routing strategy have been found. In this case the choice for a strategy is determined by financial inducement. When the requirements for control are expected to become higher in the future, and the route needs to be based on more information that is only available centrally, then it is recommended to use the centralized strategy.

In the case of the model Eurodrive it is recommended that the designed environment is adapted to make AGV control simple and its effectiveness high as a transport system. It should be the main goal to meet the requirements of the islands. With simple routes, short loading times and not really high peaks in the number of tablets, it could be interesting to use only 2 AGV's separately; one for routes from warehouse to islands and one from the islands to the oil fill & test station. It is also recommended to test the AGV control in real operation; start with a simple process of one job handling at a time. In the future it can always be expanded with more intelligence, when the requirements from the production process become higher.

ACKNOWLEDGEMENTS

This work is supported in part by Vector Aandrijftechniek BV.

REFERENCES

Bosien, A., Venzke, M., Turau, V., 2008, "A rewritable RFID environment for AGV navigation", Internatonal Workshop on Intelligent Transportation, Hamburg

Duinkerken, M.B., Ottjes, J.A., Lodewijks, G., 2006, "Comparison of routing strategies for AGV systems using simulation". Proceedings of the 2006 Winter Simulation Conference, Perrone, L.F., Wieland, F.P., Liu, J., Lawson, B.G.,Nicol, D.M., Fujimoto, R.M. eds., Monterey, California, USA, pp. 1523- 1530

Le-Anh, T., 2005, "Intelligent control of vehicle-based transport systems", Erasmus Research Institute of Management (ERIM), isbn: 978-905892-079-9

Lindeijer, D.G., 2003, "Controlling automated traffic agents", Dissertation, Delft University of Technology, Eburon, isbn 9051669488

Mes, M.R.K., van der Heijden, M.C., van Hillegersberg, J., 2008, "Design choices for agent based control of AGV's in the dough making process", J. on Decision Support Systems, Vol. 44, Iss. 4., Elsevier, Amsterdam

Nishi, T., Ando, M., Konishi, M., 2005, "Distributed route planning for multiple mobile robots using an augmented langrangian decomposition and coordination technique", IEEE International conference on Robotics and Automation Vol. 21 (6), pp. 1191-1200

Van Leeuwen, E., Veeke, H.P.M., van der Stappen, R., Lodewijks, G., "Automated Guide vehicles in a production environment", Proc. of ESM 2011, Janssens, G.K., Ramaekers, K., Caris, A., eds, Hasselt, pp. 299-305, isbn 978-90-77381-57-1

Veeke, H.P.M., Ottjes, J.A., 2000, "TOMAS: Tool for Object-oriented Modeling And Simulation", proceedings of Advanced Simulation Technology Conference (ASTC2000), Washington, D.C. pp. 76-81. The Society for Computer Simulation International (SCS), ISBN: 1-56555-199-0

Weyns, D., 2006, "An Architecture-centric Approach for Software Engineering with Situated Multiagent Systems", Dissertation, Katholieke Universiteit Leuven

Weyns, D., Schelfhout, K., Holvoet, T., Lefever, T., 2005, "Decentralized control of AGV Transportation Systems", Proc. Of AAMAS 05, ISBN:1-59593-093-0

Yoo, J.W., Sim, E.S., Cao, C., Park, J.W., 2005, "An Algorithm for deadlock avoidance in an AGV System", Int. Journal of Advanced Manufacturing Technology, 26, pp. 659-668

Forklift free factory: a case study of different transportation systems in the automotive industry

Tim Govaert, Karel Bauters, Veronique Limère, Hendrik Van Landeghem
Ghent University, Department of Industrial Management
Technologiepark 903
B-9052 Zwijnaarde, Belgium
E-mail: tim.govaert@ugent.be

KEYWORDS
simulation, electrified monorail system (EMS), automatic guided vehicle (AGV), tugger, kitting, Plant Simulation

ABSTRACT

This paper describes the results of a large project executed in the automotive industry. The goal of the project was to simulate different transportation systems (EMS, AGV and tugger) in order to compare their performances. In our simulation, we focused on investigating the following parameters: amount of loading stations, equipment and operators needed, storage place at the line,…

INTRODUCTION

The internal logistics within most production plants are currently based on forklifts. They transport the different goods throughout the factory. Forklifts are proven to be flexible because their duties can easily be adapted to changing situations and environments. But on the downside, they are also very inefficient and prone to accidents on the work floor. For all these reasons, we wanted to investigate if alternative transportation systems could be used to provide all parts to the assembly lines. First, a feasibility study was carried out. This study pointed out that only replacing the transportation system is not very cost effective. Therefore, alternative material flow and line delivery strategies needed to be investigated (Cottyn et al., 2008). Rough calculations in the feasibility study also showed which strategies were not worthwhile investigating and which were. This paper describes the results of a case in the automotive industry, where the best fitting line delivery system was designed using a detailed simulation model in Plant Simulation, a software package of Siemens PLM Software.

LITERATURE REVIEW

This paper deals with the selection of the most appropriate materials handling equipment for the supply of parts to an automotive assembly line. The Material Handling Industry of America defines material handling as follows (MHIA, 2011):

Material Handling is the movement, storage, control and protection of materials, goods and products throughout the process of manufacturing, distribution, consumption and disposal. The focus is on the methods, mechanical equipment, systems and related controls used to achieve these functions.

In contrast to logistics, material handling thus focuses not so much on the organization of flows but on the mechanical equipment, systems and related controls. Material handling contributes a big percentage into product value (Le-Anh, 2005). Tompkins et al. (2003) indicate that material handling represents between 15% and 70% of the total cost of a manufactured product, and between 20 and 50% of total operating expenses in manufacturing.

In this paper we will focus on one aspect of material handling, i.e. the movement of materials within a manufacturing plant. The main purpose of this research is the wish of our case company in the automotive manufacturing industry to evolve towards a forklift free factory. Cottyn et al (2008) explain the advantages and disadvantages of forklifts. The number of forklifts is always strongly over-dimensioned because of transport fluctuations, the performance of the forklifts is depending on the traffic in the plant and the factory lay-out and finally, because it still is manual transport, the human factor is vulnerable to social disruptions. Thus the flexibility that forklifts offer causes a great loss of efficiency. Moreover, the greatest driver to go forklift free is the issue of worker safety (Neumann et al., 2007 and Gecker, 2004). Forklift injuries are numerous and are a number one reason to avoid them on the shop floor. For this reason, a study is carried out to investigate some possible alternative Vehicle-Based Internal Transport systems (VBIT).

In Figure 1, a classification is given for internal transport equipment. We will focus on motor-driven mobile internal transport equipment for horizontal movement.

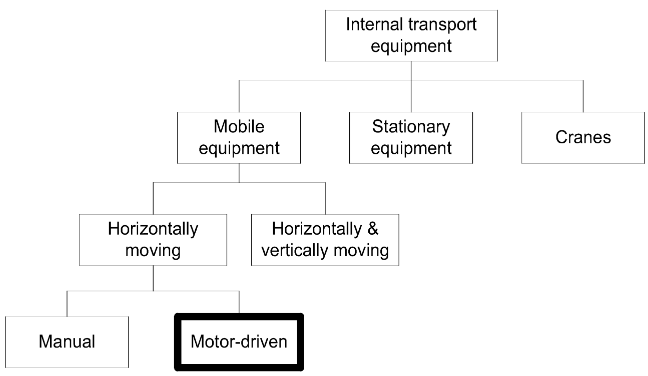


Figure 1: classification of internal transport equipment (adapted from De Koster, 1995)

	Scenario A	Scenario B	Scenario C
Transport system	EMS	AGV	Tugger
Carrier speed	1,8 m/s	0,5 m/s	1,4 m/s
Carrier length	1,5m	1,5m	11m
Spacing between carriers	0,5m	0,5m	1m
Acceleration	0,6m/s²	0,25m/s²	0,6 m/s²
Deceleration	0,6m/s²	0,5m/s²	0,6 m/s²
Loading time in supermarket	30s	30s	54s
Off-loading time in supermarket	30s	30s	54s
Off-Loading time at drop-off station	21s	30s	54s
Loading time at a separate station	67s	30s	54s
Loading time at a combined station	24s	30s	54s

Table 1: different transportation systems

Complex expert systems exist for the material handling equipment selection problem based on analytical methods, knowledge-based methods, or hybrid approaches (e.g. Mirhosseyni and Webb, 2009; Welgama and Gibson, 1995). However, in our research we want to do a parametric analysis to investigate the impact of three different systems, i.e. an Electrified Monorail System (EMS), Automatic Guided Vehicles (AGV) and tuggers. A simulation model is used to predict the behavior of the real-life situation and to generate new insights.

INTERNAL LOGISTICS FLOW

Bulk versus kit

In the situation as it is today, parts can be delivered by the supplier in two different types of packaging: blue boxes and pallets. These packages are respectively stored in the plant in the small box warehouse (SBW) and the high bay warehouse (HBW). When needed, these boxes and pallets are brought to the line, where there is an inventory of parts. This line delivery strategy is called bulk feeding.

In this paper, we investigate the possibility of bringing the parts to the line in kits. This means that, for every vehicle, a number of carts is made, containing every single part that needs to be mounted on that vehicle. These kits are then brought to the line in the sequence they need to be mounted on the vehicle. The assembly of kits is done in an area which we call the supermarket.

Transport systems

In the current situation, most line delivery is done by forklifts and partially by tuggers. In this paper we investigate if other transportation methods can be used. A feasibility study showed that a number of alternative transportation systems might bring a solution for the different problems that are linked with the use of forklifts:

- Electrified Monorail System (EMS): The Electrified Monorail System consists of an overhead track that runs along the different lines and pre-assemblies and connects them to the supermarket. On those rails a number of individual powered carriers transport the material. At each drop-off point, a drop-off station is installed. To limit the number of drop-off stations at the line, this transport system is only used in combination with full kitting.
- Automatic Guided Vehicle (AGV): The automatic guided vehicle is a mobile robot that is able to follow markers or wires in the floor. AGV's consist of a pull unit with a certain number of carts behind it.
- Tugger: Like the AGV, a tugger also consists of a pull unit and some carts behind it. The main difference with AGV's lies in the fact that tuggers are driven and loaded and unloaded manually by an operator.

Table 1 gives an overview of the different transportation systems and their main parameters used in the case study. Every EMS carrier can transport exactly one kit. A tugger can tow up to 7 kits. It will be important to foresee enough moving space for the tugger to turn in the production hall. The loading and unloading times for the EMS are obtained from the supplier. The loading and unloading times for the tugger are obtained from measurements at another company. The AGV loading times are obtained from the feasibility study.

MODEL GENERATION

For the generation of our model we have chosen to combine the commercial simulation package (Siemens Plant Design) with a free-to-use programming language (Python). More details about the simulation model are described by Govaert et al. (2009). Using this method, we were able to generate the factory lay-out in a very flexible and cost effective way.

In order to load all the kits on the transport carriers in the supermarket, a certain amount of loading stations are simulated. The kits are then transported to the assembly line by using one of the above mentioned transport systems. At the assembly line the kits are unloaded at a drop-off station. Due to the high investment cost, we will try to limit the amount of drop-off stations. The unloaded kits are then stored in a buffer. The amount of buffers indicates the amount of similar kits at the line (a similar kit contains similar parts but for a different chassis number). In our simulation both one and two buffers at the line are simulated. Per drop-off station, 7 different kits will be unloaded, which will then travel along the assembly line for several workstations.

MODEL OUTPUT

There are three kinds of output generated while running the simulation model:

- Log messages that are written to a log file
- Statistics of the objects in the simulation model after running a scenario
- Visualizations

There are four kinds of log messages:

- Msg: This entry is logged when a buffer has an empty space and sends a message to the line supplier.
- Tra: This entry is logged when a transporter is loaded and ready to transport the container to the right destination buffer.
- Buf: This entry is logged when the empty space in a buffer is refilled.
- Err: This entry is logged in case of a 'missing part'. A part is needed at a station when a new chassis arrives but the buffer of that part is empty or lacks the needed quantity.

RESULTS

In Table 1 you can see an overview of all the simulated scenarios:

EMS	Scenario MOD_A
AGV	Scenario MOD_B
Tugger	Scenario MOD_C

Table 1: overview of the simulated scenarios

For all scenarios a takt time of 10'03'' was considered.

Scenario A

In this scenario we investigate the behavior of an electrified monorail system (EMS) that only transports kits. As a consequence of mock-ups done concerning the number of drop-off points at the line, we have introduced a limited amount of drop-of points in order to limit the total cost. The kit is attached to the line and is running at the same speed of the truck.

A. Standard case

As a starting scenario we have decided to work with the following parameters:

- Amount of transporters: unlimited
- Amount of loading stations: 28
- Amount of buffers at the line: 1
- EMS speed: 1,8 m/s

For all the above mentioned parameters the working interval, upper or lower boundaries and ideal values are investigated in future paragraphs.

The results of the standard run can be found in Table 2: The first three rows in Table 2 give the average time between two logs in the model. These logs can either be a message-log (Msg), a transport-log (Tra) or a buffer-log (Buf). The last three rows give the maximum time between the logs.

Average time between Msg-Tra	1'05''
Average time between Tra-Buf	4'30''
Average time between Msg-Buf	5'35''
Maximum time between Msg-Tra	2'41''
Maximum time between Tra-Buf	6'27''
Maximum time between Msg-Buf	8'16''

Table 2: MOD_A: results of the standard run

The total duration of the simulation is 72 hours, out of which 12 hours were deleted in order to achieve a steady state model.

The maximum allowed time that can be used to replace a kit is the takt time. In our standard case, the maximum time used to replace a kit is 8'16''. As a consequence, no errors occurred. 1' 47'' is the margin that is still left for unexpected interruptions.

The loading times of the individual workstations are shown in Figure 2.

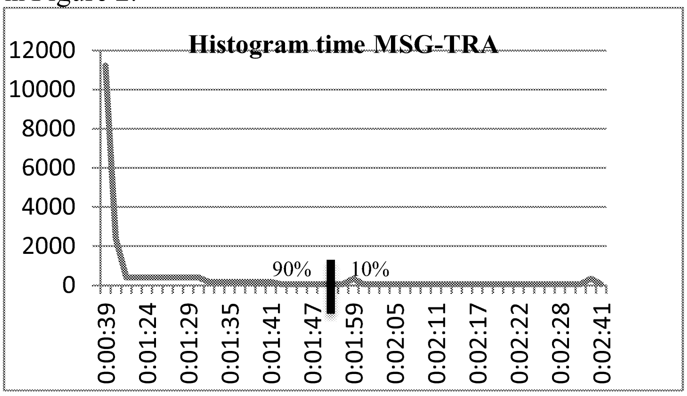


Figure 2: MOD_A: histogram of the loading times of the standard run

One can see that the majority of the kits are loaded in a very short time period. Nevertheless we also see that peaks up until 2'41'' exist. This means that in most of the cases a fast response is possible but that we will have to be careful with peak loads and corresponding waiting times.

Another interesting graph is the histogram of the amount of carriers that are in use. In Figure 3 one can see how often (y-axis) it happened that a certain amount of transporters (x-axis) are in use. The maximum amount of carriers that are used at the same moment is 113. We can also see that it is rather rare that less than 30 carriers are used at the same moment.

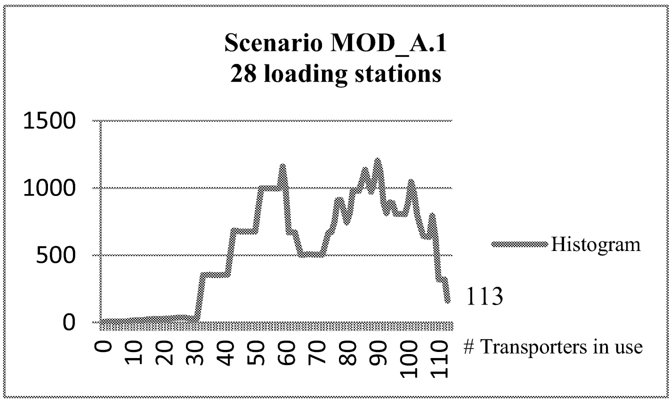


Figure 3: MOD_A: histogram of the amount of tuggers in use in the standard run

Now that we have an overview of several working limits, we will investigate the impact of varying the amount of loading stations, the amount of EMS carriers and the EMS speed.

B. Variation in the loading stations

Now that we have investigated the standard case, we will search for the working boundaries concerning the amount of loading stations. We will gradually decrease the amount of loading stations till the delivery is too slow to follow the takt time. At that moment, error messages will be generated.
We start with scenario MOD_A.1. The '1' in the scenario name refers to the amount of buffers at the assembly line.

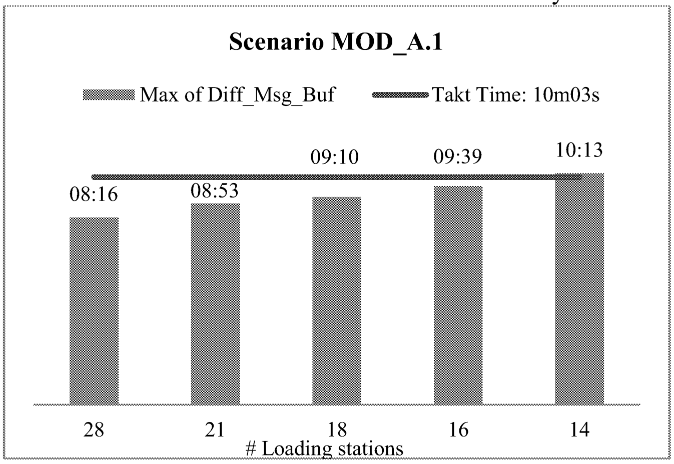


Figure 4: MOD_A.1, variation of loading stations and one buffer at the line

If we decrease the amount of loading stations we see that the replenishment time will gradually increase until the moment we are working with only 14 loading stations. At that point, the maximum replenishment time is bigger than 10'03''. Therefore the first errors occur (79 kits are delivered too late).
In a second step we will further decrease the amount of loading stations while working with a buffer of 2 kits. In this way more inventory will be stored at the line and therefore we can do the replenishment in a bigger time slot: 2 times the takt-time: 20'06''. The results of varying the amount of loading stations while having two buffers is shown in Figure 5.

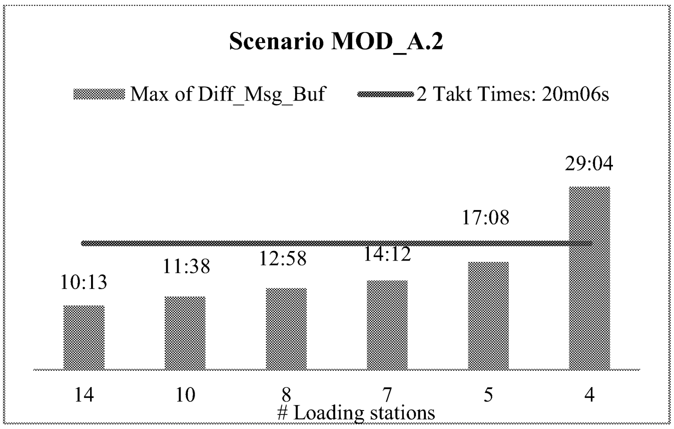


Figure 5: MOD_A.2, variation of loading stations and two buffer at the line

Now we can decrease the amount of loading stations until we only have 5 stations left. The maximum replenishment time at that point equaled 17'08'' and is lower than 20'06''. Decreasing the amount of loading stations to 4 would lead to extremely high replenishment times and a huge amount of errors: 4349 kits are not delivered on time.
As a conclusion we can say that an extra buffer helps us in overcoming the peak loads of replenishment orders that are often released. It is a trade-off between capacity at the line and capacity at the loading stations.
An overview of the spare capacity of the loading stations can be obtained by interpreting the utilization table.

# Loading stations	Utilization (%)
28	12%
21	16%
16	22%
14	25%
12	29%
10	34%
8	43%
6	57%
5	69%
4	86%

Table 3: utilization of the loading stations

C. Variation in the amount of EMS carriers

Another very interesting investigation is the importance of the amount of carriers. We will look at the impact of limiting the amount of carriers. Now the trade-off will be between the amount of carriers and the replenishment time. For all other parameters we use the settings used in the standard case (28LS = 28 loading stations).
Again, we have done the calculations for both 1 and 2 buffers at the line. We will start with 1 buffer in Figure 6.

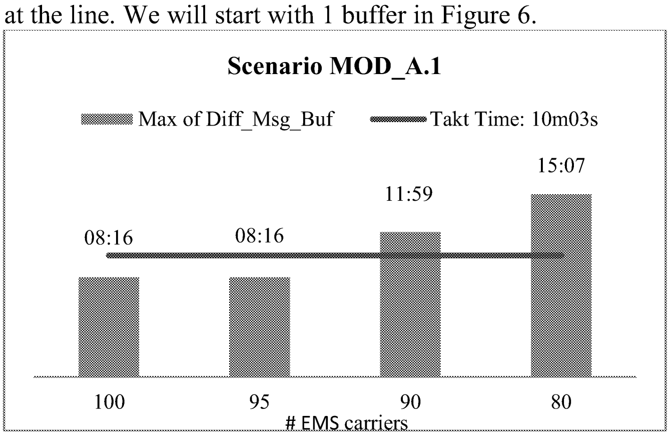


Figure 6: MOD_A.1: replenishment time versus amount of EMS carriers, one buffer at the line

In the standard scenario the maximum amount of carriers that was used at one point was 113. Now we see that we can easily decrease the amount of carriers to 95 without increasing the replenishment time. In fact, the maximum amount of kits that are requested at the same time is 91 kits. As soon as we have less than 91 carriers, the replenishment time increases drastically. More time will be needed for the replenishment and thus a second buffer will be necessary (Figure 7).

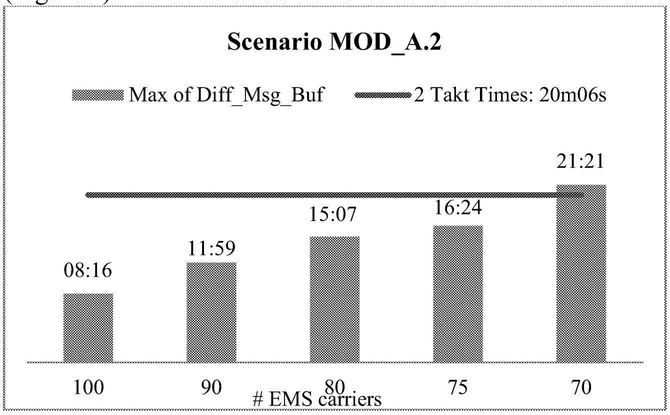


Figure 7: MOD_A.2: replenishment time versus amount of EMS carriers, two buffers at the line

We now see that all the kits can be delivered on time as long as we have 75 EMS carriers, which is less than the 91 carriers we need when we only have one buffer

D. Variation in the EMS speed

The standard case was calculated using the technical specifications obtained from the supplier. As a part of our robustness analysis we investigated the influence of a varying EMS speed (see Figure 8).

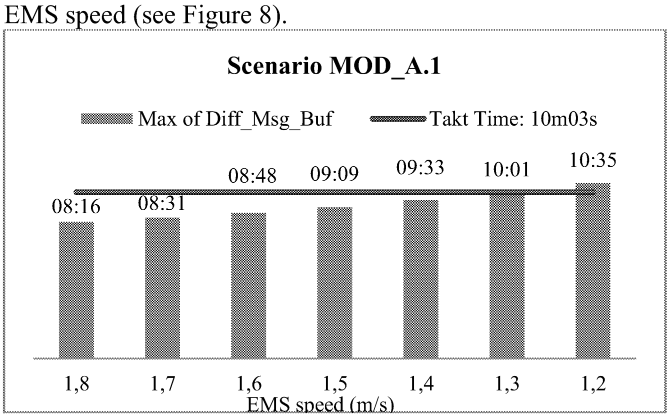


Figure 8: MOD_A.1: variation of the EMS speed

We can conclude that in the standard case the EMS speed can decrease until 1,4m/s before significant problems will appear.

E. Time-shift of the kits delivered to the same workstation

At this moment all the kits in a workstation are requested at the very same moment. As a result, big peak loads are released in the supermarket and at the loading stations. As a first step we will now release all the kits in a workstation evenly spread during the takt time. This will have two results:

- The request for kits will be more evenly spread in time
- There will be less stock at the line

As we can see in the results underneath, the impact is stunning! The amount of loading stations needed to run this scenario has decreased drastically (Figure 9). Also the amount of carriers that is needed to run the scenario is significantly less. The result for the calculations using 8 loading stations can be seen in Figure 10.

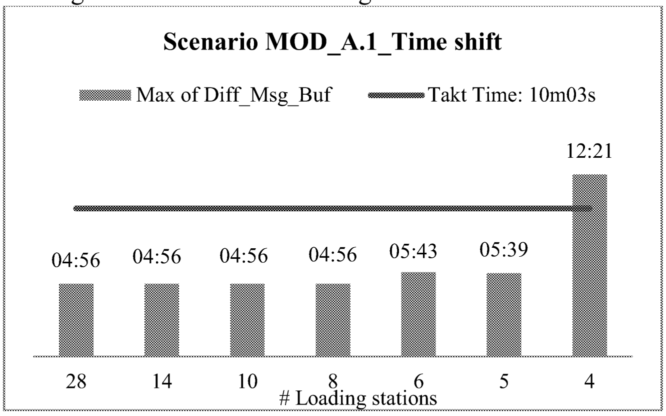


Figure 9: MOD_A.1: influence of the amount of loading stations with one buffer and a time shift

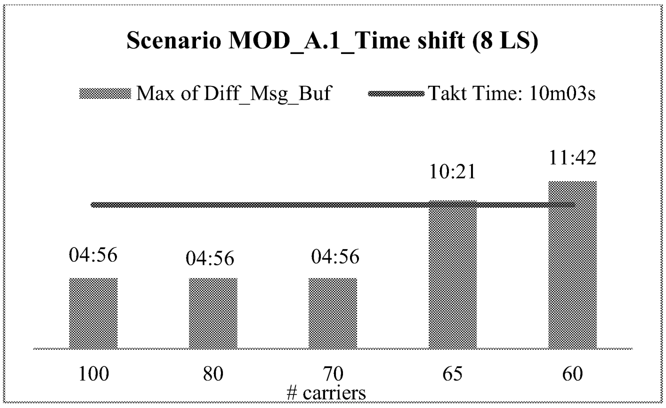


Figure 10: MOD_A.1: influence of the amount of carriers with one buffer at the line, a time shift and 8 loading stations

Under the above mentioned parameters, 70 carriers are sufficient to supply all the kits on time to the assembly line.

F. Time-shift of kits delivered to different drop-off stations

In all previous scenarios, all the drop-off stations send a request for the first kit at the same moment. Depending on the transport time to the specific drop-off stations we can request the kits earlier in order to have some extra available replenishment time. The moment we can send this ‘early message’ depends from drop-off station to drop-off station and is related to the minimum time the system needs to replenish a kit. Of course, this replenishment time is related to the distance from the supermarket to the drop-off station. Besides gaining extra time to replenish, we can also make sure the kits are requested at different times. As a consequence we are again reducing our peak loads. This reduction will lead to a decreased need of capacity at the loading stations, buffers at the line,…

The concept explained above is visible in Figure 11. On the first, black line we see the messages that are send in the simulation. The second line indicates the normal takt time. The third line indicates the total replenishment time that we have on hand if we release early messages. The forth line indicates the time that is gained by releasing this early message. The maximum of this time is equal to the minimum of the time between a transport from the loading stations and its arrival at the buffer, which depends on the individual workstation.

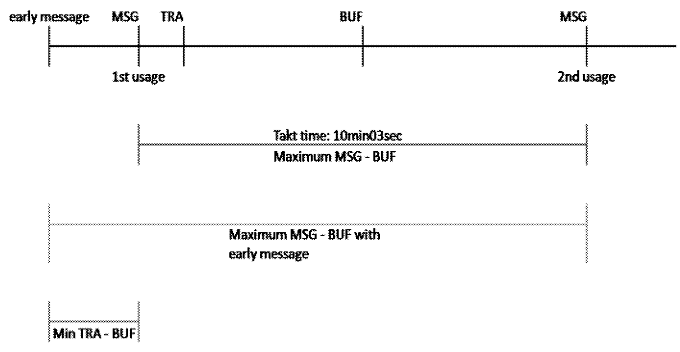


Figure 11: explanation of the early message concept

The ‘maximum time-shift for an early message’ is minimum 2’16’’ and maximum 4’06’’.

Scenario B

In this scenario, the delivery of the kits is done by Automatic Guided Vehicles. The following parameters were taken into account for the standard scenario:

- Amount of transporters: 270
- Amount of loading stations: 28
- Amount of buffers at the line: 3
- AGV speed: 0.5 m/s

After one simulation run we already saw that this scenario would not be possible due to the low speed of the AGVs.
Despite the investment in 270 AGVs and in a 3-kit-buffer, the model could not supply the kits on time. The congestion was sometimes so big that several areas got blocked for a longer amount of time.
Therefore it was concluded that AGVs are not suitable for the specific internal logistical situation at our automotive company.

Scenario C

This scenario is very similar to the previous scenarios. Again we are talking about total kitting with a limited amount of drop-off stations. The drop-off stations will be the same as in scenario A. The only difference is the mean of transportation: we now investigate the use of tuggers to transport the kits to the line.

A. Standard case

As a starting scenario we have used the following parameters:

- Amount of transporters: 30
- Amount of loading stations: 8
- Amount of buffers at the line: 1
- Tugger speed: 1,4 m/s

The results of the standard run are shown in Table 4:

Average time between Msg-Tra	0’30’’
Average time between Tra-Buf	4’33’’
Average time between Msg-Buf	5’03’’
Maximum time between Msg-Tra	0’30’’
Maximum time between Tra-Buf	6’15’’
Maximum time between Msg-Buf	6’45’’

Table 4: MOD_C: results of the standard run

If we compare these results with the EMS standard run we see that the time between the replenishment signal (Msg) and the moment the kit is loaded (Tra) is smaller in scenario C. The reason is that for the standard scenario all the tuggers can be loaded at the same moment. No time is lost due to extensive waiting. If we compare the transport time (time between Tra and Buf) we can see that both scenarios are doing almost as good. The difference between the scenarios is that scenario A is faster in transporting the kits but is slower in unloading the kits at the workstation. For scenario C all 7 kits are off-loaded at the same moment.
Each tugger is operated by an employee. Therefore it is very vital to have an overview of the amount of drivers needed to operate all the tuggers.

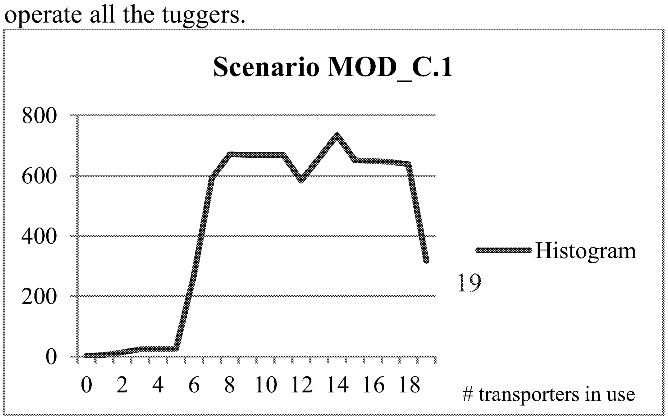


Figure 12: MOD_C: histogram of the amount of tuggers in use in the standard run

B. Variation in the amount of loading stations

The loading and off-loading of the tuggers is done by using an automatic system. Therefore, the amount of loading stations is not unrestricted. We have investigated the use of a limited amount of loading stations in the scenario of one buffer and enough tuggers (Figure 13).

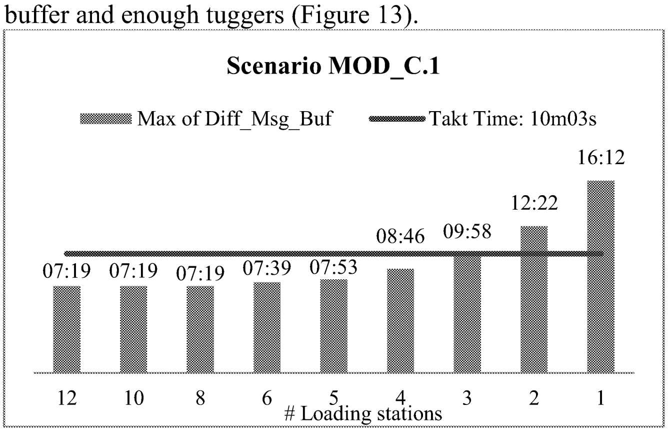


Figure 13: MOD_C.1: variation of loading stations (one buffer)

We can conclude that 4 to 5 loading stations will be enough. Here it is important to remind that one loading station consists of 7 small loading stations that can each load one kit. One big loading station loads one tugger that carries 7 kits at once.

C. Variation in the amount of tuggers

As we are using manned tuggers, a high operating cost is attached to each and every tugger that is in use. Therefore it is important to limit the amount of tuggers that are used in the model. We have researched the working limits for our scenario in relation to the amount of tuggers that are needed. The results can be seen in Figure 14

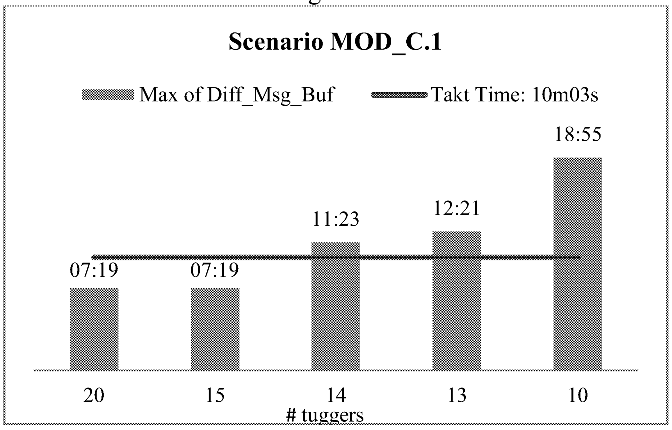


Figure 14: MOD_C.1: variation in the amount of tuggers, one buffer at the line

As we are interested in the minimum amount of tuggers needed it is interesting to see that we are able to manage the replenishment as long as we have 15 tuggers or more.
Due to the fact that staff is rather expensive, we will now simulate the same situation with 2 buffers. One can see in Figure 15 that it is possible to manage the supply with 13 tuggers instead of 15.

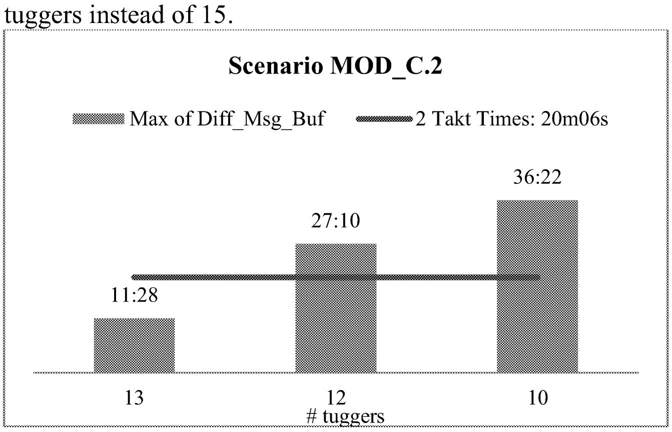


Figure 15: MOD_C.2: variation in the amount of tuggers, two buffers at the line

D. Variation in tugger speed

The average speed of a tugger is highly influenced by several factors. Human interaction, the experience of the driver and the amount of tuggers driving at the production floor are only some of the influencing factors. In order to make sure our system is robust enough to overcome these influences it is important that we execute a robustness analysis of the tugger speed. In our standard case we have calculated with a value of 1,4m/s. The maximum time needed to supply a kit was 6'45''.
We now see that the time increases step by step without exceeding the takt time of 10'02'', as long as the speed is at least 1m/s.

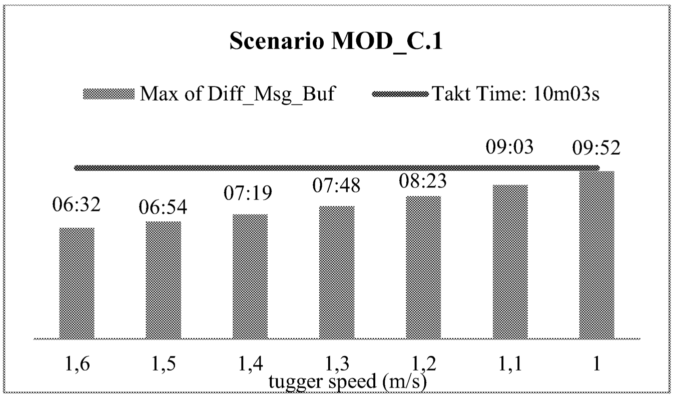


Figure 16: MOD_C.1: variation of the tugger speed

G. Time-shift of kits delivered to different drop-off stations

The same investigation will be done as for the EMS scenario. The 'maximum time-shift for an early message' is minimum 3'24'' and maximum 5'15''. We can conclude that the time-shift for tugger transport can be bigger than the time-shift for transport via EMS. This is due to the fact that transport via an EMS system is faster.

CONCLUSIONS

After introducing the bigger picture of the problem and clarifying some important definitions, the specific situation for our automotive case company was introduced and discussed. The automotive factory was simulated and different transportation methods were investigated: first we discussed the results of the EMS system, followed by the results of AGV replenishment and we concluded with the results of tugger replenishment.

The study of the implementation of alternative transport systems in an automotive company revealed that AGVs are not suitable for delivering kits to the line from a centralized warehouse. Both EMS and tuggers are a considerable alternative. The biggest opportunity for the company to decrease the investment cost is to lower the peak loads. This can be done by introducing a time-shift between the different replenishment signals.
Based on our findings the automotive company designed a business case, taking into account all investment costs, running costs and other company specific criteria.

REFERENCES

Cottyn J.; Govaert T.; and Van Landeghem H. 2008. "Alternative line delivery strategies support a forklift-free transition in a high product variety environment." In *Proceedings of the International Workshop on Harbor Maritime and Multimodal Logistics Modeling and Simulation* (Campora S. Giovanni, Italy, Sept 17-19).

De Koster, R. 1995. “Intern transportmaterieel.” In *Praktijkboek Magazijnen/Distributiecentra,* J.P. Duijker, M.B.M. de Koster and M.J. Ploos van Amstel (Eds.), Kluwer Bedrijfswetenschappen Deventer, p. 3.7-01 t/m 3.7-03

Gecker, R. 2004. “Materials handling strategies: A fork-free future.” *Inbound Logistics* 24 (12). P.52-56.

Govaert, T., Neirynck, S., Van Volsem, S., Van Landeghem, H.,2009. “Combining scripting and commercial simulation software to simulate in-plant logistics.” In *Proceedings of the 23rd European Simulation and Modeling Conference.* (Leicester, United Kingdom, October 26-29)

Le-Anh, T. 2005. "Intelligent control of vehicle-based internal transport systems," Ph.D. dissertation, ERIM, Erasmus Univ. Rotterdam, Rotterdam, The Netherlands.

Limere V. and Van Landeghem H. 2009. “Cost model for parts supply in automotive industry.” In *Proceedings of the 16th European Concurrent Engineering Conference*. (Bruges, Belgium, April 15-17), 120-125.

MHIA, 2011. Glossary, http://www.mhia.org/learning/glossary/

Mirhosseyni, S.H.L. and Webb, P. 2009. “A Hybrid Fuzzy Knowledge-Based Expert System and Genetic Algorithm for efficient selection and assignment of Material Handling Equipment” *Expert Systems with Applications* 36, 11875–11887.

Neumann, J., L. Tyler, and M. Urban. 2007. “Forklift free plants...considerations for success.” White paper, Kinetic Technologies inc. Available online (7 pages).

Tompkins, J.A., White, J.A., Bozer, Y.A. and Tanchoco, J.M.A., 2003. *Facilities planning.* John Wiley & Sons, Inc.

Welgama, P.S. and Gibson, P.R., 1995. “Hybrid knowledge based/optimization system for automated selection of materials handling system.” *Computers and Industrial Engineering* 28 (2), pp. 205–217

AUTHOR BIOGRAFY

TIM GOVAERT is scientific researcher at the department of Industrial Management at Ghent University. His research interests lie in factory logistics and in developing a method to implement continuous improvement and lean manufacturing at SMEs. Tim already completed several projects within the automotive industry.

KAREL BAUTERS is PhD student and teaching assistant at the department of Industrial Management at Ghent University. His research interests lie in lean manufacturing, production logistics and material handling.

VERONIQUE LIMERE is PhD student at the department of Industrial Management at Ghent University. Her research interests lie in facility logistics, material handling and lean manufacturing. In her PhD she is researching the organization of the flow of parts from the supplier to the use-point at the line. Recently she obtained a Doctoral Fellowship from the Intercollegiate Center for Management Science (I.C.M).

HENDRIK VAN LANDEGHEM is Professor at the department of Industrial Management at Ghent University. He is an expert in the area of logistics and their application in business processes. He advises companies in their choice and implementation of their logistic organization and production control systems. He is Fellow of the European Academy of Industrial Management (AIM) and member of the Institute of Industrial Engineering (IIE) and of the European Operations Management Association (EurOMA). He is since 2007 Fellow of the World Confederation of Productivity Science.

MATERIALS HANDLING

Simulating Value Chain and Co-operation Practice

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

KEYWORDS
Value Chain Co-operation Simulation, Supply Chain Co-operation Simulation

ABSTRACT
Co-operation within value chain management is an important and thoroughly discussed topic in strategic management and logistics. But a detailed and quantified concept as a basis for business simulation and decision support is still largely missing. This is addressed on the basis of a business survey on vertical co-operation. The results are used to create a theoretical model on value chain management co-operation. This model is supposed to show the optimal co-operation intensity in order to maximize the total value chain profit.

1. INTRODUCTION

With different emphases most definitions of supply chain management, efficient consumer response and vertical marketing systems agree on the vertical co-operation aspect within the value chain (e.g. Blanchard 2007, Peter and Donelly 2004). The existing discussion about supply chain management, efficient consumer response and vertical marketing systems has often been led independently. In business practice as well the concepts are often carried out separately due to functional organization of firms. But value chain management (VCM) is regarded as an integrating concept (Al-Mudimigh et al. 2004, Blanchard 2007, Feller et al. 2004, Gereffi et al. 2005, Kaeseler 2004, Kannegiesser 2008, Kannegiesser et al. 2009, Porter and Millar 2008, Zentes et al. 2007). Therefore a holistic view of the value chain shall be established. The added value of VCM in comparison to the other concepts is often difficult to understand, especially for practitioners (Harbert 2009, McLarty 2003, Schulz 2000). Therefore one objective of this research is to discuss which areas VCM go beyond existing concepts focusing on the co-operation aspect. This will be achieved by reviewing field research results. Furthermore the results are included into a theoretical simulation of VCM to gain further insights.

2. VALUE CHAIN CO-OPERATION THEORY

Vertical co-operation is acknowledged to be a collaboration between companies of different value adding steps in a value chain (Bauer 2003). SCM, ECR and VMS have in common their critical factor of success: They all strongly rely on vertical co-operation in the value chain. Reasons and objectives for vertical co-operation are multiple (Jespersen and Skjøtt-Larsen, 2005, Kuhn and Hellingrath 2002, Porter 1985, Rushton et al. 2006, Wang et al. 2007). Karl and Orwart (2000) list for instance "information availability, information asymmetry, opportunistic behavior, yield network benefits or the general increase of efficiency of economic transactions" (p. 382) as reasons to engage in vertical co-operation. However, since vertical co-operation is discussed in different contexts, its objectives are often not described in a concluding way (Hagenhoff 2008). In order to understand co-operation as a factor of success better, it is necessary to distinguish between different vertical co-operations that are possible in the value chain and therefore in VCM. Differences can be stated in the following areas (Baader and Montanus 2008, Bahrami 2002, Bhutta et al. 2002, Busch and Dangelmaier 2004, Diller and Metz 2006, Färber 2007, Kuhn and Hellingrath 2002, Nøkkentved 2005, Patnayakuni et al. 2006, Porter and Millar 2008, Rushton et al. 2006, Semlinger 2006, Sydow 2006, Wang et al. 2007):

- Co-operation scope: The number of co-operation partners defines whether the co-operation is considered to be a dyadic co-operation, which consists only of two partners, or a small group co-operation with three to seven partners. More than seven partners define a big group co-operation. The co-operation access is distinguished into an open and a closed co-operation. The former has no restric-tions towards partners and timing whereas the latter does not allow new co-operation partners to enter after the co-operation kick-off. Depending on the geographical coverage one finds local, regional, national, international and global co-operation. This is considered to be the co-operation expansion. It seems important to point out the difference between international and global because they are often used as synonyms. Co-operation is considered to be international when the partners collaborate across national borders. A global co-operation goes beyond that and can be regarded to be a borderless co-operation operating worldwide.
- Co-operation content: Vertical co-operation varies with respect towards the co-operation content. The co-operation content can be perceived by the end customer as the total value chain output. The four 'P' of the marketing-mix are a good approach to structure co-operation content. For example, co-operation for a product development between a manufacturer and a retailer will be regarded as product co-operation. Shared pricing decisions within the value chain will be considered as price co-operation.
- Co-operation enabler: Enablers are often used referring to the IT-technology. Depending on the IT-integration, co-operations can act and react very different-ly. Technology enables the management of "variability, risks and exceptions more effectively" (Sabri and Shaikh 2010, p. 16). In the following, enabler will only consider

technology enablers knowing that some researchers have a broader understanding of enablers (Sabri and Shaikh 2010, pp. 14).

Therefore figure 1 depicts the co-operation differentiator in a simplified way to give an overview. It implies the same impact of all three differentiators on vertical co-operation. On the basis of the value chain model by Porter the following can be assumed for differentiators (Porter 1985). The co-operation enabler and scope can be considered as the secondary co-operation processes that support the primary ones, which are summed up in the co-operation content. The co-operation content is the output of a co-operation that can be perceived by the end customers. It is assumed that within the co-operation content there are four different intensity degrees in a hierarchical set-up that are linked to the value-added steps. Therefore the information flow (medium grey arrow) concerning the content has to be on all steps as well. However the decision-making (dark grey arrow) concerning the co-operation content varies. The product is the output of co-operation amongst all value chain partners concerning e.g. time to market and quality. As the product co-operation is often only concerned with one specific product the decision-making is considered to be easier than the other three co-operation contents. The place co-operation encompasses all co-operation content that is related to logistics and distribution. It can be assumed that not all value chain partners have to decide on the place content like where to sell the product. Even though this content is named place it is considered that the product co-operation aspects are included. A promotion co-operation includes even less value chain partners in the decision-making. But it is obvious that all value chain partners need to be informed of it e.g. in order to adjust production to the increased demand. The last value-adding steps have the best knowledge about the end customer and therefore are responsible for the decision-making concerning the price content.

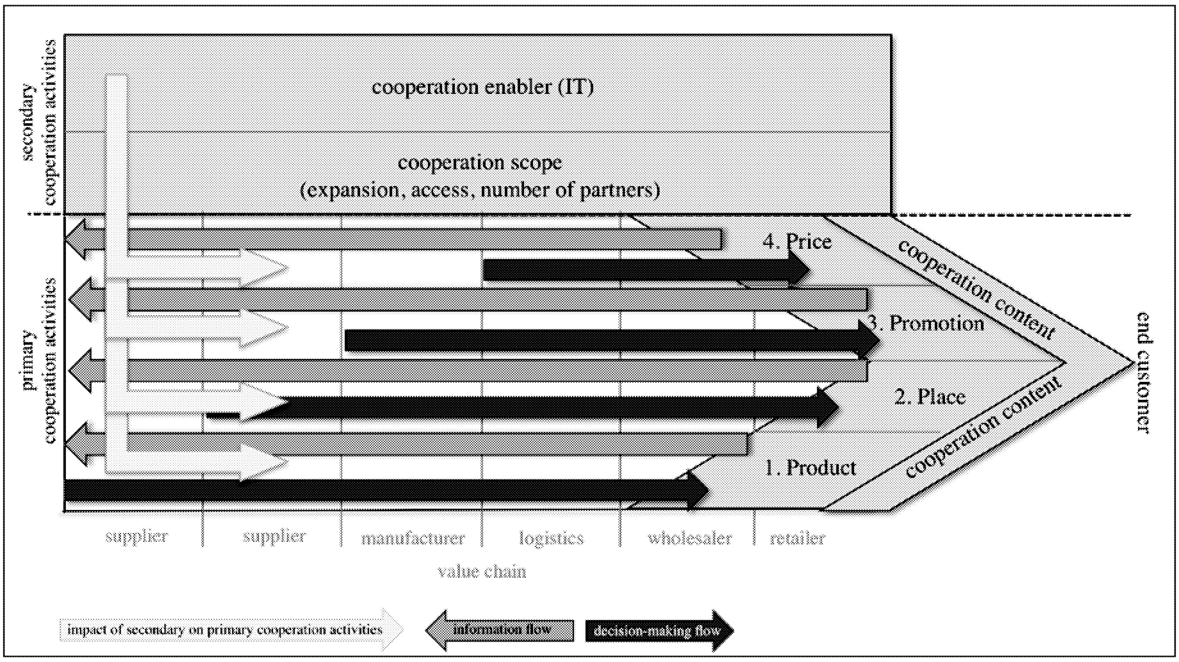


Figure 1: Co-operation in the Value Chain

3. BUSINESS PRACTICE SURVEY IN VCM

The research objective was a practical review regarding co-operation in value chains by carrying out an evaluation of six concept case on vertical co-operation. Three cases had been based on SCM, ECR and VMS literature. Furthermore three artificial concept cases had been given based on the co-operation differentiators differing in their intensity.

- Supply chain management: Seven companies in a value chain cooperate globally. The value chain can be designed via direct or indirect distribution. New companies that offer value for the unique chain can access the co-operation anytime. The co-operation objective is the optimisation of logistics through co-operative planning, forecasting and replenishment. Information is exchanged electronically.
- Low intensity concept (Low): A local co-operation involves two companies that work in an indirect channel of distribution. The co-operation is closed towards new companies. The partners exchange relevant information e.g. orders or demand changes. A special communication tool is not in place.
- Vertical marketing systems: Two companies have installed a national co-operation in an indirect channel of distribution. Their co-operation focus is a common product development and introduction, recommended retail prices, shared terms and conditions of delivery and payment and common advertisement and promotion. The co-operation is supported by IT-solutions like a common enterprise resource planning system.
- Medium intensity concept (Med.): Four companies of one value chain have founded an international co-operation. The partners are open towards integrating new partners into their co-operation. Co-operation focus is the common product development and introduction, the planning and execution of common promotion and the conjoint planning of replenishment. The information sharing of the partners takes place by using simple IT-tools like excel spreadsheets to arrange and update delivery dates.

- Efficient consumer response: Six companies of one value chain cooperate on an international level. New co-operation partners, who have a link to the value chain, can access this co-operation anytime. The co-operation contains a common product development including designing individual products for one partner. Sales promotions are commonly planned and executed. The partners agree conjointly on the assortment. Common sourcing aims at advantages like economies of scope. The use of modern logistics practices like cross docking is carried out to optimise replenishment. A common controlling makes sure the planning and decision are efficient. The partners use one common enterprise resource planning system.
- Maximum intensity concept (Max.): Co-operation involves ten companies in a value chain. New co-operation partners are admitted anytime. The partners cooperate with regard to common product development, including designing individual products for one partner that can be sold as private labels. Pricing decisions are taken together. Sales promotions are commonly planned and executed. The partners agree conjointly on the assortment. The use of modern logistics practices like cross docking is carried out to optimise re-plenishment. A common controlling makes sure the plan-ning and decision are efficient. The partners use one common enterprise resource planning system.

It is necessary to have evaluation criteria for the six concept cases. Depending on the vertical co-operation, various criteria can be found that pay respect to the uniqueness of each concept. The following eight criteria have been chosen:

- Implementation expenses: Vertical co-operation is highly dependent on its implementation expenses. High expenses for minor co-operation advantages are not efficient. But co-operation cannot succeed if the co-operation partners flinch from executing a proper implementation with the necessary expenses (Cline 2005, Eriksson and Pesämaa 2006).
- Cost cutting potential: Vertical co-operation can be used to cut costs in many ways. It can be achieved e.g. through economies of scale or the access to know-how that can increase efficiency (Daher et al. 2006; Kraege 1997).
- Risk reduction: With vertical co-operation, risks can be reduced for the co-operation partners, e.g. through sharing the risk of an investment or diversification in the competences (Kraege 1997).
- Sales growth potential: Depending on the vertical co-operation, it can imply sales growth because of access to new markets or rounding off the product range through economies of scope (Kraege 1997).
- Innovation potential: Within vertical co-operation e.g. technology transfer or common research can lead to an increased innovation potential (Hagenhoff 2008).
- Interdependence risk: In order to achieve the vertical co-operation objectives, e.g. technology sharing, the vertical co-operation is often long-term. To cut cooperative ties is not easy and can be risky with subject to the interdependence. It is important that the advantages of the vertical co-operation justify the abandonment of degrees of interdependence and independence.
- Flexibility: With increasing uncertainty due to the dynamics and the complexity of a firm's environment the need for flexibility grows. Cooperating can be a way to gain flexibility (Eriksson and Pesämaa 2006; Kraege 1997).
- Service quality: Services that improve customer satisfaction and loyalty are important differentiators and can be used to gain competitive advantages. Vertical co-operation can help to increase the offer of services and their quality.

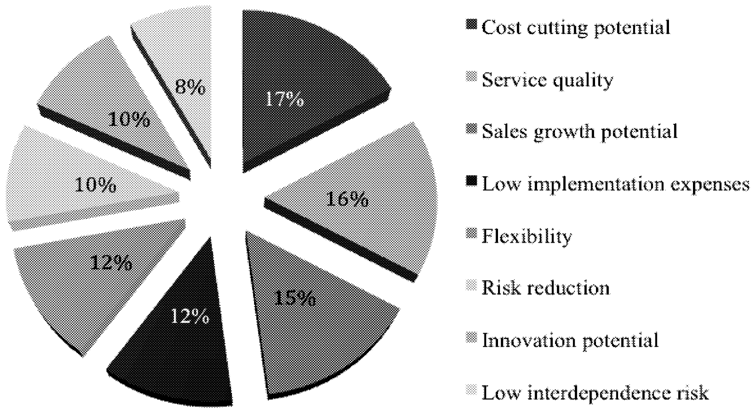


Figure 2: Business Practice Survey Results

Altogether 34 professional experts had been chosen to take part in the business practice survey. from March 2010. To avoid distorted industry results, the experts have been chosen from 16 different industries. When looking at the results, cost cutting potential (17%) seems to be the most important criterion in vertical co-operation followed by service quality (16%) and sales growth potential (15%). Low implementation expenses (12%) and flexibility (12%) as well as risk reduction (10%) and innovation potential (10%) are each regarded with the same importance. Low interdependence risk (8%) is regarded to be of lowest importance in vertical co-operation. The final score has been received through multiplication of the importance factors and the numerical value on the scale. As this result has no numerical value, the results have been put into a ranking. The medium intensity concept was the winning concept, closely followed by the maximum intensity concept. Looking at a criterion ranking the SCM concept was ranked best on the most important criteria, cost cutting potential and service quality, followed by the maximum intensity concept.

4. VALUE CHAIN CO-OPERATION SIMULATION

Literature does today not provide a model for value chain management regarding the co-operation aspect – though there is a multitude of modeling approaches regarding pricing and the connection to supply chain management in general (e.g. Pfeifer and Carraway 2000, Kuhn and Laakmann 2001, Transchel and Minner 2008, Gimpl-Heersink et al. 2008, Hansen et al. 2008, Kimms and Drechsel 2009). This quantitative model is supposed to give the best co-operation intensity that maximizes the total value chain profit, expressed as EBIT value for all companies in the chain. The model starts with the following initial situation: Co-operation in the value chain is already in place. Concerning the scope we consider the vertical coopera-tion to be a small group co-operation with international partners that is closed for others. The content of the co-operation can be related to the fields of product, place and promotion. So far only simple IT-tools are used as enablers for the co-operation. The value chain produces and sells 100,000 items of a given product with an end customer price of 13 €. Total procurement, operations, distribution and overhead costs for three companies in the value chain amount to 1,200,000 € with the basic assumption of an even distribution of these variable costs depending on the number of items sold. Therefore the EBIT in the starting situation would be 100,000 € without co-operation. In the aforementioned field research eight criteria have been used. The following three belong to the criteria that have been given the highest importance by the experts. Cost cutting potential, implementation cost and sales (earnings) growth potential will be used in the following as well to evaluate co-operation options with different intensities. The first criterion is assumed to be variable costs in nature, i.e. providing for a share of cost per item to be possibly reduced. The second criterion addresses fixed costs in nature, assuming a fixed cost investment e.g. in software or other company assets. Because service quality can overlap with sales growth potential, e.g. higher service quality might lead to higher sales volumes, it will not be explicitly used in the model but implicitly through sales growth potential.

The following *restrictions* apply to the newly designed model (in the GAMS software, 23.5.2 WIN 19281.19383 VS8 x86/MS Windows version) for the question of the optimal intensity of value chain co-operation:

- As solving algorithm a basic linear programming function was used. This inclines, that only one defined decision parameter can be introduced in order to reach an output solution in the objective function. Therefore only one of the described co-operation differentiators can be modeled at once. This model includes the differentiator co-operation enabler.
- As starting point a basic cost and earnings model (EBIT) for three value chain stages with four process areas (procurement, operations, distribution and overhead) each was used.
- The model assumes a single and static time slot without further specification, as usual in business practice it could be for example one year (business/calendar year). This implies that investment and return schemes can-not be outlined in detail, in this case we assume that invest payment (fixed cost co-operation costs) occurs in the same time period as payback.

The following *parameters* and *variables* have been used in the model in order to implement the relevant quantities:

- Parameter (scalar) p was used for the end product price of this specific value chain; p was assumed to be 13 €.
- Parameter (scalar) d was used for number of end product sold by this specific value chain; d was assumed to be 100,000 items in the calculated time period.
- Variable m was used to indicate the objective EBIT in €.
- Variable e was used in order to model the decision option of co-operation or non-co-operation regarding enablers (1 for co-operation, 0 for non-co-operation).
- Variable f was used to represent the total fixed costs in Euro incorporated by the co-operation proposal; this value is a fixed input parameter with the given value of 100,000 €.
- Variable o was used to indicate the co-operation multiplicator regarding the variable total value chain costs; in this example case it was assumed that this variable could range from 0.95 (meaning a 5 percent cost decrease induced by the co-operation proposal) to 1.00 (meaning 0 variable cost decrease due to non-co-operation).
- Variable q was used to indicate the co-operation multiplicator regarding the total value chain earnings; in this example case it was assumed that this variable could range from 1.00 (meaning no earnings increase due to non-co-operation) to 1.05 (meaning a 5 percent earnings increase due to co-operation).
- Variable v was used to represent the total variable costs of all companies in the value chain in €.

The following *functions* have been used in order to allow for the intended co-operation decision modeling:

- First the fixed co-operation cost variable [f] is calculated assuming a linear proportional equation multiplying the decision variable [e] with the as-sumed fixed investment costs of 100,000:

  (1) $f = e * 100000$

- Second the cost reduction by the co-operation investment is calculated in a first step by the cost saving percentage indicated by the variable [o] simi-larly as linear function:

  (2) $o = 0.95 + \left(0.05 * \left(1 - \frac{e}{1}\right)\right)$

- The second step uses this variable [o] in order to calculate the total variable costs [v] in the value chain in all three companies (index i) and all four cost areas (index j) as a summation:

  (3) $v = \left(\sum_{i=1}^{I} \sum_{j=1}^{J} c_{j,i} * d\right) * o$

- Third the potential earnings increase in the value chain through the coop-eration proposal regarding enablers has to be determined, represented by the variable [q]:

  (4) $q = 1.05 - \left(0.05 * \left(1 - \frac{e}{1}\right)\right)$

- Fourth and last the value chain total EBIT represented by the variable [m] is calculated by:

  (5) $m = p * d * q - v - f$

The model provides a correct solution choosing the defined higher co-operation level regarding enablers (decision variable e = 1). Obviously this depends on the defined input parameters as e.g. fixed investment costs. The increased EBIT for the whole value chain amounts to 125,000 Euro. Through expert input this research therefore established a VCM model that reflects not only the theoretical state-of-the-art but business practice as well. Though the model has to be extended in crucial parts (non-linear solving algorithm etc., see Tempelmeier et al. 2008, Wenzel et al. 2008, Färber 2007) it provides a basic structure to be used in research.

REFERENCES

Al-Mudimigh, A. S., Zairi, M., Ahmed, A. M. M. (2004). Extending the concept of supply chain: The effective management of value chains. International Journal of Production Economics, 87(3), 309-321.

Baader, A., Montanus, S. (2008). Transparency in Global Supply Chain Networks. In R. Ijioui, H. Emmerich, M. Ceyp (Eds), Strategies and Tactics in Supply Chain Event Management (pp. 3-13). Berlin, Heidelberg: Springer.

Bahrami, K. (2002). Improving Supply Chain Productivity through Horizontal Co-operation – The Case of Consumer Goods Manufacturers. In S. Seuring, M. Goldbach (Eds), Cost Management in Supply Chains (pp. 213-233). Heidelberg, New York: Physica.

Bauer, S. (2003). Perspektiven in der Organisationsgestaltung. In H.-J. Bullinger, H. J. Warnecke, E. Westkäpfer (Eds), Neue Organisationsformen in Unternehmen – Ein Handbuch für das moderne Management (pp. 93-128). 2. Ed. Berlin, Heidelberg, New York: Springer.

Bhutta, K. S., Huq, F., Maubourguet, F. (2002). Efficient Consumer Response – Increasing Efficiency through Co-operation. In S. Seuring, M. Goldbach (Eds), Cost Management in Supply Chain (pp. 218-233). Heidelberg, New York: Physica.

Blanchard, D. (2007). Supply Chain Management - Best Practices. Hoboken: Jon Wiley & Sons.

Busch, A., Dangelmaier W. (2004). Integriertes Supply Chain Management – Ein koordinationsorientierter Überblick. In A. Busch, W. Dangelmaier (Eds). Integriertes Supply Chain Management – Theorie und Praxis effektiver unternehmensübergreifender Geschäftsprozesse (pp. 3-21). 2 Ed. Wiesbaden: Gabler.

Cline, K. D. (2005). Defining the Implementation Problem – Organizational Management versus Co-operation. Journal of Public Administration Research & Theory, 10(3), 551-571.

Daher, C. E., de la Sita Silva, E. P., Fonseca, A. P. (2006). Reverse Logistics - Opportunity to Reduce Costs by Integrated Value Chain Management. Brazilian Business Review, 3(1), 57-72.

Diller, H., Metz, R. (2006). Couponing bei Procter & Gamble. In L. Müller-Hagedorn, R. Mesch (Eds), Efficient Consumer Repsonse in der Praxis – Fallstudien zu Projekten, Konzepten und Strategien (pp. 215-229). Frankfurt am Main: Deutscher Fachverlag.

Eriksson, P. E., Pesämaa, O. (2006). Modelling procurement effects on co-operation. Construction Management and Economics, 25, 893–901.

Färber, B. H. (2007). Simulationsansätze für Supply Chains. Saarbrücken: VDM.

Feller, A., Shunk, D., Callarman, T. (2006). Value Chains Versus Supply Chain. http://www.ceibs.edu/knowledge/papers/images/20060317/2847.pdf. Accessed: 27 October 2010.

Gereffi, G., Humphrey, J., Sturgeon, T. (2005). The governance of global value chains. Review of International Political Economy, 12(1), 78-104.

Gimpl-Heersink, L., Rudloff, C., Fleischmann, M., Taudes, A. (2008): Integrating Pricing and Inventory Control: Is it Worth the Effort? Business Research, 1(1), 106-124.

Hansen, A., Joseph, K., Krafft, M. (2008): Price Delegation in Sales Organizations: An Empirical Investigation. Business Research, 1(1), 94-105.

Hagenhoff, S. (2008). Innovationsmanagement für Kooperationen – Eine instrumentenorientierte Betrachtung. Habilitation. University of Göttingen.

Harbert, T. (2009). Why the Leaders Love Value Chain Management. Supply Chain Management Review, 13(8), 12-17.

Jespersen, B. D., Skjøtt-Larsen, T. (2005). Supply Chain Management – In Theory and Practice. Copenhagen: Copenhagen Business School Press.

Kaeseler, J. (2004). Value Chain Management in der Gebrauchsgüterindustrie. In H. Beckmann (Ed.), Supply Chain Management – Strategien und Entwicklungstendenzen in Spitzenunternehmen (pp. 227-260). Berlin, Heidelberg, New York: Springer.

Kannegiesser, M. (2008). Value Chain Management in the Chemical Industry – Global Value Chain Planning of Commodities. Heidelberg: Physica.

Kannegiesser, M., Günther, H. O., Beek, P. et al. (2009). Value chain management for commodities: A case study from the chemical industry. OR Spectrum, 31(1), 63-93.

Karl, H., Orwart, C. (2000). Environmental Marketing and Public Policy. In H. Folmer, H. L. Gabel (Eds), Principals of Environmental and Resource Economics – A Guide for students and decision makers (pp. 363-396). 2. Ed. Cheltenham, Northampton: Edward Elgar.

Kimms, A., Drechsel, J. (2009): Cost Sharing under Uncertainty: An Algorithmic Approach to Cooperative Interval-Valued Games. Business Research, 2(2), 206-214.

Kraege, R. (1997). Controlling strategischer Unternehmenskooperationen – Aufgaben, Instrumente und Gestaltungsempfehlungen. Mering: Hampp.

Kuhn, A., Hellingrath, H. (2002). Supply Chain Management – Optimierte Zusammenarbeit in der Wertschöpfungskette. Berlin, Heidelberg, New York: Springer.

Kuhn, A., Laakmann, F. (2001): Beherrschung großer Logistiknetze – Fragestellungen und Lösungskonzepte. Industrie Management, 17(5), 37-40.

McLarty, R. (2003). The Attitudes of British Managers in Small and Medium-Sized Enterprises to the Implementation of the Value Chain Concept. International Journal of Management, 20(4), 430-442.

Nøkkentved, C. (2005). Collaboration in e-Supply Networks. In C. An, H. Fromm (Eds), Sup-ply Chain Management on Demand – Strategies, Technologies, Applications (pp. 233-286). Berlin, Heidelberg, New York: Springer.

Patnayakuni, R., Rai, A., Seth, N. (2006). Relational Antecedents of Information Flow Integra-tion for Supply Chain Coordination. Journal of Management Information System, 23(1), 13-49.

Peter, J. P., Donelly, J. H. Jr. (2004). Marketing Management – Knowledge and Skills. 7 Ed. McGraw-Hill: New York.

Pfeifer, P. E., Carraway, R. L. (2000): Modeling Customer Relationships in Markov Chains. Journal of Interactive Marketing, 14(2), 43-55.

Porter, M. (1985). Competitive Advantage – Creating and Sustaining Superior Performance. Washington, Florence: Free Press.

Porter, M. E., Millar, V. E. (2008). How Information Gives You Competitive Advantage. In M. E. Porter (Ed.), On Competition (pp.73-96). Boston: Harvard Business School Publishing.

Rushton, A., Croucher P., Baker, P. (2006). Handbook on Logistics and Distribution Management. 3. Ed. London, Philadelphia: Kogan Page.

Sabri, E. H., Shaikh, S. N. (2010). Lean and agile value chain management – A guide to the next level of improvement. Fort Lauderdale: J. Ross Publishing.

Schulz, D. (2000). 'Value Chain' Confusion. Traffic World, 263(8), 18-19.

Semlinger, K. (2006). Effizienz und Autonomie in Zulieferungsnetzwerken – Zum strategischen Gehalt von Kooperationen. In J. Sydow (Ed.), Management von Netzwerkorganisationen – Beiträge aus der Managementforschung (pp. 29-74), 4. Ed. Wiesbaden: Gabler.

Sydow, J. (2006). Editorial – Über Netzwerke, Allianzsysteme, Verbünde, Kooperationen und Konstellationen. In J. Sydow (Ed.), Management von Netzwerkorganisationen – Beiträge aus der Managementforschung (pp. 1-7), 4. Ed. Wiesbaden: Gabler.

Tempelmeier, H. et al. (2008): Modellierung logistischer Systeme. In Arnold, D. et al.: Hand-buch Logistik (pp. 35-94), Berlin, Heidelberg: Springer.

Transchel, S., Minner, S. (2008): Coordinated Lot-sizing and Dynamic Prizing under a Supplier All-units Quantity Discount. Business Research, 1(1), 125-141.

Wang, W. Y. C., Heng, M. S. H., Chau, P. Y. K. (2007). Implementing Supply Chain Manage-ment in the New Era – A Replenishment Framework for the Supply Chain Operations Reference Model. In W. Y. C. Wang, M. S. H. Heng, P. Y. K. Chau (Eds), Supply Chain Management – Issues in the New Era of Collaboration and Competition (pp. 1-22). Hershey: Idea.

Wenzel, S., Weiß, M. Collisi-Böhmer, S., Pitsch, H., Rose, O. (2008): Qualitätskriterien für die Simulation in Produktion und Logistik. Berlin, Heidelberg: Springer.

Zentes, J., Morschett, D., Schramm-Klein, D. (2007). Strategic Retail Management – Text and International Cases. Wiesbaden: Gabler.

DEVELOPING A FLOW SIMULATION MODEL TO CALCULATE THE MINIMUM WITHDRAWAL KANBANS AND TO ANALYZE STOCK EVOLUTION WITH DIFFERENT SCHEDULING SEQUENCES.

José Villanueva Castrillón
Department of Process Engineering
Prodintec Foundation
33203, Gijón, Spain
jvc@prodintec.com

Inmaculada del Vigo García
Department of Process Engineering
Prodintec Foundation
33203, Gijón, Spain
jvc@prodintec.com

Iñigo Felgueroso Fernández-San Julián
Department of Process Engineering
Prodintec Foundation
33203, Gijón, Spain
iff@prodintec.com

Pilar Lourdes González Torre
Production Management Department
University of Oviedo
33203, Gijón, Spain
pilargt@uniovi.es

KEYWORDS
Kanban, heijunka leveling, discrete event simulation

ABSTRACT

Kanban is the method to implement a pull operation control system. The main objective of kanban is decrease stocks. Despite kanban systems being widely known as an effective method, its calculation can be difficult and the number of kanbans will be altered when the scheduling of the assembly line changes. In this paper, a simple discrete event flow simulation model is presented to calculate and minimize the number of kanbans in an assembly line. Through this model, different heijunka leveling scheduling sequences are analyzed to determine the method that allows to operate with the fewest number of kanbans, and ultimately, with less stock.

INTRODUCTION

Lean Production Methodology, based on Toyota Production System (TPS), focuses on waste elimination and resources reduction as the main objective to create a more valued product for the final consumer (Womack et al. 1996).

One of the most important principles of Lean Manufacturing is to implement a pull production system as a way to reduce inventory (Womack et al 1996). This principle is supported by one Lean technique, the kanban system, that is described by a large amount of literature and reports (Ohno 1988; Shingo 1982).

To operate with a kanban system requires an established demand. In order to be able to work with relative ease, it is necessary that the demand or production of references at issue (as far as amount and mix of product), is as uniform as possible (Cuatrecasas 2010).

On the contrary, if the rate of delivery of product to the client undergoes important variations, it would be very complicated to plan the production without incurring in delays or excessive intermediate inventories between processes.

Heijunka leveling is other Lean Manufacturing technique that allows to smooth the demand.

A real industrial line needs a flexible software to calculate the number of kanbans and the consequences of different heijunka leveling scheduling.

Simulation software based on discrete event simulations allows to represent an assembly line, to calculate the number of kanbans with an optimization application and to analyze the evolutions of stock and the evolution of the production rate.

REVIEW OF KANBAN SYSTEMS AND HEIJUNKA LEVELING

Kanban system

The kanban system is a pull planning system that regulates the internal demand by authorizing the system to manufacture products and to transport components (Womack et al. 1996). The original kanban system is an essential part of Just In Time (JIT) manufacturing or a Lean Production system.

A Kanban is a signal that triggers the production of new pieces or the movement of the finished ones (figure 1), in a pull system. A kanban is a card (in Japanese, work instruction, signal or sheet) (Schonberger 1982) which contains the necessary information to produce or move materials or finished goods, that is: the type of product to move, produce or send, as well as the amount, the moment, and the place where to move it (Liker 2003).

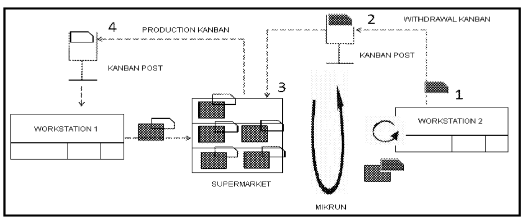


Figure 1: kanban concept.

In some occasions, not only cards, but another type of signal (empty containers, marks on the floor or on shelves, etc.) can be considered as Kanban, because all of them mark the beginning of an action (production or movement) upstream of the production processes.

In a traditional push system, the pieces move between processes immediately after they have been produced, that give rise to not regulated intermediate inventories between them, mainly due to imbalances between the cycle times of the different activities, too long changeover times, or large lot sizes, etc..

In a pull system, on the other hand, when a process is ready to consume a piece, it emits a signal (Kanban) which indicates to the previous process that it can start to produce, or to move the piece (Liker 2003).

To sum up, the most important types of kanbans are (Baudin 2004):

1. Production kanban: It indicates the type and the amount of product to make.

2. Signal kanban: When the distant orders arrives at the position indicated by the Kanban (as the order point), it will be necessary to start up the production order in the previous process.

3. Withdrawal kanban: This card must be used to retire in the previous workstation the elements necessary to make products in a later step of the process.

A pull production system controlled by kanbans has many advantages. The system is easy to understand, highly visible, reliable, interlinks the operations tightly, emphasizes operations and quality improvement and is cheap to implement (Hopp and Spearman 2000).

The object of Kanban is that a subsequence process does not remain without material to produce. The number of pieces that there are in the intermediate supermarket between a previous process and a subsequent process will be the minimum used in the subsequent process during the supply, that is to say, the time it takes the previous process to make a piece, added to the time of transporting pieces between processes. One of the equations used to calculate kanban numbers is expression 1:

$$K = Q \times LT / q \qquad (1)$$

where:

K= number of Kanban.
Q= Mean consumption of the material by time unit.
LT= Total lead time of supplying a full container, since the empty container is sent and a full container is asked for, until the full container is received passing through the period of production or provision within the previous process.
q=Container size.

Heijunka leveling

Heijunka is the leveling of production by both volume and product mix (figure 2). It does not build products according to the actual flow of customer orders, which can swing up and down wildly, but takes the total volume of orders in a period and levels them out so the same amount and mix are being made each day (Liker 2003).

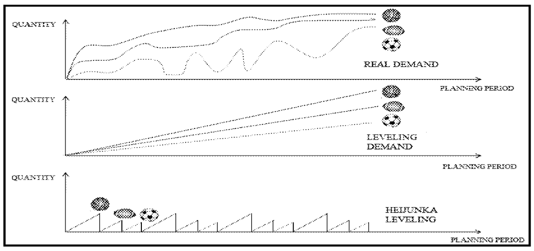


Figure 2: Heijunka leveling concept.

It is interesting to do a two phases leveling (Cuatrecasas 2010) :

1. Leveling demand (except seasonal products or services). The company's sales network must, as much as possible, avoid demand tips (avoid marketing strategies such as special promotions, that cause massive product productions at certain concrete moments).

2. Once the demand has been "established", Leveling Production (Heijunka) must be considered: Planning the mix of products and the lot sizes that simplify the "pull" flow and decrease the number of kanban in the process. The objective will be to be able to give any type of product in a short period of time (lead short time) and without having very large amounts of inventory.

To obtain the basic level schedules it is necessary to sequence small amounts of each type of product, so that repeating it n times allows for reaching the objective production after the period at issue.

For example, the monthly demands of an assembly line are:

$$Q_A = 500$$
$$Q_B = 300$$
$$Q_C = 200$$

And the basic level schedule is shown in expressions 2 to 4:

$$X = Q_A / (\text{Common maximum splitter of } Q_A, Q_B, Q_C) = 500/100 = 5\text{ A} \qquad (2)$$

$$Y = Q_B / (\text{Common maximum splitter of } Q_A, Q_B, Q_C) = 300/100 = 3\text{ B} \qquad (3)$$

$$Z = Q_C / (\text{Common maximum splitter of } Q_A, Q_B, Q_C) = 200/100 = 2\text{ C} \qquad (4)$$

In each basic level schedule, it would be necessary to produce 5 A units, 3 B units, and 2 C units.

It depends on the case (balanced activities, changeover time, etc.), but the way to produce in the considered period of time (shift, day, etc.) would be different:

1. Serial production monomodel: AAAAABBBCC, appropriate with little variety of products and significant changeover times.

2. Mixed production: ABCABCABAA (manifolds possibilities), appropriate with high variety of products and absence of times of change.

In this last case, mix production, the production scheduling will also determine the number of kanban between processes, optimizing the amount of intermediate material according to the manufacture sequence (schedule) that is programmed.

RESEARCH METHODOLOGY

The research method is based on experimental designs with simulation. The method starts with the definition of an assembly line and a product to assemble, and subsequently to model it. The research methodology used in this paper has the following steps:

1. To design a product and a manufacturing process that allows to implement kanban and heijunka leveling techniques.

2. To develop a flow simulation model using DELMIA Quest software.

3. To validate a flow simulation model.

4. To design and optimize technique so as to reduce kanban signals using OptQuest software.

5. To use the flow simulation model to compare inventory levels using a heijunka leveling scheduling and a batch scheduling.

6. To analyze the results obtained in order to know the benefits of heijunka leveling and model reliability.

7. To draw conclusions about the applicability of the model and the benefits achieved with kanban and heijunka leveling.

SIMULATION MODEL

The simulation model is a discrete event simulation model develop in Delmia Quest.

The model emulates real-world system behaviors through distributed logic, used to associate with each resource, such as routing logic, processing logic, buffering logic, push and pull production attributes, and composite processes with requirement/selection rules.

The products, processes and manufacturing routes were introduced into the model. After validating the simulation model, an optimization logic implemented in OptQuest was developed.

Product

A plastic toy dog was defined as the product for the simulation model. It is has eight parts and five configurations.

To identify the different configurations, the following notation is used:

Dog_Color1-Color2

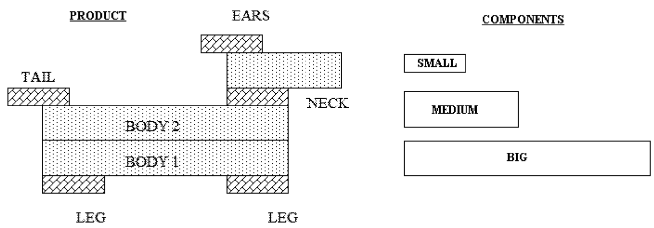


Figure 3: Product and components.

The first colour (Colour 1) is the colour of the body and of the head, and the second colour (Colour 2) is the colour of the legs, neck, tail and ears. The 5 configurations available for the model are as follows:

1. Configuration 1: Dog_Yellow-Yellow.

2. Configuration 2: Dog_White-Blue.

3. Configuration 3: Dog_Red-Black.

4. Configuration 4: Dog_Blue-Red.

5. Configuration 5: Dog_Black-White.

The following types of pieces are used in order to make the dogs (figure 3):

1. Small size: S_ Colour for legs, neck, tail and ears

2. Medium size: M_Colour for head.

3. Big size: B_Colour for the body.

In total, there are fifteen different parts, three different sizes and five different colours.

All of the parts are transported in containers that hold 120 units of the small size pieces or 60 of the other types of pieces.

Process

The process designed in this research requires three manual assembly processes, without changeover time.

Assembly station 1 assembles the legs, and a piece of the dog's body with a cycle time of 30 seconds. Assembly station 2 assembles the second piece of the dog's body, the neck and the tail with a cycle time of 30 seconds. Assembly station 3 assembles the head and the ears with a cycle time of 30 seconds. All of this data is summed up in Table 1.

Table 1: Assembly stations

Station	Parts Assembled	Time
1	Legs + Body1	30''
2	Body2 + Neck + Tail	30''
3	Head + Ears	30''

Containers with the pieces are supplied cyclically (every 2 hours) to each position. The flow is a one piece flow (lot size is 1). More details on the process are included in figure 4.

It is necessary to produce the dogs indicated in the Table 2 in an 8 hour shift.

Table 2: Production demand.

D_Yellow-Yellow	390
D_ White-Blue	210
D_ Blue-Red	180
D_ Red- Black	120
D_Black-White	60

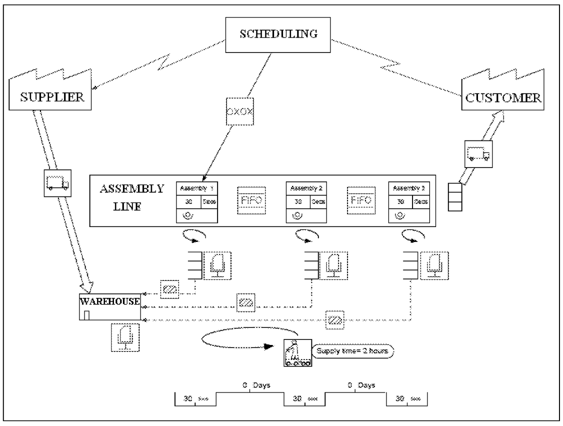


Figure 4: Value Stream Map.

Simulation Model

In the next step, a flow simulation was developed using DELMIA Quest. The model has the characteristics describes before. It has six input variables:

1. Scheduling planning
2. Kanban_ Dog_Yellow-Yellow
3. Kanban_ Dog_White-Blue
4. Kanban_ Dog_Red-Black
5. Kanban_ Dog_Blue-Red
6. Kanban_ Dog_Black-White

And it has two output analysis variables that in an assembly line will be improved. This parameters are graphed (figure 5):

1. Stock level, that it is the sum of the containers located in the three supermarkets. This parameter should be minimal, requiring less containers and less space in the plant
2. Production rate, that it is the sum of the parts produced. That should be maximum and the same in all experiments, otherwise, a stock rupture would have occurred and the number of kanban calculated was not valid.

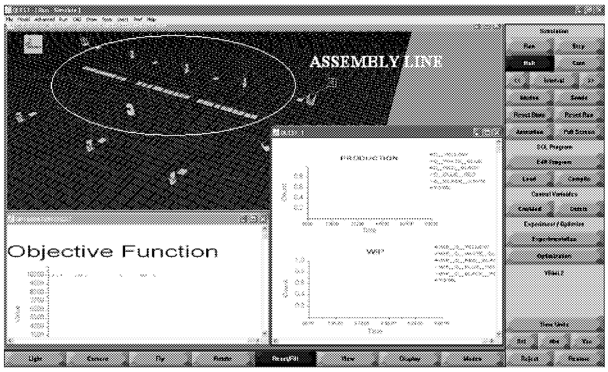


Figure 5: Simulation model.

Model Validation

To validate the model, monomodel sequences with large number of kanban were simulated, obtaining the maximum production than the line can manufacture in the simulation time.

Kanban optimization

The optimization logic is based on reducing kanban numbers and maximizing production rate. For the optimization, a simple logic was developed by using OptQuest and SCL programming language (table 3).

Like container has multiple parts, in spite of the fact there are three supermarkets with five different parts, there are only five different kanban numbers (one number for each dog). For example, for Dog_Yellow_Yellow:

$$\text{Kanban\_ Dog\_Yellow-Yellow} = \text{Kanban}_{SMALL\_YELLOW} = \text{Kanban}_{MEDIUM\_YELLOW} = \text{Kanban}_{BIG\_YELLOW}$$

Table 3: SCL program

```
Routine minimizar_stock() : Real
var
   revenue: Real
   a: integer
Begin
         ----Parts produced----
         a=part_out_count(null, get_element('Buffer_FINAL_1'))
         if(a>=957) then
                   ----Sumatory kanban number----
                   revenue =
                   get_element('CALCULO_KANBAN_1')-
                   >KANBAN_D_YELLOW+get_element('CALC
                   ULO_KANBAN_1')-
                   >KANBAN_D_WHITE_BLUE+get_element('
                   CALCULO_KANBAN_1')-
                   >KANBAN_D_RED_BLACK+get_element('C
                   ALCULO_KANBAN_1')-
                   >KANBAN_D_BLUE_RED+get_element('CA
                   LCULO_KANBAN_1')-
                   >KANBAN_D_BLACK_WHITE
         else
                   revenue = 10000
         endif
         return( revenue )
End
```

Experiments

The model developed before allows maximizing production rate with a minimum number of kanbans for different scheduling. After the model was validated, two experiments were developed:

Experiment 1

The scheduling is based on serial production monomodel (figure 6), all Dog_Yellow-Yellow parts are produced and after that, all Dog_White-Blue parts are produced and so on. The complete sequence was:

390 x Dog_Yellow-Yellow
210 x Dog_White-Blue
120 x Dog_Red-Black
180 x Dog_Blue-Red
60 x Dog_White-Black

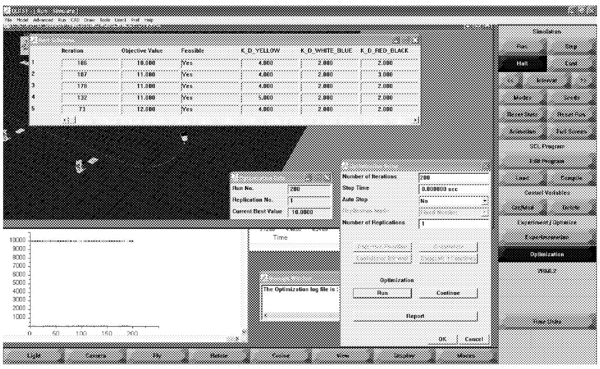


Figure 6: Experiment 1.

With this production scheduling, a minimum kanban number was calculated with 200 iterations like figure 7 shows:

Kanban_ Dog_Yellow-Yellow=4
Kanban_ Dog_White-Blue=2
Kanban_ Dog_Red-Black=2
Kanban_ Dog_Blue-Red=1
Kanban_ Dog_White-Black=1

Figure 7 presents the optimization parameters, objective function graph and the five best solutions.

Experiment 2

The scheduling in experiment 2 (figure 7) is based on heijunka leveling mixed production. The repetitive sequence was:

13 x Dog_Yellow-Yellow
7 x Dog_White-Blue
6 x Dog_Red-Black
4 x Dog_Blue-Red
2 x Dog_White-Black

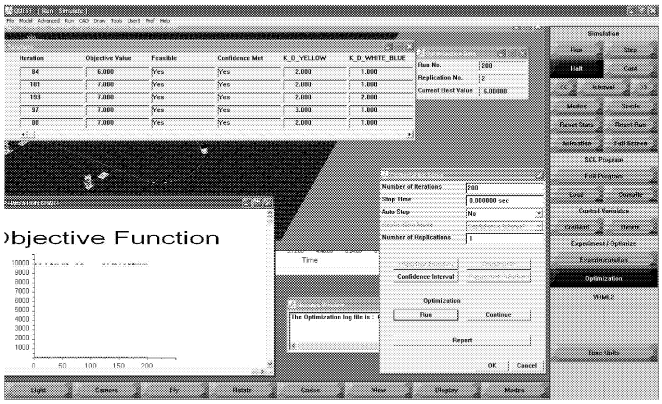


Figure 7: Experiment 2.

With this production scheduling, a minimum kanban number was calculated with 200 iteration like figure 8 shows:

Kanban_ Dog_Yellow-Yellow=2
Kanban_ Dog_White-Blue=1
Kanban_ Dog_Red-Black=1
Kanban_ Dog_Blue-Red=1
Kanban_ Dog_Black-White=1

Figure 8 presents the optimization parameters, objective function graph and the five best solutions.

These results are simulated after to ensure that this kanban number allow to obtain the maximum production in 8 hours.

ANALYSIS OF RESULTS

Both experiments carried out (1 and 2) allow comparing the production rate and stock level.

In the graphs can be seen the containers inventory and the maximum production during simulation period of the different products. In purple, the total containers and total production and in the rest of the colors, the results of each product.

Maximum stock in experiment 1 is 45 containers and total production in 8 hours is 957 parts (figure 8), while in experiment 2, the total production is the same, but maximum stock is 35 containers (figure 9), 22% minor than experiment 1.

These results prove that heijunka allows to reduce the amount of stock queued in mixed production lines.

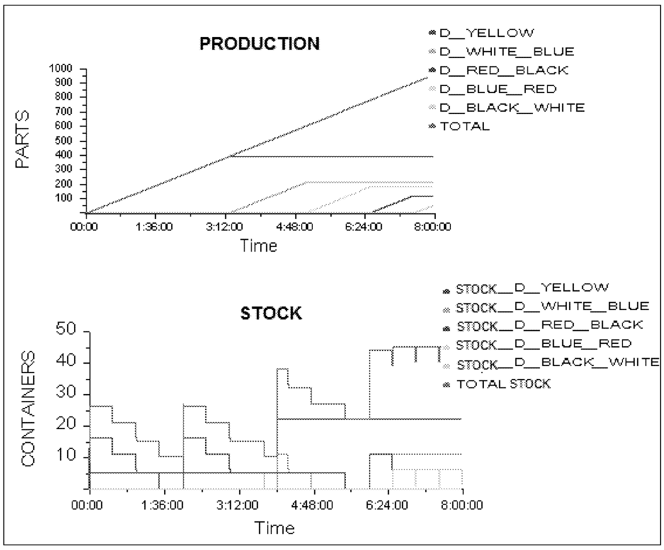


Figure 8: Monomodel scheduling production and stock.

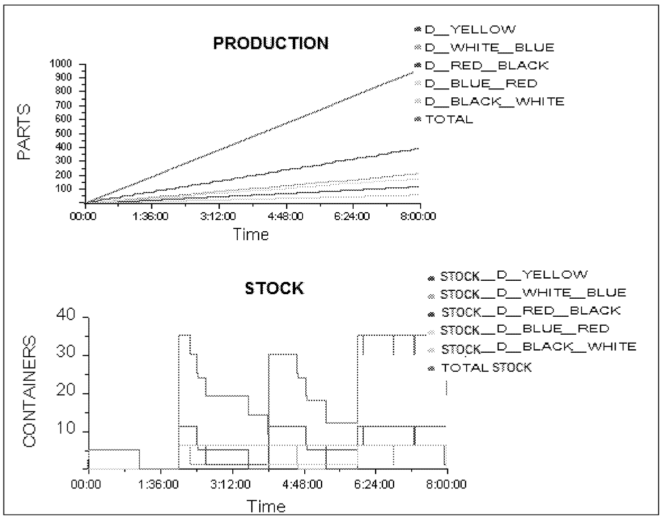


Figure 9: Mixed scheduling production and stock.

CONCLUSIONS

The main reason for carry out this research is to know, without having to implement in the real physical environment of a company, the results of different leveling scheduling implementations.

The following conclusions can be drawn the present paper:

1. The flow simulation model based on discrete event simulation allows calculating the minimum number of Kankan in an assembly line for different demands and different heijunka leveling scheduling very fast and simple.

2. Kanban systems offer a way to control inventory, eliminating waste, and reducing lead times to customers.

3. A minimum number of kanbans allows to operate with less stock, less space and to have less containers in the plant

4. Employ heijunka leveling allows to reduce the stock in the line, in the example case described in this paper, stock is reduced 22%.

This work estimates the minimum number of kanban that allows to have de maximum production and the minimum inventory in the production line. The next step on the study must include setup times to analyze the relation between production rate, stock level and total setup times.

REFERENCES

Baudin M. 2004. “Lean logistics: the nuts and bolts of delivering materials and goods”, Productivity Press, New York.

Cuatrecasas L. 2010. “Lean Management: la gestión competitive por excelencia, Profit, Barcelona.

Cuatrecasas L. 2000. “Organizacion de la producción y dirección de operaciones: sistemas actuales de gestión eficiente y competitiva, Ramón Areces, Madrid.

Hopp W., Spearman M. 2000 “Factory physics”, Irwin/McGraw-Hill, New York.

Liker, J. 2004. “The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer”, McGraw-Hill, New York.

Ohno T. 1988. “Toyota Production System: Beyond Large-scale Production”, Productivity Press, Cambridge.

Schonberger R. 1982, “Japanese manufacturing techniques”, The Free Press, New York.

Shingo S. 1982. “Study of the Toyota Production System”, Japan Management Association.

Womack P., Jones D. 1996. “Lean Thinking: Banish Waste and Create Wealth in Your Corporation”, Rawson Associates, New York.

BIOGRAPHY

JOSE VILLANUEVA was born in Asturias, Spain and studied in the University of Oviedo, where he obtained the degree of Industrial Engineering. After obtain the graduated in 2005, he start to work in Prodintec Foundation where he manage and execute Lean Manufacturing and simulation (discrete event simulations) projects for different companies. During these years he started too a research activities in the University of Oviedo, all related with Lean Manufacturing methodologies.

THE IMPACT OF PART CHARACTERISTICS ON THE DECISION OF KITTING VERSUS LINE STOCKING

Veronique Limère
Hendrik Van Landeghem
El-Houssaine Aghezzaf

Ghent University
Industrial Management
Technologiepark 903
B-9052 Ghent-Zwijnaarde, Belgium
E-mail: Veronique.Limere@ugent.be

KEYWORDS
Facility logistics, assembly line feeding, kitting, line stocking.

ABSTRACT

The increasing product variety in the automotive industry leads to an increasingly complex materials supply in automotive assembly. The high number of variant parts continues to proliferate, so bringing these parts to the line in a cost-effective manner can generate a considerable advantage over competitors. This paper discusses the generation of test instances for a computational study to investigate the link between part characteristics and the most cost-effective materials supply method, i.e. kitting or line stocking. An optimal organization of the materials supply is determined by use of a mixed-integer programming model. Results are discussed.

INTRODUCTION

Kitting as well as line stocking are common materials supply methods in industry. Line stocking parts, also commonly referred to as bulk feeding parts, are delivered to the assembly line in a bulk container as originally received from the supplier and as stored in the warehouse. Kitting parts means a subset or all of the parts that support assembly for a given end product are kitted into one container, mostly in an area called "supermarket", and delivered to the line as such. This constitutes an additional material handling step before parts are delivered to the line. Both methods have particular operational benefits and disadvantages but there is a lack of research investigating the trade-offs between both systems (Limère et al., 2011).

This paper discusses a parametric analysis that generates more insight into the impact of part characteristics on the cost-optimal decision between kitting and line stocking. It is part of a more broadly research effort by our group into this subject.

STATE OF THE ART

International scientific literature related to the choice between kitting and line stocking is scarce. Bozer and McGinnis (1992) are the first to discuss both line feeding systems simultaneously, and to emphasize the importance of gaining insight in the trade-offs between both systems. They introduce definitions in order to create a common basis for conducting research on kitting and line stocking, and they present a descriptive model to allow a quantitative comparison between both systems. Where possible we adopted their definitions and framework.

Caputo and Pelagagge (2008) and Carlsson and Hensvold (2008) build upon the model provided by Bozer and McGinnis. Both articles examine hybrid policies. However rules for the selection of the delivery strategy for individual parts are decided a priori and are not based on any model. Caputo and Pelagagge (2008) start from an ABC based classification to assign parts to line-stocking, kitting or Kanban-based supply. Carlsson and Hensvold (2008) use an Analytic Hierarchy Process technique to optimize the materials supply for multiple criteria objectives, hence a rather subjective approach.

An integrated approach to component management optimization, considering the centralization/decentralization decision of components and the right feeding policies in one comprehensive framework, is proposed by Battini et al. (2009). This research applies to multi-model assembly lines instead of high variation mixed-model assembly lines. Thus, they do not include the impact of part characteristics or study the possibility of hybrid feeding policies.

Hua and Johnson (2010) confirm that since the interest of Bozer and McGinnis in the decision choice of kitting versus line stocking in the nineties, up till now few studies have been done to get more insight in the factors that determine which of the two alternative line feeding strategies is most appropriate. Moreover they recognize the still remaining controversy about the advantages and disadvantages of both methods. This confirms both the importance and timeliness of our research.

MODEL

Limère et al. (2011) present a mixed integer programming model that assigns parts to two materials supply system alternatives to minimize total costs. The materials supply systems considered are bulk feeding (or line stocking) at the one hand, and kitting on the other hand. For the kitting option, one stationary kit per station is assumed. A *stationary kit* is delivered to a use-point and remains there until it is depleted. The model minimizes the sum of the labor costs for operator picking at the line, internal transport, the kit assembly operation and replenishment of the supermarket. Moreover, it takes into account the space constraint at the border of the line. Average part and production mix characteristics are parameters in the model. The model is a static and deterministic optimization problem with binary decision variables that assign if a part is kitted or fed in bulk to the line.

Inputs to the model are twofold. Firstly, general problem features need to be provided. These are features with regard to the layout of the plant, workstation and supermarket design, operator productivity, material equipment capacity and packaging dimensions. Secondly, a dataset with parts that need to be supplied to the line and their respective characteristics has to be provided.

The output from the model is a minimum cost assignment of each of the parts to one of both materials feeding methods and the corresponding cost structure.

The mathematical model is implemented using the modeling language AMPL 11.2, and solved with CPLEX 11.2 on an Intel Centrino Duo 1.67 GHz with 2 GB RAM memory. Run times are in the order of seconds. For details the reader is referred to the above referenced publication.

We will now discuss in more detail the input that is needed for the model and the structure of the input parameters. First of all general problem features need to be provided. These are related to:

- Plant layout
  - ➢ Distances from the warehouse to each of the work stations (D_s^p): these are needed to calculate the costs for fork lift transport of pallets.
  - ➢ Distances of the milk run tours by tuggers for the supply of boxes (D^b) and kits to the line (D^k).
  - ➢ The cost to replenish respectively one box (R^b) and one pallet (R^p) from the warehouse to the supermarket.
- Workstation layout
  - ➢ The average walking distance for an operator to pick from a kit (Δ^k).
  - ➢ The expected walking distance for an operator to pick from bulk containers (Δ_{is}^{bulk}). This distance is larger than the distance to a kit.
  - ➢ The length of available storage area along a work station (L_s). We assume the border of line is organized in one facing along the direction of the moving assembly line (x-direction). Boxes are also stacked on racks vertically (z-direction).
- Supermarket layout
  - ➢ The expected walking distance for an operator to pick from bulk containers (Δ_{is}^k).
- Operator productivity
  - ➢ The walking velocity of an operator (*OV*) (at the line or in the kitting area).
  - ➢ The average time to search for the required part from bulk stock (τ^{bulk}) (at the line or in the kitting area).
  - ➢ The hourly labor cost of an operator (*OC*) (at the line or in the kitting area).
  - ➢ The kit batch size (B^k): this is the number of kits that is assembled in a batch. If the same part number is required several times in the batch, the distance to the container of that part only has to be walked once by the operator.
- Material equipment capacity
  - ➢ Vehicle velocities for forklift trucks (V^p) and tugger trains (V^b en V^k).
  - ➢ The maximum number of units a tugger train can transport in one milk run (boxes (A^b) or kits (A^k)).
  - ➢ The expected capacity utilization of the tugger trains, given the variety in demands (ρ^b en ρ^k).
- Packaging dimensions
  - ➢ The length of respectively a box (L^b), a pallet (L^p), and a kit container (L^k) along the line.
  - ➢ The height (H^b) at which boxes are stacked at the border of the line (number of boxes).
  - ➢ The maximum weight of a single kit (w^k).

Secondly, a dataset with parts that need to be supplied to the line and their respective characteristics is required. This dataset is input in the model by reading it from an Excel file. Every part is defined by a record with the following fields:

- ➢ Part number: a unique key for each part (index *i*).
- ➢ Station: the work station to which the part needs to be supplied (index *s*).
- ➢ V_i: the part family to which part *i* belongs. A part family is a group of variant parts. Never more than one of the variant parts of the same family is assembled on an end product. Also note that parts of the same part family are always assigned to the same materials supply policy, i.e. they are either all bulk fed or either all kitted, because of practical implementation considerations.
- ➢ $|V_i|$: The cardinality of the part family to which part *i* belongs is also provided as a separate field.
- ➢ f_{is}: percentage of end products for which part i is assembled at station s (frequency of part *i* at station *s*).
- ➢ m_{is}: number of units of part *i* that will be assembled on an end product at station *s* (if that end product needs part *i*); this depends on the bill of material.
- ➢ q_{is}: yearly usage of part *i* at station *s*; this can be determined from f_{is}, m_{is} and the production *d* of the end product over the time horizon
  $q_{is} = d \times f_{is} \times m_{is}$
- ➢ w_i: weight of part *i*
- ➢ $pack_i$: supplier packaging of part *i* {Box, Pallet}
- ➢ n_i: unit-load of part *i* in its supplier packaging (number of parts)
- ➢ v_i: volume measure for part *i* (number of units of part *i* that a kit can maximally hold)
- ➢ a_i: maximum number of units of a part in one pick due to physical characteristics (weight, volume) of part *i*

CREATION OF TEST INSTANCES

To be able to extensively test the trade-offs between kitting and line stocking we want to apply the model to numerous datasets. Since no established test-bed is available for a comprehensive computational study, firstly representative datasets needed to be developed. To obtain reliable datasets we started from data from a truck manufacturing company. On the one hand, general problem features are adopted from this case study. On the other hand, part characteristics are analyzed and described by use of discrete probability distributions. Through Monte Carlo simulation (repeated random sampling) from each of the distributions, new part datasets are created. In order to generate realistic datasets, the structure of the input data needs to be understood. In the first place, each part is characterized by a number of parameters. Certain relationships between the parameters need to be satisfied. For example, there is a link between the number of units of a part that fit into a certain packaging and the weight and the volume of the part. The weight and the volume of a part will also have an influence in the first place on the choice of packaging. Aside from individual part parameters, every part belongs to a part family. The number of variant parts a customer can choose from is referred to as the cardinality of the part family. For common parts, where the customer has no choice, the part family cardinality is one. Finally, parts are to be supplied to a certain work station at the line. Fig. 1 shows this structure of the input data. It represents all the parameters, and also shows connections between parameters that are related in some way.

Visual Basic for Applications (VBA) is used to code the algorithm for creating new datasets. Monte Carlo sampling from the characterizing distributions is supplemented with conditional statements to make sure the datasets are realistic as explained above. Figure 2 describes the algorithm. Text in grey font is used for comments.

Table 1 gives an overview of the general problem features. Table 2 summarizes the ranges of the characterizing distributions from the case study.

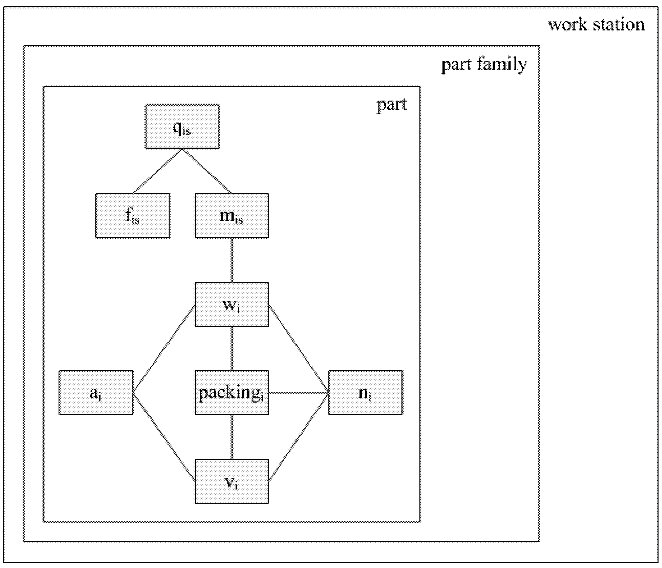


Fig. 1: Structure of the input data (Limère et al., 2011)

Algorithm – Dataset creation

Create part numbers
for each part **do**
  Assign station number to part
  Based on a distribution of the number of parts that go to a work station.
  Assign part family cardinality and part family number to part
  Based on a distribution of the number of parts that belong to a part family.
end for
for each family **do**
  Check if every part of the family is assigned to the same station
  Parts of the same family should not go to different stations
  if no **then**
    Adjust the station assignment of the latter parts in the sequence to be identical to the station assignment of the first parts
    This will not considerably influence the number of parts assigned to a station because parts are reassigned in two directions (i.e. stations get more parts if a family is broken over the station and its successor and stations get less parts if a family is broken over the station and its predecessor)
  end if
end for
for each part assigned to a station **do**
  Assign a frequency f_{is}
  Based on a distribution of the part frequencies
end for
for each family **do**
  Common parts have a frequency above 50% and the sum of the frequencies of all parts in a family is maximum 100%
  if only one part in the family (this is a common part) **do**
    while ($f_{is} < 50\%$) **do**
      Assign a new frequency f_{is} to the part
    end while
  else
    while ($\sum f_{is} > 100\%$) **and** (number of iterations < 100) **do**
      Assign a new frequency f_{is} to all parts of the family
    end while
    if ($\sum f_{is} > 100\%$) **do**
      Assign a new $f_{is} = 1/|V_i|$ to all parts of the family
    end if
  end if
end for
for each part **do**
  Assign a weight w_i
  Based on a distribution of the part weights
  Assign a supplier packaging $pack_i$
  Part weight and packaging are linked in reality. Lighter parts are more often supplied in boxes, and heavier parts in pallets. Therefore, different distributions of part packagings are used for different weight classifications.
end for
for each part assigned to a station **do**
  Assign m_{is}
  Part weight and m_{is} are linked in reality. Lighter parts are more often parts of which one needs to assemble more than one unit. Therefore, different distributions of m_{is} are used for different weight classifications.
end for
for each part **do**
  Assign a unit-load n_i
  Unit-load is linked to weight and packaging in reality. Therefore, different distributions of n_i are used for different weight/pack classifications.
  Calculate a value for v_i based on n_i and $pack_i$
  Calculate a value for a_i based on w_i and v_i
end for

Fig. 2: Algorithm for dataset creation

Table 1: General problem parameters

Parameter	Value
OV (m/h)	3600
OC (€/h)	30
Δ_{is}^{bulk} (m)	
$q_{is} > 2500$	2
$2500 \geq q_{is} > 800$	2.5
$q_{is} \leq 800$	3
τ^{bulk} (h)	0.0003
Δ^k (m)	1.5
Δ_{is}^k (m)	
$q_{is} > 2500$	2
$2500 \geq q_{is} > 800$	2.5
$q_{is} \leq 800$	3
τ^k (h)	0.0003
B^k (number of kits)	5
D_s^p (m)	[54-302]
V^p (m/h)	2880
D^b (m)	1640
V^b (m/h)	2412
A^b (number of boxes)	60
ρ^b	0.5
D^k (m)	1640
V^k (m/h)	2412
A^k (number of kits)	70
ρ^k	0.8
R^b (€)	0.2
R^p (€)	1.2
w^k (kg)	50
L^b (m)	0.8
H^b (number of boxes)	4
L^p (m)	1
L^k (m)	0.8
L_s (m)	8

It is obvious that these input distributions can be changed to analyze various kinds of problems. In this study we worked uniquely with the input parameters from Table 1 to generate results that are representative for the automotive manufacturing industry.

Table 2: Data ranges

Parameter	Value
Parts per station	[1 - 74]
Parts per family	[1 - 16]
f_{is}	[0.05 - 1]
w_i (kg)	[0.005 - 50]
%Box - %Pallet	
$w_i \geq 10kg$	[0 - 100]
$5kg \leq w_i < 10kg$	[2 - 98]
$2kg \leq w_i < 5kg$	[24 - 76]
$1kg \leq w_i < 2kg$	[48 - 52]
$0.5kg \leq w_i < 1kg$	[75 - 25]
$w_i < 0.5kg$	[94 - 6]
m_{is}	
$w_i > 5kg$	[1 - 3]
$0.5kg < w_i \leq 5kg$	[1 - 18]
$w_i \leq 0.5kg$	[1 - 34]
n_i	
Box	
$w_i > 1kg$	[1-36]
$0.5kg < w_i \leq 1kg$	[1-72]
$0.125kg < w_i \leq 0.5kg$	[1-300]
$0.015kg < w_i \leq 0.125kg$	[10-600]
$w_i \leq 0.015kg$	[1-6050]

Five test instances are created using the proposed algorithm. Some information about the datasets is summarized in Table 3.

Table 3: Input datasets

	# of parts	# of part families	# of stations
Input 1	1815	692	104
Input 2	1826	756	95
Input 3	1755	707	92
Input 4	1768	703	93
Input 5	1741	735	96

COMPUTATIONAL RESULTS

Table 4 gives a general overview of the results for the five datasets, given the general problem characteristics from the case study.

Table 4: Results

	without space constraint				**with space constraint**			
	Total cost	**% kitting**	**% kitting in stations with kitting**	**# stations with kitting**	**Total cost**	**% kitting**	**% kitting in stations with kitting**	**# stations with kitting**
Input 1	326 749	1.7%	68.9%	3 (of 104)	377 504	42.8%	61.1%	44 (of 104)
Input 2	425 269	0.0%	/	0	477 786	43.5%	64.1%	40 (of 95)
Input 3	329 421	2.5%	49.4%	3 (of 92)	367 545	36.3%	57.1%	37 (of 92)
Input 4	342 340	0.9%	94.1%	1 (of 93)	391 379	40.6%	58.9%	42 (of 93)
Input 5	391 876	3.4%	39%	6 (of 96)	444 344	39.1%	59.5%	35 (of 96)

What we can see is that initial kitting, when there is no space constraint at the line, is limited. On average around 2% should be kitted while for input dataset 2 there should not even be kitting at all. However, if we look at the proposed amount of kitting in the stations where kitting should be done, we notice that this amounts to 40% and higher, even up to 94%. This shows that the batching effects in transport are important. Once a kit is composed, additional parts can be added to that kit as 'free-riders'. This eliminates transport for the bulk containers of these parts, while no extra cost needs to be paid for transporting the kit.

When the space constraint is added to the model more kitting is proposed in the optimal solutions. On average 40% of the parts should be kitted. When we only look at the stations where kitting should be done, this increases to 60%. This increased amount of kitting is induced by the space constraint and not by the incentive to reduce costs.

To answer our research question how part characteristics influence the decision to kit or feed in bulk to the line, we analyzed the parts that are supplied to stations where at least one kit is supplied. This is done in order to mitigate the effects that station characteristics have on the solution.

First of all we investigated if the cardinality of a part family had an impact on the decision. Fig. 3 clearly shows that there is a considerable effect. Parts that are part of families with seven or more members will be kitted more frequently than parts from part families with two to six members, and the latter are still more frequently kitted than common parts, i.e. parts that are unique in their family. This important influence can be explained because variant parts belonging to part families with a high cardinality need only one space in a kit (because never more than one of the variant parts of the same family is assembled on one and the same end product), whereas much space is freed up at the line by removing multiple parts containers. As we already had noticed when looking at the general results, this once more confirms that space at the line plays an important role.

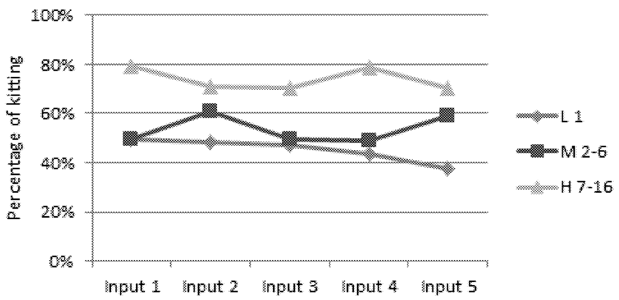


Fig. 3: The impact of family cardinality
(L=low; M=medium; H= high)

When we look at the frequencies of parts we see that parts with low frequencies have a higher tendency to be kitted (Fig. 4). This effect is clearly linked to the former effect because parts belonging to larger part families will automatically have lower frequencies. The similarity of the curves leads us to suspect that there is no effect of frequencies on the decision independent of the impact of the family cardinality.

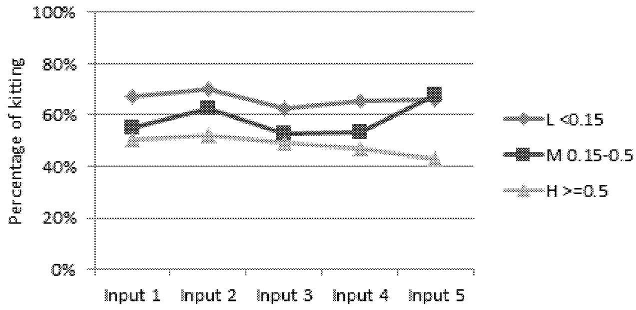


Fig. 4: The impact of part frequency
(L=low; M=medium; H= high)

Fig. 5 shows the impact of part frequencies solely for common parts (family cardinality = 1). As a matter of fact, the effect seen in Fig. 4 completely disappeared.

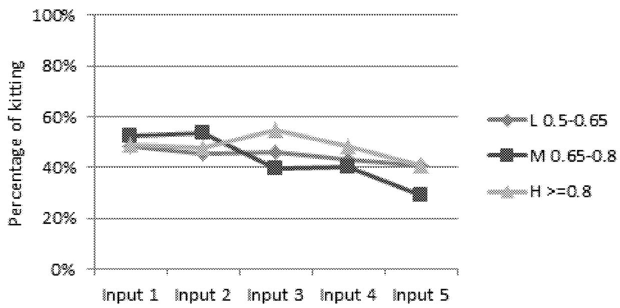


Fig. 5: The impact of part frequency
(L=low; M=medium; H= high)

Fig. 6 illustrates the impact of the original supplier packaging of a part on the likelihood of kitting. Parts supplied on pallets will be rather kitted than parts supplied in boxes. This is again caused by the space constraint at the line. Kitting is more expensive than bulk feeding and removing pallets from the border of the line will free up more space than removing boxes.

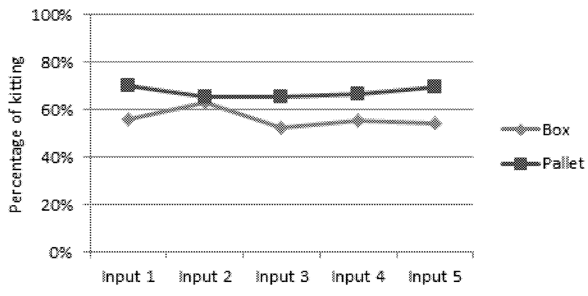


Fig. 6: The impact of supplier packaging

Physical part characteristics as weight and volume do not seem to have a major effect on the preferred line feeding method (Fig. 7 and 8).

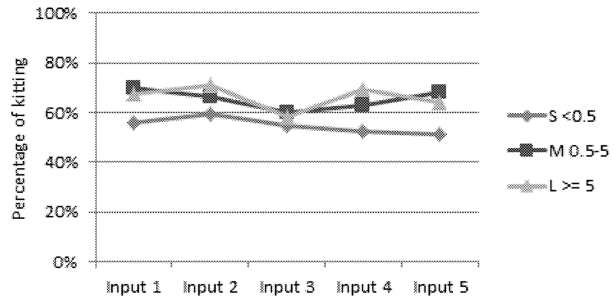


Fig. 7: The impact of weight (kg)
(S=small; M=medium; L= large)

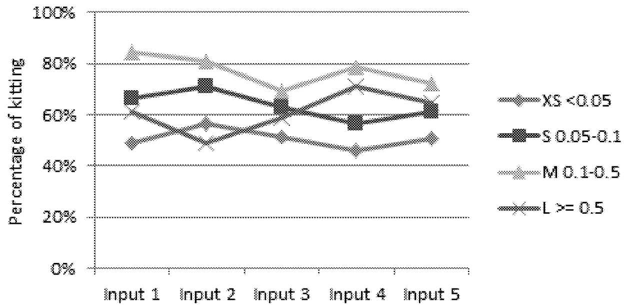


Fig. 8: The impact of volume (% of volume in kit)
(S=small; M=medium; L= large)

CONCLUSIONS AND FURTHER RESEARCH

Based on the mathematical model proposed by Limère et al. (2011), and using a newly created test bed, we were able to obtain a better understanding of the impact that part characteristics have on the decision to kit them or not.

An important conclusion that can be drawn is that for mixed-model assembly lines hybrid policies should be preferred over one method for all parts. Part characteristics do matter in the decision to kit or not to kit, and the high variety of parts in the automotive industry leads to a need for part specific decisions. Furthermore we examined which characteristics of parts are most influential in the decision. Given the current settings for the plant design, the space requirement for a part at the line seems to be a critical parameter for the decision. Physical part characteristics as weight and volume are not as determining as previously expected.

For our analysis we started from the plant design observed at a truck manufacturing company. Further research efforts will be directed towards more sensitivity testing of the model under various circumstances. We need to study if there are plant designs possible which lead to a higher preference for kitting. Kitting can also offer advantages related to ergonomic conditions for the line operators, and it can lead to reduced errors, hence improved quality. These advantages are not in the model because they are not easily quantifiable into costs, but they should be taken into account when developing a materials supply system. Moreover it will be interesting to extend the model for additional materials feeding methods as downsizing.

ACKNOWLEDGEMENT

This research was carried out with financial support from the Intercollegiate Center for Management Science (I.C.M).

REFERENCES

Battini, D., Faccio, M., Persona, A. and Sgarbossa, F., 2009. "Design of the optimal feeding policy in an assembly system." *International Journal of Production Economics*, 121, 233-254.

Bozer, Y.A. and L.F. McGinnis. 1992. "Kitting versus line stocking." *International Journal of Production Economics*, 28, 1-19.

Caputo, A.C. and Pelagagge, P.M., 2008. "Analysis and optimization of assembly line feeding policies." *In: Proceedings of the 2008 Conference on The Modern Information Technology in the Innovation Processes of the Industrial Enterprises* (Prague, Nov. 12-14).

Carlsson, O. and Hensvold, B., 2008. "Kitting in a high variation assembly line - A case study at Caterpillar BCP-E." Master thesis. Lulea University of Technology.

Hua, S.Y. and Johnson, D.J., 2010. "Research issues on factors influencing the choice of kitting versus line stocking." *International Journal of Production Research*, 48 (3), 779-800. First published on: 17 November 2008 (iFirst).

Limère, V., Van Landeghem, H., Goetschalckx, M., Aghezzaf, E.-H. and McGinnis, L. "Optimizing part feeding in the automotive assembly industry: deciding between kitting and line stocking." *International Journal of Production Research*, Under review.

BIOGRAPHY

VERONIQUE LIMERE is PhD student at the department of Industrial Management at Ghent University. Her research interests lie in facility logistics, material handling and lean manufacturing. In her PhD she is researching the organization of the flow of parts from the supplier to the use-point at the line. Recently she obtained a Doctoral Fellowship from the Intercollegiate Center for Management Science (I.C.M).

HENDRIK VAN LANDEGHEM is full professor at the department of Industrial Management at Ghent University. He is an expert in the area of logistics and their application in business processes. He advises companies in their choice and implementation of their logistic organization and production control systems. He is Fellow of the European Academy of Industrial Management (AIM) and member of the Institute of Industrial Engineering (IIE) and of the European Operations Management Association (EUROMA). He is since 2006 Fellow of the World Confederation of Productivity Science.

EL-HOUSSAINE AGHEZZAF is a professor at the department of Industrial Management at Ghent University. His interests and research fields are in logistics and supply chain management, production planning, project planning and scheduling, network design and management, and quality engineering. Prof. Aghezzaf is author of international scientific publications in Mathematical Programming, EJOR, JORS, IJPR, IJPE, …

From bulk feeding to full kitting: a practical case in the automotive industry

Karel Bauters
Tim Govaert
Veronique Limère
Hendrik Van Landeghem
University of Ghent
Department of Industrial Management
Technologiepark 903
B-9052 Zwijnaarde, Belgium
E-mail: karel.bauters@ugent.be

KEYWORDS
Parts picking, Order picking system (OPS) design, supermarket, Kitting

ABSTRACT

This paper describes a practical case study in the automotive industry, where an order picking system in a manufacturing environment was designed. We start by explaining the excel-based simulation model, the needed input parameters and a parts classification method. This is followed by a summary of the results from the case study. Based on these results, we present some general guidelines for the selection of the most appropriate order picking system, given the different problems you can face in OPS design.

GENERAL INTRODUCTION

The design of the order-picking, as presented in this paper, is part of a much more extensive research project in the automotive industry. Parts that are delivered to the factory by the supplier can be packed in two different ways: blue boxes or pallets. These are stored respectively in the small box warehouse (SBW) and the high bay warehouse (HBW). In the situation as it is today, parts are brought to the assembly line in their respective packaging directly from the warehouse. We call this material supply strategy bulk feeding. Sometimes parts are first assembled into more complex pre-assemblies which then are brought to the line and mounted on the vehicles. Figure 1 represents the current situation. In the research project, a new strategy for line delivery, full kitting, was investigated. By full kitting we mean that for each vehicle a number of kits is composed, containing every single part that needs to be mounted on the vehicle. This increases quality because less mistakes are made. Furthermore, the walking distance of the operators will decrease because all the parts can be brought as close as possible to the place of assembly. As a consequence of the kitting process, which takes place in another area, less material is stored at the line which enlarges the visibility in the plant.

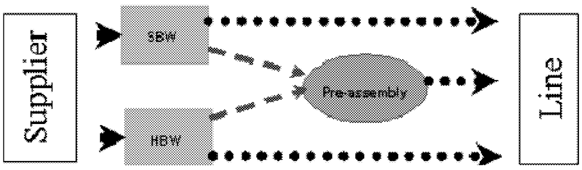


Figure 1: Material flow (bulk feeding)

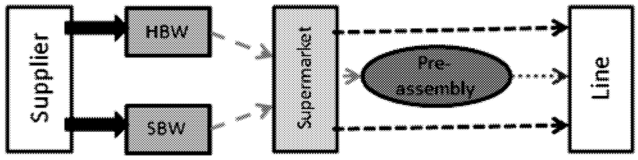


Figure 2: Material flow (full kitting)

This new line delivery strategy also entails a number of disadvantages and difficulties. Full kitting leads to smaller lot-sizes, the difficulties of point-of-use delivery, order and product customization and the need for cycle time reductions in the warehouse (De Koster et al., 2007). The current warehouse system is not able to meet these demands, therefore a new order picking system was designed. The place where all kitting actions are performed, is called the supermarket. This supermarket contains a small number of parts and is replenished by the existing warehouse system, as shown in Figure 2. This paper describes the design of this supermarket, using the design method and model presented by K. Peerlinck (Peerlinck et al., 2010).

LITERATURE REVIEW

Research on the topic of order picking is important. Whether a manual or an automated system is installed, in both cases order picking involves a lot of money. In manual systems it is the most labour-intensive operation in warehouses, and in warehouses with automated systems it is a very capital-intensive operation (Goetschalckx and Ashayeri, 1989, Drury, 1988 and Tompkins et al., 2003).
Fortunately literature on order picking systems shows that quite some research is already done on different order picking methods and operational control of the systems (de Koster et al., 2007). Research is done about various topics as layout design, storage assignment methods, routing methods, order batching, and zoning.
However, two remarks need to be taken into account. First of all, most of the existing literature focuses in the first

place on order picking systems for distribution warehouses (Baker et al., 2009; de Koster et al., 2007; Goetschalckx et al., 2001; Gu et al., 2007; Van Den Berg et al., 1999). In this research we will focus on order picking in a manufacturing warehouse. In a way these picking processes can be compared but we will also show how a picking process in a manufacturing environment differs from one in a distribution environment.

In a distribution warehouse environment orders are picked for external customers. In a manufacturing environment the picking orders result from the need to have the right materials, at the right time and in the exact amount, at the production line. However, different material supply systems can be chosen to feed parts to the line. Three main internal methods of line feeding are found in practice: bulk feeding, bulk feeding after downsizing and kitting/sequencing (Limère and Van Landeghem, 2008). The choice of line feeding will have a certain influence on the kind of order picking that needs to be done. Therefore there is a need to integrate both problems and tackle them as one.

This research has been carried out in the automotive industry. The product variety on mixed-model assembly lines in automotive assembly is an additional important factor that will have a major influence on operational results (Fisher and Ittner, 1999).

Secondly, most of the existing research focuses on selected issues when an order picking system is in operation. In our case study, the objective was not to improve the operations of an existing order picking system but to choose the appropriate order picking system in the first place. To the best of the authors' knowledge, no research exists that guides this choice in a manufacturing environment. This research is a first step into that direction.

USAGE OF THE DESIGN MODEL

In the case study, we used the excel-based simulation model described by K. Peerlinck (Peerlinck et al., 2010). The input for the model is a list of all parts used throughout the automotive factory. For each part, a number of different parameters needs to be defined:

- Part-id: a unique number for every SKU (stock keeping unit) stored in the supermarket.
- Frequency: the average number of times the part is mounted on a vehicle.
- Weight: weight per part.
- Volume: volume per part.
- Supplier embalage: the packaging in which the parts are delivered by the supplier.
- Minimal embalage: the smallest container in which a certain part can be stored.

In total, 2178 different parts are used at the assembly line and need to be stored in the supermarket. The annual production of the factory is 23173 vehicles. Given a two-shift working scheme, this leads to a takt-time of 6.28 minutes, which means that every 6.28 minutes a vehicle needs to be finished to be able to follow the market demand. Therefore, the supermarket systems also needs to pick all kits for one vehicle every 6.28 minutes. To do this, the order picking system needs to process 5820 order lines per hour. The number of kits per cart is based on the weight of the parts in the kit. An order has a maximum weight of 12 kg, resulting in an average of 7 kits per vehicle.

Classification

The model has a built-in classification structure. This function makes it possible to divide the different parts in product groups, based on the weight, volume, frequency and emballage. In the case study, product groups are only based on frequency and emballage. Two types of emballages are considered: pallets and blue boxes, which are specific to the company where the case study was carried out. These two types are subdivided in a number of subtypes with their own dimensions. The classification based on embalage types is required because different packaging types need to be handled differently: f.e. pallets are too big to be placed in vertical lift modules (VLM).

Based on the frequency, parts are divided into three subgroups: common, variant and rare parts. By changing the limits of this categorization, the model can select the range of parts that is investigated.

Order specifications

Based on the classification parameters, the parts requirement list and the takt-time of the assembly line, the following order characteristics per customer are calculated:

- Average number of orderlines/order
- Number of picks/interval
- Volume per order
- Weight per order

In this case, the customer is a work station or a group of workstations as they exist today at the assembly line. Every order is directly linked with a specific vehicle, based on the bill of material (BOM) for that vehicle. The distribution of the parts over the different kits is based on the weight and volume of the parts.

Order-picking systems

Order-picking systems are divided into two major groups: picker-to-parts and parts-to-picker systems (De Koster et al., 2007). In this case study, we investigate two different parts-to-picker systems: Vertical Lift Modules (VLM) and Horizontal Carousels (HC). These parts-to-picker systems are only taken into consideration for parts that are stored in blue boxes. For pallets we investigated the possibility of using pallet racks (picker-to-parts). For blue boxes, the possibility of using gravity racks (picker-to-parts) is also taken into account.

Table 1: Summary order picking systems

	Emballage	
	blue boxes	**pallets**
picker-to-parts	gravity racks	Pallet racks
parts-to-picker	Horizontal Carousel	AS/RS
	Vertical Lift Module	

As an extension, we also investigated the possibility of picking the parts directly from the pallets that are stored in the high-bay warehouse (HBW) using an automated storage and retrieval system (AS/RS). The order picking systems that were investigated, are summarized in Table 1.

Model Output

Once the input-file is imported, the classification is made and the order specification is done, the simulation model can start calculating the output variables for all order-picking systems that are suitable for the selected parts. The most important output variables are:

- Throughput: a throughput of 5820 order lines/hour is needed for the plant
- Required floor space: the maximum space available for the supermarket is 3360m² (96 x 35).
- Cost
- Required number of operators

The cost can be divided into two parts: running costs and investment cost. The running costs are mainly determined by the number of operators performing the picking actions in the supermarket. Since the investment is a one-off cost, we have seen that minimizing the running costs and the number of operators in particular is very important.
Based on these output variables, the optimal order-picking system for the predefined product groups can be determined. By changing the different classification parameters, the overall best solution can be defined.

DIFFICULTIES AND POSSIBLE SOLUTIONS

In this case study, we needed to investigate if it is possible to design a supermarket which is capable of supporting a full kitting line delivery strategy, given the constraints in our automotive factory. Full kitting means that for each vehicle a number of kits is composed, containing every single part that needs to be mounted on the vehicle. In this scenario, even the smallest and most common parts need to be picked in the exact amounts, implying a huge workload for the pickers in the supermarket. In the case, the location for the supermarket was already fixed. This means that the ground surface available for the supermarket, was constrained to a maximum of 3360m². Since this foreseen space is rather confined, throughout the project, the required floor surface became the most pressing limitation. The best OPS was decided to be the system with the lowest cost that is able to reach the required throughput within the surface constraint. As mentioned before, costs are directly linked to the required number of operators.

Table 2: output full kitting scenario

Throughput requirements	5820	lines/hr
SKU's	2178	
#Operators	66	
Throughput	7165	lines/hr
Ground surface	5379,28	m²
Investment cost	5105171	€
Running Cost	3960000	€

Table 2 shows the results from a first simulation run. This run simulates a full kitting scenario where only picker-to-parts order-picking systems were considered. These results show immediately that there are three major design problems for the supermarket: throughput, cost and the required ground surface. To face these problems, a number of countermeasures can be taken:

- **Case picking:** As a result of the full-kitting strategy, a lot of operating time is lost to counting parts. Case picking is bringing fast moving parts to the line in bulk, eliminating picking times partially.
- **Batching orders:** all parts for multiple vehicles are picked at the same time, reducing walking distances for the operators and workload of the automated picking systems (VLM / HC).
- **Downsizing:** downsizing is the process of transferring parts from their supplier emballages to the minimal possible emballage. This reduces the number of space consuming pallets in the supermarket.
- **Automated picking for blue boxes**: The use of automated picking systems (VLM / HC) can lead to a decrease in workload of the operators by eliminating walking distances. A VLM is also capable of storing parts at greater height, reducing the ground surface needed to install the supermarket.
- **Diminish the inventory in the supermarket:** The required ground surface can also be reduced by diminishing the inventory level in the supermarket. This countermeasure implies larger demands on the supply of parts to the supermarket.
- **Warehouse automation:** we also investigated the possibility of picking parts directly from the pallet warehouse (HBW) using an automated storage and retrieval system (AS/RS).

RESULTS

Case picking

Full kitting implies that a lot of operating time is lost to counting parts and large walking distances for the pickers in the supermarket. Therefore, we investigated the possibility of sending fast moving parts to the line in bulk by letting the model change the case-picking frequency. All parts with a lower frequency (parts/vehicle) are brought to the assembly line in blue boxes.

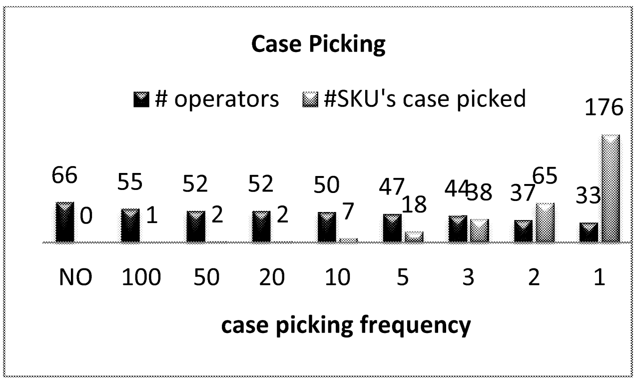


Figure 3: Effects of case-picking

As shown in Figure 3, case-picking can lead to a spectacular decrease in the required number of operators. However, there is a constraint on the case-picking frequency: blue boxes are brought to the line on the same carts as the kits. Every cart can take only one blue box. Since there are on average 7 kits per vehicle, every takt-time (6.28 minutes) a maximum of 7 carts can be transported to the line. This results in a minimal case picking frequency of 3, leading to a 33% reduction in required pickers in the supermarket (from 66 to 44 pickers). The results in Figure show that it is beneficial to change the full kitting line delivery strategy into a hybrid strategy. In this hybrid strategy, fast-moving (common) parts are still brought to the assembly line in the respective emballages, while the variant parts are put together in kits. The difference between these two strategies is explained in Figure .

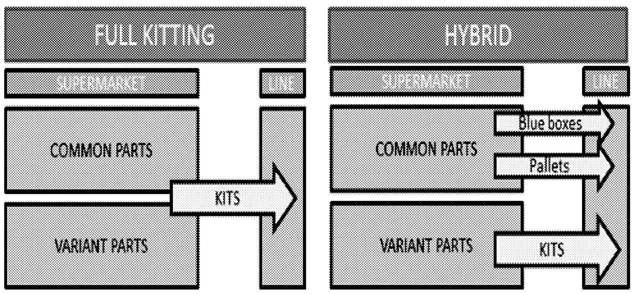


Figure 4: full kitting vs. hybrid line delivery strategy

Automated picking of blue boxes

Two different types of automated picking systems for blue boxes were examined: horizontal carousels (HC) and vertical lift modules (VLM). In general, HC's appeared to be slower and less space efficient than VLM's. Since ground surface is the most critical constraint in the case, storing slow moving parts in VLM's was a better solution in our case. Which parts are considered to be slow movers, is specified by the slow mover frequency. By changing this parameter, the model is capable of deciding which parts need to be put in VLM's. Figure 5 shows that there is a possibility to reduce the needed space as well as the required number of operators by using VLM's.

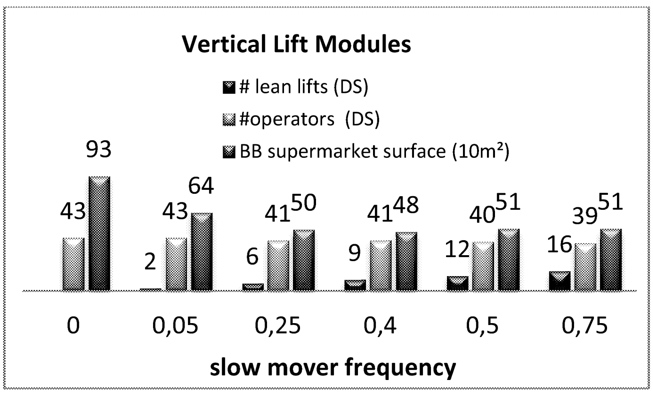


Figure 5: Results of automated picking for blue boxes

The graph also indicates that a higher degree of automation does not always result in a reduction of the required amount of operators or needed ground space. It is remarkable that the surface of the part of the supermarket in which the blue boxes are stored increases when the slow mover frequency is changed from 0.4 to 0.5. This can be explained by the limited speed of the lean lifts. The model adapts the height of the vertical lift modules to the throughput the lifts need to achieve. On average, higher VLM's will need more time to bring a box to the picker than lower lifts. If there are too many "fast-moving" products in the VLM's, it is only possible to keep up with the demand if more and smaller lifts are used. This will cause the lifts to lose the advantage of saving space by storing boxes at greater height.

Downsizing

The advantage of downsizing lies in the fact that a considerable amount of parts can be stored in smaller boxes. In our case this reduces the amount of space consuming pallets in the supermarket by 30%. The effects of transferring pallets into smaller blue boxes on the required surface is shown in Figure 6.

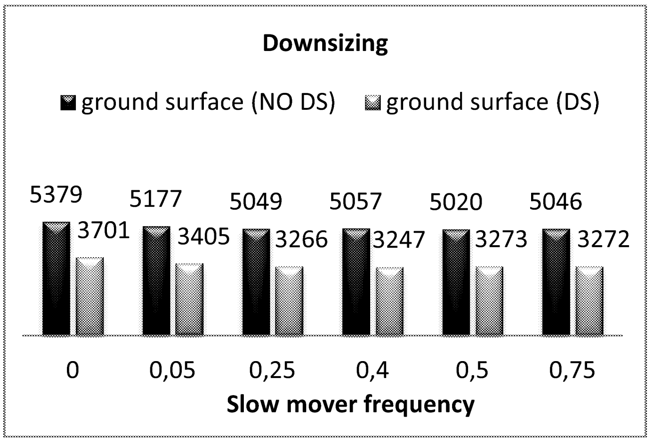


Figure 6: Effects of downsizing on supermarket surface

Since automated picking systems (VLM) can only be used for the storage and picking of blue boxes, downsizing also increases the potential degree of automation in the supermarket. In addition, the smaller supermarket surface ensures the partial elimination of walking distances for the pickers. This results in a decrease of the required amount of pickers, as shown in Figure 7.

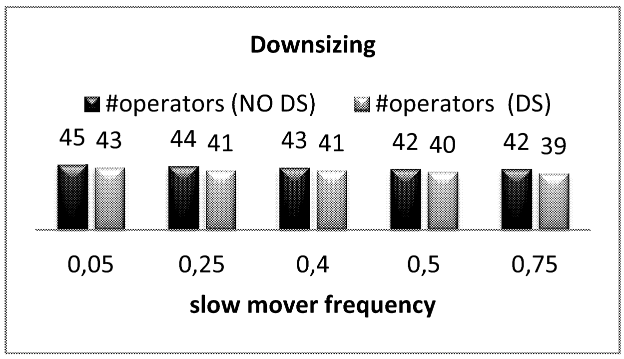


Figure 7: Effects of downsizing on required amount of pickers

Batching orders

If orders are batched, a number of finished kits will have to wait before they can be transported to the line. Therefore, a bigger buffer will be needed in the supermarket. Since the space available in the supermarket is limited, batching orders does not seem to be beneficial. In addition, batching orders would increase the lead time in the supermarket. This results in a loss in reactivity of the total system.

Diminish inventory level in supermarket

Throughout the case, a four hour inventory in the supermarket was taken into account. If it is possible to reduce this inventory level, parts can be brought closer together, resulting in smaller walking distances for the operators. By letting the model change the inventory level in the supermarket after downsizing, case-picking and automated picking for blue boxes, we obtained the graph shown in Figure 8. By lowering the inventory level in the supermarket to one hour, the required floor space was reduced by 4%. The amount of operators needed to perform all picking actions however, stayed exactly the same. On

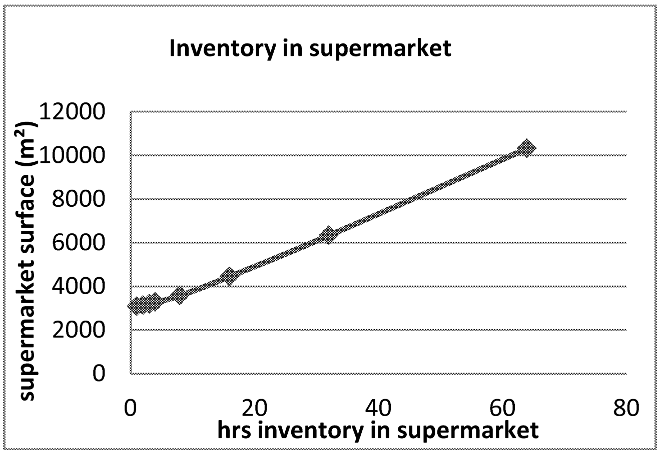


Figure 8: Effects of changing inventory level on supermarket surface

the downside, lowering the inventory level places larger demands of lead time and flexibility on the supply of the supermarket. The advantage of a small reduction in required surface doesn't outweigh these drawbacks. The possibility of eliminating the supermarket by picking the parts directly from the warehouse, was also investigated. If we could achieve a very short supplier lead time of 4 days (64 hours), more than 10.000m² of warehouse surface would be needed. This is more than the current warehouse surface. Moreover, the number of pickers should be increased by 7,5%, only to cover the extra walking distances. Therefore, in our case, eliminating the warehouse did not seem to be beneficial.

Warehouse automation

A final but far-reaching solution to get around the surface constraint, is picking all parts that are stored on pallets directly from the warehouse using an automated storage and retrieval system. By doing this, the supermarket is partially eliminated. Figure 9 shows that the number of pickers can be reduced by four if the warehouse is capable of bringing 1220 pallets to the pickers every hour. The current AS/RS in the high bay warehouse (HBW) is only reaching 200 outshoots per hour. Increasing the capacity of

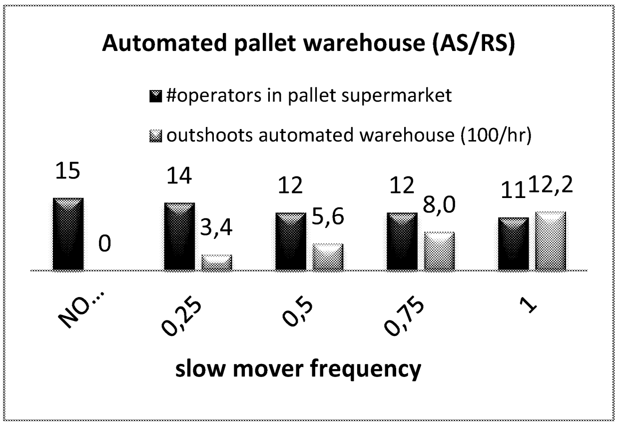


Figure 9: Effects of automated picking for pallets

Table 3: general guidelines for OPS design in a manufacturing environment

problem	possible solutions	effect	positive side-effects	negative side-effects	constraints
# operators needed	Case Picking	significant decrease	-	-	number of cases that can be transported to the line
					storage space at the line
	Batching orders	significant decrease	-	reaction time increases	Buffer size
needed ground surface	downsizing	decrease: depends on the number of SKU's that can be downsized	slight decrease in #operators	operators needed to do downsizing (if not done by supplier)	Minimal embalage (parts size)
			increases the potential number of parts that can be automated		
	VLM	decrease in supermarket for BB	decrease in # operators	cost of automation	only for blue boxes
					part frequency (VLM speed)
	Diminish inventory in supermarket	very small decrease	-	requires higher reactivity in replenishment	lead time of supply from warehouse
	Automated Warehouse	eliminates part of the pallet supermarket	remarkable decrease in #operators	cost of automation	Warehouse picking capacity
limited throughput	VLM	increased throughput by eliminating	decrease in required #operators	cost of automation	only for blue boxes
				increase in surface needed (many smaller lifts needed)	

the HBW asks for a great investment. Since we were able to fit the supermarket in the foreseen space by downsizing pallets and storing slow moving blue boxes in VLM's, extra capacity in the HBW was not necessary and could not be justified.

GENERAL CONCLUSIONS

In the case, there were three main constraints we had to take into account while designing an order-picking system for our automotive factory: cost, ground surface and the required throughput. These constraints and their importance are strongly depending on the case itself. Based on the simulations using our dataset, we established some general guidelines for the design of order-picking systems in a manufacturing environment, shown in Table . These guidelines are a first step in the direction of further and more general research in this domain.

REFERENCES

Baker, P., Canessa, M., 2009, "Warehouse design: A structured approach", *European Journal of Operational Research*, Vol. 193, pp. 425– 436.

Choe, K., Sharp, G.P., 1991, Small parts order picking: design and operation, *School of Industrial and Systems Engineering Georgia Institute of Technology*, pp. 41

Dallari, F.G., Marchet, G., Melacini, M., 2008, "Design of order picking system", *International Journal of Advanced Manufacturing Technology*, Vol. 42, pp. 1-12.

de Koster, R., Le-Duc, T., Roodbergen, K.J., 2007, "Design and control of warehouse order picking: A literature review", *European Journal of Operational Research*, Vol. 182, pp. 481-501.

Goetschalckx, M., McGinnis, L., Sharp, G., Bodner, D., Govindaraj, T., Huang, K., 2001, Development of a design methodology for warehousing systems: hierarchical framework, School of Industrial and Systems Engineering Georgia Institute of Technology Atlanta, pp. 6.

Gu, J., Goetschalckx, M., McGinnis, L.F., 2007, "Research on warehouse operation: A comprehensive review", *European Journal of Operational Research*, Vol. 177, pp. 1-21.

Meller, R.D., Klote, J.F., 2004, A throughput model for carousel/VLM pods, IIE transactions, Vol. 36, pp. 725-741.

Meller, R.D., Klote, J.F., Lilaj, E., 2007, Design of carousel orderpicking v5.2, MS Excel spreadsheet, VirginiaTech.

Van Den Berg, J.P., Zijm, W.H.M., 1999, Models for warehouse management: Classification and examples, International Journal of Production Economics, Vol. 59, pp. 519-528.

Yoon, C.S., Sharp, G.P., 1996, "A structured procedure for analysis and design of order pick systems", *IIE Transactions*, Vol. 28, No. 5, pp. 379-389.

Goetschalckx, M., and Ashayeri, J., 1989. Classification and design of order picking systems, *Logistics World* (1989) (June), pp. 99– 106

Drury, J., 1988. Towards more efficient order picking. IMM Monograph No. 1, Report, The Institute of Materials Management,Cranfield, UK.

J.A. Tompkins, J.A. White, Y.A. Bozer, E.H. Frazelle and J.M.A. Tanchoco, Facilities Planning, John Wiley & Sons, NJ (2003).

Limere V. and Van Landeghem H., 2009. Cost model for parts supply in automotive industry. In *Proceedings of the 16th European Concurrent Engineering Conference*. Eurosis, Bruges, Belgium, 120-125.

Fisher, M.L. and Ittner, C.D., 1999. The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Science*, Vol. 45, No. 6, pp. 771-786.

Peerlinck, K., Govaert, T., Van Landeghem, H., 2010. A design method for parts picking zones in a manufacturing environment. In *proceedings of the 8th Annual international Industrial Simulation Conference* (*2010*)

BIOGRAPHIES

KAREL BAUTERS is PhD student and teaching assistant at the department of Industrial Management at Ghent University. His research interests lie in lean manufacturing, production logistics and material handling.

TIM GOVAERT is scientific researcher at the department of Industrial Management at Ghent University. His research interests lie in material handling, facility logistics and lean manufacturing. Tim already completed several projects within the automotive and truck industry.

VERONIQUE LIMERE is PhD student at the department of Industrial Management at Ghent University. Her research interests lie in facility logistics, material handling and lean manufacturing. In her PhD she is researching the organization of the flow of parts from the supplier to the use-point at the line. Recently she obtained a Doctoral Fellowship from the Intercollegiate Center for Management Science (I.C.M).

HENDRIK VAN LANDEGHEM is a professor at the department of Industrial Management at Ghent University. He is an expert in the area of logistics and their application in business processes. He advises companies in their choice and implementation of their logistic organization and production control systems. He is Fellow of the European Academy of Industrial Management (AIM) and member of the Institute of Industrial Engineering (IIE) and of the European Operations Management Association (EUROMA). He is since 2007 Fellow of the World Confederation of Productivity Science.

COMPARING THE PERFORMANCE OF POLCA AND LOAD-BASED ORDER RELEASE SHOP FLOOR CONTROL METHODS USING SIMULATION

H. Van Landeghem
K. De Cock
F. Declerck
P. Verdegem

Department of Industrial Management, Ghent University
Technologiepark, 903,
B-9052 GENT, Belgium
E-mail: hendrik.vanlandeghem@ugent.be

KEYWORDS
Load Based Order Release, POLCA, production, discrete simulation.

ABSTRACT

This paper compares the behaviour of two sophisticated shop floor control methods: **Paired-cell Overlapping Loops of Cards with Authorization** (POLCA) and **Load Based Order Release** (LBOR). As data set the A-Plant from Goldratt has been selected and a simulation model is used as research method. Results show how both methods cope with limiting WIP and Average Flow Time. The paper also discusses complicating factors in the use of both methods. These findings shed new light on the applicability of these methods.

THE PURPOSE OF THIS RESEARCH

In recent years the trend to more flow-oriented manufacturing operations, driven by the Lean Manufacturing paradigm, has sparked renewed attention to methods for controlling Work in Process (WIP) on the shopfloor. WIP, after all, is the driving force behind leadtimes, whose reduction is in itself a main objective of Lean Management. Shop Floor Control is defined here as the decision when to launch production orders to the shopfloor, when to move finished orders to their next workstation, and the sequence in which orders are processed on that station. Numerous simple scheduling rules exist for this, such as Shortest Processing Time, Earliest Due Date, as well as sophisticated calculation algorithms (see Coffman and Bruno, 1976 or equivalent for an overview).

However some more sophisticated methods have been developed in the past, such as **Paired-cell Overlapping Loops of Cards with Authorization** (POLCA) and **Load Based Order Release** (LBOR). Initially they were not very widely used, a phenomena we encounter often in the conservative world of manufacturing control. However lately these methods have resurfaced, as companies that become more Lean are looking for appropriate control methods to support their effort to enhance flow on their shopfloor (Germs & Riezebos, 2008) (Pieffers & Riezebos, 2005).

In this paper we investigate though simulation the behaviour of both POLCA and LBOR on a standard test set and compare the results. We will first introduce both methods briefly, describe the data set, and the simulation, and finally discuss the results.

POLCA: PAIRED-CELL OVERLAPPING LOOPS OF CARDS WITH AUTHORIZATION

POLCA has been developed as part of the more extensive Quick Response Manufacturing strategy developed by R. Suri (1998), as a response to the limitations of Lean in High-Mix Low-Volume (HMLV) environments. QRM is based on the Time Based Competition (TBC) strategy which focuses on speed and service to provide competitive advantage (Krisnamurthy & Suri, 2003). POLCA is an advanced scheduling method that combines push and pull by respectively introducing authorization lists and POLCA cards. Authorization lists contain the orders that are ready to be worked on, because of Due Date proximity and availability of material. They are used at the first operation, and at any intermediate workstation that needs resequencing of orders (such as assembly stations). The source of information for these lists are mostly the production planning system (PPS) that converts customer orders into shop orders. Authorization lists therefore trigger the start of orders.

The flow of orders is controlled by POLCA cards. They are much like Kanban cards, but instead of controlling physical containers of material, they control quantums of capacity. A POLCA card represents say 4 hours of work on a given workstation M2 (see figure 1). Any previous workstation that wants to produce an order that subsequently needs to be processed on M2 will have to wait for a POLCA card in order to start with this order. A POLCA card therefore acts as a one-step look-ahead coordination mechanism within a loop linking two workstations that are adjacent in the parts' routing. POLCA cards look like two-colored dominoes, to identify the unique loop they are circulating in. In figure 1 the blue-orange cards link machines V3 (blue) and M2 (orange). One of the authorized orders (nr 1, greyed) waiting in front of machine V3 can be processed as soon as a V3/M2 card (nr 2) is available with enough capacity quantum. After processing (nr 3) the batch moves to the IN queue at M2 (nr 4), where it will wait until a orange-yellow POLCA card becomes available for the

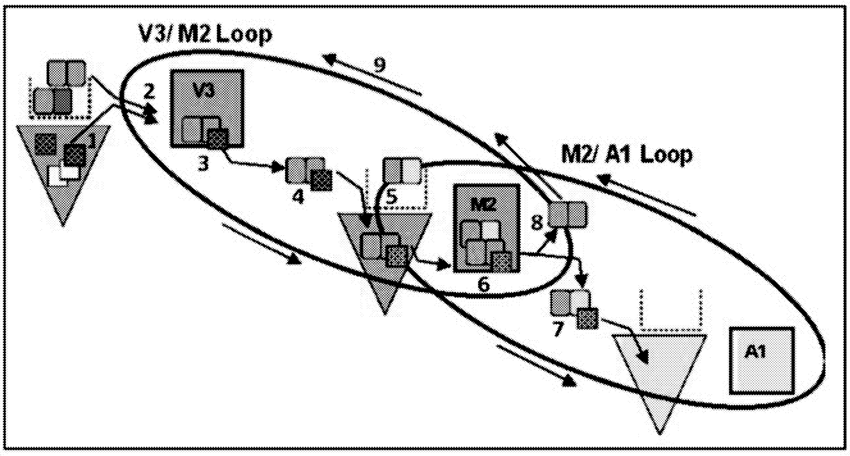


Figure 1: POLCA mechanism (Source: Pieffer & Riezebos)

M2/A1 loop (nr 5). The part will then move to M2 (nr 6), accompanied by two POLCA cards, hence the name "overlapping loops". As soon as the part has been finished on M2, the V3/M2 Polca card is released (nr 8) and returns to machine V3 to allow another part to start. Meanwhile, the part moves to the IN queue of A1 (nr 7).

Existing literature is limited for POLCA and few comparisons have been made between POLCA and other advanced scheduling methods. In multi-product environments, hybrid push-pull systems such as Polca achieve, under certain conditions, better results than pure pull or push systems (Spearman & Zazanis, 1996) (Luh et al, 2000).

The main limiting factor in POLCA is the number of loops between resources. In workshops with many machines they need to be grouped in workcells, that constitute aggregated resources, linked with POLCA loops. Cells need to contain multiple machines and should be able to handle all production steps within their scope (Pieffers & Riezebos, 2006).

The main parameter to decide is the quantum assigned to POLCA cards. Different loops i can use different quantums q_i, but of course this complicates supervision of the workshop. Too small quantums will split orders but allow for a smooth flow; too large quantums will cause WIP fluctuations and possible idle stations. We choose to work with 3 different quantums: if the largest operations times in a loop y were more than 250 minutes, the quantum q_y was set to 100 minutes. If this maximum was less than 150 minutes, q=30 minutes, and q=50 minutes for the intermediate case. For the calculation of the number of POLCA cards, formulas are proposed by Suri (1998) and by Pieffers & Riezebos (2005, 2006). They are based on the leadtime needed to process work quantums in the workstations, increased by the waiting time in the loop. The latter is approximated by a safety factor on the basic leadtime. The number of POLCA cards should be enough to achieve a reasonable stable work flow, covering at least the total lead time of one cycle. One adds safety time by adding more cards. The higher the number of POLCA cards the more WIP will be in the system, the lower the number of cards the higher the risk on idle time. The determination of the optimal number of POLCA cards remains a subject for further research as no simple rules have been published yet. In our simulation we used a safety factor of 1,8.

LBOR: LOAD-BASED ORDER RELEASE

Load-based order release stands for a method that controls flow times in a shop floor by controlling the actual input of work to each workstation (Wiendahl, 1995). It is based on the principle that in order to achieve high productivity sufficient workload should be maintained in front of workstations (Wiendahl, 1995) (Hopp & Spearman, 2001). The main parameter in LBOR is the Load Limit LL, the maximum amount of work that is allowed to be waiting in front of a workstation. From it the Loading Percentage LPG is calculated as:

$$LPG = LL \times 100 / OUT$$

the ratio of the Load Limit over the Planned Output for a given workstation and a given reference Time Period.

The LBOR method will calculate for each order the probability that in the next Time Period it will be processed by the workstation. This probability is given by the inverse of LPG, called the Conversion Factor CF, i.e.

$$CF=P(\text{order finished on WS X})= LPG(X)^{-1}$$

Thus for an order with a routing of n production steps with workload WL(i) in step i, the expected load on each of the workstations involved can be calculated as

$$E[\text{load before WS i}]=[1/LPG(i)]^{(i-1)} \text{ x}WL(i) \qquad i=1,..,n$$

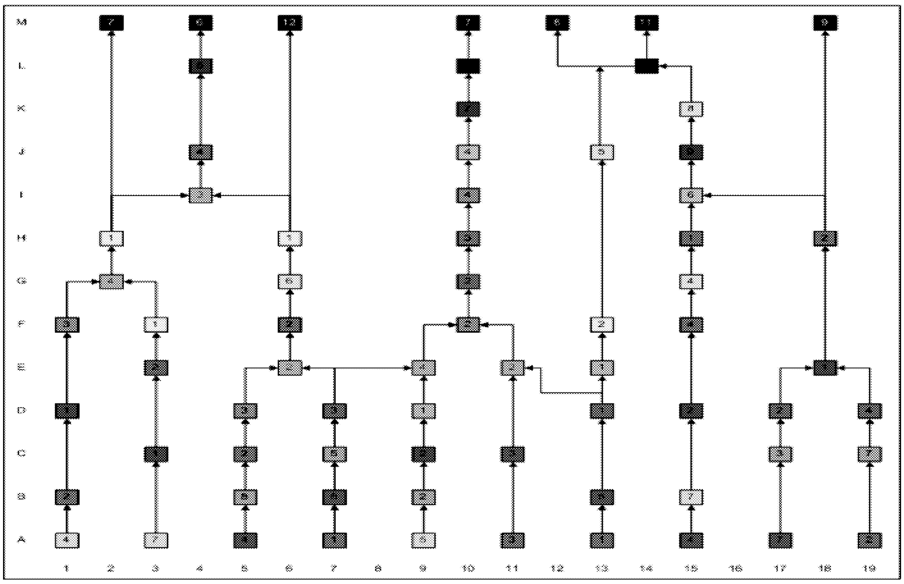


Figure 2: A-Plant routing network (Source: Goldratt)

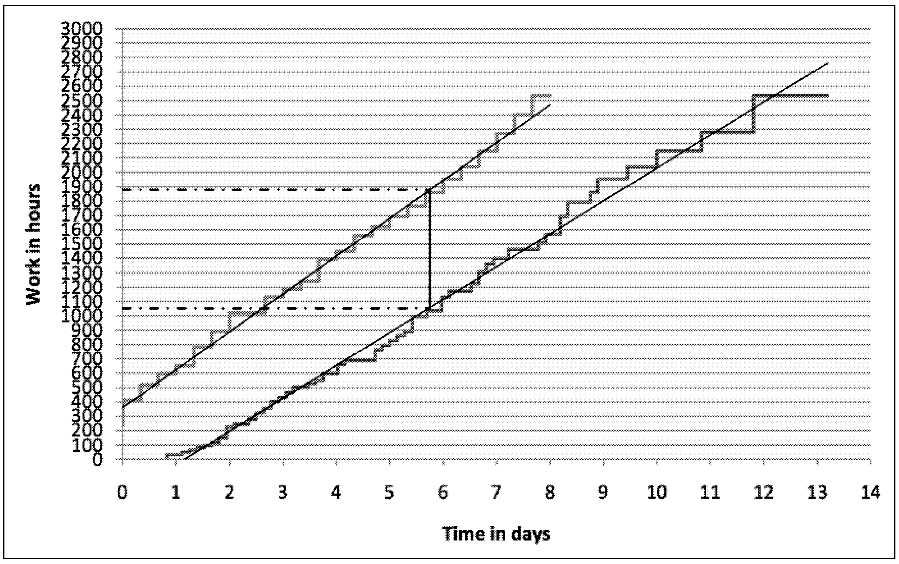


Figure 3: Input-Output curve for LBOR simulation

For example, an order with 10 hours of workload, in a system where the planned load is double the capacity (i.e. a LPG of 200% and hence Conversion Factor of 0,5) will have an expected load of 5 hours on the subsequent workstation, 2,5 hours on the next one, 1,25hrs on the next one, and so on. These loads occur in consecutive Time Periods. So choosing the Time Period too large will lead to a coarse schedule with lots of gaps in the workflow. A too small Time Period will require excessive recalculation during the production. Typically a shift or a day is used.

The actual order release method then works as follows. Backward scheduling is used to determine the latest start date (LSD) for an order. A Release Horizon is chosen that acts as release trigger. For each work order whose LSD is within the Release Horizon the expected load is calculated on each of the workstations of its routing. Orders are then actually released in the factory as long as the load limit of a work station is not exceeded. (Too) high Load Limits will increase WIP, while too low LL's will reduce productivity. The sequence in which orders are tested for release can be freely chosen, we used Earliest Due Date. The best result in the simulation was obtained with a Time period of 160 minutes, a Release Horizon of 480 minutes (one shift) and a Load Percentage of 600% (or 960 minutes), i.e. a CF=0,17.

CASE STUDY: THE A-PLANT

In order to compare these two sophisticated methods we needed a test data set that was sufficiently challenging. We choose the A-Plant case, developed by Goldrath (1994) to demonstrate his TOC software. Figure 2 shows the material flow, with each box representing a production step on a specific part type, and its color the workstation that is needed for the operation. The case contains many iterative loops (sequential steps on the same machines) and assembly steps. The problem has 16 workstations (some with multiple machines), 59 orders with due dates for 8 types of product, yielding 178 steps to process, needing 10 raw material types.

SIMULATION RESULTS

The A-plant configuration, and the POLCA and LBOR scheduling logic were implemented in Flexsim ®, a well-known simulation software from Flexsim Corp. Calculations were done in a separate excel sheet, and the simulation model read these results from the sheet. For each of the methods several runs were performed to tune each method's parameters (Declerck, 2010) (Verdegem, 2010). Figure 3 shows the input-output curve for the best case of the LBOR method.

The equivalent curve for the best POLCA run is shown in figure 4. It shows far more variability in leadtime and WIP than in figure 3. At the end quite some orders have to wait for available capacity, tapering off the input curve.

We also simulated two standard procedures: Shortest Processing Time and Earliest Due Date, without any order release mechanism (i.e. orders were all released at time 0 and allowed to progress as fast as possible), to serve as benchmarks.

Overall results are shown in table 1. We can see that both methods perform better in controlling WIP and Flow times, at the detriment of a somewhat longer makespan than the simple methods. Comparing LBOR and POLCA we see that POLCA has a slightly lower Average Flow Time, but with higher variability (figure 5). So LBOR appears to be the more robust method.

One can conclude that both methods achieve their intended purpose: controlling WIP and thereby enhancing

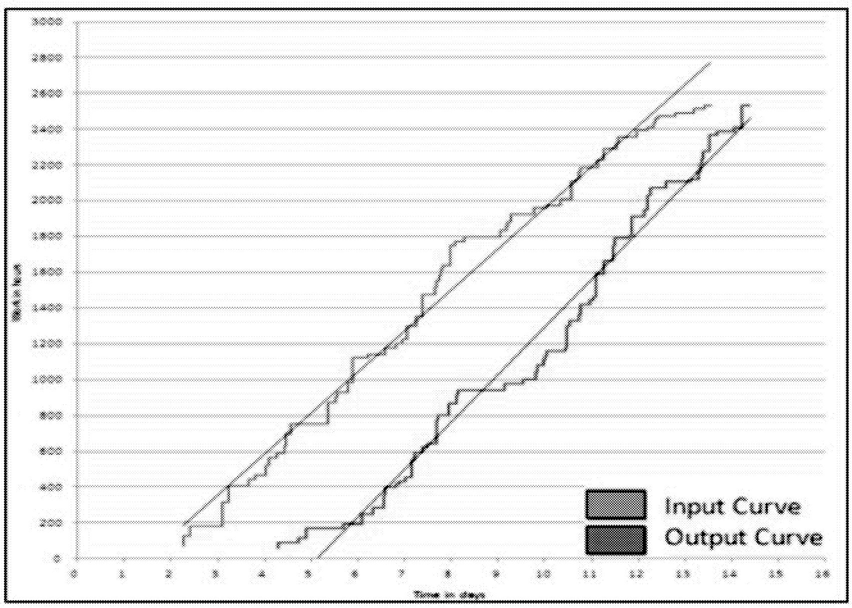


Figure 4. Input-Output curve for POLCA simulation

Table 1: Key results of the simulation study

	EDD	SPT	LBOR	POLCA
Makespan (days)	10	9,8	12,5	12,52
Average flow time (days)	6,75	6,25	2,38	2,3
Std.Dev. flow time (days)	2,1	2,67	0,76	1,11
Average WIP (hours)	1710	1520	700	713

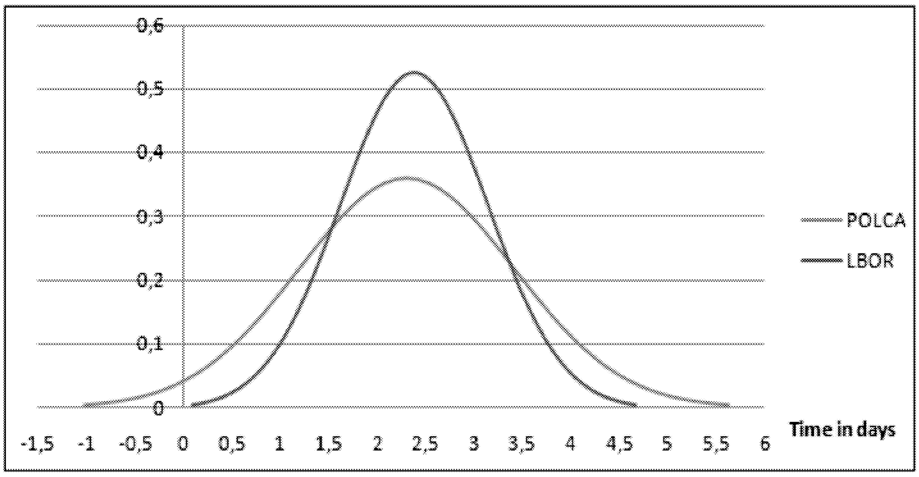


Figure 5: Flow Time Distribution of LBOR and POLCA

flow.

POLCA AND LBOR COMPARED

This then leads to the question of how these methods compare in their practical use. Both methods need careful tuning of their parameters.

Although LBOR has more complex calculations, the main parameter is the Load Limit, and the recalculation frequency is determined by the Time Period. However the LL can be different for each workstation. There are however no clear guidelines how to determine these, and for complex situations (such as the A-Plant) it is unfeasible to set up the needed expirements. Further research will be performed to look for an algorithmic approach to optimize both LL and Time Period values.

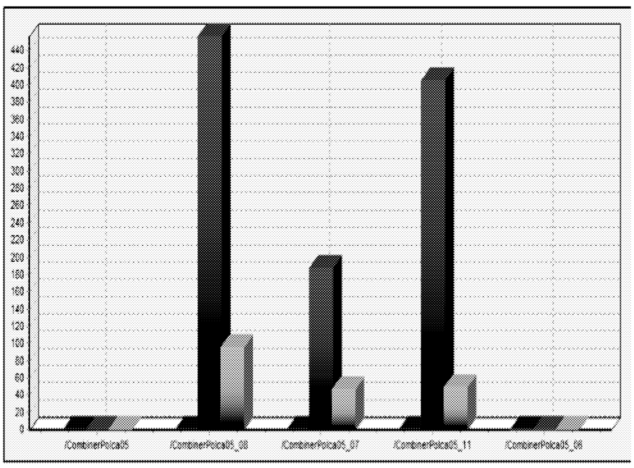


Figure 6: Min-max-avg waiting time of parts in assembly stations (POLCA)

The determination of the number of POLCA cards needs to be taken for each loop. In the A-Plant we arrived at 48 different POLCA loops (from a theoretical number of 120 possible links among 16 resources). Two factors proved to be complicating: variability of job workload and assembly stations. The former will lead to highly variable waiting times before workstations, as witnessed by figure 6.

Assembly stations cause POLCA loops to converge, requiring a card for each of the components before they can progress to the assembly station. One can use the authorization list at the stations preceding the assembly to try to synchronize their processing, but this also leads to an increase in average flow time. These stations proved to be a troublesome element in practice. Initial tests assigned too little card to assembly loops, leading to deadlocks when parts (and cards) were waiting for a final part needed in the assembly, thereby also preventing other parts to progress through the same workstation. By adding up to 10 cards to these loops, the problem was solved. However no formal rules could be established.

CONCLUSION

We analyzed the behaviour of two sophisticated shop floor control methods: POLCA and LBOR. Simulation revealed that they achieve comparable results, limiting WIP and therefore average flow time of the jobs involved. LBOR seems to be more robust with lower variability on the results. The main difficulty in practical use of these methods is their tuning, which affects performance considerably. From this viewpoint LBOR seems to be the more simple method to tune, however it is more mathematically based. This in itself can limit its appeal to (smaller) companies.

Further research will focus on methods to automatically tune (and retune) the number of POLCA cards and algorithms to adjust the Load Limit dynamically.

REFERENCES

Coffman E., Bruno J. (eds), 1976, Computer and job-shop scheduling theory, Wiley, 299 p.

Declerck, F., 2010, Comparative study of advanced scheduliing methods using simulation: Load-Based Order Release (in Dutch), *Master's thesis*, Ghent University.

Germs, R., Riezebos, J., 2008, Workload balancing capability of pull systems in MTO production, *Proceedings 15th International working seminar on Production Economics*, Innsbruck, Vol. 2

Goldratt, E., 1994, It's Not Luck, North River Press, 283 p.

Hopp, Wallace J. & Spearman, M.L., 2001, Factory Physics, McGraw-Hill

Krishnamurthy, A. & Suri, R., 2003, How to Plan and implement POLCA: A Material Control System for

High-Variety or Custom- Engineered Products. *Technical Report*, Center for Quick Response Manufacturing
Luh, P.B., Tomastik, R.N., Zhou, X., 2000, The Performance of a new Material Control Replenishment System, *Proceedings of the QRM 2000 Conference*, Society of Manufacturing Engineers.
Pieffers, J., 2005, Let's POLCA? A case study oriented research into practical applicability of POLCA as innovative material control system, (in Dutch) *Master's thesis*, Rijksuniversiteit Groningen.
Pieffers, J. & Riezebos, J., 2006, POLCA as innovative material control system, (in Dutch) *Working paper*, Rijksuniversiteit Groningen
Spearman, M.L, Zazanis, M.A., 1992, Push and pull production systems: Issues and Comparisons, Northwestern University, Evanston, Illinois
Suri, R., 1998, Quick Response Manufacturing: A Companywide Approach to Reducing Lead Times. Productivity Press Inc
Verdegem, P., 2010, Comparative study of advanced scheduliing methods using simulation: POLCA (in Dutch), *Master's thesis*, Ghent University.
H-P. Wiendahl, Load-Oriented Manufacturing Control. Berlin Heidelberg, Springer Verlag, 1995, p.368.

AUTHOR BIOGRAPHY

HENDRIK VAN LANDEGHEM is Full Professor at Ghent University. After 10 years of industrial experience he became professor in 1992. His expertise domain is in-plant logistics and production planning systems. He is author of more than 100 articles. He is member of the Flemish Academy of Belgium for Arts and Sciences, Fellow of the World Confederation of Productivity Science and member of the Institute of Industrial Engineering and EUROMA.

KURT DE COCK is a researcher at Ghent University. He specializes in ICT and simulation model building. He designed the logic of the Flexsim models in this research.

FREDERICK DECLERCK and **PIETER VERDEGEM** graduated in 2010 as Masters of Engineering in Industrial Engineering and Operations Research. They performed part of this research for their Master's thesis

ENVIRONMENTAL SIMULATION

THE OTHER GLOBAL WARMING PROBLEM

William Conley
Austin E. Cofrin School of Business
University of Wisconsin-Green Bay
Green Bay, Wisconsin 5431-7001
U.S.A.
Conleyw@uwgb.edu

KEYWORDS
Management, disaster prevention, mitigation, multi stage Monte Carlo optimization, complex systems

ABSTRACT

A brief review of the global warming problem and the political, scientific and industry debate about its causes and cures or at least its mitigation, will be presented. Additionally, the other global warming potential catastrophe of an asteroid, meteor, or comet exploding on the earth at some point in the future will be discussed with an eye toward preventing this calamity from ever happening. The world's scientific and industrial communities can play a big role in helping to remove these threats from the 21st century and perhaps all future centuries too. Specifically, presented here will be actual applied physics, mathematical and statistical calculations that may be needed to organize the solution to this other extremely dangerous global warming problem. Industry has helped solve many of the pollution and at least some of the global warming problems. It can also play a huge role in mitigating this other global warming problem. Solving these two problems may even be related from scientific and managerial viewpoints. An example of this was the exploding of a large asteroid millions of years ago that caused a great fire whose ash blotted out the sun and produced an ice age which doomed the dinosaurs.

INTRODUCTION

This theory about the dinosaurs is widely accepted and other events like this could help explain some of the variations in warming and cooling with the earth's ice age cycles. The asteroid that exploded above the ground in Siberia in 1908 (Stone, 2008) leveled a forest of many hundreds of square miles. It is disturbing that it was estimated to be a rock of less than 50 meters in length. The light from the explosion was visible over a thousand miles away. If a much larger one exploded over the earth it could set half of a continent on fire (a catastrophic global warming). These types of disasters thousands of years ago may have even influenced warming and cooling cycles.

The director of Russia's space agency, Anatoly Perminov, in (Isachenkov, 2009), is advocating an international effort of the world's space powers to deal with this problem. "People's lives are at stake. We should pay several hundred million dollars and build a system that would allow us to prevent a collision, rather than sit and wait for it to happen and kill hundreds of thousands of people," Perminov said. He wants to avoid using nuclear bombs on an asteroid in outer space, and advocates sending a rocket to intercept the asteroid Apophis which may threaten the earth in 2029 or again in 2036 (Stone, 2008).

One approach (Conley, 2010) is to put engines on other asteroids and then use these to ram into Apophis (or future threatening asteroids) and knock it off of its line toward the earth. He also says that if engine mounting on asteroids is not practical yet, we should send up rockets with big empty cargo bays, to collect small asteroids to add mass. Then have these rockets smash into the threatening asteroid to knock it off line or break it apart. So presented here will be an example where the actual momentum calculations are incorporated into an optimization problem to solve one of these asteroid deflection challenges.

This will be followed by a multivariate engine tuning optimization problem (for maximum rocket performance). Additionally, a third example dealing with probabilities of success while minimizing industrial costs in a minimum content setting will be presented. Then a further discussion of the traditional global warming (pollution, etc.) and this new global warming problem will be mentioned from a management and organizational point of view. Projects of this type are so technical and long term that getting the international management expertise is crucial to make our planet safer and more habitable in the 21st century and beyond.

A MULTIVARIATE MOMENTUM CALCULATION

An 18 billion kilogram asteroid is headed on a collision path with the earth at a velocity of 2000 meters per second. The plan is for the international asteroid defense consortium of the world's space powers to launch 20 rockets from the earth (with large empty cargo bays) and send them up to the asteroid belt to collect small asteroids to add mass. Then when the big dangerous asteroid is about 125 million miles from earth (on one of its loops) the asteroid laden-rockets will hit it with a total mass of 100 million kilograms at an average velocity of 1500 meters per second, perpendicular to the dangerous asteroid's path of motion toward the earth. Let us do the deflection calculations in millions of kilograms and hundreds of meters per second for the velocities. Also, let us assume the dangerous asteroid is heading east (so its velocity can be noted as VE). Further, we assume our rockets will attack it coming from the north and heading south (so their velocities are heading south VS).

Therefore, after the coordinated collision $18{,}000 \cdot 20 = 18100 \cdot VE$ and so $VE = (18000 \cdot 20)/18100 = 19.8895$ and $100 \cdot 15 = 18100 \cdot VS$ so $VS = (100 \cdot 15)/18100 = .08287$. So the arctan $(.08287/19.8895) = .23875$ of a degree.

Therefore, if the rockets hit the dangerous asteroid at an average speed of 1500 meters per second (and their total mass is 100 million kilograms) at a point 125 million miles from the earth, the deflection of .23875 of a degree will translate into the dangerous asteroid missing the earth by about 500,000 miles. However, gravitational attractions and whether the collision is inelastic or breaks the asteroid apart could influence the results too. The Hubble telescope photographed the aftermath of two asteroids colliding in January of 2010 (Maugh, 2010), for example. (Given the short period of time the Hubble has been in space these crashes must happen frequently and they can deflect one of the asteroids toward the earth.)

Therefore, our goal here is to solve for the speed X_i of the twenty rockets when the rockets have masses $(79+2_i)/20$ for i=1,2,3, . . . 20 and so we seek to solve $\Sigma(79+2_i)/20) X_i = 100.15$ subject to $1.2X_3 \le X_4$ and $1.3X_{14} \le X_{15}$ because rocket four and fifteen's speeds must exceed rockets three and fourteen's speeds by 20 and 30 percent respectively. Also $X_1{**}(1.03+.011) < X_2{**}(1.03+.011)2)$ because engine one is to run at less cost than engine two and additionally for $C_i = 10+.25_i$, i=1, 2, . . . 20 $\Sigma C_i X_i{**}(1.03+.011_i) <$ 15000 cost units where the left hand side of the inequality is the cost to run each rocket at speed X_i for i=1, 2, 3, . . . 20. Reviewing once again the engineers' side conditions are that the first three constraints represent a desire to run experiments where rocket four's speed is at least 20% faster than rocket one's, rocket fifteen's speed is at least 30% faster than rocket fourteen's and rocket one is to cost less to run than rocket two.

A multi stage Monte Carlo optimization (MSMCO) simulation produced a solution by stage six of the simulation (each stage drawing 25,000 sample answers) of

$X_1 = 11.85900$	$X_2 = 21.56561$
$X_3 = 15.75404$	$X_4 = 30.72025$
$X_5 = 10.06010$	$X_6 = 11.01783$
$X_7 = 3.24846$	$X_8 = 9.22081$
$X_9 = 39.76670$	$X_{10} = 1.33916$
$X_{11} = 9.89779$	$X_{12} = 18.99355$
$X_{13} = 11.62961$	$X_{14} = 12.72745$
$X_{15} = 22.85686$	$X_{16} = 31.45837$
$X_{17} = 7.56782$	$X_{18} = 14.65238$
$X_{19} = 12.49794$	$X_{20} = 5.43729$

(where these speeds (or velocities) are in hundreds of meters per second) with a total stage 6 error of 0.00000 to five decimal places. Note that the MSMCO approach draws repeated "random" samples of feasible solutions in an ever narrowing and repositioning search following a trail of better and better answers until it arrives at the exact optimal or a useful approximate solution.

AN ENGINE TUNING EXAMPLE

Most machines need adjustment or tuning. So presented here is a problem where repeated tests have been run on the stage three engine of one of the asteroid defense rockets. There are twelve control dials for the engine and each can be set for 1 through 100. After dozens of test runs the following thrust equation was fitted to the data:

$$\begin{aligned} Y = 966,588 &- X_1^2 + 146X_1 - X_2^2 + 128X_2 \\ &-X_3^2 + 42X_3 + .1X_3X_5X_8 - X_4^2 + 76X_4 \\ &-X_5^2 + 18X_5 - X_6^2 + 76X_6 - X_7^2 + 114X_7 \\ &-X_8^2 + 46X_8 - X_9^2 + 78X_9 - X_{10}^2 + 126X_{10} \\ &- X_{11}^2 + 180X_{11} - X_{12}^2 + 94X_{12} \end{aligned}$$

We seek to maximize the thrust of engine three subject to $1 \le X_i \le 100$ I=1, 2, 3, . . . 12 twelve dial settings and X1 + X2 + X3 + X4 + X5 + X6 ≤ 250 for safety reasons to prevent a possible engine overheating or exploding. Thirty "solutions" were produced quickly with the MSMCO approach (in a 16 stage maximization simulation) and three of the best are in Table 1.

Table 1: Three Engine Solutions

STAGE 16	1019716.188			
	2	1	97	7
	78	1	1	77
	1	1	43	1
Stage 16	1022048.625			
	2	1	84	1
	87	1	6	92
	1	2	1	3

Stage 16	1024285.688 = Maximum		
$X_1 = 2$	$X_2 = 11$	$X_3=83$	$X_4 = 2$
$X_5=87$	$X_6=1$	$X_7=4$	$X_8=87$
$X_9=2$	$X_{10}=25$	$X_{11}=14$	$X_{12}=7$

Therefore, as in an example, if the aerospace engineers at the engine factory set control dial 1 at 2 and control dial 2 at 11 and control dial 3 at 83 and so on, to control dial 12 at 7, the engine should produce about 1,024,285.688 pounds of thrust.

AN INTERNATIONAL MINIMUM CONTENT PROBLEM

A large asteroid is a threat to the earth at some point in the future. Therefore, the five countries who have joined their space programs in an international effort to ram this asteroid (to knock it off line) have decided on a 600 rocket mass attack. Each country has two rocket suppliers. So the management team has decided to purchase at least 10, but no more than 200, of each type of rocket to keep all the suppliers on board for future missions too. Also, at least 60 rockets in total will be purchased from each country's suppliers. Given the following estimated rocket type success patterns and their costs (with discounts for buying in quantity) how many of each of the ten rockets should be purchased for the asteroid diversion attack so that it is quite certain at least 600 rockets will hit the asteroid and yet the total rocket cost will be minimized.

Table 2: Rocket Success Rate and Cost (in millions)

	Rocket i	Estimated Success Rate	Cost Per Rocket
In General	0	.97-.02i	$(300\text{-}10i)\text{-}.5X_i$
Country A	1	.97-.02	$(300\text{-}10)\text{-}.5X_i$
	2	.97-.04	$(300\text{-}20)\text{-}.5X_i$
Country B	3	.97-.06	$(300\text{-}30)\text{-}.5X_i$
	4	.97-.08	$(300\text{-}40)\text{-}.5X_i$
Country C	5	.97-.10	$(300\text{-}50)\text{-}.5X_i$
	6	.97-.12	$(300\text{-}60)\text{-}.6X_i$
Country D	7	.97-.14	$(300\text{-}70)\text{-}.5X_i$
	8	.97-.16	$(300\text{-}80)\text{-}.5X_i$
Country E	9	.97-.18	$(300\text{-}90)\text{-}.5X_i$
	10	.97-.20	$(300\text{-}100)\text{-}.5X_i$

These success rates (risks) and cost discounts are acceptable because these are unmanned rockets with big empty cargo bays (less sophisticated except for the guidance systems).

An eight stage MSMCO minimization simulation subject to the constraints produced a solution of

	Supplier 1	Supplier 2
Country 1	$X_1=10$	$X_2=50$
Country 2	$X_3=10$	$X_4=50$
Country 3	$X_5=50$	$X_6=10$
Country 4	$X_7=10$	$X_8=147$
Country 5	$X_9=199$	$X_{10}=200$

with a minimum cost of about 109.375 billion dollars. Given the estimated success rates of the ten different rocket types, about 600 of these more than 700 rockets should hit the target (of the dangerous asteroid).

The international management committee then reran the simulation dropping the at least 60 per country constraint, because some countries are funding larger shares of this project than others. The printout revealed a solution of about 100.992 billion dollars from the following X values:

Country	One	Two
1	$X_1=13$	$X_2=17$
2	$X_3=44$	$X_4=10$
3	$X_5=11$	$X_6=16$
4	$X_7=32$	$X_8=199$
5	$X_9=200$	$X_{10}=200$

The committee is currently meeting in executive session to determine the political and economic advantages of the two optimal solutions.

SOLAR ACTIVITY PATTERNS AND ICE AGE CYCLES

Researchers (Friis-Christensen, 1993) and (Lomborg, 2004) claim that there is a high correlation between sunspot cycles and the earth's average temperature. The correlation appears to go back more than 400 years. If this is true, various types of mitigation of increased temperature (cloud whitening, planting more trees, green spaces to cool our overheated cities) might be worth pursuing.

Additionally, two geophysicists (Ryan and Pitman, 2000) present what can only be described as overwhelming scientific evidence that as a major ice age was ending about 7600 years ago, the melting and receding glaciers caused a significant rise in the ocean levels that cut a channel through what is now the Dardanelles, the Sea of Marmara, and the Bosporus (in present day Turkey) and flooded the Black Sea which was then a fresh water lake (about 100 meters below sea level). The key point is this catastrophic rise in ocean levels that flooded a civilization was not caused by greenhouse gases and modern industrialization. Therefore, mitigation strategies (building dykes and barriers) to hold back and reclaim the sea as the engineers in Venice and The Netherlands have done rather brilliantly might be important in the future. None of these facts mean that pollution is not a problem (Conley, 2008) today.

MANAGEMENT OF BOTH GLOBAL WARMING PROBLEMS

All scientific evidence accumulated over centuries indicates that the earth moves in and out of ice ages (Ryan and Pitman, 2000). Also, (Lomborg, 2004), (Calder, 1997), (Friis-Christensen and Lassen, 1991) believe that more of the recent warming of the earth's average temperature is because of a correlation with increased sunspot activity and not as much the increased carbon emissions and greenhouse gases.

However, many scientists have different viewpoints and one only needs to look at the oil well explosion in the Gulf of Mexico in 2010 and the billions of automobiles and vehicles and thousands of factories producing greenhouse gases to realize pollution (i.e. global warming by any other name) has to be dealt with (Conley, 2008).

The same is true about what to do about the dangerous asteroids. Therefore, the smart approach is to have expert managers get all of the scientific ideas relating to these problems "on the table" along with all possible mitigation strategies. Then rationally decide the best approaches to these long term problems without spending too much money and hurting the world's economies.

Additionally, management should create a "shopping list" of what it needs from industry now and in the future to solve these problems. This list could be beneficial to industry and of course vital for the world.

MULTI STAGE MONTE CARLO OPTIMIZATION

The Hungarian mathematician, John Von Neumann, first developed and popularized the regular Monte Carlo optimization technique, which consists of using a computer to randomly search the feasible solution space for the optimal solution (or a useful approximation) to difficult multivariate problems. He used the term Monte Carlo, because being a European, he associated "random searches" with the famous and glamorous gambling casino in Monte Carlo. Also, the name highlights the risk factor of not being lucky enough in the computer search to find the optimal solution.

Von Neumann was much admired for his work in both theoretical and applied computer mathematics. However, many mathematicians did point out that due to the enormous increase in the number of feasible solutions as the number of variables in an optimization problem increases, Monte Carlo optimization would probably be limited to about five or ten variables.

However, (Conley, 2008) views Von Neumann's Monte Carlo optimization "random search" as stage one of the simulation. Then in stage two, centered about the best answer so far, a second stage searches a slightly reduced region looking for better and better answers. This is followed by a stage three search, in a further reduced region, centered about the best answer so far produced in stage two. This process is repeated in one compact computer program for as many stages as necessary to find the true optimal solution or a useful approximate solution to the complex optimization problem at hand. Therefore, the name multi stage Monte Carlo optimization (MSMCO) was chosen to describe this multivariate nonlinear optimization technique. Pages 2 and 5 of (Conley, 2008) show drawings of these stages funneling into the optimal solution region.

CONCLUSION

Three multivariate optimization problems were presented along the lines of protecting the earth from explosions of dangerous asteroids, meteors, comets, etc. in our atmosphere that could cause catastrophic fires (global warming) endangering many hundreds of thousands of lives as Anatoly Priminov, head of Russia's space agency, said in (Isachenkov, 2009). Other global warming problems such as "greenhouse gases," sunspot cycles and our earth moving in an out of ice ages were also mentioned. The engineers in Venice and The Netherlands have put up dykes and barriers to protect their people from rising seas in low lying areas. This engineering technology is important for the future. Many ideas have been advanced here and in other research for controlling and/or eliminating these problems. However, because these problems require long term planning and solutions, this is definitely an international management problem where all ideas (scientific, industrial, economic and political) should be discussed and possibly short term and long term strategies for dealing with these problems must be addressed.

This is the first full century where these fields have advanced to the point that the managers can perhaps mitigate or solve these problems.

REFERENCES

Calder, N. The Manic Sun. Yelvertoft Manor England. Pilkington Press, 1997.

Conley, W.C. "We Can Save our World if Worlds Collide." Industrial Simulation Conference, ISC, 2010, Budapest, June 9-12, EUROSIS-ETI, Ghent, Belgium, pp. 45-49.

Conley, W. C. "Ecological Optimization of Pollution Control Equipment and Planning from a Simulation Perspective." International Journal of Systems Science, Vol. 1, 2008, pages 1-7.

Friis-Christensen, E., Lassen, K. "Length of Solar Cycle: an indicator of solar activity closely associated with climate." Science, Vol. 254, 1991, pages 698-700.

Friis-Christensen, E. "Solar Activity Variations and Global Temperature." Energy, Vol, 18, No. 12, 1993, pp. 273-284.

Isachenkov, V. "Russia May Send Spacecraft to Knock Away Asteroid." Associated Press, Dec. 30, 2009.

Lomborg, B. The Skeptical Environmentalist. Cambridge University Press, Cambridge, England, 2004.

Maugh, T. H. "Aftermath of Asteroid Collision Examined." Los Angeles Times, Feb.5, 2010.

Ryan, W., Pitman, W. Noah's Flood: The New Scientific Discoveries About the Event that Changed History. Touchstone Book, Simon & Schuster, New York, 2000.

Stone, R. "Target Earth." National Geographic, Washington, D.C., August 2008, pp. 134-149.

AUTHOR BIOGRAPHY

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

ECOMONFOR – A SYSTEM FOR GREENHOUSES MONITORING AND FORECASTING

Janetta Culita, Dan Stefanoiu, Alexandru Dumitrascu

„Politehnica" University of Bucharest, ROMANIA
Dept. of Automatic Control and Computer Science (www.acs.pub.ro)
E-mails: jculita@yahoo.com, danny@indinf.pub.ro, dumalex77@yahoo.com

KEYWORDS
Greenhouse, remote monitoring, distributed time series, data (pre)filtering, forecasting.

ABSTRACT
The paper describes ECOMONFOR – a monitoring and forecasting system for ecology, which consists of two sub-systems: a fixed/immobile unit, ECOMONFOR-F (that resides inside a laboratory) and a mobile unit, ECOMONFOR-M (that allows in place exploitation, outside the laboratory). ECOMONFOR was effectively integrated in an industrial application aiming real-time monitoring and control of a small greenhouse. The process yields a collection of ecological parameters that are measured by means of some wireless network of sensors and interpreted as distributed time series. Before modeling and prediction of the ecological signals, the acquired data are first organized into blocks and then preliminary processed. The forecasting results with real data are presented in the end.

1. INTRODUCTION

The environment, seen as an ecological system, constitutes a rich source of signals that encode its evolution over the time. Though, only a reduced set of them are perceived by humans and even less are correctly decoded and understood. Rapid climate changes and the negative impact of industry upon the environment require designing and employing of automatic monitoring systems over geographical areas. The general purpose of monitoring is to forecast the behavior of the ecological system, in view of life quality improvement and/or disaster anticipation (even avoidance), if possible, as soon as possible.

Ecological phenomena could be noticed either in an open or an enclosed space. There are several interesting ecological phenomena observed in an open geographical area, such as soil erosion caused by rain, the effect of the soil and water pollution, the temperature/humidity transfer between air, soil and plants. Correlations such as temperature variation and humidity transfer have to be accounted when predicting such phenomena. In a microclimate like a greenhouse, ambient temperature, humidity, dew point and solar radiation are quite correlated. On the contrary, the soil parameters are less correlated to each other, but could be correlated to ambient parameters.

The paper mainly presents an ecological monitoring and forecasting system, namely ECOMONFOR, which allows monitoring and forecasting of multi-variable ecological signals, either in small or large geographical regions. ECOMONFOR was successfully integrated in a novel application to remote monitoring and control of a small greenhouse (Dumitrascu 2010). Basically, the application aims to realize automatic watering of plants, in order to ensure suitable growth and comfort of plants. The distributed monitoring and control architecture of the ecological process interconnects several functional modules, such as: the 3-level wireless acquisition and monitoring system (which is user-friendly through interfaces like eKo-View, eko-Greenhouse or eko-Forecast) (Culita and Stefanoiu 2010); the automation control system made of PLCs and industrial communication networks; the irrigation system (consisting of two water tanks, sensors and actuators).

The article is not approaching the design of automation solution. Its main goal is to describe how the acquired data (also employed in greenhouse automatic control) enable high quality prediction to be displayed to the user. In our approach, the ecological signal prediction relies on numerical models that were previously implemented as FORWAVER, PARMA, PARMAX, KARMA predictors (Stefanoiu et al. 2008; Stefanoiu and Culita 2010). One expects that the forecasting experimental results be quite accurate, especially for those ecological data provided by the greenhouse, which are correlated to each other.

The paper is structured as follows. Section 2 introduces the distributed architecture for monitoring and control of the greenhouse. Section 3 presents the acquisition and preliminary processing of the ecological parameters provided by the greenhouse. Some improvements made to prediction quality criteria and the performances of predictors are indicated within Section 4. A conclusion and the references list complete the article.

2. MONITORING AND CONTROL SYSTEM ARCHITECTURE OF THE GREENHOUSE

The greenhouse consists of six plants, which are located in two separated laboratory rooms, in order to create different microclimates. The ambient disturbances are mainly caused by the registered traffic through both rooms (due to students and teachers movement), which generates draughts of different temperatures, direction and speed. There also are secondary sources of disturbances, which are engendered by the external environment (especially solar radiation) and natural heating/cooling of both areas.

Since the plants in our small greenhouse were long time only sporadically watered, their improper care required construction an automatic irrigation system.

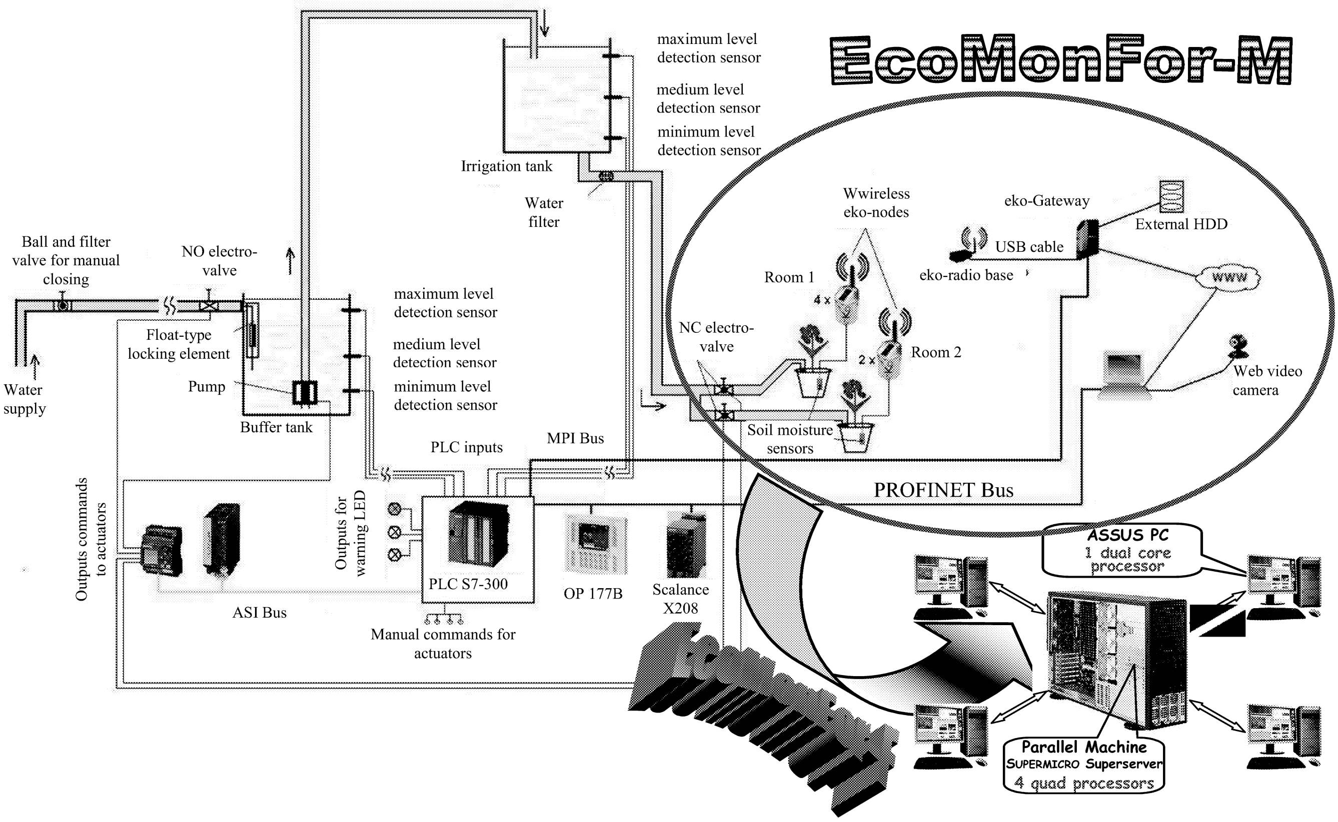


Figure 1. The complete architecture of the small greenhouse control system including *ECOMONFOR*.

Figure 1 depicts the distributed monitoring and control architecture of the greenhouse, which integrates: the automatic control system of irrigation (left side down), the irrigation system (left side up) and, mostly concerned, the ecological monitoring system ECOMONFOR (right side).

Constructively, ECOMONFOR was separated in two components: mobile, ECOMONFOR-M (circled in red) and immobile, ECOMONFOR-F (located beneath). The mobile monitoring unit is structured on three hierarchical levels as shown in the right upper side of figure 1: the set of wireless *eko-sensors*; the central (kernel) equipment of sensors network, referred to as *eko-Gateway* and a mobile computer (laptop or notebook). The last two of them are wirelessly connected to Internet, in order to enable running remote applications. Moreover, the computer fulfills the function of real-time video supervision of the whole system through some small webcam. ECOMONFOR-M is responsible with remote data acquisition and monitoring, which means it could cover an extended geographical area. It can be employed for a quick prediction of measured data, as well. The data collection supplied by the eko-Gateway is directed to ECOMONFOR-F with the aim of high quality prediction of the ecological phenomena. This transfer is suggested by the large curved arrow in the bottom of image. The core of the immobile unit consists of a parallel machine with 16 processors. This is connected via internet to an extensible computer network. The machine is hosting the group of complex algorithms for modeling, identification and forecasting of distributed ecological signals. The algorithms are: PARMA, PARMAX, KARMA and FORWAVER.

Both components of ECOMONFOR are working on the following strategy. First, the acquisition and the preliminary processing of data are accomplished. Sometimes, data provided by sensors are damaged and need to be enhanced. Some fixing operations are necessary to improve data, as shown within the next section. Visual monitoring of the greenhouse stands for the second step, which is executed in parallel with the acquisition, through the eko-Gateway. Two user friendly interfaces are available via Internet. The first one is *eko-View*, an intuitive web browser based interface that gives the user the opportunity to set and display the configuration of sensors network and thus to start monitoring and acquisition, from anywhere in the world. Moreover, several supplementary facilities are offered, in order to facilitate data handling (such as: display of graphical variations, export to the most employed programming environments, set the alerting rules etc.). An example of sensors network configuration, as displayed by eko-View, is given within the next section. The second interface is *eko-Greenhouse*, from figure 2. This is more automation oriented. Thus, its role is to help the user to directly and remotely interact with the greenhouse, via internet, by accessing the process parameters and controlling the automatic irrigation system. Technically, the main panel is based on Apache-type http protected server (users require passwords). It was built using common Web technologies: HTML, JavaScript, XML PHP. The interface is composed of two main parts: a visible one that contains graphical elements, buttons and data fields; an invisible (but active) one, which is transparent to the user, that bonds the panel to the control system (the PLC).

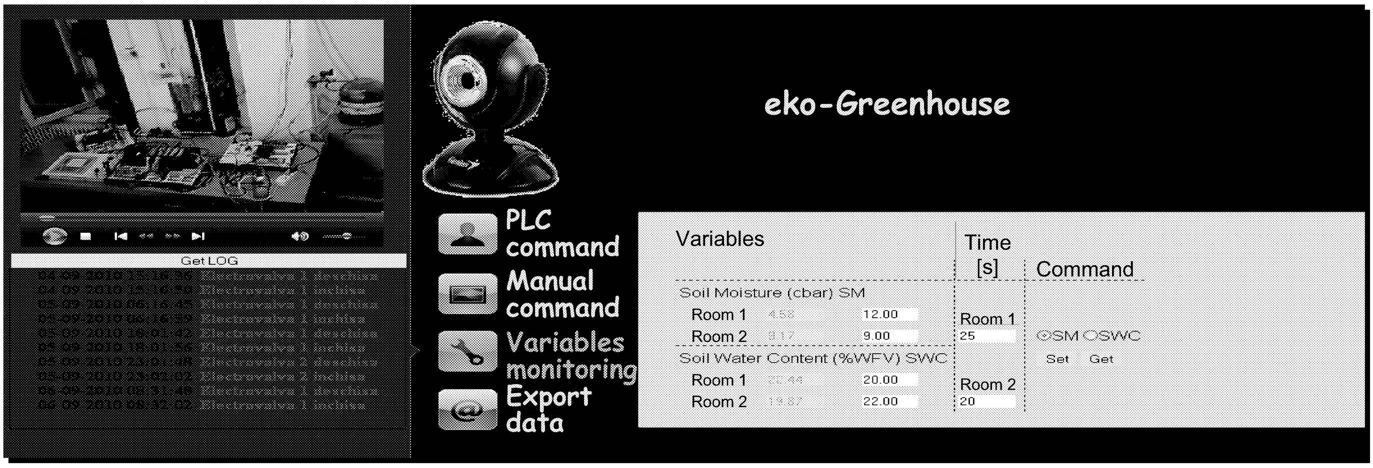


Figure 2. The web interface eko-Greenhouse, yielding the remote control.

The interface configuration displays four interesting zones. On the left side above, the visual image of the process is permanently offered through a webcam. Beneath, the results of the last 10 commands sent to actuators are completely shown. In the middle, four selection buttons are depicted. Thus, the user can: send programming commands to the control device PLC S7-300; send commands directly to actuators, in order to perform irrigation manually; display and/or set the two ecological parameters, which are employed as process outputs (soil moisture and soil water content, as figure 2 shows on the right side); export data from specific (but non standard) eko-Gateway format in a different, more comprehensible and useful format, by saving them on some external disk, for subsequent processing.

The final step of the operating strategy in ECOMONFOR system consists of data modelling, on prediction purpose, which actually is the main goal of monitoring in this context. Another user friendly graphical interface, a third one, namely the *eko-Forecast* interface, was implemented, in order to complete a forecasting experiment (as described in (Culita and Stefanoiu 2010)). It facilitates running PARMA, PARMAX, KARMA and FORWAVER predictors within FORTIS (*FORecasting of TIme Series*) simulator. The interface allows the user to initiate prediction and offers a graphical illustration of the forecasting results. Although all predictors can proceed on both units of ECOMONFOR, the faster predictors (PARMA and FORWAVER) are commonly hosted by the mobile unit, while the slower ones (PARMAX and KARMA) are usually executed on the immobile unit.

The kernel of the control subsystem resides in programmable automata CPU315F-2DP/PN of the S7-300 (which is a trademark of Siemens), in charge with data processing after being received from eko-Gateway, via MPI communication network. The human-machine interface OP 177B has two roles: to permanently display the ecological parameters (that are acquired and transmitted to automata and computers); to allow the user to set different alarm values on soil moisture parameters or irrigation intervals. The control devices S7-300 and OP 177B are connected via PROFINET bus. The effective control is achieved through the PLC LOGO!, which receives commands from the S7-300 on AS-I bus and send commands to the irrigation system components (electro-valves and pump).

For the irrigation process, two tanks of water are used, as shown in figure 1. The first one is a buffer tank (at the bottom), being directly fed from the main water supply. The second one (at the top) is employed in plants irrigation. Before the first water tank, a normal-open (NO) electro-valve was installed to interrupt the general water supply in case of emergency. The buffer tank contains some simple automatic elements, such as: a float switch to start/stop water supply, three sensors for detecting different water levels and a mini-submersible pump to fill the second tank. The irrigation of plants is performed by freefall, since the second tank is located about 3m above the plants.

In this landscape, ECOMONFOR represents an additional part of the irrigation application. On one hand, it decides the irrigation commands, indirectly, through the sensors network. On the other hand, it processes the measured data, in order to forecast them.

3. DATA ACQUISITION AND PRELIMINARY PROCESSING

As mentioned before, the greenhouse contains 6 plants placed in two different rooms. Each plant was allocated to one wireless node for acquisition and monitoring purpose. The monitoring can be carried out by using eko-View and eko-Greenhouse interfaces. Every node is capable of transmitting data from at most 4 eko-sensors, whilst each eko-sensor can measure 1 to 3 ecological parameters at the same time, but on different channels. For example, one eko-sensor can measure soil temperature and humidity; another eko-sensor provides ambient humidity, temperature and dew point; leaf wetness and solar radiation are however acquired through single, different sensors.

Figure 3 illustrates the greenhouse plants and the monitored ecological parameters for each one of them. The used acronyms are explained in Table 1, which also indicates their varying ranges and measurement units.

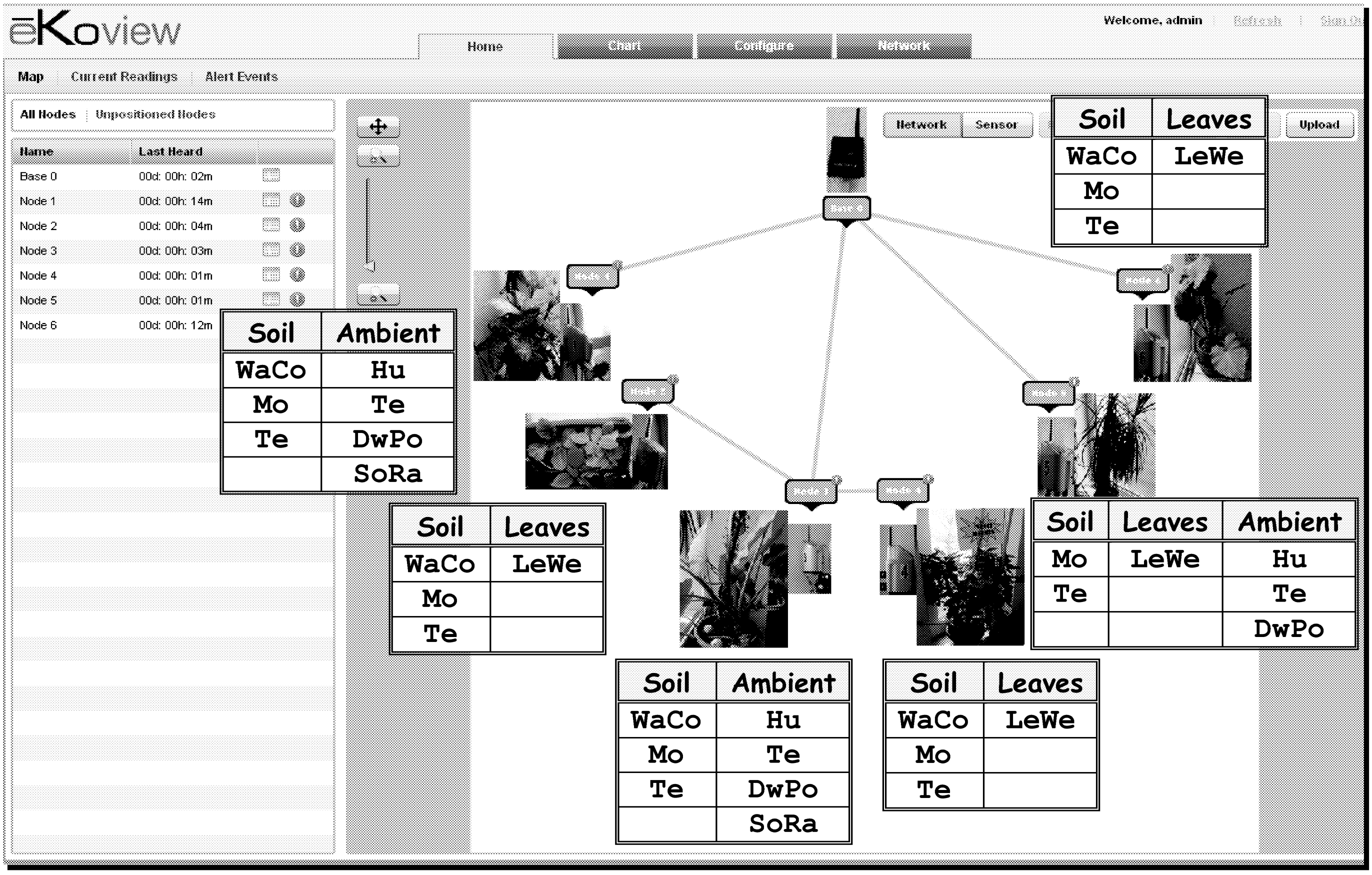


Figure 3. Synoptic map of the monitored ecological parameters inside the greenhouse.

Table 1. Ecological parameters of eko-sensors network

Soil	Leaves	Ambient
Moisture (Mo) **0 ... 240 [cbar]**	**Leaf Wetness (LeWe)** **0 ... 1024 [CntS]**	**Humidity (Hu)** **0 ... 100 [%]**
Temperature (Te) **–40 ... +65 [°C]**		**Temperature (Te)** **–40 ... +65 [°C]**
Water Content (WaCo) **0 ... 100 [%wfv]**		**Dew Point (DwPo)** **–10 ... 50 [°C]**
		Solar Radiation (SoRa) **0 ... 1800 [W/m^2]**

Figure 4 is actually built on the basis of eko-View interface, also showing eko-nodes states, connections with radio base and many other wireless network parameters that we will not describe furthermore in this paper. In our application, 21 sensors are acquiring data from 33 ecological variables all over the greenhouse.

Understanding the past and the future evolution of the ecological phenomena becomes difficult especially when the measured data are disturbed by various stochastic signals. Therefore, it is totally suitable for such data to be mathematically modeled and processed by using algorithms able to extract useful information from noise. The algorithms implemented within ECOMONFOR system are of this kind. Forecasting some ecological parameters of the greenhouse and revealing their correlations are major goals of monitoring. In order to send data to FORTIS interface, in view of prediction, the parameter values (of the same node) have to be grouped in data blocks, according to their possible correlations. For example, humidity is correlated to temperature which, in its turn, is correlated to solar radiation. It is rather difficult to presume that the soil parameters coming from different plants are correlated each other, taking into account that the plants are located differently. Instead, it seems that the ambient parameters are strongly correlated, as the greenhouse plants are positioned in rather small closed rooms. Each block corresponds to a node and contains data from 3-4 acquisition channels. It is named after an identification code including: node identity (1-6), parameter type (soil or ambient), acronyms of the measured parameters (e.g. `N3_Soil_WaCo.Mo.Te`).

Eko-sensors usually provide unsynchronized or faulty data. Therefore, some preliminary operations for improving their quality are necessary. A simple and intuitive method of obtaining synchronized data is the hourly averaging technique. Also, frequently, there could be missing samples (gaps) in data strings, on different acquisition channels, at some instants. For example, the top variation of Figure 4 (leaf wetness) exhibits important gaps. In this case, the interpolation followed by re-sampling is necessary to recover missing data. First, for isolated gaps, linear interpolation is enough, as it can be noticed from the next variation of figure 4, in the middle. Next, for consistent gaps (with more than one missing sample), autoregressive (AR) interpolation seems to be quite adequate, as shown by the bottom variation of figure 4. The interpolated values were estimated by forward and backward prediction with AR models identified by applying Levinson-Durbin Algorithm (Soderstrom and Stoica 1989).

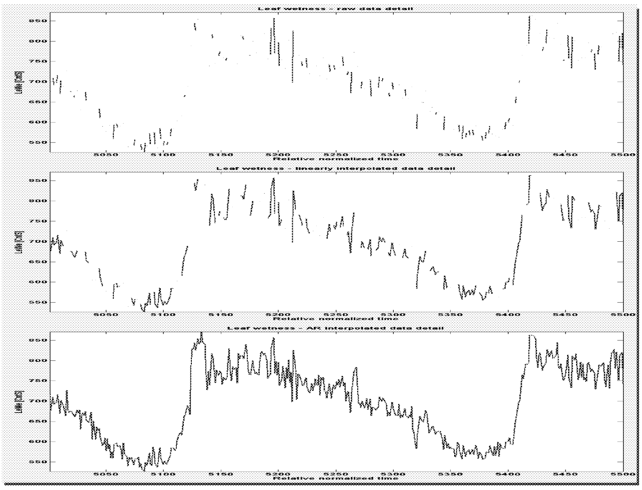


Figure 4. Leaf wetness: raw data with gaps (top); data with linear interpolation of isolated gaps (middle); completely recovered data after AR interpolation (bottom).

Another danger is over-sampling of parameters, which means gathering much more samples than necessary. This can be avoided by averaging, applied to larger duration than the sampling period. In our case, data were averaged over 3-4 hours, since the evolution of ecological phenomena is rather slow. Because of interpolation described above and due to some temporary malfunctions in eko-sensors network, data may contain important discrepancies (deviations) on short time intervals. These outstanding deviations are in general attenuated by numerical low-pass filtering. One of the best filters that can be used in this aim is of Cebyshev class and second type, thanks to its sharp cut-off frequency, together with large stop band attenuation (Proakis and Manolakis 1996). Especially in case of ecological parameters, this filter was also applied in order to refine the delimitation between the deterministic and stochastic components of prediction models, which tremendously affects models performance.

4. EXPERIMENTAL RESULTS

The automatic irrigation application intended to improve the comfort and healthy state of the plants in the greenhouse, by avoiding inappropriate watering. The interesting monitored parameters are the soil moisture (Mo) and the soil water content (WaCo). The soil moisture is inversely proportional to the soil pressure, which actually the eko-sensor provides. Thus, it is sometimes referred to as *soil pressure-moisture* (the measuring unit being the cbar). A value near 100 cbar means a very dry soil, while a value near to 0 cbar signifies a very wet soil. The automatic irrigation system became functional on the 18th of August 2010. Figure 5 (obtained through the eko-View interface) depicts the evolution of the soil pressure-moisture parameters before and after this date. Obviously, before implementing the automatic irrigation system, this parameter recorded irregular variations, amplified at the moments of manual watering. The high amplitude values indicate the necessity of watering plants. The manual watering moments are denoted by the abrupt changes in all 6 variations. After installing and starting the automatic irrigation system, the soil pressure-moisture highlighted smaller and regular variations (up to 40 cbar).

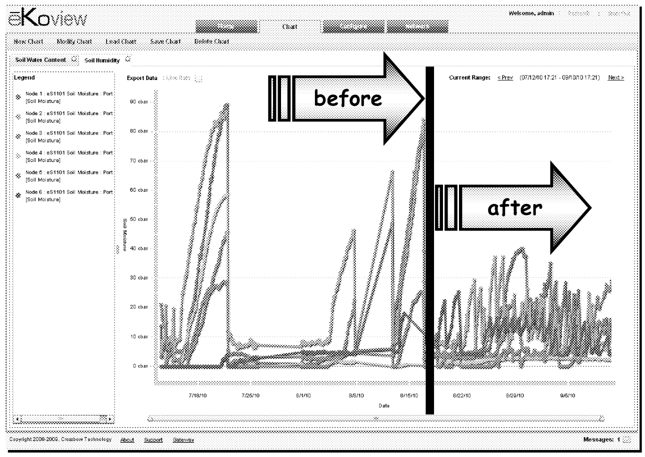


Figure 5. Soil moisture before and after the automatic irrigation system started to run.

Concerning the forecasting of greenhouse ecological system, the 4 predictors (PARMA, PARMAX, KARMA and FORWAVER) have been tested on 30 data blocks resulting from combinations of soil or ambient parameters in the synoptic map (figure 3). In order to reduce the simulation time, especially in case of PARMAX (which is the most time consuming), the ECOMONFOR-F computer network was extended to 16 PCs; the laptop of ECOMONFOR-M unit has been used as well. As already mentioned, the ecological phenomena are usually slow. Therefore, predicted values are estimated after every 8-12 hours. The simulation time for predictors varied between several minutes and a number of tens of hours, depending on their complexity, the number of analyzed ecological data and the modeling of stochastic component. Each of the 30 data files is associated to 16 graphics for every acquisition channel, coming from all four predictors. There are 4 variations for a channel, which are bond to a predictor performance: the original time series together with its optimal trend, the estimated white noise on measuring horizon; the predicted values and the most important, the *prediction quality* (PQ) (Stefanoiu and Culita 2010). Each predicted value has some probability of occurrence defined by the trusting tube. As the prediction instant goes away from the measuring horizon, the tube becomes larger and larger. This means the predicted values are less and less reliable.

Defining the PQ criterion is not an easy task, as two different predictors could lead to very close values. In our previous publications, some definitions have been proposed and tested. But none was really satisfactory. In order to increasing the capacity of making distinction between predictors, three PQ definitions could be averaged with corresponding weights.

a. PQ depending on *signal-to-noise ratio* (SNR), estimated for measure and prediction horizons:

$$\mathrm{PQ}_\alpha = \frac{100}{1+\dfrac{1}{\sqrt{\mathrm{SNR}^{Ny}}\sqrt{\mathrm{SNR}^{K}}}}\ [\%]\,, \qquad (1)$$

where $\mathrm{SNR}^{Ny} = \lambda_y^2/\lambda_e^2$ with λ_y^2 – the variance of the Ny-length acquired data and λ_e^2 – the variance of their corresponding residual white noise, whereas

$\mathrm{SNR}^K = \sigma_y^2 / \sigma_{y-\hat{y}}^2$, with σ_y^2 – the variance of data on prediction horizon and $\sigma_{y-\hat{y}}^2$ – the variance of prediction errors.

b. PQ depending on global SNR (measure and prediction horizons, all together):

$$\mathrm{PQ}_\beta = \frac{100}{1 + \dfrac{1}{\sqrt{\mathrm{SNR}^{Ny+K}}}} \; [\%], \qquad (2)$$

where $\mathrm{SNR}^{Ny+K} = \rho_y^2 / \lambda_e^2$, with ρ_y^2 – the variance of data on global horizon.

c. PQ focusing on prediction horizon only:

$$\mathrm{PQ}_\gamma = \frac{100}{1 + \mathcal{X}_{in} + \mathcal{X}_{out}} \; [\%], \qquad (3)$$

where:

$$\mathcal{X}_{in} = \sqrt{\frac{\sum_{k \in \mathcal{P}_{in}} \hat{\sigma}_k^2 \left| y[Ny+k] - \hat{y}[Ny+k] \right|^2}{\left(\sum_{k \in \mathcal{P}_{in}} \hat{\sigma}_k^2 \right) \left(\sum_{k \in \mathcal{P}_{in}} \left| y[Ny+k] - \hat{y}[Ny+k] \right|^2 \right)}}, \qquad (4)$$

$$\mathcal{X}_{out} = \sum_{k \in \mathcal{P}_{out}} \frac{\left| y[Ny+k] - \hat{y}[Ny+k] \right|}{3\hat{\sigma}_k}, \qquad (5)$$

$\mathcal{P}_{in}$ is the set of prediction instants for which the measured data lie inside the trusting tube, $\mathcal{P}_{out}$ is the set of prediction instants for which the measured data lie outside the trusting tube and $6\hat{\sigma}_k$ is the current diameter of tube (for Gaussian disturbances). The terms (4) and (5) are both depending on the distance between acquired data ($y[Ny+k]$) and predicted data ($\hat{y}[Ny+k]$). If the prediction is accurate enough, measured data are close to predicted values and $\mathcal{X}_{in}$ is quite small. Otherwise, $\mathcal{X}_{out}$ becomes non null anytime measured data exceed the trusting tube. As the sum $\mathcal{X}_{in} + \mathcal{X}_{out}$ increases, PQ_γ becomes smaller and smaller. Every time $\mathcal{X}_{out}$ has non null values, PQ_γ decreases dramatically.

Definitions (1) and (2) try to enforce the prediction model to be accurate enough on both prediction and measure horizons. Definition (3) only focuses on the local forecasting performance operation. The PQ criteria are then generated by weighted average of definitions above:

$$\mathrm{PQ} = w_\alpha \, \mathrm{PQ}_\alpha + w_\beta \, \mathrm{PQ}_\beta + w_\gamma \, \mathrm{PQ}_\gamma \qquad (6)$$

where $w_\alpha \geq 0$, $w_\beta \geq 0$ and $w_\gamma \geq 0$, are weights so that $w_\alpha + w_\beta + w_\gamma = 1$. After many forecasting simulations run with different data sets, one concluded that the weights $w_\alpha = 0.45$, $w_\beta = 0.15$ and $w_\gamma = 0.4$ allow quite a fine delimitation between predictors. As mentioned in (Culita and Stefanoiu 2010), PQ is a cost function that can only be evaluated either when selecting the optimal predictor or after performing the prediction, but not during the prediction (data being unknown on prediction horizon).

In order to demonstrate the prediction performance of EcoMonFor, the soil moisture (`Mo`) from all 6 plants has been selected. Figure 6 displays moisture variations over the greenhouse, together with their best detected trends.

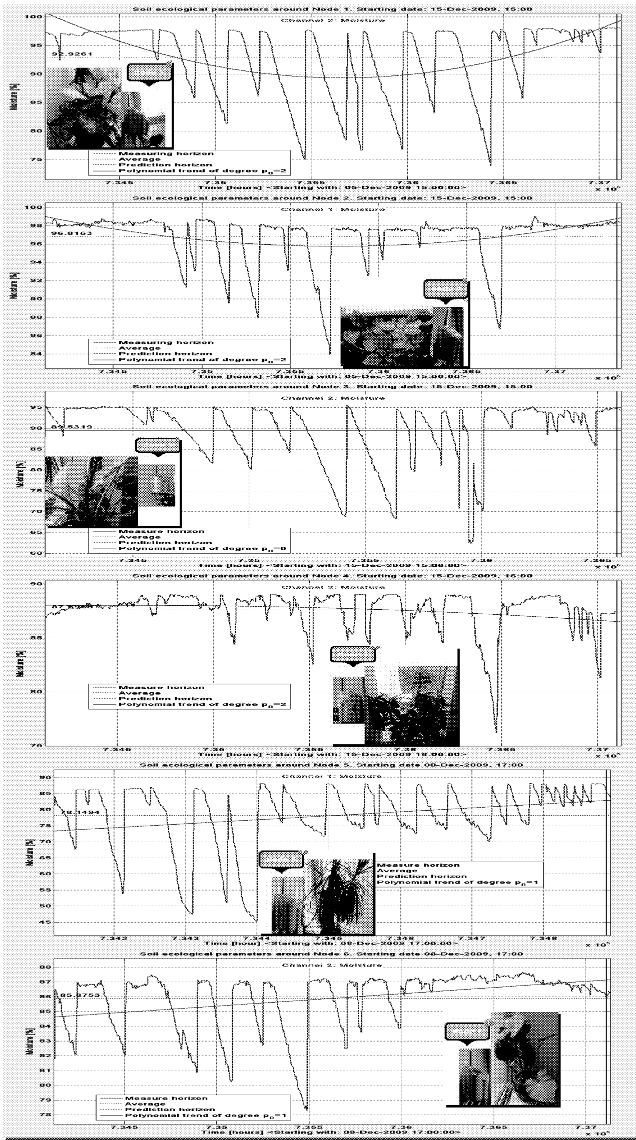


Figure 6. Soil moisture variations within the greenhouse.

Soil moisture follows, but is not identical to soil water content. Therefore, the last one is employed in automatic control of greenhouse, whereas forecasting is concerned with the first one. Figures 7–12 reveal the prediction performance for soil moisture, within each one of the 6 nodes. Best results of the 4 predictors (PARMA, PARMAX, KARMA and FORWAVER) are depicted, together with their corresponding PQ values. The soil parameters that could be correlated to moisture are pointed as well: **`WaCo`** and **`Te`**. Interestingly, the results show that there is quite a strong correlation between soil moisture and leaves wetness, as predicted values are more accurate when considering **`LeWe`**, than without accounting **`LeWe`**.

Although the predicted values are apparently very close to the real data, all variations were scaled in terms of trusting tube diameter (also drawn on all pictures). So, the PQ values may take small values, just because the tube is too wide. As a general result, PARMA is never the best, but the fastest. However, its performance is fair, with a good trade-off between speed and accuracy, which allows assigning this predictor the bronze medal. For the silver medal, FORWAVER is the righteous selection. However, like PARMA, this predictor is not accounting for correlations between parameters. It is therefore not a surprise that PARMAX receives the gold medal, in spite the long duration it needs to provide the predicted values. This time, correlations between **Mo** and the other 2 or 3 parameters helped the predictor to provide the best results in 3 nodes and the second best results in another 2 nodes. A surprise, but a deceiving one, is made by KARMA, which performed much worst than expected (not only for **Mo**, but for the other parameters as well). Excepting for the node 3, where KARMA was the best, the other results are modest. A possible explanation resides in Kalman filter over-sensitivity to the variation of internal states number. Just removing or adding one single state can dramatically modify the predicted values outside as well as inside the measure horizon. The bronze-silver-gold classification is confirmed by all tests, with different greenhouse parameters.

5. CONCLUSION

This article introduced EcoMonFor– a system that is in charge with monitoring, controlling and forecasting of a small greenhouse. This system integrates three user friendly interfaces (eko-View, eko-Greenhouse and eko-Forecast), which are implemented on a mobile or immobile computer. Various facilities are offered to the user, regarding the automatic control of greenhouse, prediction and simulation, starting from the acquired ecological parameters. The overall architecture has been designed in an open philosophy, so that many improvements could be applied. There are two main directions under consideration: extending the automatic control system by more sophisticated regulators and enhancing the collection of prediction programs with new, more accurate ones.

REFERENCES

Dumitrascu A., 2010. *Contributions to Industrial Computer Networks in Process Control*, PhD thesis, "Politehnica" University of Bucharest, Romania.

Culita J., Stefanoiu D., 2010. FORTIS – An Integrated Simulator for Distributed Time Series Forecasting, Proc. of Industrial Simulation Conference ISC-2010, Budapest, Hungary, pp. 27-33.

Proakis J.G., Manolakis D.G., 1996. *Digital Signal Processing. Principles, Algorithms and Applications.*, third edition, Prentice Hall, Upper Saddle River, New Jersey, USA.

Söderström T., Stoica P., 1989. *System Identification*, Prentice Hall, London, U.K.

Stefanoiu D., Culita J., 2010. Multivariable Prediction of Physical Data, Scientific Bulletin of PUB, Series A, Vol. 72, No. 1, pp. 95-102.

Stefanoiu D., Culita J., Ionescu F., 2008. FORWAVER – A Wavelet Based Predictor for Non Stationary Signals, Proc. of Industrial Simulation Conference ISC-2008, Lyon, France, pp. 377-381.

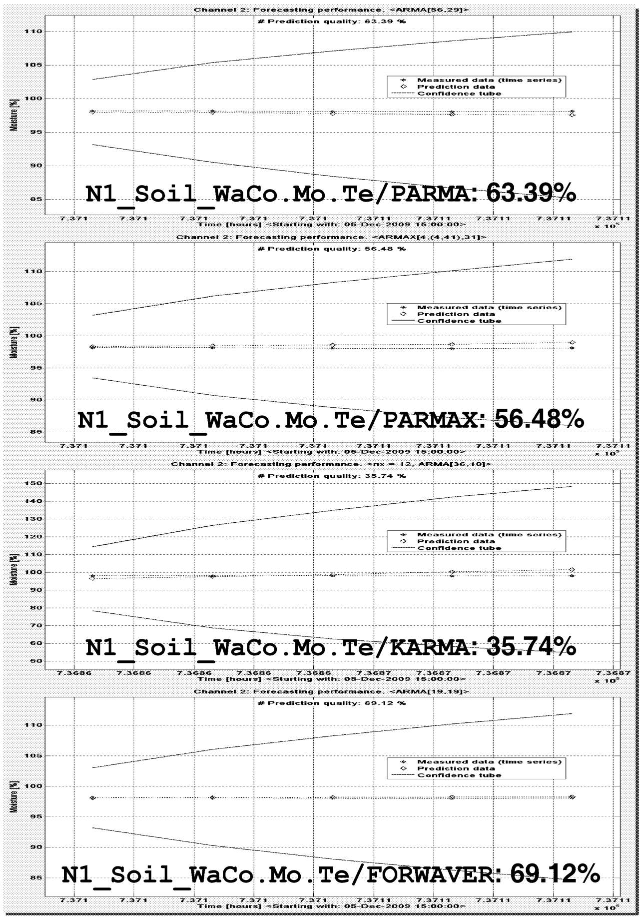


Figure 7. Forecasting performance in node 1 (**Mo**).

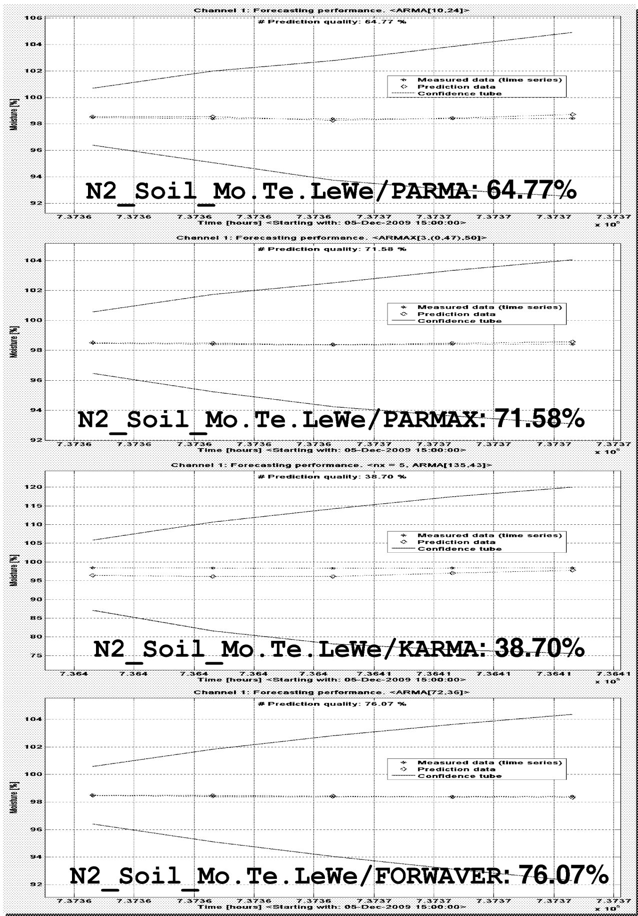


Figure 8. Forecasting performance in node 2 (**Mo**).

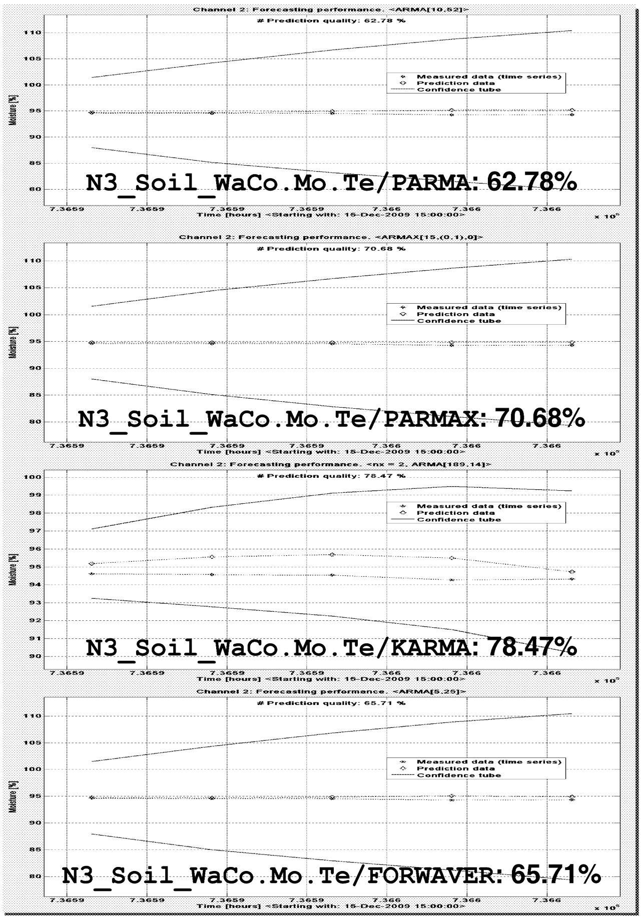


Figure 9. Forecasting performance in node 3 (**Mo**).

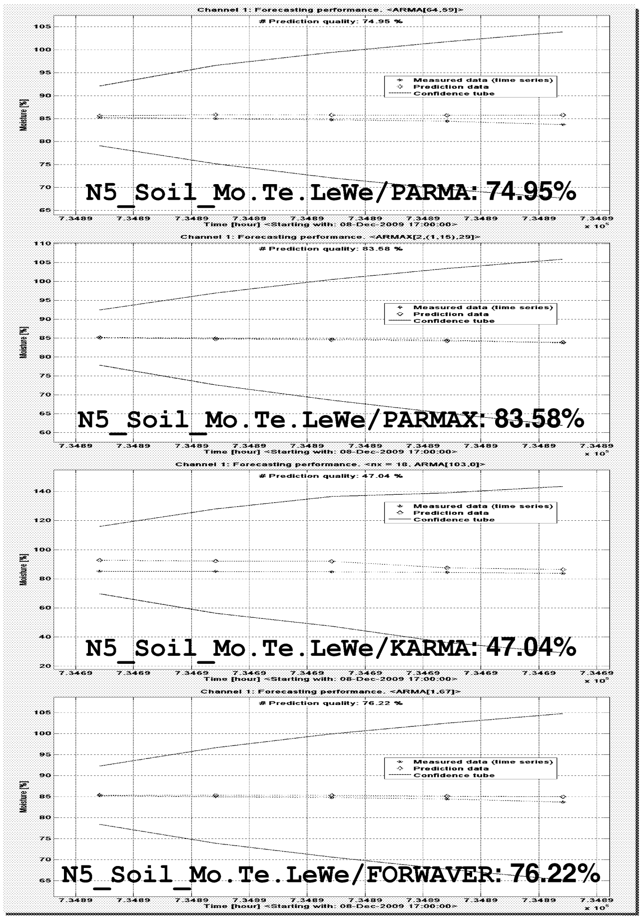


Figure 11. Forecasting performance in node 5 (**Mo**).

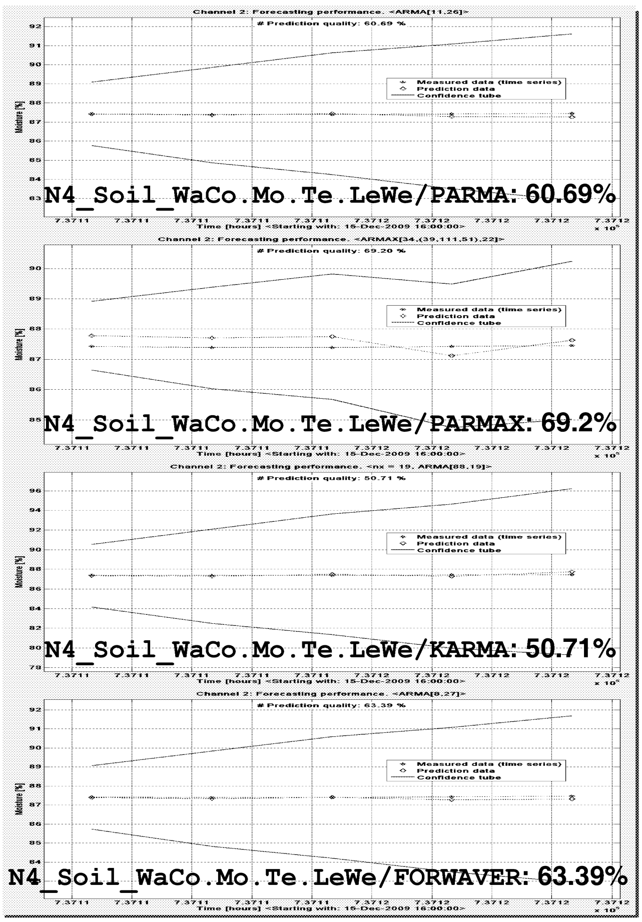


Figure 10. Forecasting performance in node 4 (**Mo**).

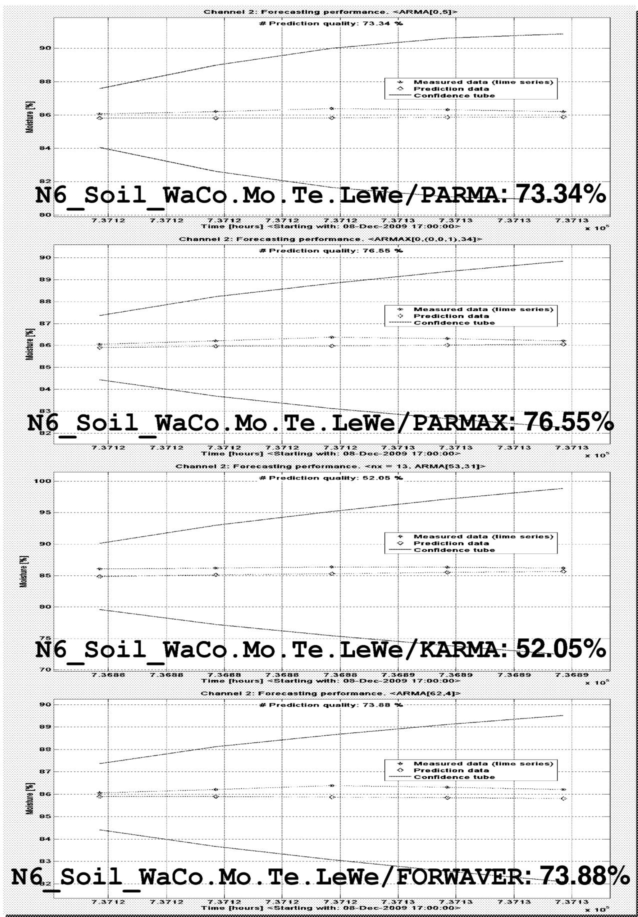


Figure 12. Forecasting performance in node 6 (**Mo**).

LATE PAPERS

TRAINING OF FAULT DIAGNOSIS SKILLS: USE OF HEURISTIC DIAGNOSTIC MODEL AND TECHNICAL SYSTEM SIMULATION

Victor M. Dozortsev and Vladimir A. Nazin
ZAO Honeywell,
Kievskaya str., 7, Moscow, Russia
Tel: +7 495 796 98 00
E-mail: victor.dozortsev@honeywell.com

KEYWORDS

Computer Systems, Decision Support Systems, Qualitative Models, Probabilistic Models, Simulation, Training.

ABSTRACT

The paper presents an upgraded approach to development of technical process operators' effective fault diagnosis skills. Computer-based decision-making support system is described based on diagnostic model which is configured and verified by means of technical process high-fidelity simulation model. This system provides cognitive similarity and motivational conformity of operator training with the real diagnostic activity. Experimental justifications of the approach and practical results of operators training by means of presented system are discussed. Some lines for future developments of the approach are proposed.

INTRODUCTION

Historically, the probabilistic approach to problem solving has been applied in two ways, as a tool for normative strategies statement, and as a way of describing the human activity in problem solving. Quite picturesque in this sense is name of Jacob Bernoulli's classic work *The Art of the Conjecture* (1713) which states the basics of probability theory *Art* along with the recipes of its practical application. At the same time, the close look at the decision-making mechanisms reveals serious gap between the strategies used in practice and the normative strategies, due to computational complexity of the latter. Apparently, the human thinks in some probabilistic terms and characteristics of the problem area during decision-making, but performs the calculations in some heuristic way. In the broad sense the interrelation of probabilistic and cognitive aspects of human decision-making is analyzed in (Brighton and Gigerenzer 2008). It is important to understand this mechanism both to describe the activity adequately and to build effective decision-making training systems.

The current work presents the research in the area of training *Technical Systems (TS)* operators *Fault Detection and Isolation (FDI, a.k.a. Diagnostic)* skill. The high importance of this particular task is emphasized by the following reasons. First of all, the diagnostic problem is essentially complicated since it requires deep operators' knowledge of the *Technical System* and the ability of swift manipulation of large amount of information about cause-and-effect relations between irregular TS events (*Faults*) and changes of dozens or even hundreds of TS parameters (*Symptoms*). These relations generally have a dynamical stochastic behavior and the symptoms manifestations are corrupted by noise. Apart from that, in the contemporary process environment the diagnostic decision must be performed quickly, and the operator must possess simple, tolerant and effective diagnostic algorithms applied to a wide field of even a priori unknown TS faults. Finally, the price to pay for poor diagnostics is substantial. For instance, by the assessments of (Duguid 2001) about a half of accidents in hydrocarbon industry happen due to operators mistakes. The latter is most often the effect of poor diagnostics.

This explains the increasing interest of research in the diagnostics mechanisms used by operators and in constructing training procedures allowing for these mechanisms.

In the paper a brief outline of current state of the area is given followed by the statement and justification of a new approach to training diagnostics skills. The approach is based on: a) cognitive modeling of diagnostic skill formation reflecting the structure of real diagnostic activity; b) development of diagnostic skills training system allowing for stochastic cause-and-effect relations and heuristic nature of operators decisions, determined by confinements of operator's cognitive abilities; c) motivational aspect of diagnostic activity; d) implementation of proposed architecture as the Computerized Diagnostics Training Systems using modern information technologies and results of high-fidelity TS simulation. The system is verified by means of psychological engineering experiment again involving "first principles" TS simulator. The areas for future research and development are outlined in the Conclusions.

STATE-OF-THE-ART IN THE AREA

One can distinguish the following approaches to FDI skills training:

Techniques demonstration is quite effective in particular tasks but meets troubles with skill transfer to other uncovered tasks.

Developing instructions. This turned out to be ineffective and sometimes even destructive, when used merely as declarative knowledge presentation. The better usage of instructions is in implicit fashion provoking operators to generate hypotheses and analyze symptoms (Schaafstal 1993);

Teaching FDI techniques also shows ambiguous results on skills transfer since operators often propose hypotheses upon some other rules beyond proposed techniques which can hardly be formulated by experts (Konradt 1995);

TS simulation. Currently, the accessibility and quality of computer simulation (Dozortsev and Kreidlin 2010) allows real TS replacement with its high-fidelity model in training, thus expecting the highest level of skills transfer. However, we share the opinion of Jonassen and Hung that the models themselves will not lead to successful training unless coupled with adequate methodic support (Jonassen and Hung 2006);

Intellectual Instructional Systems (IIS). This class of systems integrates various educational instruments such as the diagnostic problems solver, operator's actions assessor, interpreter of operator's decision strategies and others. Some developers of this type of systems declare the 80% FDI performance increase due to 20% increase of total training time devoted to IIS, apart from that, 20-30 hours of training with IIS effectively replaces the training via the real TS.

Apparently, it is the IIS framework that should be employed in modern FDI skills training. Nevertheless, practical attempts of creating valid methods for diagnostics training are not numerous. It was considered that operators' diagnostic training is inefficient for novel problems where there is no verified knowledge of symptoms. The mental load for novel problems solving is so intensive that the traditional training must be supplemented with operators' real-time support (Yoon and Hammer 1988). Such support must include action procedures in real-time situations, the feedback in cases of wrong actions and the TS state changes.

In general, one observes the lack of approaches to the problem of training diagnostic skills which would be adequate to operators' tasks complexity and the potentiality of modern information technologies.

DIAGNOSTIC SKILL FORMATION: COGNITIVE MODEL

In this paper the authors consider the *Diagnostic Problem (DP)* with regard to technical systems which state can be observed by operator through the vector of physical parameters. Each *Parameter* is considered as a single source of physical magnitude reflecting the state of some component of the TS. *The Fault* is regarded to some particular drastic change in the TS leading to sudden deviations of the TS parameters from its normal stable values (the so-called *Symptoms*).

In (Dozortsev 2009) an approach is proposed based on regulatory scheme of diagnostic skill formation and realization. The general statements of this approach are the following:

- *Diagnostics (D)* is a continuous process of decision-making comprising of stages which are different in types of necessary general skills and psychic processes involved in its realization;
- D skill training is based on general skills formation and fixation (consequences prediction and hypotheses generation) followed by its gradual integration in the D skill complex;
- The training environment must conform to the real operators' activity. The quality of obtained skills and its positive transfer are determined by the level of conformity;
- D can be considered as the process of hypotheses generation and testing. Its optimal realization is assessed by the set of formal criteria;
- Regulative structure of particular skills formation is represented by the scheme containing feedback loop(s), in which various elements of modeled TS and concepts of trainees' mental activity represent the sources of afferent signals and objects of influence.

Hypotheses generation and testing imply the proposition and verification of assumptions on the development of various TS abnormal modes. The regulative scheme of this skill formation (Figure 1) provides for selection of current hypothesis *H* followed by comparison of manifested symptoms *S* with the response S_p from the mental model under hypothetical influence. If the difference is greater than the threshold *d*, it is exposed to modification, otherwise it is considered to be acceptable and is declared as final *H**.

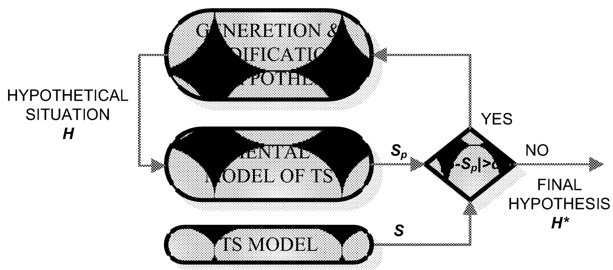


Figure 1. Hypothesis Formation Skill Cognitive Model

The particular DP contents are extremely diverse due to diversity of possible TS states, manifested symptoms and variety of influences on TS. The symptomatology presented during the training surely extends trainees' experience and promotes skill transfer to the real operators' activity. But the most essential is the adoption of the approach to problem solving. Thus, it is important that the operator is trained not only the correct influences on the mental model but rather the effective procedure of hypotheses generation and testing. The positive transfer is determined by the strengthening of the aim at DP solving.

However, it is essential that the regulative scheme on Fig.1 does not contain explicitly the mechanism of hypothesis modification in case if the manifested symptoms differ greatly from the predicted TS behavior. In the monograph (Dozortsev 2009) such mechanism was provided and analyzed in details.

It was found out that the solution of DP indeed involves the sequence of operators' inquiries of symptom values with the purpose of confirmation or modification of current hypothesis. Thus, in the cognitive aspect the diagnostic activity can be described as a closed loop scheme (Figure 2) consisting of three general elements: 1) the formation and further modification of Feasible Faults Set (FFS), i.e. the set of faults that agree with the current symptoms, 2) the decision of stopping the diagnostic search, 3) the new information inquiry based on current hypothesis and FSS.

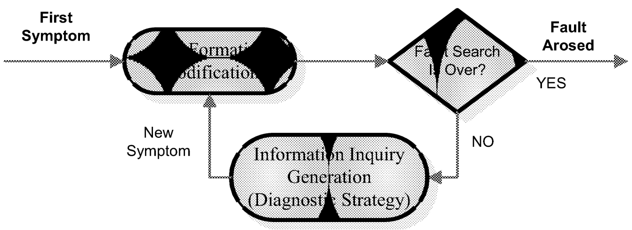


Figure 2: Generalized Cognitive Scheme of Diagnostic Problem Solving

The procedure of DP solving starts with the observation of the first symptom according to which the operator somehow forms the FFS. If the FFS is large in some sense reflecting the arisen fault ambiguity, the operator must make an inquiry of additional information about symptoms to advance in the DP solution. The criterion upon which the operator selects the TS parameter for the value observation in the search of symptom is called the *Diagnostic Strategy*. Observing more symptoms the operator modifies the FFS and once again decides whether he is sure enough to name a single Fault corresponding to observed symptoms (*The stopping Rule*) or he wants to get more information to reduce the ambiguity.

HEURISTIC DIAGNOSTIC MODEL

To support operators diagnostic decisions one needs to emphasize an important diagnostic construct, namely the *Diagnostic Model (DM)* of technical system. The DM lies in the basis of any diagnostic system which operator interacts with either performing diagnostics in real life or during training (Figure 3).

The DM is necessary for implementation of above-mentioned diagnostic sub-problems, and it is this model that must clearly demonstrate the cause-and-effect relations in the TS, which are generally stochastic, the feasible hypotheses and relative advantages of different information inquiries.

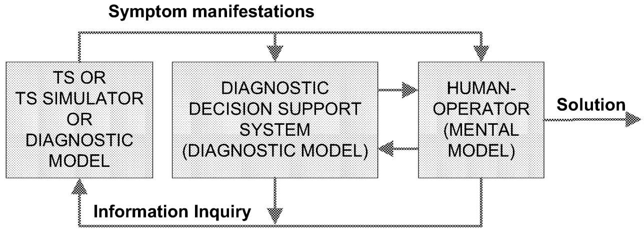


Figure 3: Training Using Diagnostic Model in the Operators' Decision Support Circuit

Three diagnostic models were considered in the paper (Dozortsev and Nazin 2009), deriving the so-called *Heuristic Model* as the most suitable one for above-mentioned purposes, which allows for probabilistic TS behavior along with the qualitative nature of human reasoning.

Fault-Symptom Matrix

We consider the *Fault-Symptom Matrix (FSM)* as the basis for diagnostic models defined as the set of a) n faults with a priori probabilities $\Pr(\hat{i}=i)=p_i$, $i=1,...,n$, $\sum_{i=1}^{n} p_i = 1$, here the $\hat{i}$ index denotes the random fault, b) m TS parameters, with the number K_j of possible discrete symptoms for j-th parameter, $j=1,...,m$. We consider the stochastic incidence of one single Fault at a time in the TS.

The relation between faults and parameters' symptoms is defined by conditional probabilities $\alpha_{i,j,k}=\Pr(\hat{k}(j)=k \mid \hat{i}=i)$ of symptom k manifestation $(k=1,...,K_j)$ in parameter j of the TS with the advent of fault i. The vector $a_{i,j}=(\alpha_{i,j,1},...,\alpha_{i,j,K_j})$, which we call as *The Symptom Vector*, in essence is the probabilities distribution of random symptom manifestation. For simplicity of the following statement we consider $K_j=3$ discrete deviation values corresponding to the value delta "decreased", "unchanged" and "increased".

Stated above is the "Stochastic" part of the *FSM* (further – *SFSM*) which should describe the objective behavior of TS in narrow diagnostic sense. The *SFSM* is used merely for DP imitations during training, i.e. picking the random fault and symptoms, which are external factors for DP solving procedure. All the internal factors represented in figure 2, i.e. FFS modification, stopping rule and information inquiry strategies, are based on the "Heuristic" part of the FSM (further – *HFSM*).

The *HFSM* should reflect the "human's" subjective understanding of TS behavior. Thus, *HFSM* cause-and-effect relations are concerned in qualitative terms.

Particularly, the symptoms are the couple $s_{i,j}$ and $sc_{i,j}$, where $s_{i,j}$ represents the *deviation* with values 0 – "unknown" or "any", 1 – "decrease", 2 – "unchanged", 3 – "increase", and $sc_{i,j}$ represents the *confidence* with values 1 – "probably" and 2 – "definitely". For arbitrary value K_j the qualitative scale of $sc_{i,j}$ must start with value $K_j - 1$ corresponding to "highest" confidence. Lower $sc_{i,j}$ values correspond to "lower" confidence up to the least value of 1.

In petrochemical industry the serious accident is a relatively rare event. Thus, in the lack of rich statistical data from the TS that could have allowed to assess the values p_i and $a_{i,j,k}$ of SFSM, we use the high-fidelity TS simulation model as the data source for HSFM further projected to SFSM.

Particularly, the faults are run in TS model and thus adopted parameter's deviations are mapped to the values of HFSM. The Experts later approve obtained results. It is assumed that there is a pre-selected scale defining the mapping of the continuous parameter value (analogue signal) deviation to HFSM couples $\langle s_{i,j}, sc_{i,j} \rangle$. Then, the SFSM values are projected from the HFSM according to the following table:

Table 1: HFSM to SFSM Symptom Values Mapping

HFSM couple $\langle s_{i,j}, sc_{i,j} \rangle$	**Symptom description**	**SFSM symptom vector** $a_{i,j} = (\alpha_{i,j,1},...,\alpha_{i,j,3})$
$\langle 1,2 \rangle$	Definitely decreased	$(1,0,0)$
$\langle 2,2 \rangle$	Definitely unchanged	$(0,1,0)$
$\langle 3,2 \rangle$	Definitely increased	$(0,0,1)$
$\langle 1,1 \rangle$	Probably decreased	$(q,(1-q),0)$
$\langle 2,1 \rangle$	Probably unchanged	$(0.5\times(1-q),$ $q,\ 0.5\times(1-q))$
$\langle 3,1 \rangle$	Probably increased	$(0,(1-q),q)$
$\langle 0,1 \rangle or \langle 0,2 \rangle$	"Unknown" or "Any deviation"	$\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$

The value of q should be selected in advance and slightly less than 1 (e.g. $q = 0.9$).

Feasible Faults Set (FFS)

The Faults *Confidence Weights (CW)* are considered as a measure of its conformity to observed symptoms during DP solving process with three qualitative values:

$cw_i = 0 \Rightarrow$ "impossible";

$cw_i = 1 \Rightarrow$ "possible";

$cw_i = 2 \Rightarrow$ "highly possible".

In the beginning of DP solving process (step $t = 0$ in the superscript) the *CWs* are initialized as $cw_i^{[0]} = 1,\ i = 1,...,n$.

Observing the symptom value $\hat{k}(j)$ on the $t = 1,2,...$ step of the diagnostic search with relevance to HFSM symptom values for parameter j in fault i, its posterior *CW* is transformed according to the following set of heuristic rules:

- The *CW* is increased only if it has not been set to null on previous steps and the observed symptom coincides exactly with the expected one.
- The *CW* is left unchanged either if the expectance of symptom is unknown (or arbitrary) or if the deviation of symptom from the expected one is within the uncertainty limit. The latter is defined by the value $K_j - sc_{i,j}$.
- The *CW* is set to null if the deviation of symptom from its expected value is out of the uncertainty limit.

Formally these rules can be written as follows:

$$cw_i^{[t]} = \begin{cases} 2, & if\ cw_i^{[t-1]} > 0\ \ and\ \ \left(\hat{k}(j) - s_{i,j}\right) = 0; \\ cw_i^{[t-1]}, & if\ s_{i,j} = 0\ or\ 0 < \left|\hat{k}(j) - s_{i,j}\right| < K_j - sc_{i,j}; \\ 0, & if\ s_{i,j} \neq 0\ \ and\ \ \left|\hat{k}(j) - s_{i,j}\right| \geq K_j - sc_{i,j}. \end{cases} \quad (1)$$

The $FFS^{[t]}$ is defined as the set of faults indices with nonzero *CWs* on the t step of the diagnostic search:

$$FFS^{[t]} = \left\{ i \in \{1,...,n\} : cw_i^{[t]} > 0 \right\}$$

In the following statements the $FFS^{[t]}$ according to (1) is considered as a functional of two independent variables: parameter j and symptom value $\hat{k}(j) = k$.

Stopping Rule

The diagnostic search stops if the set $FFS^{[t]}$ contains one single fault index which corresponds to diagnostic search solution. If this condition is not met upon inquiring all parameters, one of the faults is selected randomly from $FFS^{[t]}$ as a candidate for *"Confirmation"*, i.e. a special procedure which ether confirms the truth of that solution, or confirms its falsity thus excluding it from FFS. Therefore, any DP is solvable, since under stated conditions and algorithm (1) the picked fault i^* cannot leave the $FFS^{[t]}$.

Information Inquiry Strategies

Analyzing engineers' behavior in various kinds of searching and decision-making problems like DP the researchers discriminated two types of normative strategies of making the selection for information inquiry (Dale

1957). One of them is the MINIMAX strategy (MM), namely the criterion of minimizing the maximum possible uncertainty left upon the receipt of new information. The MM strategies relate to the kind of guaranteed behavior in decision-making since relying on the worst-case outcome in the sense of problem uncertainty measure minimization, which in our case is the Feasible Faults Set cardinality.

In considered heuristic model the MINIMAX criterion is:

$$j_{MM}^{[t]} \in Arg \min_j \max_k Card\left(FFS^{[t]}(j, \hat{k}(j) = k)\right). \quad (2)$$

The other type of strategies is related to selecting some single fault i_{hyp} as a candidate for solution and its testing by inquiring symptoms upon hypothesis fastest proof or disproof. In our case the i_{hyp} is selected among the faults with highest *CWs* on current DP solution step. The corresponding symptom inquiry criterion is as follows:

$$j_{HT}^{[t]} \in Arg \min_j Card\left(FFS^{[t]}(j, \hat{k}(j) = k_{hyp}(i_{hyp}, j))\right). \quad (3)$$

Here $k_{hyp}(i_{hyp}, j) = s_{i_{hyp},j}$ responds to indices of most probable symptoms for hypothetical fault.

MOTIVATIONAL ASPECT OF DIAGNOSIS SKILLS TRAINING

One of the most important features the effective training system must possess is the reflection of motivational structure of operator's activities during the training to his motivation in accomplishing actual diagnostic search tasks. This motivation is determined by a specific set of stressful factors experienced by the operator during practical activities. The following 3 types of TS operator's stress can be emphasized.

- *Hazardous operation stress*. The stress caused by the operation under dangerous values of some process variables (like furnace's extra high temperatures) cannot be simulated within a training session;
- *Losses expectation stress*. The feeling of responsibility for the *consequences* concerned with TS malfunctions can be simulated by calculating the overall "game" cost, which depends on the quality of accomplishing diagnosis tasks;
- *Limited time stress*. This is the most important issue of *stirring-up* the operator's activities aimed at just the improvement of diagnosis techniques rather than the motivation of the solution itself. Direct emulation of real time limits of diagnostic task accomplishment is ineffective; however, motivational likeness could be attained if the real-life lack of decision-making time will be replaced by the lack of available information in the training game.

DIAGNOST: A COMPUTER-BASED DIAGNOSTIC DECISION-MAKING SUPPORT SYSTEM

The above-stated principles were implemented in the computer training system *DIAGNOST*, the prototype of which was presented in (Dozortsev 2001). Recently, the heart module of *DIAGNOST*, i.e. its diagnostic model, was upgraded with the heuristic one. The system provides diagnostic decision-making support both on FFS formation/modification stage and on symptom inquiries assessment.

The training takes the form of a step-by-step game on a predefined sub-matrix of the FSM. Initially, the system randomly picks one of the faults and presents merely one of its symptoms to the trainee. Trainees' objective is to define the picked fault by checking the list of all feasible faults and/or inquiring symptoms for TS parameters. Each fault check or symptom inquiry costs some "time points", where the former is much more expensive than the latter. The aim is to solve the DP on receiving the minimum summary time points. Thus, the game directive focuses the operator on searching a sequence of information inquiries that provides the "quickest" solution. Such diagnostic game conforms to the motivational structure of real diagnostic activity. During the game the trainee can address to TS schematics and textual descriptions for help. Each trainee's symptom inquiry is assessed according to normative strategies upon which the trainee receives supporting comments on the optimality of the choice made.

DIAGNOST was tested in laboratory and applied environment (more than 60 installations by year 2010). Particularly, process operators of a vacuum distillation unit at one of the largest Russian refineries were trained with the help of *DIAGNOST*. The trainees' poll and the analysis of training logs showed (Figure 4) that with the number of "games" (Figure 4 abscissa) increase:

- The trainees become more 'familiar' with the unit, since the number of trainees' addresses to process diagrams and textual description decrease (curves 1 and 2 respectively);

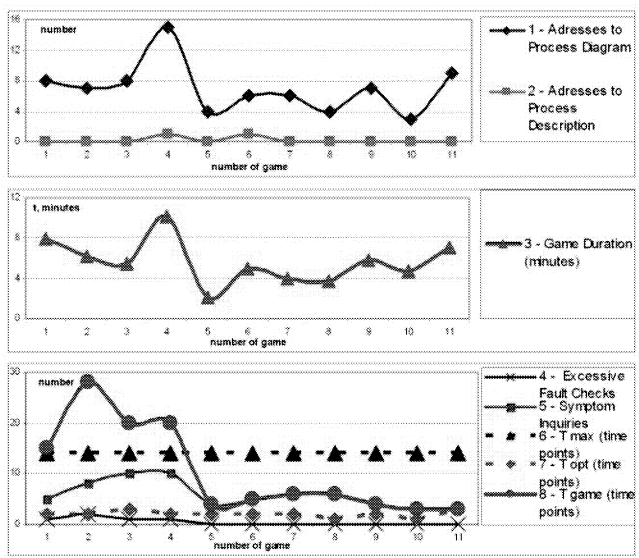


Figure 4: Results of Training with *DIAGNOST*

- Diagnostic task execution is mastered what can be seen from the decrease of problem solving duration (curve 3);

• Finally, the most important is the change in solution strategy. This is disclosed from: (i) the number of mistaken cause tests which comes to naught (curve 4) and (ii) the reduction of symptoms inquiries amount (curve 5) towards the minimax-optimal number (curve 7) with comparison to the maximum number of possible inquiries (curve 6).

It should be noted that the shape of the learning curve 8, which reflects the overall game time, complies with the classical ideas of the learning theory and includes along with the initial period of low quality also a sharp improvement, a plateau and the final segment of the developed skill.

Another research was taken to verify the effectiveness of *DIAGNOST* as a tool for preliminary training regarding the main training on the TS simulation model. Such a psychological engineering experiment involved two test groups of three graduating students of technical education. The groups were equal in sex, age and experience. The plan of experiment is presented in the following Table 2.

Table 2: *DIAGNOST* Verification Experiment Plan

	Stage 1	Stage 2	Stage 3
Experimental group	X	O_1	O_2
Control group	O_3	O_4	O_5

X stands for training on *DIAGNOST*. The $O_{1\text{-}5}$ are the mean times taken on diagnostic problems solution (averaging over stage and group).

The experimental group worked with *DIAGNOST* on the 1st stage, then solved 10 diagnostic problems on TS simulator on the 2nd stage, and after 2 days interval another 10 diagnostic problems on the 3rd stage with some new faults not experienced on the 1st and 2nd stages, thus imitating transfer from TS simulator to the real TS. The control group worked with TS simulator on all three stages solving 10 diagnostic problems on each.

Hence, each $O_{1\text{-}5}$ is averaged over 30 values what makes use of Student's criterion. The following statistically valid results were received:

• $O_2 < O_1$ *and* $O_5 < O_4 < O_3$: the learning effect which is preserved after training interruption and novel problems occurrence.

• *O1 << O3*: preliminary training with *DIAGNOST* substantially improves the first stage on TS simulator. This is the result of special effect that *DIAGNOST* gives along with the basic familiarization with the TS.

• *O1 << O4 and O1 < O5*: preliminary training with *DIAGNOST* improves the first endeavor of experimental group on TS simulator compared with the results on the 2nd and even 3rd stages of the control group. This is the main result of the experiment which confirms the usefulness of preliminary training with *DIAGNOST*.

CONCLUSIONS

An upgrade to the paradigm of computerized diagnostic skill training is proposed, implemented and tested. It is based on heuristic model representation of the technical system's fault-symptoms behavior initially obtained from high-fidelity simulation. Two basic diagnostic problem solving strategies are implemented. The learning environment provides motivation representation of real diagnostic activity realized in a step-by-step game allowing for trainees' decision-making support. The technical system simulator is again used for validation of proposed system effectiveness. Obtained results confirm the opportunity of inexpensive and non-time-consuming formation, and developing of effective diagnosis skills.

The most promising development lines are connected with:

• Comparative evaluation of mental costs for implementation of different diagnosis strategies;

• Further study of real diagnosis, selection of effective strategies and including them into reference patterns used in decision-making support;

• Incorporation of *DIAGNOST* into combined decision-support system for training other operator's skills, i.e. deviation detection, prediction, hypotheses generation, compensation planning.

REFERENCES

Brighton, H. and G.Gigerenzer. 2008. "Bayesian Brains and Cognitive Mechanisms." *The Probabilistic Mind.* Oxford, UK, 189-208.

Dale, H.C.A. 1957. "Fault-Finding in Electronic Equipment." *Ergonomics*, vol.1, no. 4, 356-385.

Dozortsev, V.M. 2001. "Diagnost: A Software for Developing Efficient Decision-Making Strategies." *Proc. "Euromedia'2001"*. Delft, Netherlands, 261-268.

Dozortsev, V.M. 2009. *Computer-based Operator Training Systems*. Moscow, Sinteg Publ. (in Russian).

Dozortsev, V.M. and V.A. Nazin. 2009. "New Approach to Training of Technical Systems Faults Diagnostic Skills: Use of Probabilistic Models." *Proc. 13th IFAC Symp. on Information Control Problems in Manufacturing*. Moscow, Russia, 701-706.

Dozortsev, V.M. and E. Yu. Kreidlin. 2010. "State-of-the-Art Automated Process Simulation Systems." *Automation and Remote Control*, vol.71, no.9, 1955-1963.

Duguid, I. 2001. "Take this Safety Database to Heart." *Chemical Engineering,* no.7. 80-84.

Jonassen, D.H. and W. Hung. 2006. "Learning to Troubleshoot: A New Theory-Based Design Architecture." *Educational Psychology Review*, vol.18, no.1, 77-114.

Konradt, U. 1995. "Strategies of Failure Diagnosis in Computer-Controlled Manufacturing Systems." *Intern. Journ. Human-Computer Studies*, vol.43, 503-521.

Schaafstal, A. 1993. "Knowledge and Strategies in Diagnostic Skill." *Ergonomics.* vol.36, no.11, 1305-1316.

Yoon, W.C. and J.M. Hammer. 1988. "Deep-Reasoning Fault Diagnosis." *IEEE Trans. SMC,* vol.18, no. 4, 659-676.

SIMULATION AS A TOOL FOR TUNING HYDROPOWER SPEED GOVERNORS

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

KEYWORDS
Modeling, simulation, tuning, permanent droop, dynamic identification

ABSTRACT

The paper is devoted to an important industrial application of the numerical simulation in the field of the electric systems power management: the fine tuning of the permanent speed droop, and the temporary speed droop of the speed governors controlling the strong hydropower units. Starting from these parameters, the designer can find the real value of the proportional gain of the PID compensator, which is the main parameter both from a stability and precision point of view. In any power system the frequency depends on the disturbances introduced by the power generators, and the consumers. The use of high power nuclear power plants, on the one hand, and the rapid increase of the wind „farms" contribution on the other hand, create new problems for the voltage and frequency control. The lastest generation of speed governors use industrial process computers to accomplish different control tasks. A lot of the power plant's peculiarities are influencing the real value of permanent or temporary droop. The dispatcher of a power plant needs to know these parameters only. The authors developed a method to find the connection between the proportional gain and the permanent droop by combining numerical simulations with experimental identification performed in the final stage of the building or refurbishing of a hydropower plant. The numerical results of the method were validated by a long series of experiments carried out on two electrohydraulic speed governors designed by the authors for Rm. Valcea Hydropower Station, in Romania.

INTRODUCTION

The operating regime of a hydraulic turbine depends on the hydraulic parameters of the hydropower plant, and the operating conditions of the electric generator. The hydraulic torque depends on the net hydraulic head H, the guide vane opening Y, and the shaft angular speed ω. For a constant net head (H=ct.) the turbine torque can be computed by the equation

$$M_t = M_0 \cdot (1 - k \cdot \omega), \tag{1}$$

and the power delivered by the turbine becomes

$$P_m = M_t \cdot \omega = M_0 \cdot \omega - k \cdot M_0 \cdot \omega^2 \tag{2}$$

The "natural" behavior of a hydraulic turbine for H=ct. and Y=ct. can be typically described by a family of characteristic curves as in Figure 1. The change of the guide vane opening "inflates" the mechanical power curve in the same direction, as it is shown in Figure 2.

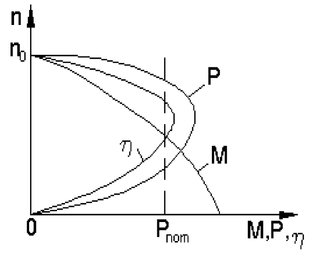


Figure 1: Typical characteristics of the hydraulic turbines

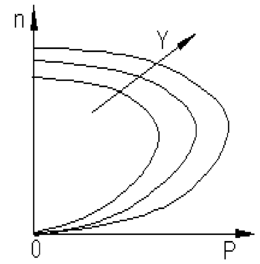


Figure 2: Turbine power curves versus guide vane opening

For a given frequency, the steady-state consumers power characteristics depend on their internal structure. The power supplied by the electric generator depends in a parabolic manner on the frequency (turbine angular speed) as shown in Figure 3. A steady-state system operation is achieved when the turbine's mechanical power equals the electric power supplied by the generator to the network and the generator internal losses. The stable operating point of the system is set by the intersection of the two curves, and is a „floating" one. Consequently, the turbine speed and the current frequency depend on the consumers power.

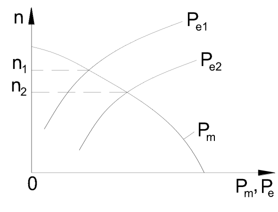


Figure 3: Steady-state characteristics of the consumers

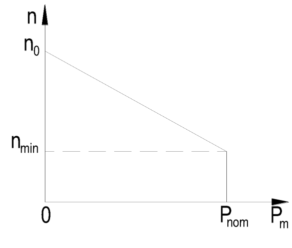


Figure 4: Linearized external characteristics of a hydraulic turbine

The great negative slope of the turbine's natural characteristics leads to a large range of frequency variation, which cannot be accepted by the modern consumers. The frequency quality control is achieved by the aid of a speed-governing system controlling all the operating modes: normal starting and stopping, connection to the network, emergency shutdown etc. For solving speed-governing problems, the external turbine mechanical characteristics n(P) can be regarded as a linear one (Anton 1) as is shown in Figure 4. The linear characteristics can be written in terms of speed or frequency:

$$n = n_0 - \frac{n_0 - n_{\min}}{P_{nom}} \cdot P \tag{3}$$

$$f = f_0 - \frac{f_0 - f_{\min}}{P_{nom}} \cdot P \tag{4}$$

The quantities

$$s_{0n} = \frac{n_0 - n_{\min}}{P_{nom}} \tag{5}$$

$$s_{0f} = \frac{f_0 - f_{\min}}{P_{nom}} \tag{6}$$

are regarded as „natural droop" with respect to the active power. The external hydropower unit characteristics becomes:

$$n = n_0 - s_{0n} \cdot P \tag{7}$$

or

$$f = f_0 - s_{0f} \cdot P \tag{8}$$

PARALEL OPERATION OF THE HYDROPOWER UNITS

Any hydropower unit may be regarded as a control system defined by the typical structure from figure 5.

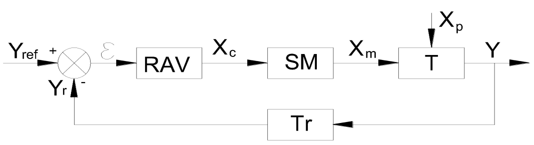


Figure 5: Block diagram of the hydropower unit

The control system is a static one if the steady-state operation is defined by a relationship between the output quantity and the disturbance:

$$\frac{\partial Y_{st}}{\partial X_p} = S \neq 0 \tag{9}$$

Here S denotes the permanent droop of the control system. The consumer active power change generates variations of the frequency, and consequently - variations of the supplied active power. In any modern power system, the hydropower units are operating in full connection because the frequency of the delivered power is the same in the supply region, which can be a county, a country or a continental zone (Figure 6).

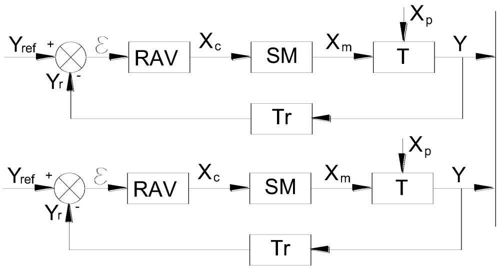


Figure 6: Paralel operation of two hydropower units

In this case, the system is disturbed by the variation of the overal active power consumption, which has to be covered by all the hydropower units by overloading or unloading them. This goal can be achieved by a static control system, introducing a smaller permanent droop than the natural one., The frequency control system can distribute the disturbance between different hydropower units by using different permanent droops. The optimal distribution depends on many operation conditions; usually, an expert system supports the decision. The natural permanent droop can be reduced to zero, turning the frequency control in a zero frequency error band. A normal value of the permanent droop is sited in the range 1%...10%. The practical manner of solving this problem depends on the compensator type. A speed governor measures the turbine speed, computes the speed error, and orders the gate opening in order to translate the external hydropower unit characteristics thus to reduce the frequency error.

SPEED GOVERNOR ARHITECTURE

According to IEC 61362 from 1998, it is to be expected that in the future, PID-controllers will remain in use for many plants for speed, power and head level control. Higher order algorithms, e.g., state control schemes will be used for the more complex system requirements. These control schemes, while necessitating more effort to implement, are justified where superior behavior with respect to the magnitude of deviations from steady state and its return to steady state can be achieved.

It is to be noted that an electronic PID-controller´s behavior can also be enhanced considerably by readily available special means, such as disturbance superposition and the feedback of secondary variables.

This in turn justifies the preference of the designers to use the PID-governor as a basis and reference for recommendations relating to system control.

The recommended ranges in parameter adjustment will suffice in all normal cases. Special conditions – extremely low inertias, extremely long penstocks – should in all cases be subjected to digital simulation and may require an extension of the recommended parameter adjustment range.

In the last three decades, the structure of the hydropower units speed-control systems was continuously improved, acccording to the customers demands, and operating security needs. The latest generation of these systems promotes the combination between the high power electrohydraulic servovalves and the Industrial Process Computers based on the Digital Signal Processors.

The main options for the governing system "architecture" depend on the size and hydropower weight in the power system, configuration of the power network, the manufacturer experience etc. At the same time, the performance requirements are stated by international standards (IEC, ASME, JSME etc.) and have to be fulfiled by all the electric energy producers, connected to the power networks. Usually, the speed of a hydraulic turbine is controlled by the system presented in Figure 7.

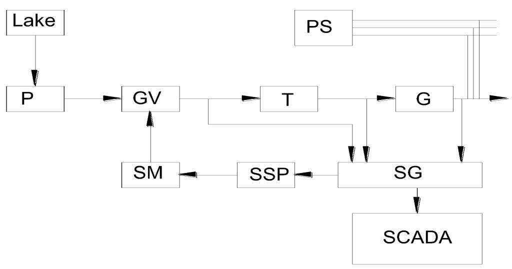


Figure 7: General structure of the control system of a hydropower unit. P: penstock; GV: guide vane; SM: servomotor; T: turbine; G: generator; SG: speed governor; PS: electric power system; SSP: servomechanism setpoint

All the hydropower units speed control systems include blocks (soubroutines) for generating the permanent and the temporary droop. These quantities are defined in Figures 8a and 8b.

The permanent droop, b_p establishes a defined relationship between the controlled variable *x*, and the relative servo-position or any other signal, in the steady-state condition, e.g. a) speed control: between frequency and servo position; b) level control: between level and servo position; c) power control: between power and frequency.

The temporary droop can be defined by the average slope of the relationship between the frequency variation and the servomotor stroke variation occuring during transients.

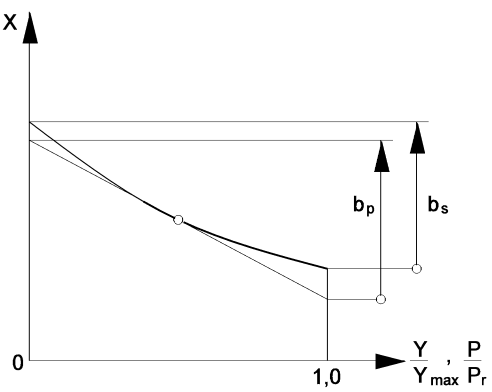


Figure 8a: Definition of the permanent droop

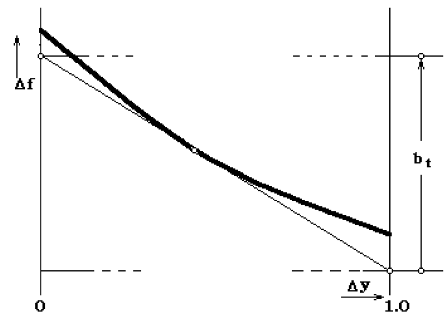


Figure 8b: Definition of the temporary droop

According (IEC 61362, 1998), the values of the permanent droop have to be computed by the following relations:

$$b_p = \frac{\frac{\Delta f}{f_n}}{\frac{\Delta y}{Y_{nom}}} \cdot 100[\%] \qquad (10)$$

$$b_p = \frac{\frac{\Delta f}{f_n}}{\frac{\Delta P}{P_{nom}}} \cdot 100\ [\%] \qquad (11)$$

The governor has to be designed to be stable in any operating mode, but it need always independent blocs to achieve the permanent and tranzient droops. Usualy, these blocks are designed in an independent manner, to avoid any undesirable connection.
The digital electrohydraulic governors are generating the two droops by special soubroutines with clear parameters which can be tuned during the hydropower unit operation.

PARAMETER IDENTIFICATION FOR INTRODUCING A PERMANENT DROOP

The energy market demands the real time change of the permanent droop by the operator, taking into account the energy supply contract, and the real hydraulic, mechanic, electric or thermal powerplant unit parameters. Different design solutions are used to generate the permanent droop block. În Figure 9 a common design solution of the speed governors based on the servomotor stroke feedback is prezented. Both permanent and temporary droop are generated by analog or digital blocks. For safety reasons, at least two displacement transducers are used for this purpose.

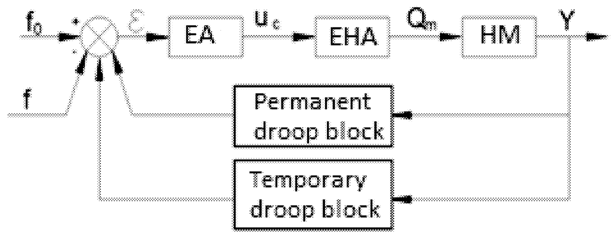


Figure 9: General structure of a speed governor

The optimal values of the tuning parameters for the permanent droop block depends on many factors. Some of them are always uncertain. The classical frequency analysis methods developed by (Borel, 1960), (Pivovarov, 1974), (Jaeger, 1977), (Catana, 1981) etc. use the linear system theory. The option regarding the operating point considered for the linearization of the differential equations describing such a complex process can create many difficulties. The financial effort needed for developing an efficient laboratory experimental research is not affordable for small manufacturers. A much more efficient method is, using numerical simulation. Some optimization methods are available for a pseudorandom search of the optimal parameters. Figure 10 presents the structure of a high accuracy, flexible digital electro hydraulic speed governor for hydro power units with two control blade systems developed by the authors (Romanian Patent No. 120101 / 2003). The speed governor contains an industrial process computer with digital signal processor (DSP) which controls two electro hydraulic servomechanisms, one for each blade system (guide vane and runner).
Two stages electro hydraulic proportional servo valves are used to control the flow of the hydraulic servomotors. The digital connection between the two servo systems is tuned for the best power efficiency in every regime. Both frequency, power and water flow can be controlled by choosing a functional key.
The main advantages of this new concept are the following: fine tuning of the digital compensator during the operation; continuous monitoring of the main hydropower plant parameters; real time fine tuning of the digital compensator for the best efficiency; high overall reliability as a result of the simplest hydromechanical structure.

The numerical simulation network of the nonlinear electrohydraulic servomechanism is presented in Figure 11. The whole speed governor simulation network is shown in Figure 12. Taking into account different kinds of nonlinearities, studied by (Wylie and Streeter, 1993) and (Souza et.al. 1999) and using a high level integration algorithm, both in SIMULINK and AMESIM, the responses like those from figure 13 are obtained.

The real behavior of the hydropower is shown in Figure 14 for the same input. Figure 15: The relation between K_{bp} and b_p for a Kaplan hydropower unit (Rm. Valcea Hydropower - Romania). The theoretical results and the practical ones are found to be in very good agreement.

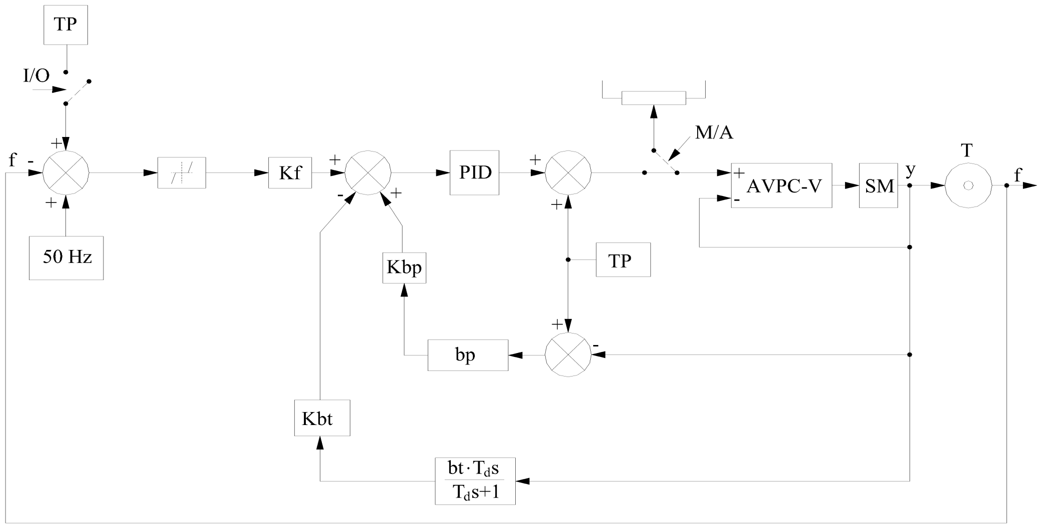


Figure 10: Block diagram of an electrohydraulic speed governor for idle or insulated operation

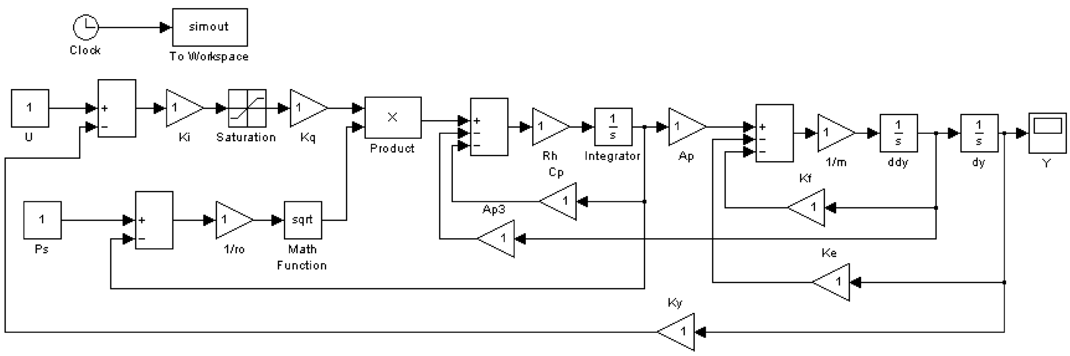


Figure 11: Block diagram for simulating the dynamic behavior of the electrohydraulic servomechanism

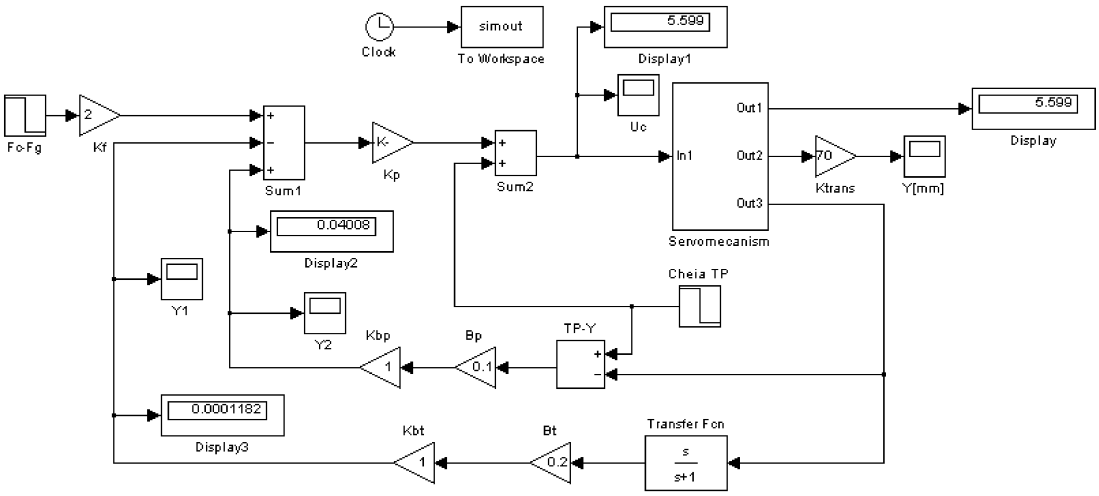


Figure 12: Block diagram for simulating the dynamic behavior of the speed governor with permanent droop

EXPERIMENTAL IDENTIFICATION

The transfer function of the turbine considering the guide vane as input variable and the rotational speed as output variable can be determined using a Pseudo Random Binary Sequence (PRBS) signal (Nicolet et al., 2007). This procedure needs a long series of rectangular small amplitude signals applied to the guide vane. The authors' option was to "inject" a frequency drop of 200 MHz at the input of the system, without serious wear on the control mechanism. The modern control panel of a speed governor is integrated in the power plant SCADA system graphic interface (figure 16). One "page" of the touch screen has to contain the real value of the permanent droop. The control program written in the Industrial Process Computer (the core of the speed governor) has to allow the change of this parameter by the serial bus of the SCADA system (RS485, Profibus, Fieldbus etc.).

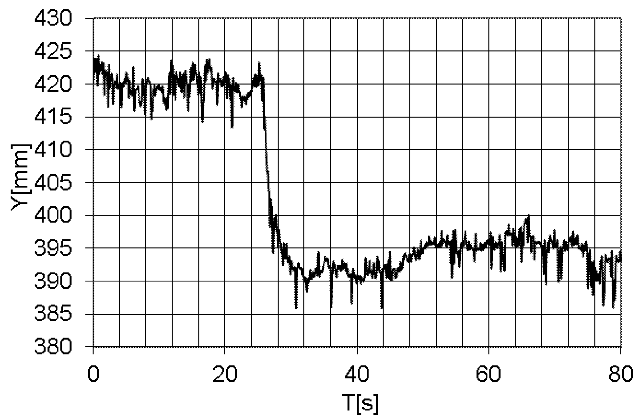


Figure 14: Real speed governor response for an imput frequency step of 200 mHz

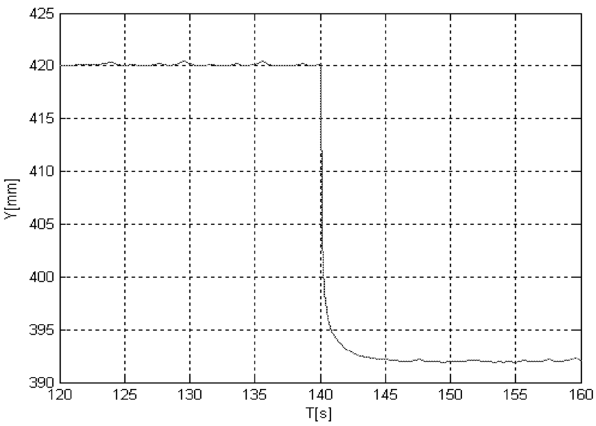


Figure 13: Simulated speed governor response for an imput frequency step of 200 mHz

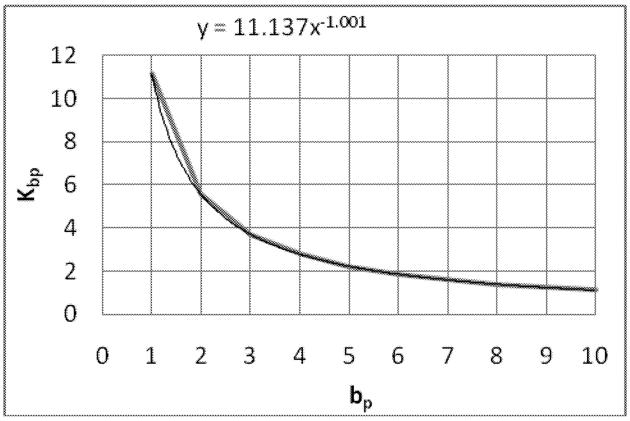


Figure 15: The relation between K_{bp} and b_p for a Kaplan hydropower unit (Rm.Valcea Hydropower Plant-Romania)

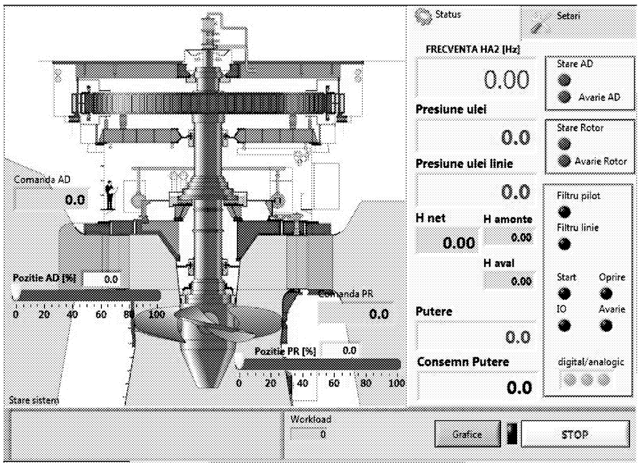


Figure 16: LabVIEW control panel of the Kaplan turbine speed governor (Rm.Valcea Hydropower Plant - Romania)

CONCLUSIONS

(Nicolet et al. 2006) have shown that it is not sufficient to validate the turbine governor parameters using a hydraulic model of the power plant, as electrical machines and power networks can considerably affect the turbine transfer function and lead to unexpected unstable operations, at least in islanded production mode. In addition, it is shown that power system stabilizers can contribute significantly to reduce speed deviations during load rejections 8. This can be built by a long series of experimental identifications only, as the author presented in this paper. Many other tests are needed for finding governor-controlled hydro generator performance. For example, the stability margin may be evaluated through its frequency response (ASME, 2005).

REFERENCES

Anton, I. 1979. “Hydraulic Turbines.” Facla Press House, Cluj (in Romanian).

ASME PTC 29-2005 (American National Standard).

Borel, L. 1960. “Stabilite de reglage des instalattions hydroelectriques, Dunod, Paris.

Catană, I. 1979. On the Stability of the Electrohidraulic Control Systems, U.P.B. Bulletin nr.1, pp.16-21.

IEC 61362. 1998. Guide to Specification of Hydraulic Turbine Control Systems (a document of The International Electrotechnical Commission).

Jaeger, R.C., 1977, “Fluid transients in hydro-electric engineering practice ", Glasgow, Blackie.

LMS International. 2010. “LMS Imagine.Lab-AMESim R10.” Roanne.

Nanu, S. N., 2003, “Researches on the Control Systems of the Hydropower Units, PhD Thesis, University “Politehnica” of Timişoara.

Nicolet, C., Greiveldinger, B., Allenbach, P., Kawkabani, B., Herou, J.-J., Simond, J.-J., Avellan, F., 2007, “Turbine speed governor parameters validation in islanded production.” In Scientific Bulletin of the “Politehnica” University of Timisoara, Transactions on Mechanics Tom 52(66), Fascicola 6.

Pivovarov,V.A. 1973. Proiectirovanie i rasciot sistem regulirovania ghidroturbin. Maşinostroenie, Leningrad.

Souza, O.H., Jr.; Barbieri, N.; Santos, A.H.M. 1999. “Study of hydraulic transients in hydropower plants through simulation of nonlinear model of penstock and hydraulic turbine model.” In IEEE Transactions on Power Systems, vol. 14, issue 4, pp. 1269 – 1272,

Vasiliu, N. and Calinoiu, C. 2003. “A new type of Electro Hydraulic Speed Governor for Kaplan Turbines. Romanian Patent No.120101.

Wylie, E. B. and Streeter, V.L., 1993, “Fluid Transients in Systems”, Prentice Hall, Englewood Cliffs, N.J.

BIOGRAPHY

Dr. Nicolae Vasiliu graduated in Hydropower Engineering from Polytechnic University of Bucharest in 1969. He became a Ph.D. in Fluid Mechanics after a research stage in Ghent State University and the Von Karman Institute in Brussels. He became state professor in 1994, leading the Fluid Control Laboratory from the Power Department the Polytechnic University of Bucharest. He worked always for the industry, as project manager or scientific advisor. In 1980 he joined the Hydraulic Control Team from the Romanian Aerospace Institute. He is working mainly in modeling, simulation, dynamic identification, remote control, and virtual instrumentation of the hydraulic and electro hydraulic control systems. Some highlights of his career include: manager, from 1996, of the ENERGY & ENVIRONMENT RESEARCH CENTRE from Polytechnic University of Bucharest; director of the ROMANIAN INNOVATION FINANCING AGENCY between 2006 and 2010. In the FLUID POWER NET INTERNATIONAL, Dr. Vasiliu currently serves as Romanian Chairman. He has received many awards for achievements in education and leadership. Some of his patents gained international awards in Geneva, Brussels, Moscow and Warsaw, and were successfully applied.

AUTHOR LISTING

AUTHOR LISTING

AUTHOR LISTING